Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 117 (2016)

From Weak to Strong LP Gaps for all CSPs

Mrinalkanti Ghosh* Madhur Tulsiani†

August 1, 2016

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 no weaker than the approximation obtained using relaxations given by $\Omega\left(\frac{\log n}{\log\log n}\right)$ levels of the Sherali-Adams hierarchy on instances of size n.

It was proved by Chan et al. [FOCS 2013] that any polynomial size LP extended formulation is no stronger than relaxations obtained by a super-constant levels of the Sherali-Adams hierarchy. Combining this with our result also implies that any polynomial size LP extended formulation is no stronger than the basic LP.

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 $\Omega(\log\log n)$ levels of the Sherali-Adams hierarchy cannot achieve an approximation better than a random assignment. We simplify their proof and strengthen the bound to $\Omega\left(\frac{\log n}{\log\log n}\right)$ levels.

1 Introduction

Given a finite alphabet $[q] = \{0,\ldots,q-1\}$ and a predicate $f : [q]^k \to \{0,1\}$, an instance of the problem MAX k-CSP(f) consists of (say) m constraints over a set of n variables $x_1,\ldots,x_n$ taking values in the set $[q]$. Each constraint C_i is of the form $f(x_{i_1}+b_{i_1},\ldots,x_{i_k}+b_{i_k})$ for some k-tuple of variables $(x_{i_1},\ldots x_{i_k})$ and $b_{i_1},\ldots,b_{i_q} \in [q]$, and the addition is taken to be modulo q. We say an assignment σ to the variables satisfying the constraint C_i if $C_i(\sigma(x_{i_1}),\ldots,\sigma(x_{i_k})) = 1$. 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., [27] 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 OPT(Φ) denote the *fraction* of constraints satisfied by an optimal assignment, and let FRAC(Φ) denote the value of the convex (LP/SDP) relaxation for the problem. Then, the

*Toyota Technological Institute at Chicago. mkghosh@ttic.edu
†Toyota Technological Institute at Chicago. madhurt@ttic.edu.

ISSN 1433-8092

performance guarantee of this algorithm is given by the integrality gap which equals the supremum of $\frac{\mathsf{FRAC}(\Phi)}{\mathsf{OPT}(\Phi)}$, 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 [22] 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 [11] proved a lower bound of $2 - o(1)$ for the integrality gap of the LP relaxations for MAX-CUT given respectively by $\Omega(\log \log n)$ and $n^{\Omega(1)}$ levels of the Sherali-Adams LP hierarchy [26]. Several follow-up works have also shown lower bounds for various other special cases of the MAX k-CSP problem, both for LP and SDP hierarchies [1, 25, 30, 24, 6, 4].

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 $\mathsf{FRAC}(\Phi) \geq c$ and $\mathsf{OPT}(\Phi) < s$. 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 tradeoff (described later) when t is a function of n, which was recently improved by Kothari et al. [19].

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 $\Omega_\varepsilon\left(\frac{\log n}{\log \log n}\right)$ 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 the basic LP, implies a similar lower bound on the integrality gap of any polynomial size extended formulation. We note that both the above results and our result apply to all f, q and all $c, s \in [0, 1]$.

Comparison with (implications of) Raghavendra's UGC hardness result. A remarkable result by Raghavendra [23] shows that a (c, s)-integrality gap instance for a "basic SDP" relaxation of MAX k-CSP(f) implies hardness of distinguishing instances Φ with $\mathsf{OPT}(\Phi) < s$ from instances with $\mathsf{OPT}(\Phi) \geq c$, assuming the Unique Games Conjecture (UGC) of Khot [14]. 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 [24], and by Khot and Saket [15], 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}.

A recent result by Khot and Saket [16] shows a connection between the integrality gaps for the basic LP and those for the basic SDP. They prove that, assuming the UGC, a (c, s) integrality gap instance for the basic LP implies an NP-hardness of distinguishing instances Φ with $\mathsf{OPT}(\Phi) \geq \Omega\left(\frac{c}{k^3 \cdot \log(q)}\right)$ from instances with $\mathsf{OPT}(\Phi) \leq 4s$. Their result also shows that a (c, s) integrality gap instance for the basic LP can be used to produce a $\left(\Omega\left(\frac{c}{k^3 \cdot \log(q)}\right), 4s\right)$ integrality gap instance for the basic SDP, and hence for $\Omega((\log \log n)^{1/4})$ levels of the mixed hierarchy.

Other related work. The power of the basic LP for solving valued CSPs *to optimality* has been studied in several previous works. These works consider 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 [28] that when the penalties are finite-valued, if the problem of finiding the optimum solution cannot be solved by the basic LP, then it is NP-hard. Kolmogorov, Thapper and Živný [18] 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 [29] shows 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 contraints implied by three levels of the Sherali-Adams hierarchy. However, the above works only consider the case when the LP gives an exact solution, and do not focus on approximation.

The techniques from [11] used in our result, were also extended by Lee [21] to prove a hardness for the Graph Pricing problem. Kenkre et al. [13] also applied these to show the optimality of a simple LP-based algorithm for Digraph Ordering.

Our results

Our main result is the following.

Theorem 1.1 *Let $f : [q]^k \to \{0, 1\}$ 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 $c_\varepsilon \cdot \frac{\log N}{\log \log N}$ levels of the Sherali-Adams hierarchy.*

Combining the above with the connection between Sherali-Adams gaps and extended formulations by Chan et al. [7] yields the following corollary. The improved tradeoff by Kothari et al. [19] gives a better exponent for $\log N$ than $3/2$.

Corollary 1.2 *Let $f : [q] \to \{0,1\}$ 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 $N \in \mathbb{N}$, there exist $(c-\varepsilon, s+\varepsilon)$ integrality gap instances of size N, for every linear extended formulation of size at most* $\exp\left(c_\varepsilon \cdot \frac{(\log N)^{3/.2}}{(\log\log N)^{1/2}}\right)$.

As an application of our methods, we also simplify and strengthen the approximation resistance results for LPs proved by Khot et al. [17]. They studied predicates $f : \{0,1\}^k \to \{0,1\}$ and provided a necessary and sufficient condition for the predicate to be strongly approximation resistant for the Sherali-Adams LP hierarchy. We say a predicate is strongly approximation resistant if for all $\varepsilon > 0$, it is hard to distinguish instances Φ for which $\left|\mathsf{OPT}(\Phi) - \mathbb{E}_{x\in\{0,1\}^k}[f(x)]\right| \leq \varepsilon$ from instances with $\mathsf{OPT}(\Phi) \geq 1-\varepsilon$. In the context of the Sherali-Adams hierarchy, they showed that when this condition is satisfied, there exist instances Φ satisfying $\left|\mathsf{OPT}(\Phi) - \mathbb{E}_{x\in\{0,1\}^k}[f(x)]\right| \leq \varepsilon$ and $\mathsf{FRAC}(\Phi) \geq 1-\varepsilon$, where $\mathsf{FRAC}(\Phi)$ 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 $f : \{0,1\}^k \to \{0,1\}$ be any predicate satisfying the condition for strong approximation resistance for LPs, given by [17]. Then for every $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that infinitely many $N \in \mathbb{N}$, there exists an instance Φ of MAX k-CSP(f) of size N, satisfying*

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

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

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

Proof overview and techniques

The gap instance. The construction of our gap instances is inspired by the construction by Khot et al. [17]. 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 [23]) 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 contraint, 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). Previous instances of gap constructions for LP and SDP hierarchies were (hyper)graphs generated according to the model $\mathcal{G}_{n,p}$. However, 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 $\mathsf{OPT}(\Phi) \leq s + \varepsilon$ 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 works 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 the induced (hyper)graph 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 (hyper)graph 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 [17] 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 [17, 12] to $O(\log \log n)$ levels. This also makes the proof a bit 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* decom-

positions) have been used extensively in the theory metric embeddings (see e.g., [20] 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 [9], 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. We then use these 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 $O\left(\frac{\log n}{\log\log n}\right)$ 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 $\{1,\ldots,n\}$. The only exception is $[q]$, where we overload this notation to denote the set $\{0,\ldots,q-1\}$, 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 or partitions of) a set S. 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 $\{0,\ldots,q-1\}$. For a predicate $f : [q]^k \to \{0,1\}$, an instance Φ of MAX k-CSP$_q$ (f) consists of a set of variables $\{x_1,\ldots,x_n\}$ and a set of constraints $C_1,\ldots,C_m$ where each constraint C_i is over a k-tuple of variables $\{x_{i_1},\ldots,x_{i_k}\}$ and is of the form*

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

for some $b_{i_1},\ldots,b_{i_k} \in [q]$, where the addition is modulo q. For an assignment $\sigma : \{x_1,\ldots,x_n\} \mapsto [q]$, let $\mathsf{sat}(\sigma)$ denote the fraction of constraints satisfied by σ. The maximum fraction of constraints that can be simultaneously satisfied is denoted by $\mathsf{OPT}(\Phi)$, i.e.

$$\mathsf{OPT}(\Phi) = \max_{\sigma:\{x_1,\ldots,x_n\}\mapsto[q]} \mathsf{sat}(\sigma).$$

For a constraint C of the above form, we use x_C to denote the tuple of variables $(x_{i_1},\ldots,x_{i_k})$ and b_C to denote the tuple $(b_{i_1},\ldots,b_{i_k})$. We then write the constraint as $f(x_C + b_C)$. We also denote by S_C the set of indices $\{i_1,\ldots,i_k\}$ 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 $\sigma : [n] \to [q]$. 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 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 *l*ocal 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.

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

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

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. Note that the all the constraints for the basic LP are implied by the relaxation obtained by level k of the Sherali-Adams hierarchy.

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

Figure 2: Basic LP relaxation for MAX k-CSP$_q$ (f)

For an LP/SDP relaxation of MAX k-CSP$_q$, and for a given instance Φ of the problem, we denote by $\mathsf{FRAC}(\Phi)$ 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 $\mathsf{FRAC}(\Phi) \geq c$ and $\mathsf{OPT}(\Phi) < s$.

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 k-hyperedge for every constraint $C \in \Phi$. We remind the reader of the familiar notions of degree, paths, and cycles for the case of (k-uniform) 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.,, $P = v_1, e_1, v_2, \ldots, v_l, e_l, v_{l+1}$, where $v_i \in V\ \forall i \in [l+1]$ and $e_i \in E\ \forall i \in [l]$. 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 $\mathcal{C} = (v_1, e_1, v_2, \ldots, v_l, e_l, v_1)$ is called a cycle of length l if the initial segment $v_1, e_1, \ldots, v_l$ is a (simple) path, $e_{l+1} \neq e_i$ for all $i \in [l]$, and $v_1 \in e_l$. For a path P (or cycle $\mathcal{C}$), we use $\mathrm{V}(P)$ (or $\mathrm{V}(\mathcal{C})$) to denote the set of vertices all the vertices that occurs in the edges, i.e., the set $\{v : (\exists i \in [h])(v \in e_i)\}$, where $e_1, \ldots, e_h$ are the edges included in P (or $\mathcal{C}$).*
- *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 [11] for the case of graphs. A stronger notion of closure was also considered by [4].

Definition 2.3 *For a given hypergraph H and $R \in \mathbb{N}$, and a set $S \subseteq \mathrm{V}(H)$, we denote by $\mathrm{cl}_R(S)$ 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.,*

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

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

3 Properties of random hypergraphs

We collect here 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 much 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 though of a 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 $m, n > 0$ and let Γ be a distribution on $[n_0]^k$. We define a distribution $\mathcal{H}_k(m, n, n_0, \Gamma)$ on k-uniform n_0-partite hypergraphs with m edges and $N = n_0 \cdot n$ vertices, divided in n_0 sets $X_1, \ldots, X_{n_0}$ of size n each. 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 $(i_1, \ldots, i_k) \in [n_0]^k$ from the distribution Γ.*
- *For all $j \in [k]$, sample v_{i_j} independently and uniformly in X_{i_j}.*
- *Add the edge $e_i = \{v_{i_1}, \ldots, v_{i_k}\}$ 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 3.5.

We will study the following metrics (similar to the ones defined in [9]) in this section:

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*

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

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

The goal of this section is to prove the following result about the local ℓ_2-embeddability of the metric ρ_μ. The proof of the theorem heavily uses results proved in [3] and [11].

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 exists $\delta > 0$, 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 $\mu \geq c \cdot \frac{\log t + \log\log n}{\log n}$, 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 (hyper)graphs. We first prove that with high probability random hypergraphs (sampled from $\mathcal{H}_k(m, n, n_0, \Gamma)$) can be modified by removing a few edges to a hypergraph whose girth is $\Omega(\log n)$ and the degree of the resulting hypergraph is bounded. The following lemma shows a possible trade-off between the degree of the hypergraph vs the number of edges required to be removed.

Lemma 3.4 *Let $H' \sim \mathcal{H}_k(m, n, n_0, \Gamma)$ be a random hypergraph with $m = \gamma \cdot n$ edges. Then for any $\varepsilon > 0$, with probability $1 - \varepsilon$ the exists a sub-hypergraph H with $V(H) = V(H')$ such that $\forall u \in V(H)$, $\deg_H(u) \leq 100 \cdot \log\left(\frac{n_0}{\varepsilon}\right) \cdot k \cdot \gamma$ and $|\mathrm{E}(H') \setminus \mathrm{E}(H)| \leq \varepsilon \cdot m$.*

Proof: By linearity of expectation, the expected degree of any vertex v in H' is at most $k \cdot \gamma$,. Let $D = 100 \cdot \log\left(\frac{n_0}{\varepsilon}\right) \cdot k \cdot \gamma$. Let S be the set of all vertices u such that $\deg_{H'}(u) > D$. Let E_S be the set of all hyperedges with 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.

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

The penultimate inequality uses the independence of the hyper-edges 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 $1 - \varepsilon$. ∎

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

Lemma 3.5 *Let $H \sim \mathcal{H}_k(m, n, n_0, \Gamma)$ be a random hypergraph and for $l \geq 2$, let $Z_l(H)$ denote the number of cycles of length at most l in H. For m, n and k such that $k^2 \cdot (m/n) > 1$, we have*

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

Proof: Let the vertices of H correspond to the set $[n_0] \times [n]$. Suppose we contract the set of $[n_0] \times \{j\}$ 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 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, 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 edge of H' is at most $\frac{mk^2}{n^2}$. Consider a given h-tuple of vertices $\mathbf{u} = (u_{i_1}, \cdots u_{i_h})$. Note that we require that edges participating in a cycle be distinct. So, the probability that $\mathbf{u}$ is part of a cycle in H', i.e., there exists distinct edges $e_j \in H'$ 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 $\left(\frac{mk^2}{n^2}\right)^h$. As a result, 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 $\frac{\left(k^2\cdot\frac{m}{n}\right)^{l+1}}{\left(k^2\cdot\frac{m}{n}\right)-1} < \left(k^2\cdot\frac{m}{n}\right)^{2l}$. ■

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 edges.

Corollary 3.6 *Let $H \sim \mathcal{H}_k(m,n,n_0,\Gamma)$ be a random hypergraph with $m = \gamma\cdot n$ for $\gamma > 1$ and $k \geq 2$. Then, there exists $\delta = \delta(\gamma) > 0$ 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 $n^{2/3}$ edges.*

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 $k^2\cdot\frac{m}{n} \geq 2$. By Lemma 3.5, $\mathbb{E}[Z_l] \leq \left(k^2\cdot\frac{m}{n}\right)^{2l}$. Since $m = \gamma\cdot n$, there exists a $g = \delta\cdot\log n$ such that $\mathbb{E}[Z_l] \leq \sqrt{n}$. By Markov's inequality, $\mathbb{P}\left[Z_l \geq n^{2/3}\right] \leq n^{-1/6}$. 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. ■

One can also extend the analysis in [3] to show that the hypergraphs are locally sparse i.e., the number of edges contained in a small set of vertices is small. For a hypergraph H and a set $S \subseteq V(H)$, we use $E(S)$ to denote the edges contained in the set S.

Definition 3.7 *We say that $S \subseteq V(H)$ is η-sparse if $|E(S)| \leq \frac{|S|}{k-1-\eta}$. We call an k-uniform hypergraph H on N vertices to be (τ,η)-sparse if all subsets $S \subset V(H), |S| \leq \tau\cdot|V(H)|$, S is η-sparse. We call H to be η-sparse if it is $(1,\eta)$-sparse, i.e., all subsets of vertices of H are sparse.*

We note here that while this notion of sparsity is a generalization of that considered in [3], it is also identical to the notions of expansion considered in works in proof complexity (see e.g., [5]) and later in works on integrality gaps [1, 6, 4]. We prove that random hypergraphs generated with our model are locally sparse:

Lemma 3.8 *Let $\eta < 1/4$ and $m = \gamma\cdot n$ for $\gamma > 1$. Then for $\tau \leq \frac{1}{n_0}\cdot\left(\frac{1}{e\cdot k^{3k}\cdot\gamma}\right)^{1/\eta}$ the following holds:*

$$\mathop{\mathbb{P}}_{H\sim\mathcal{H}_k(m,n,n_0,\Gamma)}[H \text{ is not } (\tau,\eta)\text{-sparse}] \leq 3\cdot\left(\frac{k^{3k}\cdot\gamma}{n^{\eta/4}}\right)^{1/k}.$$

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 works where η is a constant. For the proof of the lemma, we follow an approach similar to that of Lemma 3.5: we collapse the vertices of H of the form $[n_0]\times\{j\}$ to vertex $j \in [n]$ to construct H', and thus reducing the problem to random multi-hypergraph form a random multipartite hypergraph. The rest proof of the lemma is along the lines of several known proofs [1, 6] and we defer the details to Appendix A.

Charikar et al. [9] prove an analogue of Theorem 3.3 for metrics defined on locally-sparse graphs. In fact, they use a consequence of sparsity, which they call ℓ-path decomposability. To this end, we define the *incidence graph*[1] associated with a hypergraph, on which we will apply their result.

[1]This is the same notion as the constraint-variable graph considered in various works on lower bounds for CSPs.

Definition 3.9 *Let $H = (V(H), E(H))$ be a k-uniform 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 prove that for a locally sparse hypergraph H, its incidence graph G_H is also locally sparse.

Lemma 3.10 *Let H be a k-uniform (τ,η)-sparse hypergraph on N vertices with $m = \gamma \cdot n$ hyperedges. Then the incidence graph G_H is (τ',η') sparse for $\tau' = \tau/k\cdot(1+\gamma)$ and $\eta' = \eta/(1+\eta)$.*

Proof: Let $\tau' = \tau/k\cdot(1+\gamma)$ and let G_H be the incidence graph with $N + m = (1+\gamma)\cdot N$ vertices. Let G' be is the densest subgraph of G_H, among all subgraphs of size at most $\tau' \cdot (N+m)$. Let the vertex set of G' be $V' \cup E'$ where $V' \subseteq V(H)$ and $E' \subseteq E(H)$, and let the edge-set be $\mathcal{E}'$. There cannot be any isolated vertices in G' since removing those will only increase the density.

Let $S \subseteq V(H)$ be the set of all vertices contained in all edges in E' i.e., $S := \{v \in V(H) \mid \exists e \in E' \text{ s.t. } v \in e\}$. Note that $V' \subseteq S$, since there are no isolated vertices, and $E' \subseteq E(S)$, where $E(S)$ denotes the set of hyperedges contained in S.

By our choice of parameters, $|S| \le k \cdot |E'| \le k \cdot \tau' \cdot (N+m) \le \tau \cdot N$. Thus, using the sparsity of H, we have

$$\left|E'\right| \;\le\; |E(S)| \;\le\; \frac{|S|}{k-1-\eta}\,.$$

Also, since each hyperedge of E' can include at most k vertices in S, and since each edge in $\mathcal{E}'$ is incident on a vertex in V', we have

$$|S| - \left|V'\right| \;\le\; k\cdot\left|E'\right| - \left|\mathcal{E}'\right|\,.$$

Combining the two inequalities gives

$$(k-1-\eta)\cdot\left|E'\right| \;\le\; \left|V'\right| + k\cdot\left|E'\right| - \left|\mathcal{E}'\right| \quad\Rightarrow\quad \left|\mathcal{E}'\right| \;\le\; (1+\eta)\cdot\left|E'\right| + \left|V'\right|\,.$$

Hence, we get that $|\mathcal{E}'| \le \frac{|V'|+|E'|}{(1-\eta')}$ for $\eta' = \frac{\eta}{(1+\eta)}$. ∎

Charikar et al. [9] defined the following structural property of a graph.

Definition 3.11 ([9]) *A graph G is ℓ-*path decomposable *if every 2-connected subgraph G' of G, such that G' is not an edge, contains a path of length ℓ such that every vertex of the path has degree at most 2 in G'.*

The above property was also implicitly used by Arora et al. ([3]), who proved the following (see Lemma 2.12 in [3]):

Lemma 3.12 *Let $\ell > 0$ be an integer and $0 < \eta < \frac{1}{3\ell-1} < 1$. 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 and ρ_μ on H as (for $u \neq v$) :

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

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

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

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 and ρ_μ defined on H. Charikar et al. proved the following theorem (see Theorem 5.2) in [11].

Theorem 3.13 ([11]) *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 $(1-\mu)^{\ell/9} \leq \frac{\mu}{2(t+1)}$. Then for every subset S of at most t vertices there exists a mapping ψ_S from S to* 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. Given $\varepsilon>0$, from Lemma 3.4 we have that with high probability at least $1-\varepsilon/2$, there exists H_1 such that the maximum degree of H_1 is at most $D = 100\cdot\log\left(\frac{2n_0}{\varepsilon}\right)\cdot k\cdot\gamma$ with $|\mathrm{E}(H')\setminus\mathrm{E}(H_1)| \leq (\varepsilon/2)\cdot m$.

Using Corollary 3.6 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 $\mathrm{g}\geq\delta\cdot\log n$ and $|\mathrm{E}(H')\setminus\mathrm{E}(H_2)| \leq (\varepsilon/4)\cdot m$. By Lemma 3.8, there exists $\eta=\Omega_{n_0,k,\gamma,\varepsilon}(1/(\log n))$ such that 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:

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

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

Let $G=G_H$ be the incidence graph for the hypergraph H. Note that $N\leq|\mathrm{V}(G)|\leq N\cdot(1+\gamma)$ and degree of G is also bounded by D. Since a cycle in G is also a cycle in H, the girth of G is at also least $\mathrm{g}\geq\delta\cdot\log n$.

By Lemma 3.10, we have G is $(\frac{\tau}{k(1+\gamma)},\frac{\eta}{1+\eta})$-sparse. By Lemma 3.12, any subgraph of G on at most $\frac{\tau}{k(1+\gamma)}\cdot(N+m)$ vertices is ℓ-path decomposable for any $\ell\leq\min\{\mathrm{g},1/(4\eta)\}$.

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 $\exp(-\mu\ell/9) \leq \frac{\mu}{2(t+1)}$ (note that $\frac{1}{\mu} \cdot \exp(-\mu\ell/9)$ 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 $\mu_0 \leq c \cdot \frac{\log t + \log\log n}{\log n} < 1$ when $t \leq n^\theta$. Then, for any $\mu \in [\mu_0, 1)$, we have $(1-\mu)^{\ell/9} \leq \exp(-\mu\ell/9) \leq \frac{\mu}{2(t+1)}$.

We can now apply Theorem 3.13 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 $t \cdot D^{\ell+1} \leq n^{1/2} \cdot n^{1/6} = n^{2/3} \leq \frac{\tau}{k(\gamma+1)} \cdot (N+m)$. Thus, any subgraph of G on $t \cdot D^{\ell+1}$ vertices is ℓ-path decomposable.

Thus, when $\mu \geq \mu_0$, by Theorem 3.13 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 $d_G(u,v) = 2 \cdot d_H(u,v)$. ∎

4 Decompositions of hypergraphs from local geometry

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

Definition 4.1 *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*

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

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

Definition 4.2 *Let $H = (V, E)$ be a k-uniform hypergraph. 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*

$$\forall e \in E \qquad \mathop{\mathbb{P}}_{P\sim\mathcal{P}_e}[P \neq \{e\}] \;\leq\; \varepsilon\,.$$

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 (Theorem 3.2) the metrics d_μ^H and ρ_μ^H as:

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

Lemma 4.3 *Let $H=(V,E)$ be k-uniform hypergraph and let d_μ be the metric as defined above. Let H be such that for all sets $S\subseteq V$ with $|S|\leq t$, the metric induced on ρ_μ 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*

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

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

We also need the observation that the partitions produced by the above theorem are consistent, assuming the set S considered above lie 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,

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

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 $\varphi(W_T)\subseteq W_S\subseteq B(0,R_0)\subset\mathbb{R}^d$, 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 $B(S, \Delta/2)$. This does not affect the separation probability of any two points, since the only non-empty clusters are still produced by the points in $B(S, \Delta/2)$.

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 $\varphi(W_T)$. 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 $\varphi(W_T) \subseteq W_S$, we get that $P(T) = P'$ i.e., the partition P restricted to T equals P'. Thus, we get $\mathcal{P}_{S|T} = \mathcal{P}_T$. ∎

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^H_μ, we get that there exists a $\Delta_H = O(1/\mu)$ such that $\rho^H_{u,v} \leq 1/2 \Rightarrow d_H(u, v) \leq \Delta_H$. 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 $10k \cdot \sqrt{\mu \cdot t}$.

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 $\varphi(W_T) \subseteq W_S$ for some unitary transformation φ. Thus, by Claim 4.5, we get that this is a consistent partitioning scheme of order t. ∎

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 Fig. 2. In this section, we will prove Theorem 1.1. We recall the statement below.

Theorem 1.1 *Let $f : [q]^k \to \{0, 1\}$ 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 $c_\varepsilon \cdot \frac{\log N}{\log \log N}$ 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. [17] discussed in the next section. However, we describe this construction first since it's simpler. The procedure for constructing the instance Φ is described in Fig. 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 $X_1, \ldots, X_{n_0}$, 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 $S_{C_0} = \{i_1, \ldots, i_k\} \subseteq [n_0]$ denote the set of variables in this constraint.
2. For each $j \in [k]$, sample a random variable $x_{i_j} \in X_{i_j}$.
3. Add the constraint $f((x_{i_1}, \ldots, x_{i_k}) + b_{C_0})$ 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 $1 - \exp(-\Omega(n))$, $\mathsf{OPT}(\Phi) < s + \varepsilon$.*

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

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

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

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

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

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

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

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

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

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

Taking a union bound over all assignments, we get

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

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

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 $1-\exp\left(-\Omega(n)\right)$ 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 $x_i\in[n\cdot n_0]$, satisfying*

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

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

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

Each constraint $C\in\Phi$ is sampled according to some constraint $C_0\in\Phi_0$, and we take $\overline{\mathcal{D}}_{S_C} := \mathcal{D}^{(0)}_{S_{C_0}}$ for the corresponding contraint $C_0\in\Phi_0$. Also, each variable x_i for $i\in[n_0\cdot n]$, belongs to one of the sets X_j for $j\in[n_0]$, and we take $\overline{\mathcal{D}}_i := \mathcal{D}^{(0)}_j$ 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 $S_{C_0} = (j_1,\ldots,j_k)$, then $S_C = (i_1,\ldots,i_k)$ where each $i_r\in\{j_r\}\times[n]$ for all $r\in[k]$. Thus, for any $r\in[k]$,

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

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

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

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 $\exp\left(-\Omega(\varepsilon^2 \cdot m)\right)$. There exists $\gamma = O(1/\varepsilon^2)$ such that for $m \geq \gamma \cdot n$, the probability is at most $\exp(-\Omega(n))$. ∎

To construct local distributions for the Sherali-Adams hierarchy, we will consider (a slight modification) the hypergraph H corresponding to the instance Φ. We first show that distributions on edges 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 $\mathrm{cl}(U)$ includes all paths of lengths at most 1 between any two vertices in U. Thus, $E(\mathrm{cl}(U)) = \{e \in E \mid |e \cap U| \geq 2\}$. Note that Lemma 5.2 implies that edges forming a tree in H satisfy the hypothesis of Lemma 5.3 below.

Lemma 5.3 *Let $H = (V, E)$ be a k-uniform hypergraph. Let $U \subseteq V$ and let the set of edges $E(\mathrm{cl}(U))$ form a tree. For each $e \in E(\mathrm{cl}(U))$, let $\overline{\mathcal{D}}_e$ be a distribution on $[q]^e$ such that for any $u \in U$ and $e_1, e_2 \in E(\mathrm{cl}(U))$ 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 edges in $E(\mathrm{cl}(U'))$ form a subtree of $E(\mathrm{cl}(U))$, then $\overline{\mathcal{D}}_{U|U'} = \overline{\mathcal{D}}_{U'}$.*

Proof: We define the distribution by starting with an arbitrary edge and traversing the tree in an arbitrary order. Let $e_1, \ldots, e_r$ be a traversal of the edges in $E(\mathrm{cl}(U))$ such that for all i, $\left|\left(\cup_{j<i} e_j\right) \cap e_i\right| = 1$. Let $U_0 = \cup_{j<i} e_j$ be the set of vertices for which we have already sampled an assignment and let e_i be the next edge in the traversal, with u being the unique vertex in $e_i \cap U_0$. We sample an assignment to the vertices in e, conditioned on the value for the vertex u. Formally, we extend the distribution $\overline{\mathcal{D}}_{U_0}$ to $U_0 \cup e$ by taking, for any $\alpha \in [q]^{U_0 \cup e}$

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

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

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

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

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

Note that the distribution $\overline{\mathcal{D}}_{\mathrm{cl}(U)}$ and hence also the distribution $\overline{\mathcal{D}}_U$ are independent of the order in which we traverse the edges in $E(\mathrm{cl}(U))$. Also, since the above process samples each edge according to the distribution $\overline{\mathcal{D}}_e$, we have that for any $e \in E(U)$, $\overline{\mathcal{D}}_{\mathrm{cl}(U)|e} = \overline{\mathcal{D}}_e$. 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 $E(\mathrm{cl}(U'))$ forms a subtree of $E(\mathrm{cl}(U))$. Then there exists a traversal $e_1, \ldots, e_r$, and $i \in [r]$ such that $e_j \in E(\mathrm{cl}(U'))\ \forall j \leq i$ and $e_j \notin E(\mathrm{cl}(U'))\ \forall j > i$. However, the distribution defined by the partial traversal $e_1, \ldots, e_i$ is precisely $\overline{\mathcal{D}}_{\mathrm{cl}(U')}$. Thus, we get that $\overline{\mathcal{D}}_{\mathrm{cl}(U)|\,\mathrm{cl}(U')} = \overline{\mathcal{D}}_{\mathrm{cl}(U')}$ 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. Then, there is a $t = \Omega_{\varepsilon,k,n_0}\left(\frac{\log n}{\log\log n}\right)$, 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*

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

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 $\mathrm{g} = \delta \cdot \log n$. Also, there exists $\theta, c > 0$ such that 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 $\mu \geq c \cdot \frac{\log t + \log\log n}{\log n}$.

We choose $\mu = 2c \cdot \frac{\log\log n}{\log n}$ and $t = \frac{\varepsilon^2}{400k^2} \cdot \frac{1}{\mu}$ 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 $P = (U_1, \ldots, U_r)$ 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 edges 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 $\mathcal{C}(U) = \mathrm{cl}_{\Delta_H}(U)$. Also,

since the each vertex in an included path is within distance at most $\Delta_H/2$ of an endpoint, 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 $\mathrm{cl}(\mathcal{C}(U_i)) = \mathcal{C}(U_i)$ 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

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

Thus, we have

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

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 $(U_1, \ldots, U_r)$ also produces a partition T. For ease of notation, we assume that the first (say) r' clusters have non-empty intersection with S. Let $V_i = U_i \cap T$ for $1 \le i \le r'$ ($V_i = \emptyset$ for $i > r'$). Then, we have

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

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 $\{\mathcal{D}_S\}_{|S|\le t}$ has value at least $c - \varepsilon$. Recall that the value of the LP solution is given by

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

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

$$\mathop{\mathbb{P}}_{P\sim\mathcal{P}_{S_C}} [P \neq \{S_C\}] \le 10k \cdot \sqrt{\mu \cdot t} \le \frac{\varepsilon}{2}.$$

For a constraint set which is not in the deleted set C_B, if the edge 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 $C \notin C_B$,

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

Using Lemma 5.2 again, we get

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

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). ■

5.2 Integrality Gaps for resistant predicates

Let $f : \{0,1\}^k \to \{0,1\}$ be a boolean predicate and let $\rho(f) = \frac{f^{-1}(1)}{2^k}$ 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 [17] 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 show 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 strengthened [2] to $O\left(\frac{\log n}{\log\log n}\right)$ levels.

Let us first recall some useful notation defined by Khot et al. [17] before we define the notion of vanishing measure:

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

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

[2] The 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.

For a measure Λ on $\mathcal{C}(f)$, $S \subseteq [k]$, $b \in \{0,1\}^S$ and permutation $\pi : S \to S$, 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 [17] (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*

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

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 $f : \{0,1\}^k \to \{0,1\}$ 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 may $N \in \mathbb{N}$, there exists an instance Φ of MAX k-CSP(f) on N variables satisfying the following:*

- $\mathsf{OPT}(\Phi) \leq \rho(f) + \varepsilon$.
- *The optimum for the LP relaxation given by $c_\varepsilon \cdot \log\log N$ levels of Sherali-Adams hierarchy has* $\mathsf{FRAC}(\Phi) \geq 1 - O(k \cdot \sqrt{\varepsilon})$.

Combining this with our Theorem 1.1 already gives us the following stronger result:

Corollary 5.8 *Let $f : \{0,1\}^k \to \{0,1\}$ 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 may $N \in \mathbb{N}$, 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 $c_\varepsilon \cdot \frac{\log N}{\log\log N}$ levels of Sherali-Adams hierarchy has* $\mathsf{FRAC}(\Phi) \geq 1 - O(k\sqrt{\varepsilon})$.

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. [17]. One can then directly use this gap for applying Theorem 1.1 instead of going through Theorem 5.7.

Khot et al. [17] use the probabilistic construction given in Fig. 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 $\zeta_0 = \frac{1}{1-\delta} \cdot \zeta$ 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 $\forall i \in [k]\, \mathbb{E}_{\alpha\sim\nu}\left[(-1)^{\alpha_i}\right] = \zeta_i$.

Let $n_0 = \lceil \frac{1}{\varepsilon} \rceil$. Partition the interval $[0,1]$ into $n_0 + 1$ disjoint intervals $I_0, I_1, \ldots, I_{n_0}$ where $I_0 = \{0\}$ and $I_i = (i-1/n_0, i/n_0]$ for $1 \le i \le n_0$. 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_j.
- If $\zeta(j) < 0$, then negate y_j. If $\zeta(j) = 0$, then negate y_j w.p. $\frac{1}{2}$.
- 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, $|\mathsf{sat}_\Phi(\sigma) - \rho(f)| \le \varepsilon$ (see Lemma 4.4 in [17]). The proof is similar to that of of Lemma 5.1.

Additionally, we need the following claim from [17] (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 $I_0, \ldots, I_{n_0}$. 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 $t_1, \ldots, t_k \in [0,1]$ be such that for all $i \in [k]$, $|t_i - |\zeta_i|| \le \varepsilon$ for $\varepsilon < \delta/2$. Then there exists a distribution ν' on $\{0,1\}^k$ such that*

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

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 Fig. 2 for the instance given in Fig. 4. For each variable y_j coming from the set X_j for $j \in \{0, 1, \ldots, n_0\}$, we set the bias t_j of the variable to be the rightmost point of the interval I_j i.e., set $x_{(y_j,-1)} = \frac{1}{2} \cdot \left(1 - \frac{i}{n_0}\right)$ and $x_{(y_j,1)} = \frac{1}{2} \cdot \left(1 + \frac{i}{n_0}\right)$.

For each constraint C of the form $f(y_{i_1} + b_1, \ldots, y_{i_k} + b_k)$, 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 $\|\nu(C) - \nu'(C)\|_1 = O(k\varepsilon/\delta)$ and such that the biases of the *literals* satisfy $\mathbb{E}_{\alpha\sim\nu'(C)}[(-1)^{\alpha_j}] = \mathrm{sign}(\zeta_j) \cdot t_{i_j}$, 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 $\mathrm{sign}(\zeta_j) < 0$. Thus, we have $\mathbb{E}_{\alpha\sim\nu'(C)}\left[(-1)^{\alpha_j + b_j}\right] = t_{i_j}$, which is consistent with the bias given by the singleton variables $x_{(y_{i_j},1)}$ and $x_{(y_{i_j},-1)}$. We thus define the local distribution on the set S_C as $\mathcal{D}_{S_C}(\alpha) = (\nu'(C))(\alpha + b_C)$.

For all $C \in \Phi$, since $\zeta(C) \in \mathcal{C}_\delta(f)$, we have that $\mathbb{E}_{\alpha \sim \nu(C)}[f(\alpha)] \geq 1 - \delta$. Also, since $\|\nu(C) - \nu'(C)\|_1 = O(k\varepsilon/\delta)$, we get that $\mathbb{E}_{\alpha \sim \nu'(C)}[f(\alpha)] \geq 1 - \delta - O(k\varepsilon/\delta)$. Thus, we have for all $C \in \Phi$, $\mathbb{E}_{\alpha \sim \mathcal{D}_{S_C}}[f(\alpha + b_C)] \geq 1 - \delta - O(k\varepsilon/\delta)$. Taking $\delta = \sqrt{\varepsilon}$ proves the claim. ■

5.3 Lower bounds for LP extended formulations

A connection between LP integrality gaps for the Sheral-Adams hierarchy, and lower bounds on the size of LP extended formulations, was established by Chan et al. [7]. They proved the following:

Theorem 5.10 ([7]) *Let $k, q \in \mathbb{N}$ and $f : [q]^k \to \{0,1\}$ be given. Let $r : \mathbb{N} \to \mathbb{N}$ be a function such that the relaxation obtained by $r(n)$ levels of the Sherali-Adams hierarchy cannot achieve a (c,s) approximation for instances of MAX k-CSP$_q$ (f) on n variables. Then, for all large enough n, no LP extended formulation of size $n^{(r(n))^2}$ can achieve a (c,s) approximation on instances of size N, where $N \leq n^{10 \cdot r(n)}$*

Combining the above with Theorem 1.1 and taking $r(n) = \Omega\left(\frac{\log n}{\log \log n}\right)$ yields Corollary 1.2.

Corollary 1.2 *Let $f : [q] \to \{0,1\}$ 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 $N \in \mathbb{N}$, there exist $(c - \varepsilon, s + \varepsilon)$ integrality gap instances of size N, for every linear extended formulation of size at most* $\exp\left(c_\varepsilon \cdot \frac{(\log N)^{3/.2}}{(\log \log N)^{1/2}}\right)$.

Acknowledgements

We thank Chandra Chekuri, Subhash Khot and Yury Makarychev for helpful discussions, and Rishi Saket for pointers to references. This research was supported by supported by the National Science Foundation under award number CCF-1254044.

References

[1] Michael Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards strong nonapproximability results in the Lovasz-Schrijver hierarchy. In *Proceedings of the 37th ACM Symposium on Theory of Computing*, pages 294–303, 2005. 2, 11

[2] Sanjeev Arora, Béla Bollobás, and László Lovász. Proving integrality gaps without knowing the linear program. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, pages 313–322. IEEE, 2002. 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. 2, 9, 11, 12

[4] Boaz Barak, Siu On Chan, and Pravesh K. Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the 47th ACM Symposium on Theory of Computing*, pages 97–106, New York, NY, USA, 2015. ACM. 2, 8, 11

[5] Eli Ben-Sasson. *Expansion in Proof Complexity*. PhD thesis, Hebrew University, 2001. 11

[6] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP Gaps from Pairwise Independence. *Theory of Computing*, 8(1):269–289, 2012. 2, 11

[7] Siu On Chan, James Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. In *Proceedings of the 54th IEEE Symposium on Foundations of Computer Science*, pages 350–359. IEEE, 2013. 2, 3, 25

[8] Moses Charikar, Chandra Chekuri, Ashish Goel, Sudipto Guha, and Serge Plotkin. Approximating a finite metric by a small number of tree metrics. In *Proceedings of the 39th IEEE Symposium on Foundations of Computer Science*, pages 379–388, 1998. 6, 15, 16

[9] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Local global tradeoffs in metric embeddings. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, 2007. 6, 9, 11, 12

[10] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. In *Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms*, pages 62–68, 2007. 3

[11] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for Sherali-Adams relaxations. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, 2009. 2, 3, 8, 9, 13

[12] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms*, pages 53–61, 2007. 2, 5

[13] Sreyash Kenkre, Vinayaka Pandit, Manish Purohit, and Rishi Saket. On the approximability of digraph ordering. In *Algorithms - ESA 2015 - 23rd Annual European Symposium, Patras, Greece, September 14-16, 2015, Proceedings*, pages 792–803, 2015. 3

[14] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 767–775, 2002. 2

[15] Subhash Khot and Rishi Saket. SDP integrality gaps with local ℓ_1-embeddability. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, 2009. 2

[16] Subhash Khot and Rishi Saket. Approximating CSPs using LP relaxation. In *Proceedings of the 42nd International Colloquium on Automata, Languages and Programming*, 2015. 3

[17] Subhash Khot, Madhur Tulsiani, and Pratik Worah. A characterization of strong approximation resistance. In *Proceedings of the 46th ACM Symposium on Theory of Computing*, pages 634–643. ACM, 2014. 4, 5, 16, 22, 23, 24

[18] Vladimir Kolmogorov, Johan Thapper, and Stanislav Živný. The power of linear programming for general-valued CSPs. *SIAM Journal on Computing*, 44(1):1–36, 2015. 3

[19] Pravesh Kothari, Raghu Meka, and Prasad Raghavendra. Personal Communication, 2016. 2, 3

[20] Robert Krauthgamer, James R Lee, Manor Mendel, and Assaf Naor. Measured descent: A new embedding method for finite metrics. *Geometric & Functional Analysis GAFA*, 15(4):839–858, 2005. 6

[21] Euiwoong Lee. Hardness of graph pricing through generalized max-dicut. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 391–399. ACM, 2015. 3

[22] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. on Optimization*, 1(12):166–190, 1991. 2

[23] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *STOC*, pages 245–254, 2008. 2, 4

[24] Prasad Raghavendra and David Steurer. Integrality gaps for strong SDP relaxations of unique games. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, 2009. 2

[25] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, 2008. 2

[26] 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. 2

[27] Johan Håstad. On the Efficient Approximability of Constraint Satisfaction Problems. In *Surveys in Combinatorics*, volume 346, pages 201–222. Cambridge University Press, 2007. 1

[28] Johan Thapper and Stanislav Živný. The complexity of finite-valued CSPs. In *Proceedings of the 45th ACM Symposium on Theory of Computing*, pages 695–704. ACM, 2013. 3

[29] Johan Thapper and Stanislav Živný. The power of Sherali-Adams relaxations for general-valued CSPs. *arXiv preprint arXiv:1606.02577*, 2016. 3

[30] Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, 2009. 2

A Omitted proofs from Section 3 and Section 5

Lemma 3.8 *Let $\eta < 1/4$ and $m = \gamma \cdot n$ for $\gamma > 1$. Then for $\tau \leq \frac{1}{n_0} \cdot \left(\frac{1}{e \cdot k^{3k} \cdot \gamma}\right)^{1/\eta}$ the following holds:*

$$\mathop{\mathbb{P}}_{H \sim \mathcal{H}_k(m,n,n_0,\Gamma)} [H \text{ is not } (\tau,\eta)\text{-sparse}] \;\leq\; 3 \cdot \left(\frac{k^{3k} \cdot \gamma}{n^{\eta/4}}\right)^{1/k}.$$

Proof: As in the proof of Lemma 3.5, given a random hypergraph H, we construct a hypergraph H' (by contracting all the vertices in $[n_0] \times \{j\}$ to $j \in [n]$).

Consider a subset of vertices $S \subseteq \mathrm{V}(H)$ and let $S' \subseteq \mathrm{V}(H')$ be the corresponding contracted set in H'. Since each edge in H corresponds to an edge in H' (counting multiplicities), we have

$$|E(S)| \;\geq\; \frac{|S|}{k-1-\eta} \quad \Rightarrow \quad \left|E(S')\right| \;\geq\; \frac{|S|}{k-1-\eta} \;\geq\; \frac{|S'|}{k-1-\eta}\,.$$

Thus, it suffices to show that H' is (τ', η)-sparse for $\tau' = \tau \cdot n_0$, since $|S'| \leq \tau \cdot N = (\tau \cdot n_0) \cdot n$. Given any multiset in $[n]^k$, the probability that it corresponds to an edge in H' is at most $(k!) \cdot (m/n^k)$. Thus, the probability that there exists a set T of size at most $\tau' \cdot n$, containing at least $|T|/(k-1-\eta)$ edges (counting multiplicities) is at most

$$\sum_{h=1}^{\tau' \cdot n} \binom{n}{h} \cdot \binom{h^k}{r} \cdot \left(\frac{k! \cdot m}{n^k}\right)^r,$$

where $r = \frac{h}{k-1-\eta}$. Note that we also need to consider $h = 1$ as edges in H' correspond to multisets of size k, and so may not have all distinct vertices. Simplifying the above using $\binom{a}{b} \leq \left(\frac{a \cdot e}{b}\right)^b$ and $k! \leq k^k$ gives

$$\begin{aligned}
\sum_{h=1}^{\tau' \cdot n} \binom{n}{h} \cdot \binom{h^k}{r} \cdot \left(\frac{k! \cdot m}{n^k}\right)^r \;&\leq\; \sum_{h=1}^{\tau' \cdot n} \left(\frac{n \cdot e}{h}\right)^h \cdot \left(\frac{h^k \cdot e}{r}\right)^r \cdot \left(\frac{k^k \cdot m}{n^k}\right)^r \\
&=\; \sum_{h=1}^{\tau' \cdot n} \left(e^{k-\eta} \cdot (k-1-\eta) \cdot k^k \cdot \gamma \cdot \left(\frac{h}{n}\right)^\eta\right)^{h/(k-1-\eta)} \\
&\leq\; \sum_{h=1}^{\tau' \cdot n} \left(k^{3k} \cdot \gamma \cdot \left(\frac{h}{n}\right)^\eta\right)^{h/(k-1-\eta)}
\end{aligned}$$

Let $\theta = \eta/(2k)$. We divide the above summation in two parts and first consider

$$\begin{aligned}
\sum_{h=n^\theta}^{\tau' \cdot n} \left(k^{3k} \cdot \gamma \cdot \left(\frac{h}{n}\right)^\eta\right)^{h/(k-1-\eta)} \;&\leq\; \sum_{h=n^\theta}^{\tau' \cdot n} \left(k^{3k} \cdot \gamma \cdot (\tau')^\eta\right)^{n^\theta/(k-1-\eta)} \\
&\leq\; 2 \cdot \exp\left(-\frac{n^\theta}{k}\right) \\
&\leq\; 2 \cdot \frac{k}{n^\theta},
\end{aligned}$$

for $\tau' \leq \left(e \cdot k^{3k} \cdot \gamma\right)^{-1/\eta}$. Considering the first half of the summation, we get

$$\begin{aligned}
\sum_{h=1}^{n^\theta} \left(k^{3k} \cdot \gamma \cdot \left(\frac{h}{n}\right)^\eta\right)^{h/(k-1-\eta)} \;&\leq\; n^\theta \cdot \left(\frac{k^{3k} \cdot \gamma}{n^{(1-\theta)\cdot\eta}}\right)^{1/k} \\
&\leq\; \left(\frac{k^{3k} \cdot \gamma}{n^{\eta/4}}\right)^{1/k} \;=\; k^3 \cdot \gamma^{1/k} \cdot n^{-\theta/2}\,.
\end{aligned}$$

Combining the two bounds gives that the probability is at most $3k^3 \cdot \gamma^{1/k} \cdot n^{-\theta/2}$, which equals the desired bound. ∎

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 $t_1, \ldots, t_k \in [0,1]$ 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*

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

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

The biases in ν_0 satisfy the above by definition. We obtain $\overline{\nu}_j$ from $\overline{\nu}_{j-1}$ 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 set independently according to their biases in $\overline{\nu}_{j-1}$. For the j^{th} bit, we set it to $\operatorname{sign}(r_j - \zeta_j)$ (if $r_j - \zeta(j) = 0$, we can simply proceed with $\overline{\nu}_j = \overline{\nu}_{j-1}$). The biases for all except for the j^{th} bit are unchanged. For the j^{th} bit, the bias now becomes r_j if

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

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

ECCC ISSN 1433-8092

http://eccc.hpi-web.de