Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 4 (2019)

UG-hardness to NP-hardness by Losing Half

Amey Bhangale* Subhash Khot†

May 30, 2020

Abstract

The 2-to-2 Games Theorem of [KMS17, DKK+18a, DKK+18b, KMS18] implies that it is NP-hard to distinguish between Unique Games instances with assignment satisfying at least $(\frac{1}{2}-\varepsilon)$ fraction of the constraints $vs.$ no assignment satisfying more than ε fraction of the constraints, for every constant $\varepsilon > 0$. We show that the reduction can be transformed in a non-trivial way to give a stronger guarantee in the completeness case: For at least $(\frac{1}{2}-\varepsilon)$ fraction of the vertices on one side, all the constraints associated with them in the Unique Games instance can be satisfied.

We use this guarantee to convert the known UG-hardness results to NP-hardness. We show:

1. Tight inapproximability of approximating independent sets in degree d graphs within a factor of $\Omega\left(\frac{d}{\log^2 d}\right)$, where d is a constant.
2. NP-hardness of approximate the Maximum Acyclic Subgraph problem within a factor of $\frac{2}{3}+\varepsilon$, improving the previous ratio of $\frac{14}{15}+\varepsilon$ by Austrin *et al*. [AMW15].
3. For any predicate $P^{-1}(1) \subseteq [q]^k$ supporting a balanced pairwise independent distribution, given a P-CSP instance with value at least $\frac{1}{2}-\varepsilon$, it is NP-hard to satisfy more than $\frac{|P^{-1}(1)|}{q^k}+\varepsilon$ fraction of constraints.

*Weizmann Institute of Science. Research supported by Irit Dinur's ERC-CoG grant 772839. Email: amey.bhangale@weizmann.ac.il.

†Department of Computer Science, Courant Institute of Mathematical Sciences, New York University. Research supported by NSF CCF-1813438, Simons Collaboration on Algorithms and Geometry, and Simons Investigator Award. Email: khot@cs.nyu.edu

ISSN 1433-8092

1 Introduction

Unique Games Conjecture is a central open problem in computer science. It states that for a certain constraint satisfaction problem over a large alphabet, called *Unique Games* (UG), it is NP-hard to decide whether a given instance has an assignment that satisfies almost all the constraints or there is no assignment which satisfies even an ε fraction of the constraints for a very small constant $\varepsilon > 0$.

Since the formulation of the conjecture, it has found interesting connections to tight hardness of approximation results for many optimization problems [Kho02, KKMO07, KR08, Rag08, GMR08, KN09, KTW14, KV15]. One of the most notable implications is the result of Raghavendra [Rag08] which informally can be stated as follows: Assuming the NP-hardness of approximating this single CSP (Unique Games) implies tight hardness for approximating every other constraint satisfaction problem, stated in terms of the integrality gap of a certain canonical SDP.

Unique Games Conjecture is inspired by the NP-hardness of approximating a problem called Label Cover. A Label Cover instance $G = (A, B, E, \Sigma_A, \Sigma_B, \{\pi_e\}_{e\in E})$ consists of two sets of variables A and B and a bipartite graph between them with the edge set E. The variables from A take values from some alphabet Σ_A and variables from B take values from Σ_B. Every edge e in E has a d-to-1 projection constraint $\pi_e : \Sigma_A \to \Sigma_B$.[1] For an edge $e(a, b)$, a label α to a and a label β to b satisfies the edge e iff $\pi_e(\alpha) = \beta$. In this language, Unique Games is a Label Cover instance where all the constraints are 1-to-1. We denote an instance of Unique Games with $G = (A, B, E, [L], \{\pi_e\}_{e\in E})$ where $\Sigma_A = \Sigma_B = [L]$.

Given an instance of Unique Games, the goal is to find an assignment to the vertices that satisfies a *good* fraction of the edges. An instance is called ε-satisfiable if there exists an assignment $\sigma : A \cup B \to [L]$, that satisfies at least ε fraction of the edges in the graph. The Unique Games Conjecture of Khot [Kho02] states that for every $\varepsilon > 0$, there exists L such that given an Unique Games instance which is $(1-\varepsilon)$-satisfiable, it is NP-hard to find an ε-satisfiable assignment. Note that there is a polynomial time algorithm that given a 1-satisfiable instance of Unique Games, finds a 1-satisfiable assignment.

A recent series of works [KMS17, DKK+18a, DKK+18b, KMS18] implies that for a given Label Cover instance with 2-to-1 projection constraints, it is NP-hard to find an ε-satisfiable assignment even if the instance is $(1-\varepsilon)$-satisfiable for every constant $\varepsilon > 0$. This directly implies the following inapproximability for Unique Games.

Theorem 1.1. *For every constant $\varepsilon > 0$, there exists Σ such that for Unique Games instance over Σ, it is NP-hard to distinguish between the following two cases*

- *Yes Case: The instance is $(\frac{1}{2} - \varepsilon)$-satisfiable.*
- *No Case: No assignment satisfies ε fraction of the constraints.*

[1] A constraint $\pi_e : \Sigma_A \to \Sigma_B$ is called a d-to-1 projection constraint, if every $\beta \in \Sigma_B$ has exactly d pre-images.

Although we do not improve upon this theorem in terms of the inapproximability gap, we show a stronger guarantee in the Yes Case. Specifically, we show that in the Yes case, there are at least $\frac{1}{2}-\varepsilon$ fraction of the *vertices* on, say, the left side such that *all* the edges incident on them are satisfied by some assignment and also the instance is left-regular. This clearly implies the above theorem. Formally, the main theorem that we prove is as follows (See Definition 2.1 for a formal definition of Unique Games):

Theorem 1.2. *For every constant $\delta > 0$ there exists $L \in \mathbb{N}$ such that the following holds. Given an instance $G = (A, B, E, [L], \{\pi_e\}_{e\in E})$ of Unique Games, which is regular on the A side, it is NP-hard to distinguish between the following two cases:*

- *YES case: There exists a set $A' \subseteq A$ of size $(\frac{1}{2} - \delta)|A|$ and an assignment that satisfies all the edges incident on A'.*
- *NO case: Every assignment satisfies at most δ fraction of the edge constraints.*

We will denote by $\mathsf{val}(G)$ the maximum fraction, over all assignments, of the edges satisfied and $\mathsf{sval}(G)$ to be the maximum fraction, over all assignments, of the vertices in A such that all the edges incident on them are satisfied. Thus, the above theorem says that for every $\delta > 0$ there exists a label set $[L]$ such that it is NP-hard to distinguish between the cases $\mathsf{sval}(G) \geqslant \frac{1}{2} - \delta$ and $\mathsf{val}(G) \leqslant \delta$.

1.1 $(\frac{1}{2} - \varepsilon)$-satisfiable UG vs. $(1 - \varepsilon)$-satisfiable UG

Let $\varepsilon > 0$ be a very small constant. In the $(1 - \varepsilon)$-satisfiable Unique Games instance, by simple averaging argument it follows that for any satisfying assignment $\sigma : A \cup B \to [L]$, there exists $A' \subseteq A$, $|A'| \geqslant (1 - \sqrt{\varepsilon})|A|$ such that for all $v \in A'$, at least $(1 - \sqrt{\varepsilon})$ fraction of the edges incident on v are satisfied. Having such a large A' is crucial in many UG-reductions. For example, a typical k-query PCP, used in proving inapproximability of k-ary CSPs assuming the Unique Games Conjecture, samples $v \in A$ uniformly at random and k neighbors of $u_1, u_2, \ldots, u_k$ of v uniformly at random. Thus, with probability at least $(1-\sqrt{\varepsilon})(1-k\sqrt{\varepsilon}) \approx 1$ all the edges (u, v_i) are satisfied by any $(1-\varepsilon)$ satisfying assignment σ.

In contrast to this, if we take a $\frac{1}{2}$-satisfiable UG instance then the probability that all the edges (v, u_i) are satisfied is at most $\frac{1}{2^k}$ in the worst case. Therefore, in converting the known UG-hardness result to NP-hardness result using the NP-hardness of Unique Games with gap $(\frac{1}{2}-\varepsilon, \varepsilon)$, it is not always the case that we lose '*only half*' in the completeness case.

Another important property of the Unique Games instance which was used in many reductions is that in the completeness case, there are $(1 - \delta)$ fraction of the vertices on one side such that *all* the edges incident on them are satisfied i.e. $\mathsf{sval}(G) \geqslant 1 - \delta$ instead of $\mathsf{val}(G) \geqslant 1 - \delta$. For example, this property was crucial in the hardness of approximating independent sets in bounded degree graphs [AKS11] and in many other reductions [BK10, BGH+17].

As shown in [KR08], having completeness $\mathsf{val}(G) \geqslant 1-\delta$ for all sufficiently small $\delta > 0$ is equivalent to having completeness $\mathsf{sval}(G) \geqslant 1-\delta'$ for all sufficiently small $\delta' > 0$. It was crucial in the reduction that the $\mathsf{val}(G)$ is arbitrarily close to 1 for the equivalence to hold. We do not know a black-box way of showing the equivalence of $\mathsf{val}(G) = c$ and $\mathsf{sval}(G) = c$ for any $c < 1$. Thus, in order to prove Theorem 1.2 with a stronger completeness guarantee, we crucially exploit the structure of the game given by the known proofs of the 2-to-1 theorems [KMS17, DKK+18a, DKK+18b, KMS18] mentioned in the introduction.

1.2 Implications

Using Theorem 1.2, we show the following hardness results by going over the known reductions based on the Unique Games Conjecture.

Independent sets in degree d graphs The first application is approximating the maximum sized independent set in a degree d graph, where d is a large constant.

Theorem 1.3. *It is* NP*-hard (under randomized reductions) to approximate independent sets in a degree d graph within a factor of $O\left(\frac{d}{\log^2 d}\right)$, where d is constant.*

This improves the NP-hardness of approximation within a factor $O\left(\frac{d}{\log^4 d}\right)$, as shown in Chan [Cha16] as well as shows the tightness of the randomized polynomial time approximation algorithm given by Bansal *et al*. [BGG18].

Max-Acyclic Subgraph Given a directed graph $G(V, E)$, the Max-Acyclic Subgraph problem is to determine the maximum fraction of edges $E' \subseteq E$ such that removal of $E \setminus E'$ makes the graph acyclic (removes all the cycles). We can always make a graph acyclic by removing at most $\frac{1}{2}$ fraction of the edges; take any arbitrary ordering of the vertices and remove either all the forward edges or all the backward edges whichever is minimum. This gives a trivial $\frac{1}{2}$-approximation algorithm. Guruswami *et al*. [GMR08] showed this is tight by showing that assuming the Unique Games Conjecture, it is NP-hard to approximate Max-Acyclic Subgraph within a factor of $\frac{1}{2} + \varepsilon$ for all $\varepsilon > 0$. In terms of NP-hardness, Austrin *et al*. [AMW15] showed NP-hardness of approximating Max-Acyclic Subgraph within a ratio of $\frac{14}{15} + \varepsilon$, improving upon the previous bound of $\frac{65}{66} + \varepsilon$ by Newman [New01]. Our next theorem shows an improved inapproximability of $\frac{2}{3} + \varepsilon$. One interesting feature of our reduction is that it shows that in the worst case, it is NP-hard to perform better than the trivial algorithm described above on instances with value at least $\frac{3}{4}$.

Theorem 1.4. *For every constant $\varepsilon > 0$, it is NP-hard to approximate Max-Acyclic Subgraph problem within a factor of $\frac{2}{3} + \varepsilon$.*

We note that Theorem 1.1 along with the reduction from [GMR08] implies NP-hardness of Max-Acyclic Subgraph problem within a factor of $\frac{4}{5} + \varepsilon$ (See Remark 5.5 for a proof sketch). Therefore, Theorem 1.4 improves upon this bound too.

Predicates supporting balanced pairwise independent distributions The next result is approximating Max-k-CSP(P) for a predicate $P : [q]^k \to \{0,1\}$ where $P^{-1}(1)$ supports a balanced pairwise independent distribution i.e. there exists a distribution on $P^{-1}(1)$ such that 1) the marginal distribution on each coordinate is uniform and 2) the distribution is pairwise independent. Given an instance of Max-k-CSP(P), a random assignment satisfies $\frac{|P^{-1}(1)|}{q^k}$ fraction of the constraints in expectation. Austrin-Mossel [AM09] showed that assuming the Unique Games Conjecture, given a $(1-\varepsilon)$-satisfiable instance of Max-k-CSP(P), it is NP-hard to find an assignment that satisfies more than $\frac{|P^{-1}(1)|}{q^k} + \varepsilon$ fraction of the constraints for any constant $\varepsilon > 0$. This notion is called *approximation resistance* of a predicate P where the best efficient algorithm cannot perform any better than just choosing a random assignment, even if the instance is almost satisfiable. Showing approximation resistance of such predicates *unconditionally* (i.e. assuming only P$\neq$ NP) has received significant attention. For instance, Chan [Cha16] showed that a predicate P is approximation resistant if $P^{-1}(1)$ contains a subgroup that supports a balanced pairwise independent distribution. Also, Barak et al. [BCK15] showed that certain class of algorithms, namely sum-of-squares, cannot be used to perform better than the random assignment on almost satisfiable instances of Max-k-CSP(P) where P supports a balanced pairwise independent distribution.

Our main theorem shows approximation resistance of such predicates but on instances which are $\frac{1}{2}$-satisfiable. If we use Theorem 1.2 as a starting point of the reduction in [AM09], we get the following NP-hardness result.

Theorem 1.5. *If a predicate $P : [q]^k \to \{0,1\}$ supports a balanced pairwise independent distribution, then it is* NP*-hard to find a solution with value $\frac{|P^{-1}(1)|}{q^k} + \varepsilon$ if a given P-CSP instance is $(\frac{1}{2} - \varepsilon)$-satisfiable, for every constant $\varepsilon > 0$.*

Other Results Theorem 1.2 implies many more NP-hardness results in a straightforward way by going over the known reductions based on the Unique Games Conjecture, but we shall restrict ourselves to proving only the above three theorems. We only state the following important implication which follows from the result of Raghavendra [Rag08] and our main theorem. We refer to [Rag08] for the definition of (c, s) SDP integrality gap of a P-CSP instance.

Theorem 1.6. *(Informal) For all $\varepsilon > 0$, if a P-CSP has (c, s) SDP integrality gap instance, then it is* NP*-hard to distinguish between $(\frac{c}{2} - \varepsilon)$-satisfiable instances from at most $(s + \varepsilon)$-satisfiable instances.*

The reduction actually gives a stronger result; instead of completeness $(\frac{c}{2} - \varepsilon)$ one can get $(\frac{c}{2} + \frac{r}{2} - \varepsilon)$ where $r = \frac{|P^{-1}(1)|}{q^k}$ for a predicate $P : [q]^k \to \{0, 1\}$.

1.3 Overview of the proof

In this section, we give an overview of the proof of Theorem 1.2. The main idea which goes in proving Theorem 1.2 is very simple and we elaborate it next.

Let $V = \mathbb{F}_2^n$ and $\begin{bmatrix} V \\ \ell \end{bmatrix}$ denotes the set of all ℓ dimensional subspaces of V. Consider the Grassmann 2-to-1 test $\mathcal{T}_1$ from Figure 1.3 for functions $f : \begin{bmatrix} V \\ \ell \end{bmatrix} \to \mathbb{F}_2^\ell$ and $h : \begin{bmatrix} V \\ \ell-1 \end{bmatrix} \to \mathbb{F}_2^{\ell-1}$, where for a subspace L (L'), $f(L)$ $(h(L'))$ represents a linear function on the subspace, by fixing an arbitrarily chosen basis of L (L'). From the test it is clear that for every pair (L, L')

- Select a $\ell - 1$ dimensional subspace L' uniformly at random.
- Select a ℓ dimensional subspace L containing L' uniformly at random.
- Check if $f(L)_{|L'} = h(L')$.

Figure 1: 2-to-1 Test $\mathcal{T}_1$

such that $L' \subseteq L$, for every linear function β on L', there are two linear functions α_1, α_2 on L such that the test passes for any pair (α_i, β). This gives the 2-to-1 type constraints.

One way to convert a 2-to-1 test to a unique test is by choosing a random $i \in \{1, 2\}$ for every pair (L, L') such that $L' \subseteq L$ and for every linear function β on L', and adding just one accepting pair (α_i, β) where $\{(\alpha_1, \alpha_2), \beta\}$ are the original accepting assignments. This does give a unique test and if f and h are restrictions of a global linear function to the subspaces, then the test passes with probability $\approx \frac{1}{2}$. One drawback of this test is that, if we consider a bipartite graph on $\begin{bmatrix} V \\ \ell \end{bmatrix} \times \begin{bmatrix} V \\ \ell-1 \end{bmatrix}$ where two subspaces L, L' are connected iff $L' \subseteq L$, then for any global linear function we can only argue that half the edges are *satisfied* in the sense of the unique test. Note that the uniform distribution on the edges of this bipartite graph is the same as the test distribution $\mathcal{T}_1$. Hence, the similar guarantee of satisfying around half the edges stays in the final Unique Games instance created from the works of [KMS17, DKK+18a, DKK+18b, KMS18] and thus falls short of proving Theorem 1.2.

Now we convert it into a Unique Test $\mathcal{T}_2$ (Figure 1.3) with a guarantee that for around $\frac{1}{2}$ fraction of the vertices on one side of the bipartite test graph, *all* the edges incident on them are satisfied if the assignments f and h are the restrictions of a global linear function.

Towards this, we modify the domain of f. We consider two functions $f : \begin{bmatrix} V \\ \ell \end{bmatrix} \times (2^{[\ell]} \setminus \emptyset) \times \{0, 1\} \to \mathbb{F}_2^{\ell-1}$ and $h : \begin{bmatrix} V \\ \ell-1 \end{bmatrix} \to \mathbb{F}_2^{\ell-1}$. We fix an arbitrary one-to-one correspondence between the *non-zero* elements of ℓ dimension subspace and $2^{[\ell]} \setminus \emptyset$. Thus, we can now

- Select a $\ell - 1$ dimensional subspace L' uniformly at random.
- Select a ℓ dimensional subspace L containing L', $x \in L \setminus L'$ and $b \in \{0, 1\}$ uniformly at random.
- Check if $f(L, x, b)_{|L'} = h(L')$.

Figure 2: Unique Test $\mathcal{T}_2$

interpret f as defined on the tuple (L, x, b) where $x \in L \setminus \{0\}$ and $b \in \{0, 1\}$. We consider the assignments $f(L, x, b)$ and $h(L')$ as linear functions on the spaces L and L' respectively as follows: As before, we select an arbitrary basis for every $\ell - 1$ dimensional subspaces $\begin{bmatrix} V \\ \ell-1 \end{bmatrix}$. Now $f(L, x, b) \in \mathbb{F}_2^{\ell-1}$ is thought of as a linear function on L such that

1. at point x it evaluates to b and
2. the evaluations of the linear function on the subspace $L_x = \{y \in L \mid y \perp x\}$, which is an $\ell - 1$ dimensional subspace, is given by $f(L, x, b)$ (according to the already chosen basis of L_x).

As before, $h(L')$ is thought of as a linear function on L' according to the chosen basis for L'.

Consider the following bipartite graph $\left(\begin{bmatrix} V \\ \ell \end{bmatrix} \times (2^{[\ell]} \setminus \emptyset) \times \{0, 1\}, \begin{bmatrix} V \\ \ell-1 \end{bmatrix}, E\right)$ where (L, x, b) is connected to L' iff $x \notin L'$ and $L' \subseteq L$. The test distribution which we will define next will be the uniform on the edges of this graph.

We now put permutation constraint on the edges of the graph. For an edge $e \in E$ between (L, x, b) and L' we set the following unique constraint: Extend the linear function given by h on L' to a linear function $\tilde{h}_{x,b}$ on $\texttt{span}\{L', x\}$ by setting $\tilde{h}_{x,b}(x) = b$. The accepting labels for an edge e are $f(L, x, b)$ and $h(L')$ such that $\tilde{h}_{x,b}$ and f are identical when thought of as linear functions on L. Note that the constraint is one-to-one.

Fix any global linear function $g : V \to \mathbb{F}_2$. From this we define $h(L')$ as the restriction of g on L'. We define the labeling f partially by setting $f(L, x, g(x))$ as the restriction of g on L. Thus for every $(L, x, g(x))$ it is clear that all the edges are *satisfied* by the labeling h and the label $f(L, x, g(x))$. The set $\{(L, x, g(x))\}$ constitutes half fraction of one side of the bipartite test graph and hence we are done.

2 Preliminaries

We start by defining the Unique Games.

Definition 2.1 (Unique Games). *An instance $G = (A, B, E, [L], \{\pi_e\}_{e\in E})$ of the Unique Games constraint satisfaction problem consists of a bipartite graph (A, B, E), an alphabet $[L]$ and a permutation map $\pi_e : [L] \to [L]$ for every edge $e \in E$. Given a labeling $\ell : A \cup B \to [L]$, an edge $e = (u, v)$ is said to be satisfied by ℓ if $\pi_e(\ell(v)) = \ell(u)$.*

G is said to be at most δ-satisfiable *if every labeling satisfies at most a δ fraction of the edges.*

We will define the following two quantities related to the satisfiability of the Unique Games instance.

$$\mathsf{val}(G) := \max_{\sigma:A\cup B\to[L]} \{\text{fraction of edges in } G \text{ satisfied by } \sigma\} .$$

$$\mathsf{sval}(G) := \max_{\sigma:A\cup B\to[L]} \left\{ \frac{|A'|}{|A|} \,\middle|\, \forall e(u,v) \text{ s.t. } u \in A', e \text{ is satisfied by } \sigma \right\} .$$

The following is a conjecture by Khot [Kho02] which has been used to prove many *tight* inapproximability results.

Conjecture 2.2 (Unique Games Conjecture[Kho02]). *For every sufficiently small $\delta > 0$ there exists $L \in \mathbb{N}$ such that given a an instance $\mathcal{G} = (A, B, E, \{\pi_e\}_{e\in E}, [L])$ of Unique Games it is NP-hard to distinguish between the following two cases:*

- *YES case:* $\mathsf{val}(G) \geqslant 1 - \delta$.
- *NO case:* $\mathsf{val}(G) \leqslant \delta$.

For a linear subspace $L \subseteq \mathbb{F}_2^n$, the dimension of L is denoted by $\mathtt{dim}(L)$. For two subspaces $L_1, L_2 \subseteq \mathbb{F}_2^n$, we will use $\mathtt{span}(L_1, L_2)$ to denote the subspace $\{x_1 + x_2 \mid x_1 \in L_1, x_2 \in L_2\}$. We will sometimes abuse the notation and write $\mathtt{span}(x, L)$,where $x \in \mathbb{F}_2^n$, to denote $\mathtt{span}(\{0, x\}, L)$. For subspaces L_1, L_2 such that $L_1 \cap L_2 = \{0\}$, define $L_1 + L_2 := \mathtt{span}(L_1, L_2)$.

For $0 < \ell < n$, let $Gr(\mathbb{F}_2^n, \ell)$ be the set of all ℓ dimensional subspaces of $\mathbb{F}_2^n$. Similarly, for a subspace L of $\mathbb{F}_2^n$ such that $\mathtt{dim}(L) > \ell$, let $Gr(L, \ell)$ be the set of all ℓ dimensional subspaces of $\mathbb{F}_2^n$ *contained* in L.

3 The Reduction

In this section, we go over the reduction in [DKK⁺18b] from a gap 3LIN instance to a 2-to-1 Label Cover instance and then show how to reduce it to a Unique Games instance in Section 3.4. We retain most of the notations from [DKK⁺18b].

3.1 Outer Game

The starting point of the reduction is the following problem:

Definition 3.1 (REG-3LIN). *The instance (X, Eq) of* REG-3LIN *consists of n variables $X = \{x_1, x_2, \ldots, x_n\}$ taking values in $\mathbb{F}_2$ and a collection of m $\mathbb{F}_2$-linear constraints Eq where each constraint in Eq is a linear constraint on 3 variables. The instance is regular in the following ways: every equation consists of 3 distinct variables, every variable x_i appears in exactly 5 constraints and every two distinct constraints share at most one variable.*

An instance (X, Eq) is said to be t-satisfiable if there exists an assignment to X which satisfies t fraction of the constraints. We have the following theorem implied by the PCP theorem of [ALM+98, AS98, FGL+96].

Theorem 3.2. *There exists an absolute constant $s < 1$ such that for every constant $\varepsilon > 0$ it is NP-hard to distinguish between the cases when a given* REG-3LIN *instance is at least $(1-\varepsilon)$-satisfiable vs. at most s-satisfiable.*

We now define an outer 2-prover 1-round game, parameterized by $k, q \in \mathbb{Z}^+$ and $\beta \in (0, 1)$, which will be the starting point of our reduction. The verifier selects k constraints $e_1, e_2, \ldots, e_k$ from the instance (X, Eq) uniformly at random with repetition. If e_i and e_j share a variable for some $i \neq j$ then accept. Otherwise, let $x_{i,1}, x_{i,2}, x_{i,3}$ be the variables in constraint e_i. Let $X_1 = \cup_{i=1}^{k}\{x_{i,1}, x_{i,2}, x_{i,3}\}$. The verifier then selects a subset X_2 of X_1 as follows: for each $i \in [k]$, with probability $(1 - \beta)$ add $x_{i,1}, x_{i,2}, x_{i,3}$ to X_2 and with probability β, select a variable from $\{x_{i,1}, x_{i,2}, x_{i,3}\}$ uniformly at random and add it to X_2.

On top of this, the verifier selects q pair of *advice* strings (s_j, s_j^*) where $s_j \in \{0, 1\}^{X_2}$, and $s_j^* \in \{0, 1\}^{X_1}$ for $1 \leq j \leq q$ as follows : For each $j \in [q]$, select $s_j \in \{0, 1\}^{X_2}$ uniformly at random. The string s_j can be though as assigning $\mathbb{F}_2$ values to each of the variables from X_2. The string $s_j^* \in \{0, 1\}^{3k}$ is deterministically selected such that its projection on X_2 is the same as s_j and the rest of the coordinates are assigned 0.

The verifier sends $(X_1, s_1^*, s_2^*, \ldots, s_q^*)$ to prover 1 and $(X_2, s_1, s_2, \ldots, s_q)$ to prover 2. The verifier expects an assignment to variables in X_i from prover i. Finally, the verifier accepts if and only if the assignment to X_1 given by prover 1 satisfies all the equations $e_1, e_2, \ldots, e_k$ and the assignment X_2 given by prover 2 is consistent with the answer of prover 1.

Completeness: It is easy to see the completeness case. If the instance (X, Eq) is $(1 - \varepsilon)$-satisfiable then there is a provers' strategy which makes the verifier accepts with probability at least $(1 - k\epsilon)$. The strategy is to use a fixed $(1-\varepsilon)$-satisfiable assignment and answer according to it. In this case, with probability at least $(1 - k\epsilon)$, the verifier chooses k constraints which are all satisfied by the fixed assignment and hence the verifier will accept provers' answers.

Soundness: Consider the case when the instance (X, Eq) is at most s-satisfiable for $s < 1$ from Theorem 3.2. If the provers were given only X_1 and X_2 without the advice strings, then the parallel repetition theorem of Raz [Raz98] directly implies that for any provers' strategy, they can make the verifier accept with probability at most

$2^{-\Omega(\beta k)}$. This follows because in expectation, there are βk constraints out of k where prover 2 receives one variable from the constraint. It turns out that a few advice strings will not give provers any significant advantage.

To see this, for each of these βk constraints, with probability 2^{-q}, all the advice strings get assigned value 000 to the variables in the constraints and therefore does not *leak* any information, about which variable from the constraint was being sent to prover 2, to prover 1. Thus in expectation, there are $\frac{\beta k}{2^q}$ constraints vs. variable questions where prover 1 knows nothing about the which variable was being sent to the prover 2. One can then argue, by using Raz's parallel repetition theorem, that any provers' strategy can make verifier accept with probability at most $2^{-\Omega(\beta k/2^q)}$.

The soundness is formally proved in [KMS17].

Theorem 3.3 (Section 3 in [KMS17]). *If the* REG-3LIN *instance* (X, Eq) *is at most* s*-satisfiable* ($s < 1$ *from Theorem 3.2) then there is no strategy with which the provers can make the verifier accept with probability greater than* $2^{-\Omega(\beta k/2^q)}$.

Remark 3.4. *The importance of the advice strings will come later in the proof of soundness. Specifically, the proof of Theorem 3.14 (from [DKK+18b] which we use as a black-box) crucially uses the advice strings given to the provers.*

To prove our main theorem, the reduction is carried out in three steps:

$$\begin{array}{lr}
\text{Outer Game} & \\
\downarrow \text{[DKK}^+\text{18b]} & \\
G_{\texttt{unfolded}}(A, B, E, \Pi, \Sigma_A, \Sigma_B) & \text{(unfolded 2-to-1 Game)} \\
\downarrow \text{[DKK}^+\text{18b]} & \\
G_{\texttt{folded}}(\tilde{A}, B, \tilde{E}, \tilde{\Pi}, \Sigma_A, \Sigma_B) & \text{(folded 2-to-1 Game)} \\
\downarrow \text{(This work)} & \\
\mathsf{UG}_{\texttt{folded}}(\widehat{A}, B, \widehat{E}, \widehat{\Pi}, \Sigma) & \text{(Unique Game)}
\end{array}$$

The first two steps are explained in the next two subsections. These follow from [DKK+18b]. The main contribution of our work is the last step which is given in Section 3.4.

3.2 Unfolded 2-to-1 Game

In this section we reduce REG-3LIN instance (X, Eq) to an instance of 2-to-1 Label Cover instance $G_{\texttt{unfolded}} = (A, B, E, \Pi, \Sigma_A, \Sigma_B)$.

For an equation $e \in Eq$, let $\mathtt{supp}(e) = \{i_1, i_2, i_3\}$ if e is a linear constraint on $x_{i_1}, x_{i_2}, x_{i_3}$. A set of k equations $(e_1, e_2, \ldots, e_k)$ is *legitimate* if the support of equations are pairwise disjoint and for every two different equations e_i and e_j and for any $x \in e_i$ and $y \in e_j$, the pair $\{x, y\}$ does not appear in any equation in (X, Eq). Define $\mathcal{U}$ to be the following set family.

$$\mathcal{U} = \left\{ S \in \binom{[n]}{3k} \;\middle|\; S = \cup_{i=1}^{k} \mathtt{supp}(e_i) \text{ and } (e_1, e_2, \ldots, e_k) \text{ is legitimate} \right\}.$$

Note that by definition, there is a one-to-one correspondence between the set of legitimate k tuples of equations and $\mathcal{U}$. For $U \in \mathcal{U}$, let $X_U \subseteq \mathbb{F}_2^n$ be the subspace with support in U. For an equation $e \in Eq$ on $x_{i_1}, x_{i_2}, x_{i_3}$, let x_e be the vector in X_U where $x_{i_1} = x_{i_2} = x_{i_3} = 1$ and rest of the coordinates are 0. Denote by $b_e \in \mathbb{F}_2$ the RHS of the equation e. Let H_U be the span of $\{x_e \mid x_e \in X_U\}$. Finally, let $\mathcal{V}$ be the collection of all sets of variables up to size $3k$ (thought of as subsets of $[n]$). Similar to X_U, for $V \in \mathcal{V}$, let $X_V \subseteq \mathbb{F}_2^n$ be the subspace with support in V.

Vertices (A, B)**:** Let $\ell \ll k$ which we will set later. The vertex set of the game $G_{\mathtt{unfolded}}$ is defined as follows:

$$A = \{(U, L) \mid U \in \mathcal{U}, L \in Gr(X_U, \ell), L \cap H_U = \{0\}\}.$$

$$B = \{(V, L') \mid V \in \mathcal{V}, L' \in Gr(X_V, \ell - 1)\}.$$

Edges E**:** The distribution on edges are defined by the following process: Choose X_1 and X_2 as per the distribution given in the outer verifier conditioned on $X_1 \in \mathcal{U}$. Let $U = X_1$ and $V = X_2$. Choose a random subspace $L' \in Gr(X_V, \ell - 1)$ and a random $L \in Gr(X_U, \ell)$ such that $L' \subseteq L$. Output $\{(U, L), (V, L')\} \in (A, B)$.

Labels (Σ_A, Σ_B)**:** The label set $\Sigma_A = \mathbb{F}_2^{\ell}$ and the label set $\Sigma_B = \mathbb{F}_2^{\ell-1}$. A labeling $\sigma \in \Sigma_A$ to (U, L) can be thought of as a linear function $\sigma : L \to \mathbb{F}_2$. Similarly the label $\sigma' \in \Sigma_B$ to a vertex (V, L') is though of as a linear function $\sigma' : L' \to \mathbb{F}_2$. This can be done by fixing arbitrary basis of the respective spaces.

3.3 Folded 2-to-1 Game

For every assignment to the 3LIN instance, there are many vertices in the graph $G_{\mathtt{unfolded}}$ which get the same label according to strategy of labeling the vertices in $G_{\mathtt{unfolded}}$ with respect to a fixed assignment to X. So we might as well enforce this constraint on the variables in $G_{\mathtt{unfolded}}$. This is achieved by *folding*. In this section, we convert $G_{\mathtt{unfolded}}$ to the following Game $G_{\mathtt{folded}} = (\tilde{A}, B, \tilde{E}, \tilde{\Pi}, \Sigma_A, \Sigma_B)$.

Vertices $(\tilde{A}, B)$**:** Consider the following grouping of the vertices from A

$$\mathcal{C}(U_0, L_0) = \{(U, L) \in A \mid L_0 + H_U + H_{U_0} = L + H_U + H_{U_0}\}.$$

The following Lemma 3.5 says that $\mathcal{C}$ is indeed an equivalence class. The proof of the lemma crucially uses the facts that U corresponds to a *legitimate* set of equations and that the REG-3LIN instance is *regular*: namely, every equation consists of 3 distinct variables and every two distinct constraints share at most one variable.

Lemma 3.5 (Lemma 3.2 in [DKK+18b])**.** *$\mathcal{C}$ is an equivalence class: there exists an ℓ dimension subspace $R_{\mathcal{C}}$ such that for all $(U, L) \in \mathcal{C}$,*

$$H_U + L = R_{\mathcal{C}} + H_U.$$

We define the vertex set $\tilde{A}$ as follows:

$$\tilde{A} = \{\mathcal{C}(U, L) \mid (U, L) \in A\}.$$

In other words, there is a vertex for every equivalence class in $\tilde{A}$.

Edges $\tilde{E}$**:** Sample (U, L) (V, L') with respect to E. Output $\mathcal{C}(U, L)$, (V, L').

Labels (Σ_A, Σ_B)**:** The label set $\Sigma_A = \mathbb{F}_2^\ell$, a label σ to $\mathcal{C}$ can be thought of as a linear function $\sigma : R_{\mathcal{C}} \to \mathbb{F}_2$. As before, the label $\sigma' \in \Sigma_B$ to a vertex (V, L') is though of as a linear function $\sigma' : L' \to \mathbb{F}_2$.

In order to define the constraints on the edges, we need the following definitions:

Definition 3.6. *For a space $H_U + L$ such that $L \cap H_U = \{0\}$ and a linear function $\sigma : L \to \mathbb{F}_2$, the extension of σ, respecting the side conditions, to the whole space $H_U + L$ is a linear function $\beta : H_U + L \to \mathbb{F}_2$ such that for all $x_e \in X_U$, $\beta(x_e) = b_e$ and $\beta|_L = \sigma$.*

Note that there is one to one mapping from a linear function on L and its extension as all the equations in U are disjoint and hence $\{x_e \mid x_e \in X_U\}$ form a basis of the space H_U.

Definition 3.7. *Consider a label σ to a vertex $\mathcal{C}$ which is a linear function on $R_{\mathcal{C}}$. The unfolding of it to the elements of the $\mathcal{C}$ is given as follows: For $(U, L) \in \mathcal{C}$, define a linear function $\tilde{\sigma}_U : H_U + L \to \mathbb{F}_2$ such that it is equal to the extension of σ to $H_U + R_{\mathcal{C}}$ respecting the side conditions.*

The spaces $H_U + L$ and $H_U + R_{\mathcal{C}}$ are the same and hence the above definition makes sense. We are now ready to define the constraints.

Constraints $\tilde{\Pi}$: Consider linear functions $\sigma : R_{\mathcal{C}} \to \mathbb{F}_2$ and $\sigma' : L' \to \mathbb{F}_2$. A pair (σ, σ') satisfies the edge $(\mathcal{C}, (V, L')) \in \tilde{E}$, if for every $(U, L) \in \mathcal{C}$ such that $((U, L), (V, L')) \in E$, the unfolding $\tilde{\sigma}_U|_{L'} = \sigma'$.

We have the following completeness and soundness guarantee of the reduction from [DKK+18b].

Lemma 3.8 (Completeness)**.** *(Lemma 4.1 in [DKK+18b]) If the* Reg-3Lin *instance* (X, Eq) *is* $(1-\varepsilon)$*-satisfiable then there exists* $\tilde{A}' \subseteq \tilde{A}$*,* $|\tilde{A}'| \geqslant (1-k\epsilon)|\tilde{A}|$ *and a labeling to the* 2*-to-*1 *Label Cover instance* $G_{\texttt{folded}}$ *such that* all *the edges incident on* $\tilde{A}'$ *are satisfied.*

Lemma 3.9 (Soundness)**.** *(Lemma 4.2 in [DKK+18b], and [KMS18]) For all* $\delta > 0$*, there exists* $q, k \geq 1$ *and* $\beta \in (0,1)$*, such that if the* Reg-3Lin *instance* (X, Eq) *is at most* s*-satisfiable (where* s *is from Theorem 3.2) then every labeling to* $G_{\texttt{folded}}$ *satisfies at most* δ *fraction of the edges.*

3.4 Reduction to Unique Games

In this section, we convert $G_{\texttt{folded}}$ Label Cover instance to a Unique Games instance with the stronger completeness guarantee that we are after. We will reduce an instance $G_{\texttt{folded}} = (\tilde{A}, B, \tilde{E}, \tilde{\Pi}, \Sigma_A, \Sigma_B)$ to an instance of Unique Game $\mathsf{UG}_{\texttt{folded}} = (\widehat{A}, B, \widehat{E}, \widehat{\Pi}, \Sigma)$.

Vertices $(\widehat{A}, B)$: We will split each vertex $\mathcal{C} \in \tilde{A}$ into many copies. Fix an ℓ dimensional subspace $R_{\mathcal{C}}$ given by Lemma 3.5. For every $x \in R_{\mathcal{C}} \setminus \{0\}$ and $b \in \{0,1\}$ we add a copy $\mathcal{C}_{x,b}$ to $\widehat{A}$.

$$\widehat{A} = \bigcup_{\mathcal{C} \in \tilde{A}} \{\mathcal{C}_{x,b} \mid x \in R_{\mathcal{C}} \setminus \{0\}, b \in \{0,1\}\}.$$

Edges $\widehat{E}$: The distribution on the edge set $\widehat{E}$ is as follows: We first pick $((U,L),(V,L'))$ according to the distribution E. Let $(U, L) \in \mathcal{C}$. We then select $y \in (H_U + L) \setminus (H_U + L')$ and $b \in \{0,1\}$ uniformly at random. Note that $\texttt{dim}(\texttt{span}\{y, H_U\} \cap R_{\mathcal{C}}) = 1$ since $y \notin H_U$ and $y \in H_U + L = H_U + R_{\mathcal{C}}$. Let $x \in \texttt{span}\{y, H_U\} \cap R_{\mathcal{C}}$ be the non-zero vector. Output $(\mathcal{C}_{x,b}, (V, L'))$.

Claim 3.10. *x is distributed uniformly in $R_{\mathcal{C}} \setminus (H_U + L')$ conditioned on (U, V, L, L').*

Proof. We first claim that $x \in R_{\mathcal{C}} \setminus (H_U + L')$ by showing $x \notin H_U + L'$. Suppose not, then we can write $x = h + x'$ where $h \in H_U$ and $x' \in L'$. We also know that $x \in \texttt{span}\{y, H_U\} \cap R_{\mathcal{C}}$ and $R_{\mathcal{C}} \cap H_U = \{0\}$. Thus x can be written as $x = \tilde{h} + y$ where $\tilde{h} \in H_U$. This implies $h + x' = \tilde{h} + y$. In other words, $y = h + \tilde{h} + x' \in H_U + L'$, a contradiction.

Since each $y \in (H_U + L) \setminus (H_U + L')$ gives a unique non-zero $x \in \texttt{span}\{y, H_U\} \cap R_{\mathcal{C}}$, we will show that the number of $y \in (H_U + L) \setminus (H_U + L')$ which gives a fixed x is same for all $x \in R_{\mathcal{C}} \setminus (H_U + L')$ and this will prove the claim.

Fix any $\tilde{x} \in R_{\mathcal{C}} \setminus (H_U + L')$, we now claim that the set of all $y \in (H_U + L) \setminus (H_U + L')$ that gives $\tilde{x}$ is $\mathtt{span}\{\tilde{x}, H_U\} \setminus H_U$. Clearly, for any $y \notin \mathtt{span}\{\tilde{x}, H_U\} \setminus H_U$, $\tilde{x} \notin \mathtt{span}\{y, H_U\}$ and also for every $y \in \mathtt{span}\{\tilde{x}, H_U\} \setminus H_U$, $\tilde{x} \in \mathtt{span}\{y, H_U\}$. Thus, it remains to show that $\mathtt{span}\{x, H_U\} \setminus H_U \subseteq (H_U + L) \setminus (H_U + L')$ for all $x \in R_{\mathcal{C}} \setminus (H_U + L')$.

To prove the inclusion, suppose for contradiction $(\mathtt{span}\{x, H_U\} \setminus H_U) \cap (H_U + L') \neq \emptyset$. This means $x + h = \tilde{h} + v'$ for some $h, \tilde{h} \in H_U$ and $v' \in L'$. This implies $x = h + \tilde{h} + v' \in H_U + L'$ contradicting $x \in R_{\mathcal{C}} \setminus (H_U + L')$. ■

Labels Σ: The label set is $\Sigma = \mathbb{F}_2^{\ell-1}$. A label σ to $\mathcal{C}_{x,b}$ can be thought of as a linear function $\sigma : R_{\mathcal{C}} \to \mathbb{F}_2$ such that $\sigma(x) = b$. This is done by fixing an arbitrary $\ell - 1$ basis elements from the subspace $\{y \in R_{\mathcal{C}} \mid y \perp x\}$.

It is easy to see that there is a one-to-one correspondence between a label σ and a linear function $\tilde{\sigma}$ on $R_{\mathcal{C}}$. Similar to the previous case of $G_{\mathtt{folded}}$, a label from $\Sigma(= \Sigma_B)$ to a vertex (V, L') in B is interpreted as a linear function $\sigma' : L' \to \mathbb{F}_2$.

We define an analogous unfolding of label to vertices in $\widehat{A}$ to the elements of the corresponding equivalence class. Since the label sets are different, for a label σ to $\mathcal{C}_{x,b}$ (thought of as a linear function on $R_{\mathcal{C}}$ respecting $\sigma(x) = b$) we use the notation $\widehat{\sigma}_U$ to denote its unfolding to $(U, L) \in \mathcal{C}_{x,b}$.

1-to-1 Constraints $\widehat{\Pi}$: Finally the constraint $\pi_e : \Sigma \to \Sigma$ between the endpoints of an edge $e = (\mathcal{C}_{x,b}, (V, L'))$ is given as follows: Consider linear functions $\sigma : R_{\mathcal{C}} \to \mathbb{F}_2$ respecting $\sigma(x) = b$ and $\sigma' : L' \to \mathbb{F}_2$. A pair $(\sigma, \sigma') \in \pi_e$ if for every $(U, L) \in \mathcal{C}$ such that $((U, L), (V, L')) \in E$ and $\mathtt{span}\{x, H_U\} \cap L' = \{0\}$, the unfolding $\tilde{\sigma}_U$ satisfies $\tilde{\sigma}_U|_{L'} = \sigma'$.

To see that every σ' has a unique pre-image, for any linear function $\sigma' : L' \to \mathbb{F}_2$, there is a unique linear function $\sigma : R_{\mathcal{C}} \to \mathbb{F}_2$ such that $\sigma(x) = b$ satisfying the above conditions. This is because of the following claim.

Claim 3.11. *Any basis for L' along with x and $\{x_e \mid x_e \in X_U\}$ forms a basis for $H_U + R_{\mathcal{C}}$ for every $(U, L) \in \mathcal{C}$.*

Proof. Let us unwrap the conditions for putting an edge between (V, L') and $\mathcal{C}_{x,b}$. One necessary condition is that $(\mathcal{C}, (V, L'))$ should be an edge in $\tilde{E}$. By the definition of $\tilde{E}$, there exists $(U, L) \in \mathcal{C}$ such that $L' \subseteq L$. Recall, x is such that there exists $y \in (H_U + L) \setminus (H_U + L')$ such that $\mathtt{dim}(\mathtt{span}\{y, H_U\} \cap R_{\mathcal{C}}) = 1$ and $x \in \mathtt{span}(y, H_U) \cap R_{\mathcal{C}}$. Therefore $x \in (H_U + L) \setminus (H_U + L')$ and hence $\mathtt{dim}(\mathtt{span}\{x, H_U + L'\}) = k + \ell$ (as $H_U \cap L' = \{0\}$). This implies that any basis of L', basis $\{x_e \mid x_e \in X_U\}$ of H_U and x span $H_U + L$. Since by Lemma 3.5 the space $H_U + L$ is same as the space $H_U + R_{\mathcal{C}}$, the claim follows. ■

We now show the completeness and soundness of the Unique Games instance.

Lemma 3.12 (Completeness). *For all $\varepsilon > 0$, if there exists $\tilde{A}' \subseteq \tilde{A}$, $|\tilde{A}'| \geqslant (1-k\epsilon)|\tilde{A}|$ and a labeling to the 2-to-1 Label Cover instance $G_{\texttt{folded}}$ such that* all *the edges incident on $\tilde{A}'$ are satisfied then there exists $\widehat{A'} \subseteq \widehat{A}$, $|\widehat{A'}| \geqslant (\frac{1-k\epsilon}{2})|\widehat{A}|$ and a labeling to Unique Games instance $\mathsf{UG}_{\texttt{folded}}$ such that* all *the edges incident on $\widehat{A'}$ are satisfied.*

Proof. Fix a labeling $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}})$ to $G_{\texttt{folded}}$ where $\tilde{\mathcal{A}} : \tilde{A} \to \Sigma_A$ and $\tilde{\mathcal{B}} : B \to \Sigma_B$ which satisfies all the edges incident on $(1-k\epsilon)$ fraction of the vertices in $\tilde{A}$. We will construct a labeling $(\widehat{\mathcal{A}}, \widehat{\mathcal{B}})$ to the instance $\mathsf{UG}_{\texttt{folded}}$, where $\widehat{\mathcal{A}} : \widehat{A} \to \Sigma$ and $\widehat{\mathcal{B}} : B \to \Sigma$ which will satisfy all the edges adjacent to at least $\frac{(1-k\epsilon)}{2}$ fraction of vertices $\widehat{A}$ in $\mathsf{UG}_{\texttt{folded}}$.

We will set $\widehat{\mathcal{B}} = \tilde{\mathcal{B}}$. Now to assign a label to $\mathcal{C}_{x,b} \in \widehat{A}$, we look at the labeling $\sigma := \tilde{\mathcal{A}}(\mathcal{C}) \in \mathbb{F}_2^\ell$ as a linear function $\sigma : R_{\mathcal{C}} \to \mathbb{F}_2$. If $\sigma(x) = b$, we set $\widehat{\mathcal{A}}(\mathcal{C}_{x,b})$ to be the same linear function $\sigma : R_{\mathcal{C}} \to \mathbb{F}_2$ respecting $\sigma(x) = b$. Otherwise, we set $\widehat{\mathcal{A}}(\mathcal{C}_{x,b}) = \perp$. It is obvious that exactly half the vertices in $\widehat{A}$ got assigned a label in Σ.

Claim 3.13. *If the label $\tilde{\mathcal{A}}(\mathcal{C})$ to $\mathcal{C}$ satisfies all the edges incident on it, then for all $x \in R_{\mathcal{C}} \setminus \{0\}$, there exists a unique $b \in \{0,1\}$ such that the label $\widehat{\mathcal{A}}(\mathcal{C}_{x,b})$ satisfies all the edges incident on $\mathcal{C}_{x,b}$, unless $\widehat{\mathcal{A}}(\mathcal{C}_{x,b}) = \perp$.*

Proof. For convenience let $\sigma = \tilde{\mathcal{A}}(\mathcal{C})$. If we let $\Gamma(\mathcal{C}) \subseteq B$ to be the neighbors of $\mathcal{C}$ in $G_{\texttt{folded}}$, then the set of neighbors of $\mathcal{C}_{x,b}$ is a subset of $\Gamma(\mathcal{C})$. Furthermore if (V, L') is connected to $\mathcal{C}_{x,b}$ in $\mathsf{UG}_{\texttt{folded}}$ then $x \notin L'$ and $x \in R_{\mathcal{C}}$. The condition that the edge $(\mathcal{C}, (V, L'))$ is satisfied by $\tilde{\mathcal{A}}$ means that for all $(U, L) \in \mathcal{C}$ such that $L' \subseteq L$, the unfolding of σ satisfies $\tilde{\sigma}_U|_{L'} = \tilde{B}((V, L'))$. Since the unfolding of the label $\widehat{\mathcal{A}}(\mathcal{C}_{x,b})$ to $\mathcal{C}_{x,b}$ gives the same linear function $\tilde{\sigma}$, it follows that $\tilde{\sigma}_U|_{L'} = \widehat{B}((V, L'))$ for every $(U, L) \in \mathcal{C}$ and every $(V, L') \in \Gamma(\mathcal{C})$ such that $L' \subseteq L$. Therefore $\widehat{\mathcal{A}}$ satisfies all the edges incident on $\mathcal{C}_{x,b}$. ■

Let $\tilde{A}' \subseteq \tilde{A}$ be the set of vertices such that all the edges incident on them are satisfied by labeling $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}})$. By assumption $|\tilde{A}'| \geqslant (1-k\epsilon)|\tilde{A}|$. Consider the subset $\widehat{A'} \subseteq \widehat{A}$

$$\widehat{A'} = \{\mathcal{C}_{x,b} \mid \widehat{\mathcal{A}}(\mathcal{C}_{x,b}) \neq \perp, \mathcal{C} \in \tilde{A}'\}.$$

Now, $|\widehat{A'}| \geqslant \frac{1-k\epsilon}{2}|\widehat{A}|$ and from the above claim, all the edges incident on $\widehat{A'}$ are satisfied by the labeling $(\widehat{\mathcal{A}}, \widehat{\mathcal{B}})$. ■

3.5 Soundness

Let $F_U : \{L + H_U \mid L \in Gr(X_U, \ell), L \cap H_U = \{0\}\} \to \mathbb{F}_2^\ell$. $F_U[L + H_U]$ can be thought of as a linear function on $L + H_U$ respecting the side conditions. This is again by fixing arbitrary basis of L. Define $\texttt{agreement}(F_U)$ as the probability of the following event:

- Select a $\ell - 1$ dimension subspace $L' \in X_U$ such that $L' \cap H_U = \{0\}$ uniformly at random.

- Select a ℓ dimension subspaces L_1 and L_2 containing L' such that $L_1 \cap H_U = L_2 \cap H_U = \{0\}$ uniformly at random.
- Check if $F_U[L_1 + H_U]|_{L'} = F_U[L_1 + H_U]|_{L'}$.

The main technical theorem which was conjectured in [DKK+18b] and proved in [KMS18] is that if $\texttt{agreement}(F_U)$ is a constant bounded away from 0, then there is a global linear function $g : X_U \to \{0,1\}$ respecting the side conditions and a special (not too small) subset S of $\{L + H_U \mid L \in Gr(X_U, \ell), L \cap H_U = \{0\}\}$ such that for a constant fraction of elements in S, F_U agrees with g. We will not need the details of this theorem. Instead, we state the main soundness lemma from [DKK+18b] which crucially used the aforementioned structural theorem and also the advice strings as mentioned in Remark 3.4.

Theorem 3.14 (Implied by Lemma 4.1 in [DKK+18b])**.** *For every constant $\delta > 0$, there exist large enough $\ell \ll k$, $q \in \mathbb{Z}^+$ and $\beta \in (0,1)$ such that if there is an unfolded assignment $\mathcal{A} : A \to \Sigma_A$ to $G_{\texttt{unfolded}}$ such that for at least δ fraction of U,* $\texttt{agreement}(F_U) \geqslant \delta$*, then there exists a provers' strategy which makes the outer verifier accepts with probability at least p_δ, where p_δ is independent of k.*

Armed with this theorem, we are ready to prove the soundness of the Unique Games instance $\mathsf{UG}_{\texttt{folded}}$.

Lemma 3.15 (Soundness)**.** *Let $\delta > 0$ and fix $q \in \mathbb{Z}^+$ and $\beta \in (0,1)$ and $\ell \ll k$ as in Theorem 3.14. If $\mathsf{UG}_{\texttt{folded}}$ is δ-satisfiable then there exists a provers strategy which makes the outer verifier accepts with probability at least $p_{\frac{\delta^4}{2^{16}}}$.*

Proof. Fix any δ-satisfiable assignment $(\widehat{\mathcal{A}}, \widehat{\mathcal{B}})$, $\widehat{\mathcal{A}} : \widehat{A} \to \Sigma$, $\widehat{\mathcal{B}} : \widehat{B} \to \Sigma$ to the Unique Games instance $\mathsf{UG}_{\texttt{folded}}$. We first get a randomized labeling $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}})$ to $G_{\texttt{folded}}$ where $\tilde{\mathcal{A}} : \tilde{A} \to \Sigma_A$ and $\tilde{\mathcal{B}} : B \to \Sigma_B$ as follows: We will keep $\tilde{\mathcal{B}} = \widehat{\mathcal{B}}$. For every $\mathcal{C} \in \tilde{A}$, we pick a random $x \in R_{\mathcal{C}}$ and $b \in \{0,1\}$ and set $\tilde{\mathcal{A}}(\mathcal{C}) = \widehat{\mathcal{A}}(\mathcal{C}_{x,b})$. We now unfold the assignment $\tilde{\mathcal{A}}$ to $\mathcal{A}$. Define $F_U[L + H_U] = \mathcal{A}(U, L)$ for every $L \in Gr(X_U, \ell)$.

Let $p(U)$ denote the probability that an edge in $\mathsf{UG}_{\texttt{folded}}$ is satisfied conditioned on U. Consider U such that $p(U) \geqslant \frac{\delta}{2}$. By an averaging argument, there are at least $\frac{\delta}{2}$ fraction of U such that $p(U) \geqslant \frac{\delta}{2}$.

Claim 3.16. $\mathbf{E}_{F_U}[\texttt{agreement}(F_U)] \geqslant \frac{p(U)^4}{2^{11}} - o_k(1)$.

Proof. Define a randomized assignment $F'_U[L']$ as follows: Select a random $V \subseteq U$ conditioned on the event that $L' \subseteq X_V$. Set $F'_U[L'] = \widehat{\mathcal{B}}(V, L')$. Consider the following two distributions:

<u>Distribution $\mathcal{D}_U$:</u>

- Select $\mathbf{V}$ uniformly at random from $\{V \mid (U,V) \in E\}$

- Select $\mathbf{L}'$ uniformly at random from $Gr(X_V, \ell - 1)$
- Select $\mathbf{L}$ uniformly at random from $\{L \mid L \in Gr(X_U, \ell) \text{ and } L' \subseteq L\}$
- Let $\mathcal{C}$ be the equivalence class such that $(U, L) \in \mathcal{C}$, select $\mathbf{x} \sim R_{\mathcal{C}}$ as in the edge distribution $\widehat{E}$.
- Select $\mathbf{b} \in \{0, 1\}$ uniformly at random.

Distribution $\mathcal{D}'_U$:

- Select $\mathbf{L}'$ uniformly at random from $Gr(X_U, \ell - 1)$
- Select $\mathbf{V}$ uniformly at random from $\{V \mid (U, V) \in E \text{ and } L' \in Gr(X_V, \ell - 1)\}$
- Select $\mathbf{L}$ uniformly at random from $\{L \mid L \in Gr(X_U, \ell) \text{ and } L' \subseteq L\}$
- Let $\mathcal{C}$ be the equivalence class such that $(U, L) \in \mathcal{C}$, select $\mathbf{x} \sim R_{\mathcal{C}}$ as in the edge distribution $\widehat{E}$.
- Select $\mathbf{b} \in \{0, 1\}$ uniformly at random.

We have the following lemma from [DKK+18b].

Lemma 3.17 (Lemma 3.6 in [DKK+18b])**.** *Consider the two marginal distributions on the pair (V, L'), one with respect to $\mathcal{D}_U$ and another with respect to $\mathcal{D}'_U$. If $2^\ell \beta \leq \frac{1}{8}$, then the statistical distance between the two distributions is at most $\beta\sqrt{k} \cdot 2^{\ell+4}$.*

In the distribution $\mathcal{D}_U$, there is always a constraint between $\mathcal{C}_{x,b}$ and (V, L') in $\mathsf{UG}_{\texttt{folded}}$. Moreover, the distribution of $(\mathcal{C}_{x,b}, (V, L'))$ is same as the edge distribution $\widehat{E}$. Therefore

$$p(U) = \Pr_{\mathcal{D}_U}\left[\widehat{\mathcal{A}}(\mathcal{C}_{x,b}), \widehat{\mathcal{B}}(V, L') \text{ satisfy the edge } (\mathcal{C}_{x,b}, (V, L'))\right].$$

Rewriting the above equality,

$$p(U) = \Pr_{\mathcal{D}_U}\left[\widehat{\sigma}_U|_{L'} = \widehat{\mathcal{B}}(V, L') \mid \sigma = \widehat{\mathcal{A}}(\mathcal{C}_{x,b})\right].$$

Using Claim 3.10, the distribution of $F_U[L + H_U]$, conditioned on $x \in R_{\mathcal{C}} \setminus (H_U + L')$, is same as the distribution $\widehat{\mathcal{A}}(\mathcal{C}_{x,b})$ (with appropriate unfolding of it) chosen with respect to $\mathcal{D}_U$. As $|R_{\mathcal{C}} \setminus (H_U + L')| = |R_{\mathcal{C}}|/2$ for a random $x \in R_{\mathcal{C}}$, the event $x \in R_{\mathcal{C}} \setminus (H_U + L')$ happens with probability $\frac{1}{2}$. Since we pick an uniformly random $x \in R_{\mathcal{C}}$ while defining $\tilde{\mathcal{A}}(\mathcal{C})$, which in turn defines $F_U[L + H_U]$, we have

$$\frac{p(U)}{2} \leqslant \mathop{\mathbf{E}}_{F_U} \Pr_{\mathcal{D}_U}\left[F_U[L + H_U]|_{L'} = \widehat{\mathcal{B}}(V, L')\right],$$

Now,

$$\Pr_{\mathcal{D}_U}\left[F_U[L+H_U]|_{L'} = \widehat{\mathcal{B}}(V,L')\right] \approx \Pr_{\mathcal{D}'_U}\left[F_U[L+H_U]|_{L'} = \widehat{\mathcal{B}}(V,L')\right].$$

follows from the closeness of distributions $\mathcal{D}_U$ and $\mathcal{D}'_U$ on (V,L') given by Lemma 3.17 by setting $\beta \ll \frac{1}{\sqrt{k}}$ (this setting of β is consistent with the setting of β in Theorem 3.14). Conditioned on L' the distribution of (V,L') in $\mathcal{D}'_U$ is same as the distribution we used to assign $F'_U[L']$ and therefore we get

$$\frac{p(U)}{2} - o_k(1) \leqslant \mathop{\mathbf{E}}_{F_U} \Pr_{L' \subseteq L}\left[F_U[L+H_U]|_{L'} = F'_U[L']\right].$$

Let E_1 be the event that $\frac{p(U)}{4} \leqslant \Pr_{L'\subseteq L}\left[F_U[L+H_U]|_{L'} = F'_U[L']\right]$, by averaging argument $\Pr[E_1] \geqslant \frac{P(U)}{4}$. We now fix an F_U for which E_1 occurs. By an averaging argument, there are at least $\frac{p(U)}{8}$ fraction of $L' \in Gr(X_U, \ell-1)$ such that $\Pr_{L\supseteq L'}\left[F_U[L+H_U]|_{L'} = F'_U[L']\right] \geqslant \frac{p(U)}{8}$. For each of such L' we have,

$$\begin{aligned}\Pr_{L_1,L_2\supseteq L'}\left[F_U[L_1+H_U] = F_U[L_2+H_U]\right] &= \Pr_{L_1,L_2\supseteq L'}\left[F_U[L_1+H_U]|_{L'} = F_U[L_2+H_U]|_{L'} = F'_U[L']\right] \\ &\geqslant \frac{p(U)^2}{2^6} - o_k(1).\end{aligned}$$

Thus overall, we get

$$\Pr_{L_1,L_2\supseteq L'}\left[F_U[L_1+H_U] = F_U[L_2+H_U] \mid E_1\right] \geqslant \frac{p(U)^3}{2^9} - o_k(1).$$

Hence,

$$\mathop{\mathbf{E}}_{F_U}[\mathtt{agreement}(F_U)] \geqslant \Pr[E_1]\cdot \Pr_{L_1,L_2\supseteq L'}\left[F_U[L_1+H_U] = F_U[L_2+H_U] \mid E_1\right] \geqslant \frac{p(U)^4}{2^{11}} - o_k(1).$$

■

There are at least $\frac{\delta}{2}$ fraction of U such that $p(U) \geqslant \frac{\delta}{2}$. This means for at least $\frac{\delta}{2}$ fraction of U, $\mathbf{E}[\mathtt{agreement}(F_U)] \geqslant \frac{\delta^4}{2^{15}} - o_k(1)$ using the previous claim. Thus, again by an averaging argument, there exists a fixed $\{F_U \mid U \in \mathcal{U}\}$, coming from unfolding of some assignment $\tilde{\mathcal{A}}$, such that for at least $\frac{\delta^4}{2^{16}}$ fraction of U, we have $\mathtt{agreement}(F_U) \geqslant \frac{\delta^4}{2^{16}}$. The Lemma now follows from Theorem 3.14. ■

We now prove the main theorem.

Proof of Theorem 1.2 Fix $\delta > 0$. We let q, β and $\ell \ll k$ be as given in the setting of Theorem 3.14. Firstly, if we look the the marginal distribution of the edge distribution on $\widehat{A}$ then it is uniform and hence the instance is left-regular.[2] Now, starting with an instance of (X, Eq) we have the following two guarantees of the reduction:

1. If the instance (X, Eq) is $(1 - \frac{2\delta}{k})$-satisfiable then by Lemma 3.8 and Lemma 3.12, the Unique Games instance $\mathsf{UG}_{\texttt{folded}}$ has a property that for at least $(\frac{1}{2} - \delta)$ fraction of the vertices in $\widehat{A}$, all the edges incident on them are satisfied.

2. Consider the other case in which the instance (X, Eq) is at most s-satisfiable where $s < 1$. If the Unique Games instance $\mathsf{UG}_{\texttt{folded}}$ is has a δ-satisfying assignment, then by Lemma 3.15 there is a provers' strategy which can make the outer verifier accepts with probability at least $p_{\frac{\delta^4}{2^{16}}} \gg 2^{-\Omega(\beta k/2^q)}$ for large enough k. This contradicts Theorem 3.3 and hence in this case, $\mathsf{UG}_{\texttt{folded}}$ has no assignment which satisfies δ fraction of the edges.

Since by Theorem 3.2 distinguishing between a given instance (X, Eq) being at least $(1 - \frac{2\delta}{k})$-satisfiable or at most s-satisfiable is NP-hard, this proves our main theorem. ■

4 Independent set in degree d graphs

We consider a weighted graph $H = (V, E)$ where the sum of all the weights of the vertices is 1 and also sum of all weights of the edges is also 1. For $S \subseteq V$, we will denote the total weight of vertices in S by $w(S)$.

Definition 4.1. *A graph H is (δ, ε)-dense if for every $S \subseteq V(H)$ with $w(S) \geqslant \delta$, the total weight of edges inside S is at least ε.*

For $\rho \in [-1, 1]$ and $\beta \in [0, 1]$, the quantity $\Gamma_\rho(\beta)$ is defined as:

$$\Gamma_\rho(\beta) := \Pr[X \leq \phi^{-1}(\beta) \wedge Y \leq \phi^{-1}(\beta)],$$

where X and Y are jointly distributed normal Gaussian random variables with co-variance ρ and ϕ is the cumulative density function of a normal Gaussian random variable.

We will prove the following theorem.

Theorem 4.2. *Fix $\varepsilon > 0, p \in \left(0, \frac{1}{2}\right]$, then for all sufficiently small $\delta > 0$, there exists a polynomial time reduction from an instance of a left-regular Unique Games $G(A, B, E, [L], \{\pi_e\}_{e \in E})$ to a graph H such that*

1. *If $\mathsf{sval}(G) \geqslant c$, then there is an independent set of weight $c \cdot p$ in H.*

[2]The edges have weights, but it can be made an unweighted left-regular instance by adding multiple edges proportional to its weight with the same constraint.

Let $G(A, B, E, [L], \{\pi_e\}_{e \in E})$ be an instance of Unique Games. The distribution of edges in H is as follows:

- Select $u \in B$ uniformly at random.
- Select its two neighbors v_1 and v_2 uniformly at random. Let π_1 and π_2 are the constraints between (u, v_1) and (u, v_2) respectively.
- Select $x, y \in \{0,1\}^L$, such that for each $i \in [L]$, (x_i, y_i) are sampled independently from the distribution $\mathcal{D}$.
- Output an edge $(v_1, x \circ \pi_1), (v_2, y \circ \pi_2)$.

Figure 3: Reduction from UG to Independent Set from [AKS11].

2. *If* $\mathsf{val}(G) \leqslant \delta$*, then* H *is* $(\beta, \Gamma_\rho(\beta) - \varepsilon)$ *dense for every* $\beta \in [0,1]$ *and* $\rho = -\frac{p}{p-1}$.

The reduction is exactly the same as the one in [AKS11]. We will only show the complete case (1) here. The soundness is proved in [AKS11]. This theorem will imply Theorem 1.3 using a randomized sparsification technique of [AKS11] to convert the weighted graph into a bounded degree unweighted graph.

4.1 The AKS reduction

Consider the distribution $\mathcal{D}$ on $(a, b) \in \{0,1\}^2$ such that $\Pr[a = b = 1] = 0$ and each bit is p-biased i.e. $\Pr[b = 1] = \Pr[b = 1] = p$. For a string $x \in \{0,1\}^L$ and a permutation $\pi : [L] \to [L]$, define $x \circ \pi \in \{0,1\}^L$ as $(x \circ \pi)_i = x_{\pi(i)}$ for all $i \in [L]$.

Let $G(A, B, E, [L], \{\pi_e\}_{e \in E})$ be an instance of Unique Games which is regular on the A side. We convert it into a weighted graph H. The vertex set is $A \times \{0,1\}^L$. Weight of a vertex (v, x) where $v \in A$ and $x \in \{0,1\}^L$ is $\frac{\mu_p(x)}{|A|}$, where $\mu_p(x) := p^{|x|}(1-p)^{L-|x|}$. The edge distribution is given in Figure 3.

Lemma 4.3 (Completeness)**.** *If* $\mathsf{sval}(G) \geqslant c$*, then there is an independent set in* H *of weight* $c \cdot p$.

Proof. Fix an assignment $\ell : A \cup B \to \Sigma$ which gives $\mathsf{sval}(G) \geqslant c$. Let $A' \subseteq A$ be the set of vertices such all the edges incident on A' are satisfied by ℓ, we know that $|A'| \geqslant c \cdot |A|$. Consider the following subset of vertices in H.

$$I = \{(v, x) \mid v \in A', x_{\ell(v)} = 1\}.$$

Firstly, the weight of set I is $c \cdot p$. We show that I is in fact an independent set in H. Suppose for contradiction, there exists an edge $(v_1, x), (v_2, y)$ in H and both of its endpoints are in I.

Let u be the common neighbor of v_1, v_2 (one such u must exist). If we let π_1 and π_2 be the permutation constraints between (u, v_1) and (u, v_2) then the conditions for being an edge implies that $(x_{\pi_1(\ell(u))}, y_{\pi_2(\ell(u))})$ should have a support in $\mathcal{D}$. Since all the edges incident on A' are satisfied, $\pi_i(\ell(u)) = \ell(v_i)$ for $i \in \{1, 2\}$. Therefore,$(x_{\ell(v_1)}, y_{\ell(v_2)})$ is also supported in $\mathcal{D}$ and hence both cannot be 1 which implies that both cannot belong to I. ■

Lemma 4.4 (Soundness [AKS11])**.** *For every constant $\varepsilon > 0$, if H is not $(\beta, \Gamma_\rho(\beta) - \varepsilon)$-dense for some $\beta \in [0, 1]$ and $\rho = -\frac{p}{p-1}$, then G is δ-satisfiable for $\delta := \delta(\varepsilon, p) > 0$.*

Lemma 4.3 and Lemma 4.4 prove Theorem 4.2.

5 Maximum Acyclic Subgraph

In this section we state the reduction from [GMR08] and analyze the completeness case. Given a directed graph $H = (V, E)$, we will denote by $\mathtt{Val}(H)$ the fraction of edges in the maximum sized acyclic subgraph of H. We need the following definition.

Definition 5.1. *A t-ordering of a directed graph $H = (V, E)$ consists of a map $O : V \to [t]$. The value of a t-ordering O is given by*

$$\mathtt{Val}_t(O) = \Pr_{(a,b)\in E}[O(a) < O(b)] + \frac{1}{2} \cdot \Pr_{(a,b)\in E}[O(a) = O(b)].$$

Define $\mathtt{Val}_t(H)$ as:

$$\mathtt{Val}_t(H) = \max_O \mathtt{Val}_t(O).$$

The following lemma [GMR08] will be crucial in the reduction from Unique Games to Maximum Acyclic Subgraph.

Lemma 5.2 ([GMR08])**.** *Given $\eta > 0$ and a positive integer t, for every sufficiently large m, there exists a weighted directed acyclic graph $H(V, E)$ on m vertices along with a of distribution $\mathcal{D}$ on the orderings $\{O : V \to [m]\}$ such that:*

1. *For every $u \in V$ and $i \in [m]$, $\Pr_{O\sim\mathcal{D}}[O(u) = i] = \frac{1}{m}$.*
2. *For every directed edge $(a \to b)$, $\Pr_{O\sim\mathcal{D}}[O(a) < O(b)] \geqslant 1 - \eta$.*
3. *$\mathtt{Val}_t(H) \leqslant \frac{1}{2} + \eta$.*

The reduction is given in Figure 4. For a string $x \in [q]^L$ and a permutation $\pi : [L] \to [L]$, define $x \circ \pi \in [q]^L$ such that $(x \circ \pi)_i = x_{\pi(i)}$ for all $i \in [L]$.

Lemma 5.3. *(Completeness) For small enough $\varepsilon, \eta > 0$, if the Unique Games instance G is has $\mathsf{sval}(G) \geqslant c$ then $\mathtt{Val}(\mathcal{G}) \geqslant c \cdot (1 - 2\epsilon)(1 - \eta) + (1 - c) \cdot \left(\frac{1}{2} - \frac{1}{2m}\right)$*

Let $G(A, B, E, [L], \{\pi_e\}_{e \in E})$ be an instance of Unique Games. Fix a graph $H([m], E_H)$ from Lemma 5.2 with parameters $\eta > 0$ and $t \in \mathbb{Z}^+$, along with the distribution $\mathcal{D}$. Construct a weighted directed graph $\mathcal{G}$ on $B \times [m]^L$ with the following distribution on the edges:

- Select $u \in A$ uniformly at random.
- Select its two neighbors v_1 and v_2 uniformly at random. Let π_1 and π_2 are the constraints between (u, v_1) and (u, v_2) respectively.
- Pick an edge $e = (a, b) \in E_H$ at random from the graph H.
- Select $x, y \in [m]^L$, such that for each $i \in [L]$, (x_i, y_i) are sampled independently as follows:
  - sample $O \sim \mathcal{D}$, set $x_i = O(a)$ and $y_i = O(b)$.
- Perturb x and y as follows: for each $i \in [L]$, with probability $(1-\varepsilon)$, set $\tilde{x}_i = x_i$, with probability ε set $\tilde{x}_i$ to be uniformly at random from $[m]$. Do the same thing for y independently to get $\tilde{y}$.
- Output a directed edge $(v_1, \tilde{x} \circ \pi_1) \to (v_2, \tilde{y} \circ \pi_2)$.

Figure 4: Reduction from UG to Max-Acyclic Graph from [GMR08].

Proof. Fix an assignment $\ell : A \cup B \to \Sigma$ which gives $\mathsf{sval}(G) \geqslant c$. Let $A' \subseteq A$ be the set of vertices such that its edges are satisfied by ℓ, we know that $|A'| \geqslant c \cdot |A|$. Consider the following m ordering $\mathcal{O} : B \times [m]^L \to [m]$ of the vertices of $\mathcal{G}$: $\mathcal{O}(v, x) = x_{\ell(v)}$. We will show that $\mathtt{Val}_m(\mathcal{O}) \geqslant c(1-2\varepsilon)(1-\eta) + (1-c) \cdot \left(\frac{1}{2} - \frac{1}{2m}\right)$. This will prove the lemma.

$$\begin{aligned}
\mathtt{Val}(\mathcal{G}) \geqslant \mathtt{Val}_m(\mathcal{O}) &\geqslant \Pr[\mathcal{O}((v_1, \tilde{x} \circ \pi_1) < \mathcal{O}(v_2, \tilde{y} \circ \pi_2)] \\
&= \Pr[\tilde{x}_{\pi_1(\ell(v_1))} < \tilde{y}_{\pi_2(\ell(v_2))}] \\
&\geqslant c \cdot \Pr[\tilde{x}_{\pi_1(\ell(v_1))} < \tilde{y}_{\pi_2(\ell(v_2))} \mid u \in A'] \\
&\quad + (1-c) \cdot \Pr[\tilde{x}_{\pi_1(\ell(v_1))} < \tilde{y}_{\pi_2(\ell(v_2))} \mid u \notin A']. \qquad (1)
\end{aligned}$$

Now, if $u \in A'$ then $\pi_1(\ell(v_1)) = \pi_2(\ell(v_2)) = \ell(u)$ and hence,

$$\begin{aligned}
\Pr[\tilde{x}_{\pi_1(\ell(v_1))} < \tilde{y}_{\pi_2(\ell(v_2))} \mid u \in A'] &= \Pr[\tilde{x}_{\ell(u)} < \tilde{y}_{\ell(u)}] \\
&\geqslant (1-2\epsilon) \cdot \mathop{\mathbf{E}}_{(a,b) \in E_H} \Pr_{O \sim \mathcal{D}}[O(a) < O(b)] \\
&\geqslant (1-2\epsilon)(1-\eta). \qquad (2)
\end{aligned}$$

Now, we can lower bound $\Pr[\tilde{x}_{\pi_1(\ell(v_1))} < \tilde{y}_{\pi_2(\ell(v_2))} \mid u \notin A']$ by $(1-2\epsilon)(1-\eta)$ as above if $\pi_1(\ell(v_1)) = \pi_2(\ell(v_2))$. If $\pi_1(\ell(v_1)) \neq \pi_2(\ell(v_2))$ then $\tilde{x}_{\pi_1(\ell(v_1))}$ and $\tilde{y}_{\pi_1(\ell(v_1))}$ are uncorrelated and are distributed uniformly in $[m]$. Therefore, $\Pr[\tilde{x}_{\pi_1(\ell(v_1))} < \tilde{y}_{\pi_2(\ell(v_2))} \mid u \notin A'] = \frac{\binom{m}{2}}{m^2} = \frac{1}{2} - \frac{1}{2m}$. Thus, for small enough ε and η, we can lower bound

$$\Pr[\tilde{x}_{\pi_1(\ell(v_1))} < \tilde{y}_{\pi_2(\ell(v_2))} \mid u \notin A'] \geqslant \min\left\{(1-2\epsilon)(1-\eta), \frac{1}{2} - \frac{1}{2m}\right\} \geqslant \frac{1}{2} - \frac{1}{2m}. \quad (3)$$

Plugging (2) and (3) into (1), we get

$$\mathtt{Val}(\mathcal{G}) \geqslant c \cdot (1-2\epsilon)(1-\eta) + (1-c) \cdot \left(\frac{1}{2} - \frac{1}{2m}\right).$$

■

The following soundness of the reduction is shown in [GMR08].

Lemma 5.4. *(Soundness)[GMR08] If the Unique Games instance G has $\mathsf{val}(G) \leqslant \delta$ then $\mathtt{Val}(\mathcal{G}) \leqslant \frac{1}{2} + \eta + o_t(1) + \delta'$, where $\delta' \to 0$ as $\delta \to 0$.*

Proof of Theorem 1.4 For every $\varepsilon' < 0$, setting $\varepsilon, \eta, \delta > 0$ small enough constants and m large enough, in the completeness case we have a maximum acyclic subgraph of size at least $\frac{c}{2} + \frac{1}{2} - \varepsilon'$, whereas in the soundness case it is at most $\frac{1}{2} + \varepsilon'$. Since by Theorem 1.2, it is NP-hard to distinguish between $\mathsf{sval}(G) \geqslant \frac{1}{2} - \delta$ and $\mathsf{val}(G) \leqslant \delta$ we get that it is NP-hard to approximate the Maximum Acyclic Subgraph problem within a factor of $\frac{1/2+\varepsilon'}{1/4+1/2-\varepsilon'-\delta/2} \approx \frac{2}{3}$.

Remark 5.5. *Instead of $\mathsf{sval}(G) = \frac{1}{2}$, if we only have $\mathsf{val}(G) = \frac{1}{2}$, then the same construction and the labeling from Lemma 5.3 gives $\mathtt{Val}(\mathcal{G}) \geqslant \frac{5}{8}$. To see this, fix an assignment $\ell : A \cup B \to \Sigma$ which gives $\mathsf{val}(G) \geqslant \frac{1}{2}$. Let α_u denote the fraction of edges attached to u that are satisfied by ℓ. Therefore, we have $\mathsf{val}(G) = \mathbf{E}_{u \in A}[\alpha_u] = \frac{1}{2}$. Using a similar analysis as in the completeness case, we get $\mathtt{Val}(\mathcal{G}) \geqslant \mathbf{E}_{u \in A}[\alpha_u^2 \cdot (1-2\epsilon)] + \mathbf{E}_{u \in A}[(1-\alpha_u^2) \cdot \frac{1}{2}] \geqslant (1-2\epsilon)\,\mathbf{E}[\frac{1}{2} + \frac{\alpha_u^2}{2}]$. By Cauchy-Schwartz inequality $\mathbf{E}[\alpha_u^2] \geqslant (\mathbf{E}[\alpha_u])^2 = \frac{1}{4}$ and hence $\mathtt{Val}(\mathcal{G}) \geqslant (1-2\epsilon) \cdot \frac{5}{8}$. This along with the soundness lemma gives the NP-hardness of $\frac{4}{5}$.*

6 Predicates supporting Pairwise Independence

In this section, we prove Theorem 1.5.

6.1 The Austrin-Mossel reduction

Let $\mathcal{D}$ be a distribution on $P^{-1}(1)$ which is balanced and pairwise independent. For a string $x \in [q]^L$ and a permutation $\pi : [L] \to [L]$, define $x \circ \pi \in [q]^L$ such that $(x \circ \pi)_i = x_{\pi(i)}$ for all $i \in [L]$.

Let $G(A, B, E, [L], \{\pi_e\}_{e \in E})$ be an instance of Unique Games.

- Select $u \in A$ uniformly at random.
- Select k neighbors $\{v_1, v_2, \ldots, v_k\}$ of u uniformly at random. Let π_i be the constraints between (u, v_j) for all $j \in [k]$.
- Select $x^1, x^2, \ldots, x^k \in [q]^L$, such that for each $i \in [L]$ sample $(x_i^1, x_i^2, \ldots, x_i^k)$ independently as follows:
  - with probability $(1 - \varepsilon)$, $(x_i^1, x_i^2, \ldots, x_i^k)$ is sampled from the distribution $\mathcal{D}$.
  - with probability ε, $(x_i^1, x_i^2, \ldots, x_i^k)$ is sampled from $[q]^k$ uniformly at random.
- Output $((v_1, x^1 \circ \pi_1), (v_2, x^2 \circ \pi_2), \ldots, (v_k, x^k \circ \pi_k))$.

Figure 5: Reduction from UG to a P-CSP instance $\mathcal{I}$ from [AM09].

Let $G(A, B, E, [L], \{\pi_e\}_{e \in E})$ be an instance of Unique Games. We convert it into a P-CSP instance $\mathcal{I}$ as follows. The variable set is $B \times [q]^L$. The variable sets are *folded* in the sense that for every assignment $f : B \times [q]^L \to [q]$ to the variables, we enforce that for every $v \in B$, $x \in [q]^L$ and $\alpha \in [q]$,

$$f(v, x + \alpha^L) = f(v, x) + \alpha,$$

where additions are $(\texttt{mod}\ q)$.

The distribution on the constraints is given in Figure 5:

Lemma 6.1 (Completeness)**.** *If* $\mathsf{sval}(G) \geqslant c$*, the* $\mathcal{I}$ *is* $(c - \varepsilon)$*-satisfiable.*

Proof. Fix an assignment $\ell : A \cup B \to \Sigma$ which gives $\mathsf{sval}(G) \geqslant c$. Let $A' \subseteq A$ be the set of vertices such that all the edges incident on A' are satisfied by ℓ, we know that $|A'| \geqslant c \cdot |A|$. Thus with probability c, $u \in A'$ and all edges attached to it are satisfied by ℓ. Consider the following assignment f to the variables of $\mathcal{I}$: For a variable (v, x), we assign $f(v, x) = x_{\ell(v)}$.

Conditioned on $u \in A'$, we will show that $(f(v_1, x^1 \circ \pi_1), f(v_2, x^2 \circ \pi_2), \ldots, f(v_k, x^k \circ \pi_k)) \in P^{-1}(1)$ with probability $(1 - \varepsilon)$ and this will prove the lemma. Now, $(f(v_2, x^2 \circ \pi_2), \ldots, f(v_k, x^k \circ \pi_k)))$ is same as $((x^1 \circ \pi_1)_{\ell(v_1)}, (x^2 \circ \pi_2)_{\ell(v_2)}, \ldots, (x^k \circ \pi_k)_{\ell(v_k)})$, which in turns equals $(x^1_{\pi_1(\ell(v_1))}, x^2_{\pi_2(\ell(v_2))}, \ldots, x^k_{\pi_k(\ell(v_k))})$. Since ℓ satisfies all the edges (u, v_i), we have that for all $j \in [k]$, $\pi_j(\ell(v_j)) = \ell(u) =: i$ for some $i \in [L]$. Therefore we get $(x_i^1, x_i^2, \ldots, x_i^k)$, and according to the distribution, it belongs to $P^{-1}(1)$ with probability $(1 - \varepsilon)$. ■

We have the following soundness of the reduction.

Lemma 6.2 (Soundness [AM09]). *If the instance $\mathcal{I}$ is $\left(\frac{P^{-1}(1)}{q^k} + \eta\right)$-satisfiable, then G is $\delta := \delta(\eta, \varepsilon, k, q) > 0$ satisfiable.*

The completeness and soundness of the reduction, along with our main theorem, imply Theorem 1.5.

Acknowledgements. We would like to thank all the anonymous reviewers for their comments which helped in improving the presentation significantly.

References

[AKS11] Per Austrin, Subhash Khot, and Muli Safra. Inapproximability of Vertex Cover and Independent Set in Bounded Degree Graphs. *Theory of Computing*, 7(3):27–43, 2011.

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof Verification and the Hardness of Approximation Problems. *Journal of the ACM*, 45(3):501–555, May 1998. (Preliminary version in *33rd FOCS*, 1992).

[AM09] Per Austrin and Elchanan Mossel. Approximation Resistant Predicates from Pairwise Independence. *computational complexity*, 18(2):249–271, Jun 2009.

[AMW15] Per Austrin, Rajsekar Manokaran, and Cenny Wenner. On the NP-Hardness of Approximating Ordering-Constraint Satisfaction Problems. *Theory of Computing*, 11(10):257–283, 2015.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic Checking of Proofs: A New Characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998. (Preliminary version in *33rd FOCS*, 1992).

[BCK15] Boaz Barak, Siu On Chan, and Pravesh K Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 97–106. ACM, 2015.

[BGG18] N. Bansal, A. Gupta, and G. Guruganesh. On the Lovász Theta Function for Independent Sets in Sparse Graphs. *SIAM Journal on Computing*, 47(3):1039–1055, 2018.

[BGH+17] Amey Bhangale, Rajiv Gandhi, Mohammad Taghi Hajiaghayi, Rohit Khandekar, and Guy Kortsarz. Bi-Covering: Covering Edges with Two Small Subsets of Vertices. *SIAM J. Discrete Math.*, 31(4):2626–2646, 2017.

[BK10] Nikhil Bansal and Subhash Khot. Inapproximability of Hypergraph Vertex Cover and Applications to Scheduling Problems. In *Automata, Languages and Programming, 37th International Colloquium, ICALP 2010, Bordeaux, France, July 6-10, 2010, Proceedings, Part I*, pages 250–261, 2010.

[Cha16] Siu On Chan. Approximation resistance from pairwise-independent subgroups. *Journal of the ACM (JACM)*, 63(3):27, 2016.

[DKK+18a] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. On non-optimally expanding sets in Grassmann graphs. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 940–951. ACM, 2018.

[DKK+18b] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018*, pages 376–389, 2018.

[FGL+96] Uriel Feige, Shafi Goldwasser, Laszlo Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM (JACM)*, 43(2):268–292, 1996.

[GMR08] Venkatesan Guruswami, Rajsekar Manokaran, and Prasad Raghavendra. Beating the random ordering is hard: Inapproximability of maximum acyclic subgraph. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 573–582. IEEE, 2008.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing, STOC 2002*, pages 767–775. ACM, 2002.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[KMS17] Subhash Khot, Dor Minzer, and Muli Safra. On independent sets, 2-to-2 games, and Grassmann graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017*, pages 576–589, 2017.

[KMS18] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom Sets in Grassmann Graph have Near-Perfect Expansion. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018*, pages 592–601, 2018.

[KN09] Subhash Khot and Assaf Naor. Approximate kernel clustering. *Mathematika*, 55(1-2):129–165, 2009.

[KR08] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2- ε. *Journal of Computer and System Sciences*, 74(3):335–349, 2008.

[KTW14] Subhash Khot, Madhur Tulsiani, and Pratik Worah. A characterization of strong approximation resistance. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 634–643. ACM, 2014.

[KV15] Subhash A Khot and Nisheeth K Vishnoi. The Unique Games Conjecture, Integrality Gap for Cut Problems and Embeddability of Negative-Type Metrics into ℓ_1. *Journal of the ACM (JACM)*, 62(1):8, 2015.

[New01] Alantha Newman. The maximum acyclic subgraph problem and degree-3 graphs. In *Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques*, pages 147–158. Springer, 2001.

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 245–254. ACM, 2008.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il