

Optimal Inapproximability of Satisfiable k -LIN over Non-Abelian Groups

Amey Bhangale*

Subhash Khot[†]

Abstract

A seminal result of Håstad [Hås01] shows that for $k \geq 3$, it is NP-hard to find an assignment that satisfies $\frac{1}{|G|} + \varepsilon$ fraction of the constraints of a given k -LIN instance over an abelian group, even if there is an assignment that satisfies $(1 - \varepsilon)$ fraction of the constraints, for any constant $\varepsilon > 0$. Engebretsen, Holmerin, and Russell [EHR04] later showed that the same hardness result holds for k -LIN instances over any finite non-abelian group.

Unlike the abelian case, where we can efficiently find a solution if the instance is satisfiable, in the non-abelian case, it is NP-complete to decide if a given system of linear equations is satisfiable or not, as shown by Goldmann and Russell [GR02].

Surprisingly, for certain non-abelian groups G , given a satisfiable k -LIN instance over G , one can in fact do better than just outputting a random assignment using a simple but clever algorithm. The approximation factor achieved by this algorithm varies with the underlying group. In this paper, we show that this algorithm is *optimal* by proving a tight hardness of approximation of a satisfiable k -LIN instance over *any* non-abelian G , assuming $\mathbf{P} \neq \mathbf{NP}$. As a corollary, we also get 3-query probabilistically checkable proofs with perfect completeness over large alphabets with improved soundness.

Our proof crucially uses structural properties captured by Gowers' notion of *quasi-randomness* for non-abelian groups [Gow08].

1 Introduction

Constraint satisfaction problems (CSPs) are the most fundamental problems in computer science. One of the simplest such CSPs that we know how to solve is a system of k -LIN equations over some abelian group. More generally, an instance of Max- k -LIN over a group $G = (G, \cdot)$, not necessarily abelian, consists of a set of variables $x_1, x_2, \dots, x_n$ and a set of constraints $C_1, C_2, \dots, C_m$. Each C_i is a linear equation involving k variables, for example $a_1 \cdot x_{i_1} \cdot a_2 \cdot x_{i_2} \cdot \dots \cdot a_k \cdot x_{i_k} = b$, for some group elements $a_1, a_2, \dots, a_k, b \in G$. The task is to find an assignment to the variables that satisfies as many constraints as possible.

For *any* abelian group G , if there is a perfect solution to the given Max- k -LIN instance over G , then it can be found efficiently [GR02] in polynomial time by solving modular linear systems. A

*Department of Computer Science and Engineering, University of California, Riverside, CA, USA. Email: ameyrb@ucr.edu

[†]Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, NY, USA. Email: khot@cs.nyu.edu

given instance is *almost* satisfiable if there exists an assignment that satisfies $(1 - \varepsilon)$ -fraction of the constraints for a small constant $\varepsilon > 0$. If the given instance of Max- k -LIN over an abelian group G is almost satisfiable, then Håstad [Hås01] showed that it is **NP**-hard to even find an assignment that satisfies a $\frac{1}{|G|} + \varepsilon$ fraction of the constraints for every constant $\varepsilon > 0$. In other words, one cannot do significantly better than just outputting a random assignment.

The situation changes completely if the instance is a set of linear equations over a non-abelian group. In this case, Goldmann and Russell [GR02] showed that even the problem of deciding if a given instance is satisfiable or not is **NP**-complete for *every* non-abelian group.

An algorithm (folklore): It turns out that for certain groups, one can do much better than assigning a random assignment when the instance is satisfiable. Given an instance ϕ over G , consider an instance ϕ' over $H = G/[G, G]$ where $[G, G]$ is a commutator subgroup of G , i.e., the subgroup generated by the elements $\{g^{-1}h^{-1}gh \mid g, h \in G\}$. The instance ϕ' is the same as ϕ except that all group constants are replaced by their equivalence class in $G/[G, G]$. The important property of this quotient group H is that it is an abelian group. Since ϕ has a satisfying assignment over G , ϕ' has a satisfying assignment over H . Hence, we can find the satisfying assignment σ of ϕ' in polynomial time. The solution σ is an assignment of cosets of $[G, G]$ to the variables. We construct a random assignment to ϕ such that for every variable x , we select a random group element from $\sigma(x)$ and assign it to x . It is easy to see that each constraint is satisfied with a probability equal to $\frac{1}{|[G, G]|}$. Thus, this gives an assignment that satisfies $\frac{1}{|[G, G]|}$ fraction of the constraints in expectation. Therefore, if $[G, G]$ is a proper subgroup of G , then we get an algorithm that does better than the random assignment threshold.

If the instance is almost satisfiable, then it is not clear how to modify the above algorithm for almost satisfiable instances to get better than $\frac{1}{|G|}$ approximation. In fact, for almost satisfiable instances over any non-abelian group, Engebretsen, Holmerin, and Russell [EHR04] showed that it is **NP**-hard to do better than outputting a random assignment.

This leaves an intriguing question of finding the correct approximation threshold for satisfiable instances over non-abelian groups. In this paper, we show that the above-described algorithm for satisfiable instances over non-abelian groups is the best one can hope for. More specifically, we prove the following theorem.

Theorem 1.1. *For any constant $\varepsilon > 0$, given a satisfiable instance of Max-3-LIN over a finite non-abelian group G , it is **NP**-hard to find an assignment that satisfies $\frac{1}{|[G, G]|} + \varepsilon$ fraction of the constraints.*

The theorem can be extended to Max- k -LIN for any $k \geq 3$ to imply a similar hardness result for all Max- k -LIN problems over G .

If G is a *perfect group*, i.e., $[G, G] = G$, then Theorem 1.1 implies an NP-hardness of approximating satisfiable Max-3-CSP instances over an alphabet of size $q = |G|$ to within a factor of $\frac{1}{q} + \varepsilon$, for every constant $\varepsilon > 0$. As a direct consequence, we get improved soundness of 3-query probabilistically checkable proofs (PCPs) with perfect completeness over large alphabets. Since

PCPs are not the main focus of this paper, we refer interested readers to the book by Arora and Barak [AB09, Chapter 18] to see the relation between PCPs and CSPs.

Corollary 1.2. *For infinitely many $q \in \mathbb{Z}^+$, any language in **NP** is decided by a nonadaptive PCP with answers from a domain of size q that queries three positions in the proof, has perfect completeness and soundness $\frac{1}{q} + \varepsilon$ for any constant $\varepsilon > 0$.*

This improves a result by Engebretsen and Holmerin [EH05] where they constructed PCPs with soundness $\frac{1}{q} + \frac{1}{q^2} + \varepsilon$, and also a result by Tang [Tan09] in which they showed a conditional result with soundness $\frac{1}{q} + \frac{1}{q^2} - \frac{1}{q^3} + \varepsilon$, for any constant $\varepsilon > 0$.¹

1.1 Techniques

We assume some familiarity with the Fourier analysis of functions over abelian groups (for instance, Chapter 8 of Ryan O’Donnell’s book [O’D14]). Throughout the section, $\varepsilon > 0$ is an arbitrarily small constant and $\delta(\varepsilon) > 0$ decays with ε . We only discuss 3-LIN here, however the argument is similar for k -LIN in general.

Given the body of work on hardness of approximation – where many reductions begin from the Label Cover problem (see Definition 2.1) – a central ingredient in these reductions is the use of *dictatorship tests*. A function $f : G^n \rightarrow G$ is a dictator function if it depends only on one variable in the following way: $f(x_1, x_2, \dots, x_n) = x_i$ for some $i \in [n]$. On the other hand, we have functions which are *far* from dictator functions.

To understand a notion of distance from a dictator function, which is useful for a reduction to work, we first consider a few reasonable notions and finally arrive at the right one. Consider a scalar representation $h : G^n \rightarrow \mathbb{C}$ of a G -valued function f by having a one-to-one map $\tau : G \rightarrow \{1, 2, \dots, |G|\}$ and defining $h(g) = \tau(f(g))$. Define the influence of the i^{th} coordinate of a function h to be the expected variance of h when the i^{th} coordinate is sampled uniformly from G , with all other coordinates held fixed. The expectation is taken over uniformly random choices of the other coordinates. Analytically, in terms of the Fourier coefficients of h , the i^{th} influence is equal to the following quantity:

$$\mathbf{Inf}_i(h) := \sum_{\alpha: \alpha_i \neq 0} |\hat{h}(\alpha)|^2.$$

Thus, the i^{th} dictator function has $\mathbf{Inf}_i(h) = \Omega_G(1)$ as a function of G independent of n , and every other influences 0. At the first attempt, it might make sense to define functions that are close to dictators as those with a coordinate that has a large influence. However, for a prime p and $G \cong \mathbb{Z}_p$, note that there are linear functions $\ell_S = \sum_{i \in S} \beta_i x_i$ where $S \subseteq [n]$ such that ℓ_S has all variables $i \in S$ with influences $\Omega(1)$. We would like to isolate these functions with large $|S|$ from dictator functions. This motivates us to define a more refined notion of a low-degree influence of a variable i . For a function $h : G^n \rightarrow \mathbb{C}$, define its degree- d influence of the i^{th} coordinate as

$$\mathbf{Inf}_i^{\leq d}(h) := \sum_{\alpha: \alpha_i \neq 0 \wedge |\alpha| \leq d} |\hat{h}(\alpha)|^2,$$

¹The theorem in [EH05] holds for every $q \geq 3$, and the theorem in [Tan09] holds for every $q \geq 4$. Our theorem holds for q such that there is a perfect group of order q .

where $|\alpha|$ is the number of non-zero coordinates of α . Thus, for the scalar encoding of the i^{th} dictator function, its low-degree ($d = 1$) influence of the coordinate i is $\Omega(1)$ (and the rest of the influences are 0). We consider a function to be far from any dictator function if all the low-degree influences, for some $d = O(1)$ that is independent of n , of the function are small, say at most ε .

Although the above definition is the right definition for most reductions, for the hardness results of Max- k -LIN, we work with an even weaker notion of the distance. We consider the following definition. We always assume the function is *folded*, i.e., it satisfies $f(ax) = af(x)$ for all $a \in G$ and $x \in G^n$.² A function is far from dictator functions if for every α such that $|\alpha| = O(1)$ which is independent of n , $|\hat{h}(\alpha)|^2 \leq \varepsilon$. In other words, all *low-degree* Fourier coefficients of h have small weights. Note that this notion isolates ℓ_S with large $|S|$ from dictators.

A (non-adaptive) dictatorship test queries the function f at a few locations and, based on the values it sees, decides whether the function is a dictator function or far from it. The choice of the accepting criterion for the test is tightly connected to the specific constraint satisfaction problem (CSP) for which we want to show an NP-hardness result. Furthermore, the gap between the test passing probability in the *completeness case* (when f is a dictator function) and the *soundness case* (when f is far from any dictator function) translates into the inapproximability factor of the NP-hardness of the CSP.³

1.1.1 Abelian Groups

Let us look at a candidate dictatorship test where the predicate is a linear equation in 3 variables over an abelian group $G \cong \mathbb{Z}_q$. Here, 0 is the identity element of G .

- Select $x, y \sim G^n$ uniformly at random.
- Set $z = x + y$.
- Check if $f(x) + f(y) = f(z)$.

It is clear that if f is an i^{th} dictator, then the test passes with probability 1. The non-trivial thing is to analyze the test passing probability if f is far from any dictator. It is easy to see that there are functions that are far from dictators, and still the test passes with probability 1 on them. A family of such functions is the linear functions of the form $\ell_S = \sum_{i \in S} \beta_i x_i$ where $S \subseteq [n]$ and $\beta_i \in G \setminus \{0\}$, with large $|S|$. It is not hard to see that these functions pass the test with probability 1. In fact, Blum, Luby, and Rubinfeld [BLR93] showed that this is a good test for the linear functions (instead of the dictator functions).

To prove the hardness of the approximation for Max-3-LIN over abelian groups, one needs to design a test in which ℓ_S , for large $|S|$, passes with only a small probability. To design such a test, Håstad [Hås01] introduced the so-called *noise* to each coordinate. The modified test is as follows:

²Such a condition can be enforced in the reduction because we allow group constants in each constraint.

³This is not totally correct as one has to overcome other important issues when such a test is used in the actual reduction, starting with the Label Cover instance.

- Select $x, y \sim G^n$ uniformly at random.
- Set $z = x + y$.
- For each $i \in [n]$, resample (x_i, y_i, z_i) from G^3 uniformly at random, with probability ε .
- Check if $f(x) + f(y) = f(z)$.

This noise takes care of the earlier mentioned counterexamples, i.e., functions ℓ_S for large S now pass the test with probability roughly $\frac{1}{|G|}$. In general, Håstad [Hås01] showed that if f is far from dictator functions, then the test passes with probability at most $\frac{1}{|G|} + \delta$ for a small constant $\delta > 0$. The proof of this statement uses Fourier analysis over abelian groups. Note that this bound is optimal as even a random function passes the test with probability $\frac{1}{|G|}$.

However, now the guarantee in the completeness case is no longer the same. We only get that dictator functions pass this test with probability $(1 - \varepsilon) + \frac{\varepsilon}{|G|}$ (instead of 1). This gap in the test passing probability is translated into the NP-hardness of 3-LIN over abelian groups and coincidentally, in this abelian case, the NP-hardness result is *optimal*. More precisely, given a system of linear equations over an abelian group G , where each equation involves 3 variables, it is NP-hard to distinguish between the cases when there exists an assignment that satisfies at least $(1 - \varepsilon)$ -fraction of the constraints vs. no assignment can satisfy more than $\frac{1}{|G|} + \delta$ fraction of the constraints.

1.1.2 Non-Abelian Groups

We now look into the non-abelian case. Since we would like to design a test which passes with probability 1 (or $(1 - \varepsilon)$) in the completeness case, there is a natural generalization of the above-mentioned tests to a non-abelian group G . Here we denote the group operation by the symbol $\cdot$ and the identity element of G by 1_G .

We first describe the test with completeness $(1 - \varepsilon)$, which is similar to the test over the abelian group with noise that we described earlier.

- Select $x, y \sim G^n$ uniformly at random.
- For each $i \in [n]$, set $z_i = y_i^{-1} \cdot x_i^{-1}$.
- For each $i \in [n]$, resample (x_i, y_i, z_i) from G^3 uniformly at random, with probability ε .
- Check if $f(x) \cdot f(y) \cdot f(z) = 1_G$.

The analysis of this test is implicit in the work of Engebretsen, Holmerin, and Russell [EHR04]. Firstly, it is easy to see that the dictator functions pass this test with probability $(1 - \varepsilon) + \frac{\varepsilon}{|G|}$. The soundness of this test is analyzed in [EHR04] where the authors show that in the soundness

case, the test passes with probability at most $\frac{1}{|G|} + \delta$ for a small constant $\delta > 0$. Their proof goes via Fourier analysis over non-abelian groups. As in the abelian case, in this case also, it can be shown that the noise takes care of *high degree* Fourier terms.⁴ This implies NP-hardness result of approximating 3-LIN instances over a non-abelian group, which is similar to the abelian case.

Although the proof of the soundness of the test in [EHR04] uses representation theory and Fourier analysis of functions on non-abelian groups, the proof now follows from the more general statement called the *invariance principle* of Mossel [Mos10]. The distribution on the tuple (x, y, z) is a product distribution $\mu^{\otimes n}$ where μ is a distribution on (x_i, y_i, z_i) (note that for all i it is the same distribution). Since we add noise to each coordinate with some non-zero probability, the distribution μ is *connected* and hence we can easily take care of high-degree functions in the analysis. Furthermore, the distribution μ is pairwise independent. These two conditions, along with the fact that f has small low-degree influences, imply that the test passes with probability at most $\frac{1}{|G|} + \delta$. This statement is implicit in the result of Austrin and Mossel [AM09].

We now state the (obvious) dictatorship test with perfect completeness.

- Select $x, y \sim G^n$ uniformly at random.
- For each $i \in [n]$, set $z_i = y_i^{-1} \cdot x_i^{-1}$.
- Check if $f(x) \cdot f(y) \cdot f(z) = 1_G$.

Our main contribution is the soundness analysis of the above dictatorship test over non-abelian groups *without noise*. Note that we cannot use the invariance principle-based techniques in this case, as the distribution does not satisfy the condition of *connectedness*.

Proof overview. Our proof of the soundness analysis is inspired by the magic that was discovered by Gowers [Gow08] to show that there are non-abelian groups where the size of any product free set is sublinear in $|G|$.⁵ Gowers' trick applied only to quasi-random groups,⁶ as the goal was to obtain an $o(|G|)$ bound on product-free sets, whereas our reduction applies to every non-abelian group.

The trick is elegantly captured by the following inequality due to Babai, Nikolov, and Pyber [BNP08]. For any functions $f, g : G \rightarrow \mathbb{C}$ with at least one of f, g having mean zero:

$$\|f * g\|_{L^2(G)} \leq \frac{1}{\sqrt{D}} \|f\|_{L^2(G)} \|g\|_{L^2(G)}, \quad (1)$$

where D is the smallest dimension of a non-trivial representation of G ⁷ while a trivial application of the Cauchy-Schwarz inequality gives an upper bound of $\|f\|_{L^2(G)} \|g\|_{L^2(G)}$. Thus, Equation (1) has a multiplicative improvement of a factor $\frac{1}{\sqrt{D}}$ over a trivial upper bound.

⁴Although we have not formally defined a degree of a Fourier coefficient of a function in this non-abelian setting, think of it as the number of non-trivial irreducible representations in α (See Proposition 2.29).

⁵unlike the abelian case where one can always find, in this case, a 'sum-free' set of size $\Omega(|G|)$.

⁶A group G is called *quasi-random* if the smallest dimension of any non-trivial representation of G is large.

⁷see Definition 2.14 and Definition 2.19 for the definitions of $\|\cdot\|_{L^2(G)}$ and the convolution operator $*$, and Section 2.2.1 for representation theory;

Returning to analyzing the soundness of our test, its analysis boils down to analyzing the following expression.

$$\begin{aligned} |\mathbf{E}[g_1(x)g_2(y)g_3(z)]| &= |\mathbf{E}[(g_1 * g_2 * g_3)(1_{G^n})]| \\ &\leq \sum_{\alpha \in \text{Irrep}(G^n)} \dim(\alpha) \cdot \|\hat{g}_1(\alpha)\|_{\text{HS}} \cdot \|\hat{g}_2(\alpha)\|_{\text{HS}} \cdot \|\hat{g}_3(\alpha)\|_{\text{HS}}, \end{aligned} \quad (2)$$

where $g_i : G^n \rightarrow \mathbb{C}$ are bounded balanced functions obtained from f (so $\|g_i\|_2 \leq 1$). Here, $\hat{g}_i(\alpha)$ denotes the Fourier coefficient of g_i associated with the irreducible representation α of G^n , and $\|\cdot\|_{\text{HS}}$ denotes the Hilbert–Schmidt norm of a matrix. The stated inequality follows by expanding $(g_1 * g_2 * g_3)$ in the Fourier basis and applying the triangle inequality.

Once we have this expression, similar to Equation (1), it is easy to bound the terms with large $\dim(\alpha)$: By applying the Cauchy-Schwarz inequality and using Parseval’s identity, we can show the following:

$$\sum_{\dim(\alpha) \geq D} \dim(\alpha) \cdot \|\hat{g}_1(\alpha)\|_{\text{HS}} \cdot \|\hat{g}_2(\alpha)\|_{\text{HS}} \cdot \|\hat{g}_3(\alpha)\|_{\text{HS}} \leq \frac{1}{\sqrt{D}} \|g_1\|_2 \cdot \|g_2\|_2 \cdot \|g_3\|_2 \leq \frac{1}{\sqrt{D}}. \quad (3)$$

Thus, we can effectively bound the *higher dimensional* terms in the expression. Therefore, if the original expectation is δ , then we essentially get

$$\sum_{\dim(\alpha) \leq D} \dim(\alpha) \cdot \|\hat{g}_1(\alpha)\|_{\text{HS}} \cdot \|\hat{g}_2(\alpha)\|_{\text{HS}} \cdot \|\hat{g}_3(\alpha)\|_{\text{HS}} = \Omega(\delta),$$

for $D = \Theta(1/\delta^2)$. Taking the maximum $\|\hat{g}_1(\alpha)\|_{\text{HS}}$ out in the summation, the remaining sum can be upper-bounded by 1. Therefore, we get that there exists an α such that the dimension of α is at most D and $\|\hat{g}_1(\alpha)\|_{\text{HS}} = \Omega(\delta)$.

Our analysis shows that if the test passes with probability greater than $\frac{1}{|G,G|} + \delta$, then there exists an α such that $\dim(\alpha) \leq O_\delta(1)$ and $\|\hat{f}(\alpha)\|_{\text{HS}} \approx \delta$, for some function $\hat{f} : G^n \rightarrow \mathbb{C}$ derived from f . Note that this conclusion is different from what we had aimed for, i.e., concluding that there exists a *low-degree* Fourier coefficient with large magnitude. However, we show that such a bound is enough to carry out the actual soundness analysis of the reduction.

Since we do not introduce noise to each coordinate, there are functions with the Fourier mass concentrated on a large dimensional α that pass this test with probability $\frac{1}{|G,G|}$. Thus, our analysis of the test is also optimal.

Although the dictatorship test works, there are many complications that arise when we compose this test with the Label Cover instance. We briefly discuss three issues here and elaborate on the last one in this proof overview:

1. As observed before, in the soundness analysis, we conclude that there is a low-dimensional Fourier coefficient whose norm is large, if the test passes with non-trivial probability. However, there are terms with dimension 1 but with a high degree, which are problematic for the final *decoding strategy* in our reduction. In [EHR04], these problematic terms were handled by adding noise, a technique similar to the Abelian case. In our case, we do not have the

noise. However, in Lemma 2.30, we observe a stronger property of the folded functions.⁸ Namely, if f is folded, then the function $\rho(f(x))_{ij}$, where ρ is any irreducible representation of G of dimension at least 2, has all Fourier coefficients with dimension 1 zero. Thus, we can focus on terms with dimensions at least 2.

2. Our decoding strategy is different from that in [EHR04]. The decoding strategy in [EHR04] is based on non-empty low-degree Fourier coefficients, which was similar to Håstad's decoding strategy. In our reduction, we slightly changed the decoding strategy — it is based on Fourier coefficients whose dimension is at least 2, but they can be of high degree. This condition is forced on us by the way we handle the higher-dimensional terms. Fortunately, the decoding strategy works without much trouble.
3. Finally, due to the d -to-1 nature of the projection constraints in the LABEL-COVER instance, the three queries tend to be more correlated to each other. Thus, one has to take care of many potential scenarios in the actual reduction when the error term can be large. We handle this by starting with a variant of the LABEL-COVER problem called the *layered* LABEL-COVER whose NP-hardness was shown in [DGKR05]. The reason we had to work with a layered LABEL-COVER is somewhat technical and here, we try to shed some light on it.

Controlling the terms with large dimension. In the dictatorship test, large representation dimension gives us a useful saving almost automatically. After composing the test with Label Cover, however, a large-dimensional representation can break into small-dimensional pieces, and we must find another way to control their contribution.

To see where this difficulty arises, fix a d -to-1 projection $\pi : [dn] \rightarrow [n]$ and two folded tables $f_1 : G^{dn} \rightarrow G$ and $f_2 : G^n \rightarrow G$. Choose independent uniform inputs $\mathbf{a} \in G^{dn}$ and $\mathbf{b} \in G^n$, set $c_i = b_{\pi(i)}^{-1} a_i^{-1}$, and check whether

$$f_1(\mathbf{a}) \cdot f_2(\mathbf{b}) \cdot f_1(\mathbf{c}) = 1_G.$$

Consistent dictators pass this test with probability 1: if $f_1(\mathbf{x}) = x_i$ and $f_2(\mathbf{y}) = y_{\pi(i)}$, the three answers multiply to the identity.

Each queried input is still marginally uniform on its own domain. The difficulty comes from the lifted tuple $\mathbf{b} \circ \pi$, which is uniform on the subgroup

$$S_\pi := \{\mathbf{x} \circ \pi : \mathbf{x} \in G^n\} \leq G^{dn}.$$

The Fourier calculation therefore involves restricting representations of G^{dn} to S_π . An irreducible representation α of large dimension may become reducible under this restriction and contain irreducible constituents β of very small dimension. Thus, the dimension of α alone no longer gives the saving we used in the dictatorship test.

Here is the expression that captures this obstruction. Let $g_1 : G^{dn} \rightarrow \mathbb{C}$ and $g_2 : G^n \rightarrow \mathbb{C}$ be the 1-bounded matrix-coefficient functions derived from the two tables. For $\alpha \in \text{Irrep}(G^{dn})$ and $\beta \in \text{Irrep}(G^n)$, let $P_{\alpha, \pi, \beta}$ be the orthogonal projection onto the sum of all copies of β in the restriction of α to $S_\pi \cong G^n$, with the zero projection if β does not occur. Multiplying $\hat{g}_1(\alpha)$ by

⁸A function f is called folded if $f(c \bullet x) = c \bullet f(x)$ for all $x \in G^n$ and $c \in G$

this projection projects each of its columns onto that subspace. After applying Cauchy–Schwarz, the constituents with $\dim(\beta) > D$ contribute at most $1/D$ to the bound on the squared high-dimensional contribution. The remaining quantity is

$$\Delta := \sum_{\substack{\beta \in \text{Irrep}(G^n) \\ \dim(\beta) \leq D}} \|\hat{g}_2(\beta)\|_{\text{HS}}^2 \sum_{\substack{\alpha \in \text{Irrep}(G^{dn}) \\ \dim(\alpha) \geq D}} \dim(\alpha) \|P_{\alpha, \pi, \beta} \hat{g}_1(\alpha)\|_{\text{HS}}^2. \quad (4)$$

The inner sum measures how much Fourier mass from large-dimensional α 's lands in copies of β . The outer factor measures how strongly the other table uses that same β . Consequently, Δ can be large when the two tables have substantial Fourier mass on matching small-dimensional pieces, even though the original α 's have large dimension.

Our way of handling this possibility is to use many layers. We start from the layered Label Cover construction of [DGKR05], in which projections along a path compose consistently. Sample a uniform path $p = (v_1, \dots, v_\ell)$ through the layers, choose a uniform pair $k < k'$, and apply the test above to the tables at v_k and $v_{k'}$. The structural properties of the construction ensure that this samples a uniform constraint between the chosen layers. Writing $\Delta_{p,k,k'}$ for the corresponding quantity in (4), we show that a large average $\mathbb{E}_{p,k,k'}[\Delta_{p,k,k'}]$ would give a labeling satisfying a nontrivial fraction of the Label Cover constraints.

The geometric idea is simple: a vector has only a bounded amount of squared norm to distribute among mutually orthogonal subspaces. Fix a path and an initial layer k . As we move to later layers, the corresponding subgroups of G^{R_k} form a decreasing chain. We can therefore choose the individual irreducible-copy subspaces successively, so that each block at a later layer lies inside a block at every earlier layer. This lets us follow the same Fourier directions through all the later layers.

Suppose the average contribution is at least $\eta > 0$. Representations β for which $\|\hat{g}_2(\beta)\|_{\text{HS}}^2$ is much smaller than η account for only a small part of this contribution. Call the remaining representations *heavy*. By Parseval, there are only $O_\eta(1)$ heavy representations at each vertex. If the subspaces corresponding to heavy representations at different later layers were always orthogonal, their contributions would have to share the bounded Fourier mass of the initial table. With sufficiently many layers, their average would then be small. A quantitative version of this observation shows that, for some fixed pair of later layers, heavy components overlap on a nontrivial fraction of paths.

Finally, this overlap gives consistent labels. For a representation $\beta = \bigotimes_j \rho_j$, consider the coordinates j for which $\dim(\rho_j) \geq 2$. Since $\dim(\beta) \leq D$, there are at most $\log_2 D$ such coordinates. Folding ensures that this set is nonempty for every heavy representation used in the argument. Moreover, when a later block lies inside an earlier block, the later representation is a constituent of the restriction of the earlier one. Every coordinate of dimension at least two in the later representation must therefore be the image, under the Label Cover projection, of such a coordinate in the earlier representation.

Taking the union of these coordinate sets over the heavy representations gives each vertex a short list of candidate labels. The overlapping components guarantee that these lists contain a consistent pair on a nontrivial fraction of constraints. Choosing labels independently and uniformly from the lists therefore satisfies an $\Omega_{\eta,D}(1)$ fraction of the constraints between two layers, contradicting the soundness of the layered Label Cover instance.

2 Preliminaries

2.1 Label Cover

We start by defining the LABEL-COVER problem.

Definition 2.1 (LABEL-COVER). An instance $\mathcal{H} = (\mathcal{U}, \mathcal{V}, E, [L], [R], \{\pi_e\}_{e \in E})$ of the LABEL-COVER constraint satisfaction problem consists of a bi-regular bipartite graph $(\mathcal{U}, \mathcal{V}, E)$, two alphabets $[L]$ and $[R]$ and a surjective projection map $\pi_e : [R] \rightarrow [L]$ for every edge $e \in E$. Given a labeling $\ell_{\mathcal{U}} : \mathcal{U} \rightarrow [L], \ell_{\mathcal{V}} : \mathcal{V} \rightarrow [R]$, an edge $e = (u, v)$ is said to be satisfied by ℓ if $\pi_e(\ell_{\mathcal{V}}(v)) = \ell_{\mathcal{U}}(u)$.

$\mathcal{H}$ is said to be satisfiable if there exists a labeling that satisfies all the edges. $\mathcal{H}$ is said to be at most δ -satisfiable if every labeling satisfies at most a δ fraction of the edges.

The hardness of LABEL-COVER stated below follows from the PCP Theorem [AS98, ALM⁺98, FGL⁺96] and Raz's Parallel Repetition Theorem [Raz98].

Theorem 2.2 (Hardness of LABEL-COVER). For every $r \in \mathbb{N}$, there is a deterministic $n^{O(r)}$ -time reduction from a 3-SAT instance of size n to an instance $\mathcal{H} = (\mathcal{U}, \mathcal{V}, E, [L], [R], \{\pi_e\}_{e \in E})$ of LABEL-COVER with the following properties:

1. $|\mathcal{U}|, |\mathcal{V}| \leq n^{O(r)}; L, R \leq 2^{O(r)}; \mathcal{H}$ is bi-regular with degrees bounded by $2^{O(r)}$.
2. There is a constant $s_0 \in (0, 1)$ such that,
 - YES Case: If the 3-SAT instance is satisfiable, then $\mathcal{H}$ is satisfiable.
 - NO Case: If the 3-SAT instance is unsatisfiable, then $\mathcal{H}$ is at most $2^{-s_0 r}$ -satisfiable.

For our hardness result, we need the following variant of the LABEL-COVER problem.

Definition 2.3 (Layered Label Cover). An ℓ -layered Label Cover instance, given by $\mathcal{H} = (\mathcal{V} = \{V_1, \dots, V_\ell\}, \{\Pi_{i,j}\}_{1 \leq i < j \leq \ell}, \{[R_i]\}_{i \in [\ell]})$ consists of ℓ sets of vertices $\mathcal{V} = \{V_1, \dots, V_\ell\}$. The label set of vertices in layer i is denoted by $[R_i]$. Every pair of layers $1 \leq i < j \leq \ell$ has a set of constraints $\Pi_{i,j}$ between the vertices in V_i and V_j . The constraint between $v \in V_i$ and $u \in V_j$ (if one exists in $\Pi_{i,j}$) is denoted by $\pi_{vu} : [R_i] \rightarrow [R_j]$. Moreover, every constraint between a pair of vertices is a projection constraint – for every assignment $k \in [R_i]$ to $v \in V_i$ there is a unique assignment to $u \in V_j$ that satisfies the constraint π_{vu} .

We have the following NP-hardness result from [DGKR05], [DRS05], which we use as a starting point in proving Theorem 1.1. The additional structural properties mentioned in the theorem statement can be derived easily from their construction. For completeness, we include the proof of this in the appendix.

Theorem 2.4 ([DGKR05], [DRS05]). For any constant parameters $\ell \geq 2, r \in \mathbb{N}$ the following problem is NP-hard. Given an ℓ -layered Label Cover instance $\mathcal{H} = (\mathcal{V} = \{V_1, \dots, V_\ell\}, \{\Pi_{i,j}\}_{1 \leq i < j \leq \ell}, \{[R_i]\}_{i \in [\ell]})$ where all variable ranges $[R_i]$ are of size $R_i \leq 2^{O(\ell r)}$, distinguish between the following two cases:

Completeness. There is an assignment satisfying all the constraints of the Label Cover instance. In this case, we say that $\mathcal{H}$ is fully satisfiable.

Soundness. For every $1 \leq i < j \leq \ell$, no assignment satisfies more than a 2^{-r} fraction of the set of constraints $\Pi_{i,j}$ between layers i and j . In this case, we say that $\mathcal{H}$ is at most 2^{-r} -satisfiable.

Furthermore, the above hardness result also holds for instances with the following additional structure.

1. **Transitivity of the constraints.** The set of constraints $\Pi_{i,i+1}$ for $1 \leq i \leq \ell - 1$ define all the other constraints by transitivity. In other words, fix any three layers $1 \leq i < j < k \leq \ell$. For any two constraints $\pi_{vu} \in \Pi_{i,j}$ and $\pi_{uw} \in \Pi_{j,k}$, there is a constraint $\pi_{vw} \in \Pi_{i,k}$ such that $\pi_{vw}(a) = \pi_{uw}(\pi_{vu}(a))$ for all $a \in [R_i]$.
2. **Regularity of the constraints.** The constraint graph $G_{i,j}$ is biregular for every $1 \leq i < j \leq \ell$. Moreover, the following distributions on constraints are identical:
 - (a) Choose a layer pair $i < j$ uniformly, then a constraint uniformly from $\Pi_{i,j}$.
 - (b) Choose a full consecutive-layer chain $(x_1, \dots, x_\ell)$ uniformly, independently choose a layer pair $i < j$ uniformly, and output its induced constraint (x_i, x_j) .
3. **Regular projection constraints.** There exists a constant d such that $R_{i-1} = dR_i$ for all $2 \leq i \leq \ell$ and each constraint $\pi_{vu} \in \Pi_{i,j}$ is a $d^{(j-i)}$ -to-1 projection constraint.

2.2 Fourier analysis

In this section, we give a brief overview of the representation theory of non-abelian groups and Fourier analysis over non-abelian groups. For a more comprehensive understanding, we refer the reader to the book by Terras [Ter99]. We state many propositions in the following subsection, and the proofs of these propositions can be found in the same book [Ter99].

2.2.1 Representation Theory

In this paper, we only consider non-abelian groups which are *finite*. Let $G = (G, \cdot)$ be a finite non-abelian group. The identity element of a group is denoted by 1_G .

Definition 2.5. A representation (V, ρ) of G is a vector space V together with a group homomorphism $\rho : G \rightarrow \text{GL}(V)$ from G to the group $\text{GL}(V)$ of invertible $\mathbb{C}$ -linear transformations from V to V . The dimension of the vector space V is denoted by $\dim(\rho)$.

For convenience, we just use the letter ρ to denote a representation of G and use ρ_V to denote the underlying vector space. We view a representation $\rho(\cdot)$ as its corresponding matrix of the linear transformation. Thus $\rho(\cdot)_{ij}$ is used to denote the $(i, j)^{\text{th}}$ entry of that matrix. We always work with representations that are unitary. There is one representation which is obvious – just map everything to $1 \in \mathbb{C}$. This representation is called the *trivial representation*, which has dimension 1. We will denote the trivial representation by $\{1\}$.

Definition 2.6. Let ρ and τ be representations of G . An isomorphism from ρ_V to τ_V is an invertible linear transformation $\phi : \rho_V \rightarrow \tau_V$ such that

$$\phi \circ \rho(g) = \tau(g) \circ \phi,$$

for all $g \in G$. We say that ρ_V and τ_V are isomorphic and write $\rho_V \cong \tau_V$ if there exists an isomorphism from ρ_V to τ_V .

Definition 2.7. Let ρ be a representation of G . A vector subspace $W \subset \rho_V$ is G -invariant if $\rho(g)w \in W$ for all $g \in G$ and $w \in W$.

If a representation (V, ρ) has a G -invariant subspace W other than $\{0\}$ and V itself, then the action on W itself is a representation of G . This leads to the following important definition of irreducible representations.

Definition 2.8. A representation ρ of G is irreducible if $\rho_V \neq \{0\}$ and ρ_V has no G -invariant subspaces other than $\{0\}$ and ρ_V .

We will denote the set of all irreducible representations of G up to isomorphism by $\text{Irrep}(G)$.

Fact 2.9. Let G be a group and H be any subgroup of G . If $\rho \in \text{Irrep}(G)$ then ρ restricted to H is also a (not necessarily irreducible) representation of H .

Definition 2.10. The tensor product of two representations ρ and τ of a group G is the representation $\rho \otimes \tau$ on $\rho_V \otimes \tau_V$ defined by the condition

$$(\rho \otimes \tau)(g)(v \otimes w) = \rho(g)(v) \otimes \tau(g)(w),$$

and extended to all vectors in $\rho_V \otimes \tau_V$ by linearity.

Definition 2.11. The direct sum of two representations ρ and τ is the space $\rho_V \oplus \tau_V$ with the block-diagonal action $\rho \oplus \tau$ of G .

If the representation is not irreducible, then by an appropriate change of basis ρ can be converted into a block diagonal matrix with blocks corresponding to the invariant subspaces. Thus, any representation can be completely decomposed into a direct sum of irreducible representations of G , by applying an appropriate unitary transformation. Note that this decomposition is *unique* (the invariant subspaces and change of basis are not unique; only the irreducible isomorphism classes and their multiplicities are unique). We use the following notation to denote the decomposition of a reducible representation: If ρ is a reducible representation of G then $\rho \cong \bigoplus_i n_i \rho_i$, where for each i we have *distinct* $\rho_i \in \text{Irrep}(G)$ and n_i denotes the multiplicity of ρ_i in the decomposition. It will be convenient to think of this representation as a block diagonal matrix with ρ_i as the blocks along the diagonal with multiplicity n_i .

We will use the following proposition which follows from an appropriate change of basis of the underlying space.

Proposition 2.12. Let $A \leq B \leq G$, and let $\rho : G \rightarrow U(V)$ be a finite-dimensional unitary representation. There are orthogonal decompositions

$$V = \bigoplus_{i=1}^r W_i, \quad W_i = \bigoplus_{j=1}^{s_i} V_{i,j},$$

such that each W_i is B -invariant and the representation of B on W_i is irreducible, while each $V_{i,j}$ is A -invariant and the representation of A on $V_{i,j}$ is irreducible.

An orthonormal basis adapted to these decompositions gives a single unitary change of basis U for which

$$U\rho(b)U^* = \bigoplus_{i=1}^r M_i(b) \quad (b \in B), \quad U\rho(a)U^* = \bigoplus_{i=1}^r \bigoplus_{j=1}^{s_i} N_{i,j}(a) \quad (a \in A).$$

Here each block M_i is the action on one irreducible B -summand W_i , and each block $N_{i,j}$ is the action on one irreducible A -summand $V_{i,j}$. Equivalent summands remain separate blocks; their indices count individual copies, not isomorphism classes.

The following proposition gives Schur orthogonality for matrix coefficients with respect to a symmetric bilinear form. Matrix coefficients belonging to non-isomorphic irreducible representations are orthogonal. For a fixed irreducible representation, the pairing is zero unless the corresponding indices are reversed, in which case it equals the reciprocal of the dimension of the representation.

Proposition 2.13. *If ρ and τ are two non-isomorphic irreducible representations of G then for any i, j, k, ℓ we have*

$$\langle (\rho)_{ij} \mid (\tau)_{k\ell} \rangle_G = 0, \quad (5)$$

where $\langle f_1 \mid f_2 \rangle_G := \frac{1}{|G|} \sum_{g \in G} f_1(g) f_2(g^{-1})$ (called a “symmetric bilinear form”). Also,

$$\langle (\rho)_{ij} \mid (\rho)_{k\ell} \rangle_G = \frac{\delta_{i\ell} \delta_{jk}}{\dim(\rho)}, \quad (6)$$

where δ_{ij} is the Kronecker delta function which is 1 if $i = j$ and 0 otherwise.

2.2.2 Fourier analysis on non-abelian groups

In this paper, we will be interested in studying $L^2(G)$, the space of functions from a finite group G to the complex numbers $\mathbb{C}$.

Definition 2.14. *Define the inner product $\langle \cdot, \cdot \rangle_{L^2(G)}$ on $L^2(G)$ by*

$$\langle f, g \rangle_{L^2(G)} = \mathbf{E}_{x \in G} [f(x) \overline{g(x)}].$$

We can define a character for every representation of a group.

Definition 2.15. *The character of a representation ρ is the function $\chi_\rho : G \rightarrow \mathbb{C}$ defined by $\chi_\rho(g) = \text{tr}(\rho(g))$.*

The following proposition shows that the characters corresponding to the irreducible representations of a group are orthogonal to each other.

Proposition 2.16 (Orthogonality of characters). *For $\rho, \tau \in \text{Irrep}(G)$, we have*

$$\frac{1}{|G|} \sum_{g \in G} \chi_\rho(g) \overline{\chi_\tau(g)} = \begin{cases} 1 & \text{if } \rho \cong \tau, \\ 0 & \text{otherwise.} \end{cases}$$

We use Proposition 2.13 many times in the proof. For convenience, we note an important identity that follows from Proposition 2.13 (by setting τ to be the trivial representation $\{1\}$).

Proposition 2.17. *If $\rho \in \text{Irrep}(G) \setminus \{1\}$, then $\sum_{g \in G} \rho(g) = 0$.*

We have the following proposition. It also shows that the maximum dimension of any irreducible representation of G is at most $\sqrt{|G|}$.

Proposition 2.18.

$$\sum_{\rho \in \text{Irrep}(G)} \dim(\rho) \chi_{\rho}(g) = \begin{cases} |G| & \text{if } g = 1_G, \\ 0 & \text{otherwise.} \end{cases}$$

This implies the following:

$$\sum_{\rho \in \text{Irrep}(G)} \dim(\rho)^2 = |G|.$$

Definition 2.19. *For two functions $f, g \in L^2(G)$ their convolution $f * g \in L^2(G)$ is defined as*

$$(f * g)(x) := \mathbf{E}_{y \in G} [f(y)g(y^{-1}x)].$$

For an abelian group, any function $f : G \rightarrow \mathbb{C}$ can be written as a linear combination of characters, i.e., the characters span the whole space $L^2(G)$. However, for non-abelian groups, characters of irreducible representations form an orthonormal basis only for the set of *class functions* – maps which are constant on *conjugacy classes*. A conjugacy class in G is a nonempty subset H of G such that the following two conditions hold: Given any $x, y \in H$, there exists $g \in G$ such that $gxg^{-1} = y$, and if $x \in H$ and $g \in G$ then $gxg^{-1} \in H$. Since this is an equivalence relation, any group is a collection of disjoint conjugacy classes.

As in the abelian case, we can understand operations like inner product, convolution, etc., using the Fourier transform, which is defined as follows.

Definition 2.20. *For a function $f \in L^2(G)$, define the Fourier transform of f to be the element $\hat{f} \in \prod_{\rho \in \text{Irrep}(G)} \text{End } \rho_V$ given by*

$$\hat{f}(\rho) = \mathbf{E}_{x \in G} [f(x)\rho(x)] \in \text{End } \rho_V.$$

Here $\rho : G \rightarrow \text{GL}(\rho_V)$ is an irreducible representation acting on the complex vector space ρ_V , and $\text{End}(\rho_V)$ denotes the space of $\mathbb{C}$ -linear endomorphisms of ρ_V (equivalently, the algebra of all linear operators $\rho_V \rightarrow \rho_V$).

Definition 2.21. *Let V be a finite-dimensional complex inner product space. Define an inner product $\langle \cdot, \cdot \rangle_{\text{End } V}$ on $\text{End } V$ by*

$$\langle A, B \rangle_{\text{End } V} = \text{tr}(AB^*).$$

We can now state the Fourier inversion theorem.

Proposition 2.22 (Fourier inversion theorem). *For $f \in L^2(G)$ we have*

$$f(x) = \sum_{\rho \in \text{Irrep}(G)} \dim(\rho) \cdot \langle \hat{f}(\rho), \rho(x) \rangle_{\text{End } \rho_V}.$$

We have the following simple identities (See [Ter99] for the proofs).

Proposition 2.23 (Plancherel's identity). For $f, g \in L^2(G)$

$$\langle f, g \rangle_{L^2(G)} = \sum_{\rho \in \text{Irrep}(G)} \dim(\rho) \cdot \langle \hat{f}(\rho), \hat{g}(\rho) \rangle_{\text{End } \rho_V}.$$

Proposition 2.24 (Parseval's identity). For $f \in L^2(G)$

$$\mathbf{E}_{x \in G} [|f(x)|^2] = \sum_{\rho \in \text{Irrep}(G)} \dim(\rho) \cdot \|\hat{f}(\rho)\|_{\text{HS}}^2,$$

where $A \in \text{End } V$, $\|A\|_{\text{HS}} := \sqrt{\langle A, A \rangle_{\text{End } V}} = \sqrt{\text{tr}(AA^*)} = \sqrt{\sum_{ij} |A_{ij}|^2}$.

Note that the norm $\|\cdot\|_{\text{HS}}$ satisfies submultiplicativity.

Claim 2.25. $\|AB\|_{\text{HS}} \leq \|A\|_{\text{HS}} \cdot \|B\|_{\text{HS}}$.

Proof. $\|AB\|_{\text{HS}}^2 = \sum_{ij} |(AB)_{ij}|^2 \leq \sum_{ij} (\sum_k |A_{ik} B_{kj}|)^2$. Using the Cauchy-Schwarz inequality on the inner sum,

$$\|AB\|_{\text{HS}}^2 \leq \sum_{ij} \left(\sum_k |A_{ik}|^2 \right) \left(\sum_\ell |B_{\ell j}|^2 \right) = \sum_{ijk\ell} |A_{ik}|^2 |B_{\ell j}|^2 = \left(\sum_{ik} |A_{ik}|^2 \right) \left(\sum_{\ell j} |B_{\ell j}|^2 \right) = \|A\|_{\text{HS}}^2 \cdot \|B\|_{\text{HS}}^2.$$

□

Claim 2.26. Let A be any matrix and U be any unitary matrix, then $\|UA\|_{\text{HS}} = \|AU\|_{\text{HS}} = \|A\|_{\text{HS}}$.

Proof. Using the identity that $\text{tr}(AB) = \text{tr}(BA)$, we have

$$\|UA\|_{\text{HS}}^2 = \text{tr}(UAA^*U^*) = \text{tr}(AA^*) = \|A\|_{\text{HS}}^2.$$

Similarly,

$$\|AU\|_{\text{HS}}^2 = \text{tr}(AUU^*A^*) = \text{tr}(AA^*) = \|A\|_{\text{HS}}^2.$$

□

Proposition 2.27 (Convolution theorem). For $f, g \in L^2(G)$ we have

$$f \hat{*} g(\rho) = \hat{f}(\rho) \hat{g}(\rho).$$

The following claim shows that the character functions always come in 'pairs' with respect to the complex conjugation.

Claim 2.28. Let G be any non-abelian group. For every $\rho \in \text{Irrep}(G)$, such that $\dim(\rho) = 1$, there exists $\tilde{\rho} \in \text{Irrep}(G)$ with $\dim(\tilde{\rho}) = 1$ such that

$$\chi_\rho(g) = \overline{\chi_{\tilde{\rho}}(g)}, \quad \forall g \in G.$$

Proof. We claim that the set of characters corresponding to dimension 1 irreducible representations of G forms a group under point-wise multiplication. This will be enough to show the claim.

Let $G' = G/[G, G]$ be the abelian quotient group. Assume ρ is a degree 1 representation of G . Then it satisfies $\rho(a)\rho(b) = \rho(ab)$ for all $a, b \in G$. Define a map $\Gamma_\rho : G' \rightarrow \mathbb{C}$ as $\Gamma_\rho(g') = \rho(g)$ where $g' = g[G, G]$. This is a well-defined map as

$$\rho(aba^{-1}b^{-1}) = \rho(a)\rho(b)\rho(a^{-1})\rho(b^{-1}) = \rho(a)\rho(a^{-1})\rho(b)\rho(b^{-1}) = 1.$$

Thus, the map ρ is constant on every coset of $[G, G]$ and hence Γ_ρ is well defined. The set of all $\{\Gamma_\rho \mid \rho \in \text{Irrep}(G), \dim(\rho) = 1\}$ is the set of all the multiplicative characters of the abelian group G' and hence forms a group under coordinate-wise multiplication. There is a one-to-one correspondence between the coordinate-wise multiplicative action of Γ_ρ 's and ρ 's. Thus, $\{\chi_\rho \mid \rho \in \text{Irrep}(G), \dim(\rho) = 1\}$ forms a group under point-wise multiplication. $\square$

2.3 Functions on G^n

For any non-abelian group G and $n \geq 1$, we have a group G^n where the group operation is defined coordinate-wise. The irreducible representations of G^n are precisely those representations obtained by taking tensor products of n irreducible representations of G .

Proposition 2.29 ([Ter99]). *The set of irreducible representations of G^n is given by*

$$\text{Irrep}(G^n) = \{\alpha \mid \alpha = \otimes_{i \in [n]} \rho_i \text{ where } \rho_i \in \text{Irrep}(G)\}.$$

We denote α by the corresponding tuple $(\rho_1, \rho_2, \dots, \rho_n)$. We define the weight of a representation $\alpha = (\rho_1, \rho_2, \dots, \rho_n)$ (denoted by $|\alpha|$) to be the number of non-trivial representations in $(\rho_1, \rho_2, \dots, \rho_n)$.

We will be working with functions $f : G^n \rightarrow G$ which are *folded*. f is said to be folded if $f(c\mathbf{x}) = cf(\mathbf{x})$ for all $c \in G$ and $\mathbf{x} \in G^n$. The following claim shows that for all functions $g(\mathbf{x}) := \rho(f(\mathbf{x}))_{ij}$ where $\dim(\rho) \geq 2$ and $1 \leq i, j \leq \dim(\rho)$, all the Fourier coefficients corresponding to representations of dimension 1 are zero if f is folded.

Lemma 2.30. *Let $f : G^n \rightarrow G$ be any folded function and $g(\mathbf{x}) := \rho(f(\mathbf{x}))_{ij}$ where $\rho \in \text{Irrep}(G), \dim(\rho) \geq 2$ and $1 \leq i, j \leq \dim(\rho)$. If α is any representation of G^n such that $\dim(\alpha) = 1$, then $\hat{g}(\alpha) = 0$.*

Proof. Recall, for any $\mathbf{x} \in G^n$, $\alpha(\mathbf{x})$ is a scalar as $\dim(\alpha) = 1$. f is folded which means that $f(c\mathbf{x}) = cf(\mathbf{x})$ for all $c \in G$ and $\mathbf{x} \in G^n$. For $c \in G$, let $\mathbf{c} \in G^n$ be such that every coordinate of $\mathbf{c}$ is c . Since $\rho(\cdot)$ has dimension at least 2, in the following analysis, we use $[\rho(\cdot)]$ to denote the matrix of linear transformation for clarity.

$$\begin{aligned} \hat{g}(\alpha) &= \mathbf{E}_{\mathbf{x}}[g(\mathbf{x})\alpha(\mathbf{x})] = \mathbf{E}_{\mathbf{x}}[[\rho(f(\mathbf{x}))]_{ij} \cdot \alpha(\mathbf{x})] \\ &= \frac{1}{|G|} \mathbf{E}_{\mathbf{x}} \left[\sum_{c \in G} [\rho(f(c\mathbf{x}))]_{ij} \cdot \alpha(c\mathbf{x}) \right] \\ &= \frac{1}{|G|} \mathbf{E}_{\mathbf{x}} \left[\sum_{c \in G} [\rho(cf(\mathbf{x}))]_{ij} \cdot \alpha(c)\alpha(\mathbf{x}) \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{|G|} \mathbf{E}_x \left[\sum_{c \in G} (\alpha(c) [\rho(c)] \cdot [\rho(f(x))])_{ij} \alpha(x) \right] \\
&= \frac{1}{|G|} \mathbf{E}_x \left[\left(\left(\sum_{c \in G} \alpha(c) [\rho(c)] \right) \cdot [\rho(f(x))] \right)_{ij} \alpha(x) \right].
\end{aligned}$$

Now, for $\alpha \in \text{Irrep}(G^n)$, let $\tilde{\alpha} \in \text{Irrep}(G^n)$ be the dimension 1 representation satisfying the condition in Claim 2.28. We have:

$$\begin{aligned}
\sum_{c \in G} \alpha(c) [\rho(c)] &= \sum_{c \in G} \overline{\tilde{\alpha}(c)} \cdot [\rho(c)] \\
&= \sum_{c \in G} \tilde{\alpha}(c^{-1}) \cdot [\rho(c)] \\
&= \sum_{c \in G} (\otimes_{i=1}^n \tilde{\alpha}_i(c^{-1})) \cdot [\rho(c)] \\
&= \sum_{c \in G} \tau(c^{-1}) \cdot [\rho(c)] & \dim(\tau) = 1 \\
&= 0, & \text{(Using Proposition 2.13)}
\end{aligned}$$

where in the second last step, we used the fact that the product of dimension 1 representations $(\otimes_{i=1}^n \tilde{\alpha}_i)$ of G is itself a dimension 1 representation (τ) of G . Therefore, $\hat{g}(\alpha) = 0$. □

Fix any surjective projection map $\pi : [R] \rightarrow [L]$ for some $R \geq L$. Consider the following subgroup of G^R given by the elements

$$\{(x \circ \pi) \in G^R \mid x \in G^L\},$$

where $(x \circ \pi)_i = x_{\pi(i)}$. Let us denote this group by $\pi(G^R)$. Note that this group is isomorphic to G^L . Thus, any representation $\alpha \in \text{Irrep}(G^R)$ (which is a representation of G^L using Fact 2.9) can be decomposed into irreducible representations of a subgroup of G^R isomorphic to G^L .

2.4 Notations

Whenever possible, we use the notation α, β to denote the representations of the group G^n and ρ, τ for the group G . Also, we use bold letters $\mathbf{x}, \mathbf{c}$ to denote the elements of G^n .

For a representation $\alpha \in \text{Irrep}(G^n)$ where $\alpha = \otimes_{i=1}^n \rho_i$, we use the notation $\dim_{\geq k}(\alpha)$ to denote the number of $i \in [n]$ such that $\dim(\rho_i) \geq k$.

3 Warm-up: Dictatorship Test

In this section, we analyze the dictatorship test, which involves checking some linear equation over a non-abelian group. The analysis will highlight a few important differences between our test and the linearity test over abelian groups.

Fix a non-abelian group G . Let $f : G^n \rightarrow G$ be a function. A function is called a dictator function if it is of the form $f(x) = x_i$ for some $i \in [n]$. We use $\cdot$ to denote the group operation. Consider the following 3-query dictatorship test for f :

1. Sample $\mathbf{a} = (a_1, a_2, \dots, a_n)$ from G^n uniformly at random.
2. Sample $\mathbf{b} = (b_1, b_2, \dots, b_n)$ from G^n uniformly at random.
3. Calculate $\mathbf{c} = (c_1, c_2, \dots, c_n)$ such that $c_i = b_i^{-1} a_i^{-1}$.
4. Check if $f(\mathbf{a}) \cdot f(\mathbf{b}) \cdot f(\mathbf{c}) = 1_G$.

Completeness is trivial: If f is an i^{th} dictator function, i.e., $f(x_1, x_2, \dots, x_n) = x_i$, then the test passes with probability 1. This is because we are essentially checking if $a_i \cdot b_i \cdot c_i = 1_G$, which is always true by the definition of c_i .

We analyze the soundness of the test. The following lemma says that if the test passes with some non-trivial probability, then it must be the case that f (or a minor variation of f) has a low-dimensional Fourier coefficient whose Hilbert-Schmidt norm is large. The actual conclusion is somewhat stronger than this. In the next section, we will show that such a conclusion can be used to analyze the soundness of the final reduction (which is also presented in the next section).

Lemma 3.1. *Assume f is folded. For all $\varepsilon > 0$ and $\frac{1}{2} > \delta > 0$, if f passes the test with probability $\frac{1}{|G, G|} + \varepsilon$, then there exist $\rho \in \text{Irrep}(G)$ with $\dim(\rho) \geq 2$ and $1 \leq i, j \leq \dim(\rho)$ such that for $h(\mathbf{x}) := \rho(f(\mathbf{x}))_{ij}$,*

$$\max_{\substack{\alpha \in \text{Irrep}(G^n), \\ \dim(\alpha) \geq 2, \\ \dim_{\geq 2}(\alpha) < \frac{1}{2\delta^2}}} \|\hat{h}(\alpha)\|_{\text{HS}} \geq \frac{\varepsilon}{|G|} - \delta.$$

Proof. Using Proposition 2.18, the probability that the test passes can be expressed as follows:

$$\begin{aligned} \Pr[\text{Test passes}] &= \frac{1}{|G|} \sum_{\rho \in \text{Irrep}(G)} \dim(\rho) \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [\chi_\rho(f(\mathbf{a}) \cdot f(\mathbf{b}) \cdot f(\mathbf{c}))] \\ &= \frac{1}{|G|} \sum_{\substack{\rho \in \text{Irrep}(G), \\ \dim(\rho)=1}} \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [\chi_\rho(f(\mathbf{a}) \cdot f(\mathbf{b}) \cdot f(\mathbf{c}))] \\ &\quad + \frac{1}{|G|} \sum_{\substack{\rho \in \text{Irrep}(G), \\ \dim(\rho) \geq 2}} \dim(\rho) \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [\chi_\rho(f(\mathbf{a}) \cdot f(\mathbf{b}) \cdot f(\mathbf{c}))]. \end{aligned}$$

In the first summation, for any $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) = 1$, using the multiplicativity of the characters, we have

$$\begin{aligned} |\chi_\rho(f(\mathbf{a}) \cdot f(\mathbf{b}) \cdot f(\mathbf{c}))| &= |\chi_\rho(f(\mathbf{a})) \chi_\rho(f(\mathbf{b})) \chi_\rho(f(\mathbf{c}))| \\ &\leq |\chi_\rho(f(\mathbf{a}))| \cdot |\chi_\rho(f(\mathbf{b}))| \cdot |\chi_\rho(f(\mathbf{c}))| \\ &= 1. \end{aligned} \quad (\text{unitary representations})$$

As the number of dimension 1 representations of a group G is equal to the size of the quotient $G/[G, G]$, we get

$$\Pr[\text{Test passes}] \leq \frac{1}{|[G, G]|} + \frac{1}{|G|} \sum_{\substack{\rho \in \text{Irrep}(G), \\ \dim(\rho) \geq 2}} \dim(\rho) \mathbf{E}_{a, b, c} [\chi_\rho(f(a) \cdot f(b) \cdot f(c))]. \quad (7)$$

Now, fix any $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) \geq 2$. For $1 \leq i, j \leq \dim(\rho)$, let $g_{ij} : G^n \rightarrow \mathbb{C}$ be defined as $g_{ij}(x) := \rho(f(x))_{ij}$. Using the definition of characters, we have

$$\begin{aligned} \mathbf{E}_{a, b, c} [\chi_\rho(f(a) \cdot f(b) \cdot f(c))] &= \mathbf{E}_{a, b, c} [\text{tr}(\rho(f(a) \cdot f(b) \cdot f(c)))] \\ (\rho \text{ is a homomorphism}) &= \mathbf{E}_{a, b, c} [\text{tr}(\rho(f(a)) \cdot \rho(f(b)) \cdot \rho(f(c)))] \\ &= \mathbf{E}_{a, b, c} \left[\sum_{1 \leq i, j, k \leq \dim(\rho)} \rho(f(a))_{ij} \cdot \rho(f(b))_{jk} \cdot \rho(f(c))_{ki} \right] \\ &= \sum_{1 \leq i, j, k \leq \dim(\rho)} \mathbf{E}_{a, b, c} [g_{ij}(a) g_{jk}(b) g_{ki}(c)] \\ &= \sum_{1 \leq i, j, k \leq \dim(\rho)} (g_{ij} * g_{jk} * g_{ki})(1_{G^n}). \end{aligned} \quad (8)$$

Since we assume that the test passes with probability $\frac{1}{|[G, G]|} + \varepsilon$, from Equation (7) and Equation (8) (and using $\dim(\rho) \leq \sqrt{|G|}$), we conclude that there exists ρ with $\dim(\rho) \geq 2$ and $1 \leq i, j, k \leq \dim(\rho)$ such that

$$|(g_{ij} * g_{jk} * g_{ki})(1_{G^n})| \geq \frac{\varepsilon}{|G|}.$$

We now analyze the term $(g_{ij} * g_{jk} * g_{ki})(1_{G^n})$ for a fixed (i, j, k) . For ease of notation, we write $h_1 := g_{ij}$, $h_2 := g_{jk}$ and $h_3 := g_{ki}$.

$$\begin{aligned} \frac{\varepsilon}{|G|} &\leq |(g_{ij} * g_{jk} * g_{ki})(1_{G^n})| = |(h_1 * h_2 * h_3)(1_{G^n})| \\ &= \left| \sum_{\alpha \in \text{Irrep}(G^n)} \dim(\alpha) \cdot \text{tr}(h_1 * \widehat{h_2} * h_3(\alpha)) \right| \\ &\leq \sum_{\alpha \in \text{Irrep}(G^n)} \dim(\alpha) \cdot |\text{tr}(\hat{h}_1(\alpha) \hat{h}_2(\alpha) \hat{h}_3(\alpha))| \\ &= \sum_{\alpha \in \text{Irrep}(G^n)} \dim(\alpha) \cdot \left| \langle \hat{h}_1(\alpha) \hat{h}_2(\alpha), \hat{h}_3(\alpha)^* \rangle_{\text{End } \alpha_V} \right| \\ &\leq \sum_{\alpha \in \text{Irrep}(G^n)} \dim(\alpha) \cdot \|\hat{h}_1(\alpha) \hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)^*\|_{\text{HS}}. \end{aligned}$$

We now use Lemma 2.30 to conclude that for all $1 \leq r \leq 3$, $\hat{h}_r(\alpha) = 0$ if $\dim(\alpha) = 1$. Using this,

we continue as follows:

$$\begin{aligned}
|(h_1 * h_2 * h_3)(1_{G^n})| &\leq \sum_{\substack{\alpha \in \text{Irrep}(G^n), \\ \dim(\alpha) \geq 2}} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)^*\|_{\text{HS}} \\
&= \sum_{\substack{\alpha \in \text{Irrep}(G^n), \\ \dim(\alpha) \geq 2}} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)\|_{\text{HS}}.
\end{aligned}$$

Let $D := \frac{1}{2\delta^2}$. Now, we split the sum into two parts $|(h_1 * h_2 * h_3)(1_{G^n})| \leq \Theta_{\text{low}} + \Theta_{\text{high}}$ where

$$\Theta_{\text{low}} = \sum_{\substack{\alpha \in \text{Irrep}(G^n), \\ \dim(\alpha) \geq 2, \\ \dim_{\geq 2}(\alpha) < D}} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)\|_{\text{HS}},$$

and

$$\Theta_{\text{high}} = \sum_{\substack{\alpha \in \text{Irrep}(G^n), \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)\|_{\text{HS}}.$$

3.1 Bounding higher order terms

In this section, we show that the high-degree terms can be upper bounded by a small constant, even though the three queries are perfectly correlated.

We bound Θ_{high} as follows:

$$\begin{aligned}
\Theta_{\text{high}} &= \sum_{\substack{\alpha_r \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)\|_{\text{HS}} \\
&\leq \sum_{\substack{\alpha_r \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\|_{\text{HS}} \|\hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)\|_{\text{HS}} \quad (\text{Claim 2.25}) \\
&\leq \frac{1}{\sqrt{2D}} \sum_{\substack{\alpha_r \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha)^{3/2} \cdot \|\hat{h}_1(\alpha)\|_{\text{HS}} \|\hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)\|_{\text{HS}}.
\end{aligned}$$

Here, we used the fact that all the representations α of G^n with $\dim_{\geq 2}(\alpha) \geq D$ have dimensions at least $2^D \geq 2D$. At this point, we would like to highlight the primary source of effectively bounding the higher-order terms. It is the size of $\dim(\alpha)$ in the summation. In Gowers' [Gow08] proof, a similar expression appears in the analysis, with the same condition that all the representations in the summation have large dimension. It is in some sense the main difference between the abelian and the non-abelian setting (both in this work and Gowers'), similar to the Equation (1) mentioned in the introduction.

Now, using Cauchy-Schwarz inequality,

$$\Theta_{\text{high}} \leq \frac{1}{\sqrt{2D}} \sum_{\substack{\alpha_r \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha)^{3/2} \cdot \|\hat{h}_1(\alpha)\|_{\text{HS}} \|\hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)\|_{\text{HS}}$$

$$\begin{aligned}
&\leq \frac{1}{\sqrt{2D}} \left(\sum_{\substack{\alpha_i \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\|_{\text{HS}}^2 \right)^{1/2} \cdot \left(\sum_{\substack{\alpha_i \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha)^2 \cdot \|\hat{h}_2(\alpha)\|_{\text{HS}}^2 \|\hat{h}_3(\alpha)\|_{\text{HS}}^2 \right)^{1/2} \\
&\leq \frac{1}{\sqrt{2D}} \left(\sum_{\substack{\alpha_i \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\|_{\text{HS}}^2 \right)^{1/2} \cdot \\
&\quad \left(\left(\sum_{\substack{\alpha_i \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha) \cdot \|\hat{h}_2(\alpha)\|_{\text{HS}}^2 \right) \cdot \left(\sum_{\substack{\alpha_i \\ \dim_{\geq 2}(\alpha) \geq D}} \dim(\alpha) \cdot \|\hat{h}_3(\alpha)\|_{\text{HS}}^2 \right) \right)^{1/2} \\
&\leq \frac{1}{\sqrt{2D}} \cdot \|h_1\|_{L^2(G^n)} \cdot \|h_2\|_{L^2(G^n)} \cdot \|h_3\|_{L^2(G^n)}. \quad (\text{Using Proposition 2.24})
\end{aligned}$$

Since h_1 was defined as $h_1(x) = \rho(f(x))_{ij}$ where $\rho \in \text{Irrep}(G)$, $|h_1(x)| \leq 1$. The same is true for h_2 and h_3 , and hence all the norms are bounded by 1. Therefore,

$$\Theta_{\text{high}} \leq \frac{1}{\sqrt{2D}} = \delta.$$

3.2 Bounding lower order terms

It remains to show that Θ_{low} is related to the Fourier mass of h_3 on the low-dimensional representations.

$$\begin{aligned}
\Theta_{\text{low}} &= \sum_{\substack{\alpha_i \\ \dim(\alpha) \geq 2, \\ \dim_{\geq 2}(\alpha) < D}} \dim(\alpha) \cdot \|\hat{h}_1(\alpha) \hat{h}_2(\alpha)\|_{\text{HS}} \|\hat{h}_3(\alpha)\|_{\text{HS}} \\
&\leq \max_{\substack{\alpha_i \\ \dim(\alpha) \geq 2, \\ \dim_{\geq 2}(\alpha) < D}} \|\hat{h}_3(\alpha)\|_{\text{HS}} \cdot \left(\sum_{\alpha} \dim(\alpha) \cdot \|\hat{h}_1(\alpha) \hat{h}_2(\alpha)\|_{\text{HS}} \right).
\end{aligned}$$

We can upper bound the summation by 1 using the Cauchy-Schwarz inequality as follows:

$$\begin{aligned}
\sum_{\alpha} \dim(\alpha) \cdot \|\hat{h}_1(\alpha) \hat{h}_2(\alpha)\|_{\text{HS}} &\leq \sum_{\alpha} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\|_{\text{HS}} \|\hat{h}_2(\alpha)\|_{\text{HS}} \\
&\leq \left(\sum_{\alpha} \dim(\alpha) \cdot \|\hat{h}_1(\alpha)\|_{\text{HS}}^2 \right)^{1/2} \cdot \left(\sum_{\alpha} \dim(\alpha) \cdot \|\hat{h}_2(\alpha)\|_{\text{HS}}^2 \right)^{1/2} \\
&= \|h_1\|_2 \cdot \|h_2\|_2 \quad (\text{Proposition 2.24}) \\
&\leq 1.
\end{aligned}$$

Here, the last inequality uses the fact that $|h_1(\mathbf{x})|, |h_2(\mathbf{x})| \leq 1$ for all $\mathbf{x} \in G^n$. Using the upper bound on Θ_{high} , we have $\Theta_{\text{low}} \geq \frac{\varepsilon}{|G|} - \delta$. Therefore, we get

$$\max_{\substack{\alpha, \\ \dim(\alpha) \geq 2, \\ \dim_{\geq 2}(\alpha) < D}} \|\hat{h}_3(\alpha)\|_{\text{HS}} \geq \left(\frac{\varepsilon}{|G|} - \delta \right).$$

□

4 Main Reduction

In this section, we prove Theorem 1.1. We give a reduction from an instance of a Layered LABEL-COVER, $\mathcal{H} = (\mathcal{V} = \{V_1, V_2, \dots, V_\ell\}, \{\Pi_{t,t'}\}_{1 \leq t < t' \leq \ell}, \{[R_t]\}_{t \in [\ell]})$ as in Definition 2.3, to a 3-LIN instance $\mathcal{I}$ over a non-abelian group G . Towards obtaining the $1 \text{ vs. } \frac{1}{|G|} + \delta$ hardness of Max-3-LIN for $\delta \in (0, \frac{1}{2})$, we will use the following setting of ℓ and r in Theorem 2.4

$$2^{-r} \leq \min \left\{ \frac{\delta^{10}}{(2|G|)^{100}}, \frac{\delta^2}{10|G|^{10C}} \right\}, \quad \ell \geq \left(\frac{8|G|^3}{\delta} \right)^4,$$

where $C := \frac{8|G|^6}{\delta^2}$.

The reduction is given in Figure 1.

The value of the instance $\text{val}(\mathcal{I})$ is the maximum probability that the above test is satisfied, where the maximum is over all folded functions $\{f_v\}_{v \in \cup_t V_t}$.

4.1 Analysis

Lemma 4.1 (Completeness). *If $\mathcal{H}$ is a satisfiable instance of the Layered LABEL-COVER, then $\text{val}(\mathcal{I}) = 1$.*

Proof. Fix a satisfying assignment $\{\sigma_t : V_t \rightarrow [R_t]\}_{t \in [\ell]}$ of $\mathcal{H}$. Consider the long code encodings of the labeling σ_t : for every $t \in [\ell]$ and $v \in V_t$, set $f_v(\mathbf{x}) = x_{\sigma_t(v)}$. We show that this assignment to $\mathcal{I}$ satisfies all the constraints. For the constraint $\pi_{vu} \in \Pi_{t,t'}$, the LHS of the test becomes

$$\begin{aligned} f_v(\mathbf{a}) \cdot f_u(\mathbf{b}) \cdot f_v(\mathbf{c}) &= a_{\sigma_t(v)} \cdot b_{\sigma_{t'}(u)} \cdot c_{\sigma_t(v)} \\ &= a_{\sigma_t(v)} \cdot b_{\sigma_{t'}(u)} \cdot (b \circ \pi_{vu})_{\sigma_t(v)}^{-1} \cdot a_{\sigma_t(v)}^{-1} \\ &= a_{\sigma_t(v)} \cdot b_{\sigma_{t'}(u)} \cdot b_{\pi_{vu}(\sigma_t(v))}^{-1} \cdot a_{\sigma_t(v)}^{-1} \\ &= a_{\sigma_t(v)} \cdot b_{\sigma_{t'}(u)} \cdot b_{\sigma_{t'}(u)}^{-1} \cdot a_{\sigma_t(v)}^{-1} \quad (\pi_{vu}(\sigma_t(v)) = \sigma_{t'}(u)) \\ &= 1_G. \end{aligned}$$

Therefore, the assignment to the 3-LIN instance $\mathcal{I}$ satisfies all the constraints. □

We now prove the main soundness lemma which is as follows

The set of variables of $\mathcal{I}$ is $\bigcup_{t \in [\ell]} V_t \times G^{[R_t]}$. Any assignment to the instance $\mathcal{I}$ is given by a set of functions $f_v : G^{[R_t]} \rightarrow G$ for each $v \in V_t$ for all $t \in [\ell]$. We further enforce that these functions are *folded* (the set $G^{[R_t]}$ can be thought of as G^{R_t-1} variables and their literals by the following association: every $\mathbf{x} \in G^{[R_t]}$ such that $x_1 = 1_G$ is a variable, and its g -literal is given by $g\mathbf{x}$).

The distribution of the 3-LIN constraints in $\mathcal{I}$ is given by the following procedure:

1. Pick a uniformly random pair (t, t') satisfying $1 \leq t < t' \leq \ell$.
2. Choose a constraint $\pi_{vu} \in \Pi_{t,t'}$ of $\mathcal{H}$ uniformly at random.
3. Sample $\mathbf{a} = (a_1, a_2, \dots, a_{R_t})$ from G^{R_t} uniformly at random.
4. Sample $\mathbf{b} = (b_1, b_2, \dots, b_{R_{t'}})$ from $G^{R_{t'}}$ uniformly at random.
5. Let $\mathbf{c} = (c_1, c_2, \dots, c_{R_t})$ be such that $c_i = (b \circ \pi_{vu})_i^{-1} \cdot a_i^{-1}$, here $\mathbf{x} \circ \pi \in G^{R_t}$ is the string defined as $(\mathbf{x} \circ \pi)_i := x_{\pi(i)}$ for $i \in [R_t]$.
6. Output the constraint $f_v(\mathbf{a}) \cdot f_u(\mathbf{b}) \cdot f_v(\mathbf{c}) = 1_G$.

Figure 1: Reduction from a Layered LABEL-COVER instance $\mathcal{H} = (\mathcal{V} = \{V_t\}_{t \in [\ell]}, \{\Pi_{t,t'}\}_{1 \leq t < t' \leq \ell}, \{[R_t]\}_{t \in [\ell]})$ to a Max-3-LIN $\mathcal{I}$ over G .

Lemma 4.2 (Soundness). *If $\mathcal{H}$ is at most 2^{-r} -satisfiable, then $\text{val}(\mathcal{I}) \leq \frac{1}{|[G,G]|} + \delta$.*

Before we see the proof, we first conclude that this is enough to prove the main theorem of this paper.

Proof of Theorem 1.1. Lemma 4.1 and Lemma 4.2 above, along with the **NP**-hardness of the Layered LABEL-COVER from Theorem 2.4 imply our main theorem, Theorem 1.1, for any constant $\varepsilon > 0$. $\square$

Proof of Lemma 4.2. Fix any assignment to the instance $\mathcal{I}$ given by a set of functions $f_v : G^{R_t} \rightarrow G$ for each $t \in [\ell]$ and $v \in V_t$. Using Proposition 2.18, the value of the instance for this assignment is given by:

$$\mathbf{E}_{1 \leq t < t' \leq \ell} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \mathbf{E}_{a,b,c} \left[\frac{1}{|G|} \sum_{\rho \in \text{Irrep}(G)} \dim(\rho) \chi_\rho(f_v(a) \cdot f_u(b) \cdot f_v(c)) \right]$$

For any $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) = 1$, we have

$$\begin{aligned} |\chi_\rho(f_v(a) \cdot f_u(b) \cdot f_v(c))| &= |\chi_\rho(f_v(a)) \cdot \chi_\rho(f_u(b)) \cdot \chi_\rho(f_v(c))| \\ &\leq |\chi_\rho(f_v(a))| \cdot |\chi_\rho(f_u(b))| \cdot |\chi_\rho(f_v(c))| \\ &= 1. \end{aligned} \quad (\text{unitary representations})$$

As the number of dimension 1 representations of a group G is equal to the size of the quotient group $G/[G,G]$, we get

$$\begin{aligned} \text{val}(\mathcal{I}) &\leq \frac{1}{|[G,G]|} + \frac{1}{|G|} \left| \mathbf{E}_{1 \leq t < t' \leq \ell} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\sum_{\substack{\rho \in \text{Irrep}(G), \\ \dim(\rho) \geq 2}} \dim(\rho) \mathbf{E}_{a,b,c} [\chi_\rho(f_v(a) \cdot f_u(b) \cdot f_v(c))] \right] \right| \\ &\leq \frac{1}{|[G,G]|} + \sum_{\substack{\rho \in \text{Irrep}(G), \\ \dim(\rho) \geq 2}} \left| \mathbf{E}_{1 \leq t < t' \leq \ell} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \mathbf{E}_{a,b,c} [\chi_\rho(f_v(a) \cdot f_u(b) \cdot f_v(c))] \right|. \end{aligned}$$

The lemma follows from the following Claim 4.3. $\square$

Claim 4.3. *If $\mathcal{H}$ is at most 2^{-r} -satisfiable, then for every $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) \geq 2$,*

$$\left| \mathbf{E}_{1 \leq t < t' \leq \ell} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \mathbf{E}_{a,b,c} [\chi_\rho(f_v(a) \cdot f_u(b) \cdot f_v(c))] \right| \leq \frac{\delta}{|G|}.$$

Proof. Fix any $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) \geq 2$. Let

$$\Gamma := \mathbf{E}_{1 \leq t < t' \leq \ell} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \mathbf{E}_{a,b,c} [\chi_\rho(f_v(a) \cdot f_u(b) \cdot f_v(c))]$$

$$= \mathbf{E}_{1 \leq t < t' \leq \ell} [\Gamma_{t,t'}].$$

where

$$\Gamma_{t,t'} := \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [\chi_\rho(f_v(\mathbf{a}) \cdot f_u(\mathbf{b}) \cdot f_v(\mathbf{c}))].$$

Fix $1 \leq t < t' \leq \ell$ and $\pi_{vu} \in \Pi_{t,t'}$. We first look at the inner expectation. For $1 \leq p, q \leq \dim(\rho)$, let $g_{pq} : G^{R_t} \rightarrow \mathbb{C}$ be defined as $g_{pq}(\mathbf{x}) := \rho(f_v(\mathbf{x}))_{pq}$. Also, let $h_{pq} : G^{R_{t'}} \rightarrow \mathbb{C}$ be defined as $h_{pq}(\mathbf{y}) := \rho(f_u(\mathbf{y}))_{pq}$. We have

$$\begin{aligned} \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [\chi_\rho(f_v(\mathbf{a}) f_u(\mathbf{b}) f_v(\mathbf{c}))] &= \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [\text{tr}(\rho(f_v(\mathbf{a}) \cdot f_u(\mathbf{b}) \cdot f_v(\mathbf{c})))] \\ (\rho \text{ is a homomorphism}) \quad &= \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [\text{tr}(\rho(f_v(\mathbf{a})) \cdot \rho(f_u(\mathbf{b})) \cdot \rho(f_v(\mathbf{c})))] \\ &= \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} \left[\sum_{1 \leq p, q, r \leq \dim(\rho)} \rho(f_v(\mathbf{a}))_{pq} \cdot \rho(f_u(\mathbf{b}))_{qr} \cdot \rho(f_v(\mathbf{c}))_{rp} \right] \\ &= \sum_{1 \leq p, q, r \leq \dim(\rho)} \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [g_{pq}(\mathbf{a}) \cdot h_{qr}(\mathbf{b}) \cdot g_{rp}(\mathbf{c})]. \end{aligned}$$

We now analyze the term $\Theta_{p,q,r}^{(v,u)} := \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [g_{pq}(\mathbf{a}) h_{qr}(\mathbf{b}) g_{rp}(\mathbf{c})]$ for a fixed (p, q, r) . For ease of notation, we write $g' := g_{pq}$, $h := h_{qr}$ and $g := g_{rp}$. Also, we use π for π_{vu} . The following identity follows from Proposition 2.27 and we will use it to simplify $\Theta_{p,q,r}^{(v,u)}$.

$$\mathbf{E}_{\mathbf{a}, \mathbf{b}} [g'(\mathbf{a}) \cdot g((\mathbf{b} \circ \pi)^{-1} \cdot \mathbf{a}^{-1})] = \mathbf{E}_{\mathbf{b}} [(g * g')((\mathbf{b} \circ \pi)^{-1})].$$

We now bound the expectation as follows.

$$\begin{aligned} \Theta_{p,q,r}^{(v,u)} &= \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [g_{pq}(\mathbf{a}) \cdot h_{qr}(\mathbf{b}) \cdot g_{rp}(\mathbf{c})] \\ &= \mathbf{E}_{\mathbf{a}, \mathbf{b}, \mathbf{c}} [g'(\mathbf{a}) \cdot h(\mathbf{b}) \cdot g((\mathbf{b} \circ \pi)^{-1} \cdot \mathbf{a}^{-1})] \\ &= \mathbf{E}_{\mathbf{b}} [(g * g')(\mathbf{b}^{-1} \circ \pi) \cdot h(\mathbf{b})] \\ &= \mathbf{E}_{\mathbf{b}} [(g * g')(\mathbf{b} \circ \pi) \cdot h(\mathbf{b}^{-1})] \\ &= \mathbf{E}_{\mathbf{b}} \left[\left(\sum_{\alpha} \dim(\alpha) \text{tr}(\hat{g}(\alpha) \hat{g}'(\alpha) \alpha(\mathbf{b} \circ \pi)) \right) \cdot \left(\sum_{\beta} \dim(\beta) \text{tr}(\hat{h}(\beta) \beta(\mathbf{b}^{-1})) \right) \right] \\ &= \sum_{\substack{\alpha, \beta, \\ \dim(\alpha), \dim(\beta) \geq 2}} \underbrace{\dim(\alpha) \dim(\beta) \mathbf{E}_{\mathbf{b}} [\text{tr}(\hat{g}(\alpha) \hat{g}'(\alpha) \alpha(\mathbf{b} \circ \pi)) \cdot \text{tr}(\hat{h}(\beta) \beta(\mathbf{b}^{-1}))]}_{\text{Term}^{(v,u)}(\alpha, \beta)}, \end{aligned}$$

where the last step uses the fact that the functions g, g' and h satisfy the condition of Lemma 2.30 and hence $\hat{g}(\alpha) = 0$ if $\dim(\alpha) = 1$ (same for $\hat{g}'(\alpha)$ and $\hat{h}(\beta)$).

We now break the sum into two parts for $C := \frac{8|G|^6}{\delta^2}$:

$$\Theta_{p,q,r}^{(v,u)}(\text{low}) := \sum_{\substack{\alpha, \beta, \\ \dim(\alpha), \dim(\beta) \geq 2, \\ \dim_{\geq 2}(\alpha) \leq C}} \mathbf{Term}^{(v,u)}(\alpha, \beta), \quad \Theta_{p,q,r}^{(v,u)}(\text{high}) := \sum_{\substack{\alpha, \beta, \\ \dim(\alpha), \dim(\beta) \geq 2, \\ \dim_{\geq 2}(\alpha) > C}} \mathbf{Term}^{(v,u)}(\alpha, \beta).$$

Recall, $\dim_{\geq 2}(\alpha)$ denotes the number of representations in $\alpha = (\rho_1, \rho_2, \dots, \rho_{R_t})$ which are of dimensions at least 2. With these notations, we have

$$\Gamma_{t,t'} := \sum_{p,q,r} \left(\mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\Theta_{p,q,r}^{(v,u)}(\text{low}) \right] + \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\Theta_{p,q,r}^{(v,u)}(\text{high}) \right] \right).$$

The upper bound on the absolute value of $\Gamma := \mathbf{E}_{1 \leq t < t' \leq \ell} [\Gamma_{t,t'}]$ follows from Claim 4.4, Claim 4.5 and triangle inequality (and also noting that p, q and r take at most $\sqrt{|G|}$ distinct values). $\square$

Claim 4.4. *If $\mathcal{H}$ is at most $\frac{\delta^2}{10|G|^{100}}$ -satisfiable, then for every $1 \leq t < t' \leq \ell$, $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) \geq 2$, and every $1 \leq p, q, r \leq \dim(\rho)$,*

$$\left| \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\Theta_{p,q,r}^{(v,u)}(\text{low}) \right] \right| \leq \frac{\delta}{2|G|^3}.$$

Claim 4.5. *Suppose $\mathcal{H}$ is at most $\frac{\delta^{10}}{(2|G|)^{100}}$ -satisfiable. Then for every $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) \geq 2$ and $1 \leq p, q, r \leq \dim(\rho)$,*

$$\left| \mathbf{E}_{1 \leq t < t' \leq \ell} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\Theta_{p,q,r}^{(v,u)}(\text{high}) \right] \right| \leq \frac{\delta}{2|G|^3}.$$

4.1.1 Bounding the low terms

Claim 4.6. *(Restatement of Claim 4.4) If $\mathcal{H}$ is at most $\frac{\delta^2}{10|G|^{100}}$ -satisfiable, then for every $1 \leq t < t' \leq \ell$, $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) \geq 2$, and every $1 \leq p, q, r \leq \dim(\rho)$,*

$$\left| \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\Theta_{p,q,r}^{(v,u)}(\text{low}) \right] \right| \leq \frac{\delta}{2|G|^3}.$$

Proof. Fix any $1 \leq t < t' \leq \ell$ and $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) \geq 2$. Assume towards contradiction that there exists $1 \leq p, q, r \leq \dim(\rho)$ such that

$$\mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\left| \Theta_{p,q,r}^{(v,u)}(\text{low}) \right| \right] \geq \left| \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\Theta_{p,q,r}^{(v,u)}(\text{low}) \right] \right| > \frac{\delta}{2|G|^3}.$$

We show that in this case, $\mathcal{H}$ has a $> \frac{\delta^2}{10|G|^{100}}$ satisfying assignment which is a contradiction. Consider the term $\mathbf{Term}^{(v,u)}(\alpha, \beta)$ when $\alpha = (\rho_1, \rho_2, \dots, \rho_{R_t})$ and $\beta = (\tau_1, \tau_2, \dots, \tau_{R_{t'}})$.

$$\begin{aligned}
\mathbf{Term}^{(v,u)}(\alpha, \beta) &= \dim(\alpha) \dim(\beta) \mathbf{E}_{\mathbf{b}} \left[\text{tr}(\hat{g}(\alpha) \hat{g}'(\alpha) \alpha(\mathbf{b} \circ \pi)) \cdot \text{tr}(\hat{h}(\beta) \beta(\mathbf{b}^{-1})) \right] \\
&= \dim(\alpha) \dim(\beta) \mathbf{E}_{\mathbf{b}} \left[\sum_{i,k} (\hat{g}(\alpha) \hat{g}'(\alpha))_{ik} \alpha(\mathbf{b} \circ \pi)_{ki} \cdot \sum_{i',k'} \hat{h}(\beta)_{i'k'} \beta(\mathbf{b}^{-1})_{k'i'} \right] \\
&= \dim(\alpha) \dim(\beta) \mathbf{E}_{\mathbf{b}} \left[\sum_{\substack{i,k \\ i',k'}} (\hat{g}(\alpha) \hat{g}'(\alpha))_{ik} \alpha(\mathbf{b} \circ \pi)_{ki} \hat{h}(\beta)_{i'k'} \beta(\mathbf{b}^{-1})_{k'i'} \right] \\
&= \dim(\alpha) \dim(\beta) \sum_{\substack{i,k \\ i',k'}} (\hat{g}(\alpha) \hat{g}'(\alpha))_{ik} \hat{h}(\beta)_{i'k'} \mathbf{E}_{\mathbf{b}} \left[\alpha(\mathbf{b} \circ \pi)_{ki} \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right],
\end{aligned}$$

where (i, k) are the tuples $i = (i_1, i_2, \dots, i_{R_t})$ and $k = (k_1, k_2, \dots, k_{R_t})$ such that for all $\ell \in [R_t]$, $1 \leq i_\ell, k_\ell \leq \dim(\rho_\ell)$. Similarly, (i', k') are the tuples $i' = (i'_1, i'_2, \dots, i'_{R_{t'}})$ and $k' = (k'_1, k'_2, \dots, k'_{R_{t'}})$ such that for all $\ell' \in [R_{t'}]$, $1 \leq i'_{\ell'}, k'_{\ell'} \leq \dim(\tau_{\ell'})$. Now,

$$\begin{aligned}
\mathbf{E}_{\mathbf{b}} \left[\alpha(\mathbf{b} \circ \pi)_{ki} \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right] &= \mathbf{E}_{\mathbf{b}} \left[\prod_{\ell'=1}^{R_{t'}} \left(\tau_{\ell'}(b_{\ell'}^{-1})_{k'_{\ell'} i'_{\ell'}} \cdot \prod_{\ell \in \pi^{-1}(\ell')} \rho_\ell(b_{\ell'})_{k_\ell i_\ell} \right) \right] \\
&= \prod_{\ell'=1}^{R_{t'}} \mathbf{E}_{\mathbf{b}} \left[\tau_{\ell'}(b_{\ell'}^{-1})_{k'_{\ell'} i'_{\ell'}} \cdot \prod_{\ell \in \pi^{-1}(\ell')} \rho_\ell(b_{\ell'})_{k_\ell i_\ell} \right].
\end{aligned}$$

Now suppose there exists an ℓ' such that for all $\ell \in \pi^{-1}(\ell')$, $\dim(\rho_\ell) = 1$. Since the product of dimension 1 representations is also a dimension 1 representation, for the expectation to be nonzero, $\dim(\tau_{\ell'})$ must be 1, by Proposition 2.13. Thus, if $\dim(\beta) \geq 2$, then there must exist an ℓ' such that $\dim(\tau_{\ell'}) \geq 2$ and an $\ell \in \pi^{-1}(\ell')$ such that $\dim(\rho_\ell) \geq 2$ (again, for the expectation to be non-zero). Therefore, we conclude that the terms that are nonzero in the expression $\Theta_{p,q,r}^{(v,u)}(\text{low})$ are all (α, β) such that for all $\ell' \in [R_{t'}]$ whenever $\dim(\tau_{\ell'}) \geq 2$, there exists an $\ell \in \pi^{-1}(\ell')$ such that $\dim(\rho_\ell) \geq 2$. Let us define $\pi_{\geq 2}(\alpha) = \{\ell' \in [R_{t'}] \mid \exists \ell \in \pi^{-1}(\ell'), \dim(\rho_\ell) \geq 2\}$ and $\beta_{\geq 2} = \{\ell' \in [R_{t'}] \mid \dim(\tau_{\ell'}) \geq 2\}$. We get

$$\Theta_{p,q,r}^{(v,u)}(\text{low}) = \sum_{\substack{\alpha, \beta, \\ \dim(\alpha) \geq 2, \dim(\beta) \geq 2, \\ \dim_{\geq 2}(\alpha) \leq C, \\ \beta_{\geq 2} \subseteq \pi_{\geq 2}(\alpha)}} \dim(\alpha) \dim(\beta) \sum_{i,k} (\hat{g}(\alpha) \hat{g}'(\alpha))_{ik} \sum_{i',k'} \hat{h}(\beta)_{i'k'} \mathbf{E}_{\mathbf{b}} \left[\alpha(\mathbf{b} \circ \pi)_{ki} \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right].$$

Now, let $F_\alpha^{ik} : G^{R_{t'}} \rightarrow \mathbb{C}$ be the following function:

$$F_\alpha^{ik}(\mathbf{b}) := \alpha(\mathbf{b}^{-1} \circ \pi)_{ki}.$$

Note that,

$$\sum_k \|F_\alpha^{ik}\|_2^2 = \sum_k \mathbf{E}_{\mathbf{b}} [|\alpha(\mathbf{b}^{-1} \circ \pi)_{ki}|^2] = \mathbf{E}_{\mathbf{b}} \sum_k |\alpha(\mathbf{b}^{-1} \circ \pi)_{ki}|^2 = 1, \quad (9)$$

where the last equality uses the fact that the sum expression is exactly the squared norm of the column i of representation α , which is 1 ($\alpha(\cdot)$ is unitary). We now analyze the expectation.

$$\begin{aligned}
\mathbf{E}_{\mathbf{b}} \left[\alpha(\mathbf{b} \circ \pi)_{ki} \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right] &= \mathbf{E}_{\mathbf{b}} \left[F_{\alpha}^{ik}(\mathbf{b}^{-1}) \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right] \\
&= \mathbf{E}_{\mathbf{b}} \left[\sum_{\beta'} \dim(\beta') \operatorname{tr}(\hat{F}_{\alpha}^{ik}(\beta') \beta'(\mathbf{b}^{-1})^*) \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right] \\
&= \mathbf{E}_{\mathbf{b}} \left[\sum_{\beta'} \dim(\beta') \operatorname{tr}(\hat{F}_{\alpha}^{ik}(\beta') \beta'(\mathbf{b})) \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right] \\
&= \mathbf{E}_{\mathbf{b}} \left[\sum_{\beta'} \dim(\beta') \sum_{i'', k''} \hat{F}_{\alpha}^{ik}(\beta')_{i'', k''} \beta'(\mathbf{b})_{k'', i''} \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right] \\
&= \sum_{\beta'} \dim(\beta') \sum_{i'', k''} \hat{F}_{\alpha}^{ik}(\beta')_{i'', k''} \mathbf{E}_{\mathbf{b}} \left[\beta'(\mathbf{b})_{k'', i''} \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right].
\end{aligned}$$

By Proposition 2.13, the expectation is $1/\dim(\beta')$ if $\beta' = \beta$, $i'' = k'$ and $k'' = i'$; otherwise, it is zero. Therefore,

$$\mathbf{E}_{\mathbf{b}} \left[\alpha(\mathbf{b} \circ \pi)_{ki} \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right] = \hat{F}_{\alpha}^{ik}(\beta)_{k', i'}.$$

Plugging this into $\Theta_{p,q,r}^{(v,u)}(\text{low})$, we get (suppressing the conditions on α, β in the summation for now)

$$\begin{aligned}
\left| \Theta_{p,q,r}^{(v,u)}(\text{low}) \right|^2 &= \left| \sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \sum_{\substack{i, k, \\ i', k'}} (\hat{g}(\alpha) \hat{g}'(\alpha))_{ik} \hat{h}(\beta)_{i'k'} \hat{F}_{\alpha}^{ik}(\beta)_{k', i'} \right|^2 \\
&= \left| \sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \sum_{\substack{i, j, k, \\ i', k'}} \hat{g}(\alpha)_{ij} \hat{g}'(\alpha)_{jk} \hat{h}(\beta)_{i'k'} \hat{F}_{\alpha}^{ik}(\beta)_{k', i'} \right|^2 \\
&\leq \left(\sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \sum_{\substack{i, j, k, \\ i', k'}} |\hat{g}'(\alpha)_{jk}|^2 |\hat{h}(\beta)_{i'k'}|^2 \right) \left(\sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \sum_{\substack{i, j, k, \\ i', k'}} |\hat{g}(\alpha)_{ij}|^2 |\hat{F}_{\alpha}^{ik}(\beta)_{k', i'}|^2 \right).
\end{aligned}$$

We can bound the second term as follows.

$$\begin{aligned}
\left(\sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \sum_{\substack{i, j, k, \\ i', k'}} |\hat{g}(\alpha)_{ij}|^2 |\hat{F}_{\alpha}^{ik}(\beta)_{k', i'}|^2 \right) &= \left(\sum_{\alpha} \dim(\alpha) \sum_{i, j} |\hat{g}(\alpha)_{ij}|^2 \sum_k \sum_{\beta} \dim(\beta) \sum_{i', k'} |\hat{F}_{\alpha}^{ik}(\beta)_{k', i'}|^2 \right) \\
&= \left(\sum_{\alpha} \dim(\alpha) \sum_{i, j} |\hat{g}(\alpha)_{ij}|^2 \sum_k \|F_{\alpha}^{ik}\|_2^2 \right)
\end{aligned}$$

$$\begin{aligned}
&= \left(\sum_{\alpha} \dim(\alpha) \sum_{i,j} |\hat{g}(\alpha)_{ij}|^2 \right) \quad (\text{Using Equation (9)}) \\
&= \|g\|_2^2 \leq 1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left| \Theta_{p,q,r}^{(v,u)}(\text{low}) \right|^2 &\leq \sum_{\substack{\alpha, \beta, \\ \dim(\alpha), \dim(\beta) \geq 2, \\ \dim_{\geq 2}(\alpha) \leq C, \\ \beta_{\geq 2} \subseteq \pi_{\geq 2}(\alpha)}} \dim(\alpha) \dim(\beta) \sum_{\substack{i,j,k, \\ i',k'}} |\hat{g}'(\alpha)_{jk}|^2 |\hat{h}(\beta)_{i'k'}|^2 \\
&\leq |G|^C \sum_{\substack{\alpha, \beta, \\ \dim(\alpha), \dim(\beta) \geq 2, \\ \dim_{\geq 2}(\alpha) \leq C, \\ \beta_{\geq 2} \subseteq \pi_{\geq 2}(\alpha)}} \dim(\alpha) \dim(\beta) \sum_{\substack{j,k, \\ i',k'}} |\hat{g}'(\alpha)_{jk}|^2 |\hat{h}(\beta)_{i'k'}|^2 \\
&= |G|^C \sum_{\substack{\alpha, \beta, \\ \dim(\alpha), \dim(\beta) \geq 2, \\ \dim_{\geq 2}(\alpha) \leq C, \\ \beta_{\geq 2} \subseteq \pi_{\geq 2}(\alpha)}} \dim(\alpha) \dim(\beta) \|\hat{g}'(\alpha)\|_{\text{HS}}^2 \|\hat{h}(\beta)\|_{\text{HS}}^2.
\end{aligned}$$

Here, in the second inequality, the omitted sum over i contributes $\dim(\alpha)$, while $\dim_{\geq 2}(\alpha) \leq C$ implies $\dim(\alpha) \leq \sqrt{|G|}^C \leq |G|^C$. We now show that if $\Theta_{p,q,r}^{(v,u)}(\text{low})$ is large for a typical $\pi_{vu} \in \Pi_{t,t'}$, then it can be used to get a good labeling for the Label Cover instance $\mathcal{H}$.

Randomized labeling. Consider the following randomized labeling. For each $v \in V_t$, consider $g_{pq} : G^{R_t} \rightarrow \mathbb{C}$ which is defined as $g_{pq}(x) := \rho(f_v(x))_{pq}$. Select $\alpha = (\rho_1, \rho_2, \dots, \rho_{R_t})$ with probability $\dim(\alpha) \|\hat{g}_{pq}(\alpha)\|_{\text{HS}}^2$.⁹ Select a uniformly random $\ell_v \in [R_t]$ such that $\dim(\rho_{\ell_v}) \geq 2$ and assign the label ℓ_v to v . If there is no such index, then assign $\perp$.

For each $u \in V_{t'}$, let $h_{qr} : G^{R_{t'}} \rightarrow \mathbb{C}$ be defined as $h_{qr}(y) := \rho(f_u(y))_{qr}$. Select $\beta = (\tau_1, \tau_2, \dots, \tau_{R_{t'}})$ with probability $\dim(\beta) \|\hat{h}_{qr}(\beta)\|_{\text{HS}}^2$. Select a uniformly random $\ell_u \in [R_{t'}]$ such that $\dim(\tau_{\ell_u}) \geq 2$ and assign the label ℓ_u to u . If there is no such index, then assign $\perp$.

Now fix a constraint π_{vu} . The probability, p_{vu} , that this edge is satisfied by the randomized labeling, is lower bounded by:

$$\begin{aligned}
p_{vu} &\geq \sum_{\substack{\alpha, \beta, \\ \dim(\alpha), \dim(\beta) \geq 2, \\ \dim_{\geq 2}(\alpha) \leq C, \\ \beta_{\geq 2} \subseteq \pi_{\geq 2}(\alpha)}} \dim(\alpha) \dim(\beta) \|\hat{g}'(\alpha)\|_{\text{HS}}^2 \|\hat{h}(\beta)\|_{\text{HS}}^2 \cdot \frac{1}{\dim_{\geq 2}(\alpha)} \\
&\geq \frac{1}{C} \sum_{\substack{\alpha, \beta, \\ \dim(\alpha), \dim(\beta) \geq 2, \\ \dim_{\geq 2}(\alpha) \leq C, \\ \beta_{\geq 2} \subseteq \pi_{\geq 2}(\alpha)}} \dim(\alpha) \dim(\beta) \|\hat{g}'(\alpha)\|_{\text{HS}}^2 \|\hat{h}(\beta)\|_{\text{HS}}^2
\end{aligned}$$

⁹As $\|g_{pq}\|_2^2 \leq 1$, by Parseval's identity, the sum of these probabilities is at most 1. Select $\perp$ with the remaining probability to make it a distribution.

$$\geq \frac{1}{C \cdot |G|^C} \left| \Theta_{p,q,r}^{(v,u)}(\text{low}) \right|^2.$$

Therefore, the expected fraction of constraints between the layers t and t' satisfied by the randomized labeling is lower bounded by

$$\begin{aligned} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} [p_{vu}] &\geq \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\frac{1}{C \cdot |G|^C} \left| \Theta_{p,q,r}^{(v,u)}(\text{low}) \right|^2 \right] \\ &= \frac{1}{C \cdot |G|^C} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\left| \Theta_{p,q,r}^{(v,u)}(\text{low}) \right|^2 \right] \\ &\geq \frac{1}{C \cdot |G|^C} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\left| \Theta_{p,q,r}^{(v,u)}(\text{low}) \right| \right]^2 && \text{(Using convexity)} \\ &\geq \frac{1}{C \cdot |G|^C} \cdot \frac{\delta^2}{4|G|^6} \\ &> \frac{\delta^2}{10|G|^{10C}}. \end{aligned}$$

Since the expected fraction of the edges that are satisfied is strictly greater than $\frac{\delta^2}{10|G|^{10C}}$, by conditional expectation, there exists a labeling to the Label Cover instance $\mathcal{H}$ that satisfies more than $\frac{\delta^2}{10|G|^{10C}}$ fraction of the edges between the layers t and t' , which is a contradiction. $\square$

4.1.2 Bounding the high terms

We now show the following claim:

Claim 4.7. (Restatement of Claim 4.5) Suppose $\mathcal{H}$ is at most $\frac{\delta^{10}}{(2|G|)^{100}}$ -satisfiable. Then for every $\rho \in \text{Irrep}(G)$ such that $\dim(\rho) \geq 2$ and $1 \leq p, q, r \leq \dim(\rho)$,

$$\left| \mathbf{E}_{1 \leq t < t' \leq \ell} \mathbf{E}_{\pi_{vu} \in \Pi_{t,t'}} \left[\Theta_{p,q,r}^{(v,u)}(\text{high}) \right] \right| \leq \frac{\delta}{2|G|^3}.$$

Proof. We begin by giving an alternate but equivalent way of sampling the constraint π_{vu} in the expectation from the claim. Sample a random path $\mathbf{p} = (v_1, v_2, \dots, v_\ell)$ as follows. Sample $v_1 \in V_1$ u.a.r. and then for each $2 \leq i \leq \ell$, sample a uniformly random neighbor v_i of v_{i-1} from the graph $G_{i-1,i}$. Next, sample a uniformly random pair (t, t') such that $1 \leq t < t' \leq \ell$ and output the constraint $\pi_{v_t v_{t'}}$. Note that by the property (2) from Theorem 2.4, the above distribution is identical to the distribution of first sampling a random pair (t, t') and then sampling a random constraint from the graph $G_{t,t'}$. Our goal is to show the following bound

$$\Delta := \left| \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_{1 \leq t < t' \leq \ell} \left[\Theta_{p,q,r}^{(v_t, v_{t'})}(\text{high}) \right] \right| \leq \frac{\delta}{2|G|^3}.$$

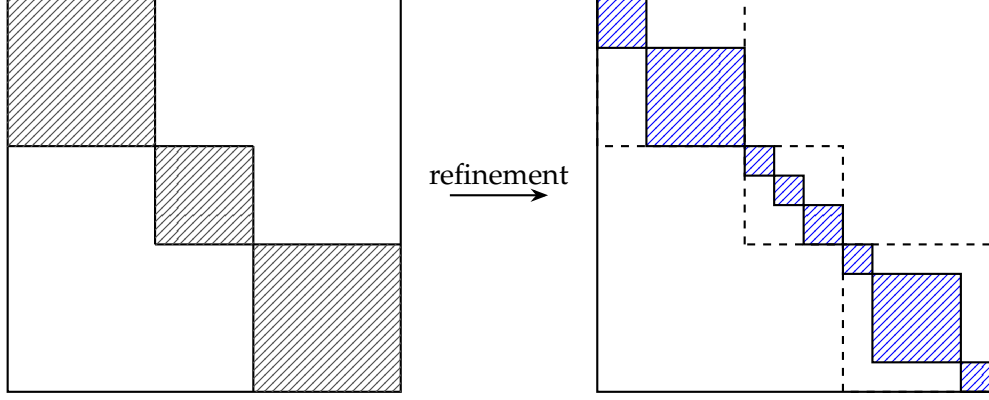


Figure 2: Refinement of a block diagonal matrix

In what follows, we fix a path $p = (v_1, v_2, \dots, v_\ell)$ and the layer t and simplify the expression inside the expectation for this path and the layer t . Recall,

$$\Theta_{p,q,r}^{(v_t, v_{t'})}(\text{high}) := \sum_{\substack{\alpha, \beta, \\ \dim(\alpha), \dim(\beta) \geq 2, \\ \dim_{\geq 2}(\alpha) > C}} \mathbf{Term}^{(v_t, v_{t'})}(\alpha, \beta).$$

For notational convenience, we suppress the conditions on α, β and simply write the sum over pairs α, β . We will analyze the complete sum with the extra conditions on α, β , once we simplify the expression.

Let $U(p, \alpha, t)$ be a transformation (i.e., change of basis) which takes a representation $\alpha(\cdot)$ and converts it into a direct sum of irreducible representations of $\pi_{v_t v_\ell}(G^{R_t}) := \{x \circ \pi_{v_t v_\ell} \mid x \in G^{R_t}\}$ which is a subgroup of G^{R_t} isomorphic to the group G^{R_t} . In other words, if we let $U = U(p, \alpha, t)$, then $U\alpha(x \circ \pi_{v_t v_\ell})U^*$ is a block diagonal matrix for each $x \in G^{R_t}$ with the same block structure.

Although the irreducible isomorphism classes and their multiplicities in $\alpha(\cdot)$ are uniquely determined, the choice of $U(p, \alpha, t)$ is not unique. For our purposes, we choose $U(p, \alpha, t)$ with the following additional structure which will simplify things towards the end of the analysis. We choose $U(p, \alpha, t)$ such that the following holds for every $t < j < \ell$: The block matrices $U\alpha(x \circ \pi_{v_t v_{j+1}})U^*$ corresponding to $x \in G^{R_{j+1}}$ are a *refinement* of the block matrices $U\alpha(x \circ \pi_{v_t v_j})U^*$ corresponding to $x \in G^{R_j}$, as depicted in Figure 2. Note that this is possible from Proposition 2.12 as $\{x \circ \pi_{v_t v_{j+1}} \mid x \in G^{R_{j+1}}\}$ is a subgroup of $\{x \circ \pi_{v_t v_j} \mid x \in G^{R_j}\}$.

For ease of notation, we denote this unitary matrix, which depends on p, α, t , by U . We have

$$\begin{aligned} \text{tr}(\hat{g}(\alpha)\hat{g}'(\alpha)\alpha(\mathbf{b} \circ \pi_{v_t v_{t'}})) &= \text{tr}(U\hat{g}(\alpha)\hat{g}'(\alpha)\alpha(\mathbf{b} \circ \pi_{v_t v_{t'}})U^*) \quad (\text{cyclic property of tr, and } UU^* = I) \\ &= \text{tr}(U\hat{g}(\alpha)\hat{g}'(\alpha)U^*U\alpha(\mathbf{b} \circ \pi_{v_t v_{t'}})U^*). \end{aligned} \quad (10)$$

In this last expression, $U\alpha(\mathbf{b} \circ \pi_{v_t v_{t'}})U^*$ is a block diagonal matrix.

To simplify the notation further, we are going to suppress the use of U and write:

$$U\hat{g}(\alpha) =: A(\alpha), \quad \hat{g}'(\alpha)U^* =: A'(\alpha) \quad \text{and} \quad U\alpha(\mathbf{b} \circ \pi_{v_t v_{t'}})U^* =: B_{t,t'}(\alpha)(\mathbf{b}).$$

Note that by Claim 2.26, the $\|\cdot\|_{\text{HS}}$ of the matrices are preserved, i.e., $\|A(\alpha)\|_{\text{HS}} = \|\hat{g}(\alpha)\|_{\text{HS}}$ and $\|A'(\alpha)\|_{\text{HS}} = \|\hat{g}'(\alpha)\|_{\text{HS}}$. Coming back to the task of simplifying the expression:

$$\begin{aligned}
\Theta_{p,q,r}^{(v_t, v_{t'})}(\text{high}) &= \sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \mathbf{E}_{\mathbf{b}} \left[\text{tr}(\hat{g}(\alpha) \hat{g}'(\alpha) \alpha(\mathbf{b} \circ \pi)) \cdot \text{tr}(\hat{h}(\beta) \beta(\mathbf{b}^{-1})) \right] \\
&= \sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \mathbf{E}_{\mathbf{b}} \left[\text{tr}(U \hat{g}(\alpha) \hat{g}'(\alpha) U^* U \alpha(\mathbf{b} \circ \pi) U^*) \cdot \text{tr}(\hat{h}(\beta) \beta(\mathbf{b}^{-1})) \right] \\
&= \sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \mathbf{E}_{\mathbf{b}} \left[\text{tr}(A(\alpha) A'(\alpha) B_{t,t'}(\alpha)(\mathbf{b})) \cdot \text{tr}(\hat{h}(\beta) \beta(\mathbf{b}^{-1})) \right] \\
&= \sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \mathbf{E}_{\mathbf{b}} \left[\sum_{i,k} (A(\alpha) A'(\alpha))_{ik} \cdot (B_{t,t'}(\alpha)(\mathbf{b}))_{ki} \cdot \sum_{i',k'} \hat{h}(\beta)_{i'k'} \beta(\mathbf{b}^{-1})_{k'i'} \right] \\
&= \sum_{\alpha, \beta} \dim(\alpha) \dim(\beta) \sum_{i,k} (A(\alpha) A'(\alpha))_{ik} \cdot \sum_{i',k'} \hat{h}(\beta)_{i'k'} \mathbf{E}_{\mathbf{b}} \left[(B_{t,t'}(\alpha)(\mathbf{b}))_{ki} \cdot \beta(\mathbf{b}^{-1})_{k'i'} \right]
\end{aligned}$$

Let $\beta_{U, \alpha, t', i}$ denote the irreducible representation of $G^{R_{t'}}$ present in the i^{th} column/row of the block matrix $B_{t,t'}(\alpha)(\cdot)$. Index the individual irreducible blocks of $B_{t,t'}(\alpha)$ by q , and write I_q for the coordinate set of block q . For each block choose its representative $\beta_q \in \text{Irrep}(G^{R_{t'}})$ and a unitary intertwiner S_q so that

$$B_{t,t'}(\alpha)(\mathbf{b})|_{I_q} = S_q \beta_q(\mathbf{b}) S_q^*.$$

The intertwiner may depend on the block and on t' ; it does not change U . Define $K_q := S_q \hat{h}(\beta_q) S_q^*$.

The matrix-coefficient orthogonality in Proposition 2.13, applied in the β_q basis, now yields

$$\dim(\beta) \mathbf{E}_{\mathbf{b}} \left[(B_{t,t'}(\alpha)(\mathbf{b}))_{ki} \text{tr}(\hat{h}(\beta) \beta(\mathbf{b}^{-1})) \right] = \begin{cases} (K_q)_{ki}, & i, k \in I_q \text{ and } \beta = \beta_q, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, a block entry may pair with several matrix entries in the chosen β_q basis; the conjugated coefficient K_q accounts for all of them. If $q(k)$ denotes the unique block containing k ,

$$\Theta_{p,q,r}^{(v_t, v_{t'})}(\text{high}) = \sum_{\substack{\alpha \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \sum_{j,k} A'(\alpha)_{jk} \sum_{i \in I_{q(k)}} A(\alpha)_{ij} (K_{q(k)})_{ki}.$$

By Parseval and the unitary invariance of the Hilbert–Schmidt norm, $\sum_{\alpha} \dim(\alpha) \|A'(\alpha)\|_{\text{HS}}^2 \leq \|g'\|_2^2 \leq 1$. Apply Cauchy–Schwarz to the last display, and for each block put $a_{qj} := (A(\alpha)_{ij})_{i \in I_q}$. Then

$$\begin{aligned}
|\Theta_{p,q,r}^{(v_t, v_{t'})}(\text{high})|^2 &\leq \sum_{\substack{\alpha \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \sum_{j,q} \sum_{k \in I_q} \left| \sum_{i \in I_q} A(\alpha)_{ij} (K_q)_{ki} \right|^2 \\
&= \sum_{\substack{\alpha \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \sum_{j,q} \|K_q a_{qj}\|_2^2
\end{aligned}$$

$$\leq \sum_{\substack{\alpha \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \sum_{j,q} \|K_q\|_{\text{HS}}^2 \sum_{i \in I_q} |A(\alpha)_{ij}|^2.$$

Since, using Claim 2.26, $\|K_q\|_{\text{HS}} = \|\hat{h}(\beta_q)\|_{\text{HS}}$ and $\beta_q = \beta_{U,\alpha,t',i}$ for $i \in I_q$, we get

$$|\Theta_{p,q,r}^{(v_t, v_{t'})}(\text{high})|^2 \leq \sum_{\substack{\alpha \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \sum_{i,j} |(U\hat{g}(\alpha))_{ij}|^2 \cdot \|\hat{h}(\beta_{U,\alpha,t',i})\|_{\text{HS}}^2. \quad (11)$$

We now come back to the expectation that we need to bound. In the following analysis, we split the expectation on t, t' and write those separately (they are still sampled as a uniformly random pair $1 \leq t < t' \leq \ell$). Applying the Cauchy-Schwarz inequality and using the upper bound from (11) above, we get

$$\begin{aligned} \Delta^2 &= \left| \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_{1 \leq t < t' \leq \ell} \left[\Theta_{p,q,r}^{(v_t, v_{t'})}(\text{high}) \right] \right|^2 \\ &\leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\left| \Theta_{p,q,r}^{(v_t, v_{t'})}(\text{high}) \right|^2 \right] \\ &\leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\alpha} \dim(\alpha) \sum_{i,j} |(U\hat{g}(\alpha))_{ij}|^2 \cdot \|\hat{h}(\beta_{U,\alpha,t',i})\|_{\text{HS}}^2 \right] \\ &= \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\alpha} \dim(\alpha) \sum_i \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}(\beta_{U,\alpha,t',i})\|_{\text{HS}}^2 \right], \end{aligned}$$

where we denote the ℓ_2 norm of the i^{th} row of the matrix $U\hat{g}(\alpha)$ by $\|(U\hat{g}(\alpha))_i\|_2$.

By taking a closer look at the expression above, it seems that there can be multiple scenarios when the expression is large. For instance, it might happen that some $\beta_{U,\alpha,t',i}$'s for which $\|\hat{h}(\beta_{U,\alpha,t',i})\|_{\text{HS}}^2$ is large can have small dimension, and in this case, we will not be able to get the advantage that we saw in the dictatorship test. Surprisingly, we will show next that if the expression is large because of $\beta_{U,\alpha,t',i}$'s that have small dimensions, then this will give a strategy to assign labels to the LABEL-COVER instance that will satisfy a non-trivial fraction of the constraints.

We first bound the expression when the β s have a large dimension using the trick we saw in the dictatorship test. Towards this, we recall $C = \frac{8|G|^6}{\delta^2}$.

$$\begin{aligned} \Delta^2 &\leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\alpha} \dim(\alpha) \sum_i \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}(\beta_{U,\alpha,t',i})\|_{\text{HS}}^2 \right] \\ &= \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\alpha} \dim(\alpha) \left(\sum_{\substack{i \\ \dim(\beta_{U,\alpha,t',i}) \leq C}} \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}(\beta_{U,\alpha,t',i})\|_{\text{HS}}^2 \right) \right] + \end{aligned}$$

$$\begin{aligned}
& \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\alpha} \dim(\alpha) \left(\sum_{\substack{i| \\ \dim(\beta_{U, \alpha, t', i}) > C}} \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \right) \right] \\
&= \Delta_1 + \Delta_2,
\end{aligned}$$

where Δ_1 is the former expectation and Δ_2 is the latter expectation. Δ_2 can be upper-bounded as follows.

$$\begin{aligned}
\Delta_2 &= \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\alpha} \dim(\alpha) \left(\sum_{\substack{i| \\ \dim(\beta_{U, \alpha, t', i}) > C}} \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \right) \right] \\
&\leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\alpha} \dim(\alpha) \|\hat{g}(\alpha)\|_{\text{HS}}^2 \left(\sum_{\substack{\beta| \\ \exists i \text{ s.t. } \beta = \beta_{U, \alpha, t', i} \\ \dim(\beta) > C}} \|\hat{h}(\beta)\|_{\text{HS}}^2 \right) \right] \\
&\leq \frac{1}{C} \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\alpha} \dim(\alpha) \|\hat{g}(\alpha)\|_{\text{HS}}^2 \left(\sum_{\beta} \dim(\beta) \|\hat{h}(\beta)\|_{\text{HS}}^2 \right) \right] \\
&\leq \frac{1}{C} \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} [\|\hat{g}\|_2^2 \|\hat{h}\|_2^2] \leq \frac{1}{C}.
\end{aligned}$$

Therefore, we have,

$$\Delta^2 \leq \Delta_1 + \frac{1}{C}.$$

Thus, we are left with bounding Δ_1 . In what follows, we make the function h explicit by writing its subscript. Recall, $h_v(\mathbf{y}) := \rho(f_v(\mathbf{y}))_{qr}$ for a vertex $v \in \mathcal{V}$. We have,

$$\Delta_1 = \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\substack{\alpha| \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \left(\sum_{\substack{i| \\ \dim(\beta_{U, \alpha, t', i}) \leq C}} \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \right) \right]. \quad (12)$$

In order to intuitively make sense of the above expression, here is an alternate way to view the above upper bound. Let α_V be the complex vector space on which α acts. Considering $\alpha(\cdot)$ as a representation of $G^{R_{t'}}$, let β_V be the span of invariant subspaces of α_V , corresponding to (all copies of) the irreducible representation β of $G^{R_{t'}}$. For a vector $w \in \alpha_V$, denote the projection of w onto the subspace β_V by $P_{\beta_V}(w)$. The following is an alternate description of the quantity Δ_1 .

$$\Delta_1 = \mathbf{E}_{\mathbf{p}} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\substack{\beta| \\ \dim(\beta) \leq C}} \|\hat{h}_{v_{t'}}(\beta)\|_{\text{HS}}^2 \sum_{\substack{\alpha| \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \sum_{1 \leq j \leq \dim(\alpha)} \|P_{\beta_V}(j^{\text{th}} \text{ column of } \hat{g}(\alpha))\|_2^2 \right]. \quad (13)$$

To see why this is the same expression as above, fix p, t and t' . For simplicity, assume that $U = I$ and fix a representation $\beta \in \text{Irrep}(G^{R_{t'}})$ and a representation $\alpha \in \text{Irrep}(G^{R_t})$. Now in the expression (12), if we let $\beta_V = \text{span}(\{e_{\ell_1}, e_{\ell_2}, \dots, e_{\ell_m}\})$ be the invariant space of α_V corresponding to (all copies of) the representation β , then $\|\hat{h}_{v_{t'}}(\beta)\|_{\text{HS}}^2$ gets multiplied by the following terms corresponding to α with $\dim_{\geq 2}(\alpha) > C$

$$\begin{aligned} \dim(\alpha) \sum_{k \in [m]} \|\hat{g}(\alpha)_{\ell_k}\|_2^2 &= \dim(\alpha) \sum_{k \in [m]} \sum_{1 \leq j \leq \dim(\alpha)} |\hat{g}(\alpha)_{\ell_k, j}|^2 \\ &= \dim(\alpha) \sum_{1 \leq j \leq \dim(\alpha)} \sum_{k \in [m]} |\hat{g}(\alpha)_{\ell_k, j}|^2 \\ &= \dim(\alpha) \sum_{1 \leq j \leq \dim(\alpha)} \|P_{\beta_V}(j^{\text{th}} \text{ column of } \hat{g}(\alpha))\|_2^2, \end{aligned}$$

as required.

Looking at the expressions (12) and (13), we make a few observations.

1. The quantity inside the expectations is upper-bounded by $\|h_{v_{t'}}\|_2^2 \|g\|_2^2 \leq 1$,
2. Looking at (12), for a fixed α , the invariant spaces β_V are mutually orthogonal for different β 's. Furthermore, β 's for which $\|\hat{h}_{v_{t'}}(\beta)\|_{\text{HS}}^2$ are small do not contribute much to Δ_1 (see (14) below). Now, suppose Δ_1 is large, say at least $\eta > 0$. It then follows from (13) that for at most $O_\eta(1)$ β 's, $\|\hat{h}_{v_{t'}}(\beta)\|_{\text{HS}}^2 \geq \Omega_\eta(1)$, and the contribution from the second sum is $\Omega_\eta(1)$ for these β 's; Call these β 's *heavy*.
3. Finally, if Δ_1 is large, then for a constant fraction of paths $p = (v_1, v_2, \dots, v_\ell)$ and $t \in [\ell]$, for constant fraction of pairs $t < t_1 < t_2 \leq \ell$, the span of invariant spaces for the heavy β 's for v_{t_1} must not be mutually orthogonal to the span of invariant spaces for the heavy β 's for v_{t_2} . Otherwise, in the second sum in (13), we get projections of columns of $\hat{g}(\alpha)$ on mutually orthogonal subspaces as we vary t' , and thus the expectation over t' would make Δ_1 small. As the dimension of β 's under consideration is small, this allows us to deduce consistent labeling for a constant fraction of pairs of layers.

We make the above intuition formal in the following claim.

Claim 4.8. *Let $\eta = \frac{\delta^2}{8|G|^6}$. If $\mathcal{H}$ is less than $\left(\frac{\eta}{2 \log_2 C}\right)^2 \cdot \frac{\eta^2}{64}$ -satisfiable, then $\Delta_1 \leq \eta$.*

Proof. Suppose $\Delta_1 \geq \eta$. From (12), we have

$$\begin{aligned} \eta &\leq \mathbf{E}_{p=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\substack{\alpha \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \left(\sum_{\substack{i \\ \dim(\beta_{U, \alpha, t', i}) \leq C}} \|(U \hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \right) \right] \\ &= \mathbf{E}_{p=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\substack{\alpha \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \left(\sum_i \|(U \hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \cdot \mathbf{1}_{\dim(\beta_{U, \alpha, t', i}) \leq C} \right) \right] \end{aligned}$$

$$\leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\substack{\alpha \mid \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \left(\sum_i \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \cdot \mathbf{1}_{\substack{\dim(\beta_{U, \alpha, t', i}) \leq C \wedge \\ \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \geq \eta/2}} \right) \right] + \frac{\eta}{2}.$$

The last step follows from the following bound

$$\begin{aligned} & \sum_{\substack{\alpha \mid \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \left(\sum_i \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \cdot \mathbf{1}_{\dim(\beta_{U, \alpha, t', i}) \leq C \wedge \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 < \eta/2} \right) \\ & \leq \sum_{\substack{\alpha \mid \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \left(\sum_i \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \frac{\eta}{2} \right) \\ & = \frac{\eta}{2} \sum_{\substack{\alpha \mid \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \left(\sum_i \|(U\hat{g}(\alpha))_i\|_2^2 \right) \\ & = \frac{\eta}{2} \sum_{\substack{\alpha \mid \\ \dim_{\geq 2}(\alpha) > C}} \dim(\alpha) \|U\hat{g}(\alpha)\|_{\text{HS}}^2 \\ & \leq \frac{\eta}{2} \|\mathbf{g}\|_2^2 \leq \frac{\eta}{2}. \end{aligned} \tag{14}$$

Therefore, we have

$$\begin{aligned} \frac{\eta}{2} & \leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \mathbf{E}_{t'} \left[\sum_{\alpha} \dim(\alpha) \left(\sum_i \|(U\hat{g}(\alpha))_i\|_2^2 \cdot \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \cdot \mathbf{1}_{\dim(\beta_{U, \alpha, t', i}) \leq C \wedge \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \geq \eta/2} \right) \right] \\ & \leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \left[\max_{\alpha, i} \left\{ \mathbf{E}_{t'} \left[\|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \cdot \mathbf{1}_{\dim(\beta_{U, \alpha, t', i}) \leq C \wedge \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \geq \eta/2} \right] \right\} \|\mathbf{g}\|_2^2 \right] \\ & \leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \mathbf{E}_t \left[\max_{\alpha, i} \left\{ \mathbf{E}_{t'} \left[\|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \cdot \mathbf{1}_{\dim(\beta_{U, \alpha, t', i}) \leq C \wedge \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \geq \eta/2} \right] \right\} \right] \end{aligned}$$

To reiterate, in the last expression, α varies over all the irreducible representations of G^{R_t} and $1 \leq i \leq \dim(\alpha)$. As the quantity inside the expectation is at most 1, the above implies that there exists a small enough $t^* \leq \ell - \sqrt{\ell}$ such that,

$$\begin{aligned} \frac{\eta}{4} & \leq \frac{\eta}{2} - \frac{1}{\sqrt{\ell}} \leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \left[\max_{\alpha, i} \left\{ \mathbf{E}_{t' > t^*} \left[\|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \cdot \mathbf{1}_{\dim(\beta_{U, \alpha, t', i}) \leq C \wedge \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \geq \eta/2} \right] \right\} \right] \\ & \leq \mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \left[\max_{\alpha, i} \left\{ \mathbf{E}_{t' > t^*} \left[\mathbf{1}_{\dim(\beta_{U, \alpha, t', i}) \leq C \wedge \|\hat{h}_{v_{t'}}(\beta_{U, \alpha, t', i})\|_{\text{HS}}^2 \geq \eta/2} \right] \right\} \right], \end{aligned}$$

where we used the fact that $\sqrt{\ell} \geq \frac{8}{\eta}$. In the above expressions, α refers to a representation of $G^{R_{t^*}}$ and $\beta_{U, \alpha, t', i}$ is the representation in the i^{th} row of the block diagonal matrix $U(\mathbf{p}, \alpha, t^*)\alpha(\cdot)U^*(\mathbf{p}, \alpha, t^*)$.

Now suppose for a path $p = (v_1, v_2, \dots, v_\ell)$ for all α and i there exists at most one $t' > t^*$ such that the condition

$$\mathbf{1}_{\dim(\beta_{U,\alpha,t',i}) \leq C \wedge \|\hat{h}_{v_{t'}}(\beta_{U,\alpha,t',i})\|_{\text{HS}}^2 \geq \eta/2}$$

holds. In this case, the quantity inside the expectation is at most $\frac{1}{\ell - t^*} \leq \frac{1}{\sqrt{\ell}}$. Using this, and the fact that $\sqrt{\ell} \geq \frac{8}{\eta}$, we get

$$\frac{\eta}{8} \leq \frac{\eta}{4} - \frac{1}{\sqrt{\ell}} \leq \mathbf{E}_{p=(v_1, v_2, \dots, v_\ell)} \left[\mathbf{1}_{\dim(\beta_{U,\alpha,t_1,i}) \leq C \wedge \|\hat{h}_{v_{t_1}}(\beta_{U,\alpha,t_1,i})\|_{\text{HS}}^2 \geq \eta/2} \cdot \mathbf{1}_{\dim(\beta_{U,\alpha,t_2,i}) \leq C \wedge \|\hat{h}_{v_{t_2}}(\beta_{U,\alpha,t_2,i})\|_{\text{HS}}^2 \geq \eta/2} \right]$$

By Jensen's inequality and averaging over pairs, there exists $t_2^* > t_1^* > t^*$ such that

$$\frac{\eta^2}{64} \leq \mathbf{E}_{p=(v_1, v_2, \dots, v_\ell)} \left[\exists(\alpha, i), \mathbf{1}_{\dim(\beta_{U,\alpha,t_1^*,i}) \leq C \wedge \|\hat{h}_{v_{t_1^*}}(\beta_{U,\alpha,t_1^*,i})\|_{\text{HS}}^2 \geq \eta/2} \cdot \mathbf{1}_{\dim(\beta_{U,\alpha,t_2^*,i}) \leq C \wedge \|\hat{h}_{v_{t_2^*}}(\beta_{U,\alpha,t_2^*,i})\|_{\text{HS}}^2 \geq \eta/2} \right].$$

Define $\text{supp}_{\geq 2}(\beta) := \{q : \dim(\beta_q) \geq 2\}$. We now fix the above three layers $t_2^* > t_1^* > t^*$ and consider the following two events when sampling a random path $p = (v_1, v_2, \dots, v_\ell)$:

$\mathcal{E}_1$. $\exists(\alpha, i)$, such that $(\dim(\beta_{U,\alpha,t_1^*,i}) \leq C \wedge \|\hat{h}_{v_{t_1^*}}(\beta_{U,\alpha,t_1^*,i})\|_{\text{HS}}^2 \geq \eta/2)$ and

$$(\dim(\beta_{U,\alpha,t_2^*,i}) \leq C \wedge \|\hat{h}_{v_{t_2^*}}(\beta_{U,\alpha,t_2^*,i})\|_{\text{HS}}^2 \geq \eta/2)$$

$\mathcal{E}_2$. $\exists \beta_1, \beta_2$ such that $\dim(\beta_1) \leq C, \dim(\beta_2) \leq C, \|\hat{h}_{v_{t_1^*}}(\beta_1)\|_{\text{HS}}^2 \geq \eta/2, \|\hat{h}_{v_{t_2^*}}(\beta_2)\|_{\text{HS}}^2 \geq \eta/2, \pi_{v_{t_1^*}v_{t_2^*}}(\text{supp}_{\geq 2}(\beta_1)) \cap \text{supp}_{\geq 2}(\beta_2) \neq \emptyset$

We now show that the event $\mathcal{E}_1$ implies the event $\mathcal{E}_2$. Suppose the event $\mathcal{E}_1$ holds with (α, i) . Let $\beta_1 = \beta_{U,\alpha,t_1^*,i}$ and $\beta_2 = \beta_{U,\alpha,t_2^*,i}$ from the event $\mathcal{E}_1$ and furthermore $\beta_1 = (\rho_1, \rho_2, \dots, \rho_{R_{t_1^*}})$ and $\beta_2 = (\tau_1, \tau_2, \dots, \tau_{R_{t_2^*}})$. We show that this (β_1, β_2) satisfies the event $\mathcal{E}_2$. For this, we need to look at the i^{th} row of the block matrices $U\alpha(\mathbf{b} \circ \pi_{v_{t_1^*}v_{t_2^*}})U^*$ for all $\mathbf{b} \in G^{R_{t_2^*}}$ and $U\alpha(\mathbf{a} \circ \pi_{v_{t_1^*}v_{t_2^*}})U^*$ for all $\mathbf{a} \in G^{R_{t_1^*}}$. The i^{th} row of the block matrices $U\alpha(\mathbf{a} \circ \pi_{v_{t_1^*}v_{t_2^*}})U^*$ corresponds to the representation β_1 and i^{th} row of the block matrices $U\alpha(\mathbf{b} \circ \pi_{v_{t_1^*}v_{t_2^*}})U^*$ correspond to the representation β_2 . In other words, if we write β_1 as a direct sum of representations of the subgroup $\pi_{v_{t_1^*}v_{t_2^*}}(G^{R_{t_1^*}})$, then β_2 appears as one of the representations in the direct sum β_1 . For the blocks $\pi_{v_{t_1^*}v_{t_2^*}}^{-1}(i)$ where $i \in [R_{t_2^*}]$ such that $\otimes_{j \in \pi_{v_{t_1^*}v_{t_2^*}}^{-1}(i)} \rho_j$ has dimension 1, the corresponding τ_i will also have dimension 1. It then follows that for every $i \in [R_{t_2^*}]$ such that $\dim(\tau_i) \geq 2$, we have $i \in \pi_{v_{t_1^*}v_{t_2^*}}(\text{supp}_{\geq 2}(\beta_1))$ (one such i exists as $\dim_{\geq 2}(\beta_2) \neq 0$ from Lemma 2.30). Thus $\mathcal{E}_1 \implies \mathcal{E}_2$. Therefore, we have

$$\mathbf{E}_{p=(v_1, v_2, \dots, v_\ell)} [\mathbf{1}(\mathcal{E}_2)] \geq \frac{\eta^2}{64}.$$

Define the following label set for a vertex $v \in V_{t_1^*} \cup V_{t_2^*}$

$$\mathcal{L}[v] = \bigcup \left\{ \text{supp}_{\geq 2}(\beta) \mid \dim(\beta) \leq C \text{ and } \|\hat{h}_v(\beta)\|_{\text{HS}}^2 \geq \frac{\eta}{2} \right\}.$$

As we know that

$$1 \geq \|h_v\|_2^2 = \sum_{\beta} \dim(\beta) \|\hat{h}_v(\beta)\|_{\text{HS}}^2,$$

we have $|\mathcal{L}[v]| \leq \frac{2\log_2 C}{\eta}$ since each β with $\dim(\beta) \leq C$ contributes at most $\log_2 C$ to the set $\mathcal{L}[v]$ and there can be at most $\frac{2}{\eta}$ β s such that $\|\hat{h}_v(\beta)\|_{\text{HS}}^2 \geq \eta/2$.

Consider the following strategy to assign labels to vertices from layers $V_{t_1^*}$ and $V_{t_2^*}$. For each $v \in V_{t_1^*} \cup V_{t_2^*}$, assign a random label from $\mathcal{L}[v]$ to v independently if $\mathcal{L}[v]$ is non-empty; otherwise, assign a random label from the respective set.

The expected fraction of the constraints in $\Pi_{t_1^*, t_2^*}$ satisfied by the above random assignment is lower bounded by the following expression. Note that when $\mathcal{E}_2$ happens then $1 \leq |\mathcal{L}[v_{t_1^*}]|, |\mathcal{L}[v_{t_2^*}]|$; when $\mathbf{1}(\mathcal{E}_2) \equiv 0$, we use the convention that the quantity inside the expectation below is 0.

$$\mathbf{E}_{\mathbf{p}=(v_1, v_2, \dots, v_\ell)} \left[\frac{1}{|\mathcal{L}[v_{t_1^*}]||\mathcal{L}[v_{t_2^*}]|} \mathbf{1}(\mathcal{E}_2) \right] \geq \left(\frac{\eta}{2\log_2 C} \right)^2 \cdot \frac{\eta^2}{64}.$$

For our choice of parameters, $\frac{\delta^{10}}{(2|G|)^{100}} < \left(\frac{\eta}{2\log_2 C} \right)^2 \cdot \frac{\eta^2}{64}$. Since $\mathcal{H}$ is at most $\frac{\delta^{10}}{(2|G|)^{100}}$ -satisfiable, we arrive at a contradiction and therefore $\Delta_1 \leq \frac{\delta^2}{8|G|^6}$ as required. $\square$

Thus, we have

$$\Delta^2 \leq \Delta_1 + \frac{1}{C} \leq \frac{\delta^2}{8|G|^6} + \frac{\delta^2}{8|G|^6} = \frac{\delta^2}{4|G|^6},$$

as required. $\square$

Remark 4.9. We note that the analysis from Section 4.1.2 to bound the high terms is independent of the dimension of α . Therefore, the same analysis would also work to control the low terms. However, we keep the analysis of bounding the low terms currently presented in the paper as is since it is much simpler.

5 Conclusion

In this paper, we showed the optimal inapproximability of Max- k -LIN over every non-abelian group G . The underlying predicate for this constraint satisfaction problem for $k = 3$ is given by the following map, $P_G : G^3 \rightarrow \{0, 1\}$ where $P_G^{-1}(1) = \{(a, b, c) \mid a \cdot b \cdot c = 1_G\}$. The optimal approximation algorithm for this CSP uses the following *abelian embedding* property of the predicate into an abelian group. Consider the map $\alpha : G \rightarrow H$ into the abelian group $H := G/[G, G]$ where $\alpha(g)$ is $g[G, G]$. For every $(a, b, c) \in P_G^{-1}(1)$, we have $\alpha(a) + \alpha(b) + \alpha(c) = 0_H$ where H is the abelian group $G/[G, G]$. Thus, every accepting assignment of the predicate satisfies a linear equation over an abelian group. This property was used towards getting the optimal approximation algorithm with approximation ratio $\frac{1}{|[G, G]|}$ for Max-3-LIN.

Since the publication of this paper, in the subsequent works [BKM22, BKM23, BKM24, BKM25], the notion of abelian embedding of a predicate is used to characterize the optimal approximation

ratio of satisfiable CSPs for certain 3-ary predicates. We believe that *abelian embeddability* is one of the key properties towards getting an optimal approximation algorithm for satisfiable CSPs.

Acknowledgment. We sincerely thank the reviewer of the JACM who pointed out a mistake in the conference version of the paper.

References

- [AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.
- [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 (JACM)*, 45(3):501–555, 1998.
- [AM09] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009.
- [AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM (JACM)*, 45(1):70–122, 1998.
- [BKM22] Amey Bhangale, Subhash Khot, and Dor Minzer. On Approximability of Satisfiable k -CSPs: I. In *54th Annual Symposium on Theory of Computing (STOC’22)*, pages 976–988, New York, NY, USA, 2022. ACM.
- [BKM23] Amey Bhangale, Subhash Khot, and Dor Minzer. On Approximability of Satisfiable K -CSPs: II. In *55th Annual Symposium on Theory of Computing (STOC’23)*, pages 632–642, New York, NY, USA, 2023. ACM.
- [BKM24] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable k -csps: Iv. In *56th Annual Symposium on Theory of Computing (STOC’24)*, pages 1423–1434, New York, NY, USA, 2024. ACM.
- [BKM25] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable csps: V. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC ’25*, page 62–71, New York, NY, USA, 2025. Association for Computing Machinery.
- [BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993.
- [BNP08] László Babai, Nikolay Nikolov, and László Pyber. Product growth and mixing in finite groups. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 248–257, 2008.
- [DGKR05] Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. *SIAM J. Comput.*, 34(5):1129–1146, 2005.

- [DRS05] Irit Dinur, Oded Regev, and Clifford D. Smyth. The hardness of 3-uniform hypergraph coloring. *Comb.*, 25(5):519–535, 2005.
- [EH05] Lars Engebretsen and Jonas Holmerin. Three-query pcps with perfect completeness over non-boolean domains. *Random Structures & Algorithms*, 27(1):46–75, 2005.
- [EHR04] Lars Engebretsen, Jonas Holmerin, and Alexander Russell. Inapproximability results for equations over finite groups. *Theoretical Computer Science*, 312(1):17–45, 2004.
- [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.
- [Gow08] W.T. Gowers. Quasirandom groups. *Combinatorics, Probability and Computing*, 17(3):363–387, 2008.
- [GR02] Mikael Goldmann and Alexander Russell. The complexity of solving equations over finite groups. *Information and Computation*, 178(1):253–262, 2002.
- [Gur04] Venkatesan Guruswami. Inapproximability results for set splitting and satisfiability problems with no mixed clauses. *Algorithmica*, 38(3):451–469, 2004.
- [Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001.
- [Mos10] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010.
- [O’D14] Ryan O’Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.
- [Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.
- [Tan09] Linqing Tang. Conditional hardness of approximating satisfiable Max 3CSP-q. In *International Symposium on Algorithms and Computation*, pages 923–932. Springer, 2009.
- [Ter99] Audrey Terras. *Fourier analysis on finite groups and applications*. Number 43 in London Mathematical Society Student Texts. Cambridge University Press, 1999.

A Proof of Theorem 2.4

Theorem A.1 (Restatement of Theorem 2.4). *For any constant parameters $\ell \geq 2, r \in \mathbb{N}$ the following problem is NP-hard. Given an ℓ -layered Label Cover instance $\mathcal{H} = (\mathcal{V} = \{V_1, \dots, V_\ell\}, \{\Pi_{i,j}\}_{1 \leq i < j \leq \ell}, \{[R_i]\}_{i \in [\ell]})$ where all variable ranges $[R_i]$ are of size $R_i \leq 2^{O(\ell r)}$, distinguish between the following two cases:*

Completeness. *There is an assignment satisfying all the constraints of the Label Cover instance. In this case, we say that $\mathcal{H}$ is fully satisfiable.*

Soundness. For every $1 \leq i < j \leq \ell$, no assignment satisfies more than a 2^{-r} fraction of the set of constraints $\Pi_{i,j}$ between layers i and j . In this case, we say that $\mathcal{H}$ is at most 2^{-r} -satisfiable.

Furthermore, the above hardness result also holds for instances with the following additional structure.

1. *Transitivity of the constraints.* The set of constraints $\Pi_{i,i+1}$ for $1 \leq i \leq \ell - 1$ define all the other constraints by transitivity. In other words, fix any three layers $1 \leq i < j < k \leq \ell$. For any two constraints $\pi_{vu} \in \Pi_{i,j}$ and $\pi_{uw} \in \Pi_{j,k}$, there is a constraint $\pi_{vw} \in \Pi_{i,k}$ such that $\pi_{vw}(a) = \pi_{uw}(\pi_{vu}(a))$ for all $a \in [R_i]$.
2. *Regularity of the constraints.* The constraint graph $G_{i,j}$ is biregular for every $1 \leq i < j \leq \ell$. Moreover, the following distributions on constraints are identical:
 - (a) Choose a layer pair $i < j$ uniformly, then a constraint uniformly from $\Pi_{i,j}$.
 - (b) Choose a full consecutive-layer chain $(x_1, \dots, x_\ell)$ uniformly, independently choose a layer pair $i < j$ uniformly, and output its induced constraint (x_i, x_j) .
3. *Regular projection constraints.* There exists a constant d such that $R_{i-1} = dR_i$ for all $2 \leq i \leq \ell$ and each constraint $\pi_{vu} \in \Pi_{i,j}$ is a $d^{(j-i)}$ -to-1 projection constraint.

Proof. We use the bounded-occurrence, perfect-completeness gap version of NAE-3SAT as a starting point: there are absolute constants $D \geq 1$ and $\eta > 0$ for which it is NP-hard to distinguish satisfiable formulas from formulas of value at most $1 - \eta$, with each clause containing three distinct variables and each variable occurring between 1 and D times. This follows, for example, from the hardness reduction for MAX 3-SET SPLITTING in [Gur04]. The reduction starts from the constraint satisfaction problem MAXSNE₄, whose hardness of approximation is established in [Hås01], even under a bounded-occurrence restriction. Applying the reduction to these restricted instances yields a NAE-3SAT instance with maximum degree bounded by an absolute constant.

We make the instance regular, i.e., every variable appears in the same number of constraints. Let F be the above NAE-3SAT instance, with n variables, m clauses, and occurrence counts a_x for the variable x . Take three disjoint copies $F^{(1)}, F^{(2)}, F^{(3)}$. For every original variable x , add $D - a_x$ copies of the clause

$$\text{NAE}(x^{(1)}, x^{(2)}, x^{(3)}).$$

Call the resulting formula F_{reg} . Each of its $3n$ variables occurs exactly D times, and its number of clauses is

$$3m + \sum_x (D - a_x) = nD.$$

If F is satisfiable by σ , use σ on the first two copies and its Boolean complement on the third. NAE clauses are invariant under complementing all variables, and each added clause has values $(\sigma(x), \sigma(x), 1 - \sigma(x))$. Thus F_{reg} is satisfiable. In the NO case, every assignment violates at least $3\eta m$ clauses among the three original copies. Since $n \leq 3m$, the violated fraction in F_{reg} is at least

$$\frac{3\eta m}{nD} \geq \frac{\eta}{D}.$$

Form the clause–variable Label Cover instance $G_0 = (Y_0, Z_0, E_0)$ of F_{reg} . A left label is one of the six satisfying local assignments of its NAE clause; a right label is a Boolean value. Each incidence constraint projects a local assignment to the value of the corresponding underlying variable. For either prescribed Boolean value, exactly three of the six satisfying local assignments extend it. Hence every projection is exactly 3-to-1, including when the literal is negated.

The graph is simple and $(3, D)$ -biregular. Completeness is perfect. In the NO case, any assignment to the right vertices violates at least an η/D fraction of the clauses. At each such clause, every satisfying local assignment disagrees with the right assignment on at least one of its three incidences. Therefore

$$\text{val}(G_0) \leq 1 - \gamma, \quad \gamma := \frac{\eta}{3D} > 0.$$

Apply t independent parallel repetitions. Since the initial alphabets and gap are absolute constants, Raz's theorem [Raz98] gives an absolute $c > 0$ such that the repeated value is at most 2^{-ct} in the NO case. Choose $t = O(r)$ large enough that this is at most 2^{-r} . Denote the repeated instance by $G = (Y, Z, E)$ and its alphabets by

$$A = [6]^t, \quad B = \{0, 1\}^t.$$

Parallel repetition preserves the biregularity: the two degrees are

$$\Delta_Y = 3^t, \quad \Delta_Z = D^t.$$

Every repeated projection $\rho_{yz} : A \rightarrow B$ is exactly $d := 3^t$ to one. Consequently $|A| = d|B|$, both alphabets have size $2^{O(r)}$, completeness is perfect, and the NO-case value is at most 2^{-r} . This establishes the required small-soundness, biregular base Label Cover.

We now reduce the Label Cover instance to a Layered Label Cover instance.

The Layered Label Cover construction. Following the product construction of [DGKR05], put

$$V_i = Z^i \times Y^{\ell-i}, \quad \mathcal{A}_i = B^i \times A^{\ell-i}, \quad R_i = |\mathcal{A}_i|.$$

Fix bijections between $\mathcal{A}_i$ and $[R_i]$. For $i < j$, include a constraint from $v \in V_i$ to $u \in V_j$ precisely when

$$\begin{cases} v_p = u_p, & p \leq i \text{ or } p > j, \\ (v_p, u_p) \in E, & i < p \leq j. \end{cases}$$

For $a = (a_1, \dots, a_\ell) \in \mathcal{A}_i$, define

$$(\pi_{vu}(a))_p = \begin{cases} \rho_{v_p u_p}(a_p), & i < p \leq j, \\ a_p, & p \leq i \text{ or } p > j. \end{cases}$$

The completeness and soundness of the construction are given in [DGKR05]. Here, we verify the structural properties mentioned in the theorem statement.

1. Suppose $(v, u) \in \Pi_{i,j}$ and $(u, w) \in \Pi_{j,k}$. For $i < p \leq j$, $u_p = w_p$; for $j < p \leq k$, $v_p = u_p$; outside $(i, k]$, all three coordinates agree. The required base edges therefore exist throughout

$(i, k]$, so $(v, w) \in \Pi_{i,k}$. The two maps change disjoint coordinate intervals, giving $\pi_{vw} = \pi_{uw} \circ \pi_{vu}$. Conversely, an edge $(v, w) \in \Pi_{i,k}$ has the unique intermediate vertex

$$u = (w_1, \dots, w_j, v_{j+1}, \dots, v_\ell) \in V_j.$$

Iterating gives its unique consecutive-layer factorization.

2. For a fixed $v \in V_i$, its neighbors in V_j are obtained by independently choosing a base neighbor for each of $v_{i+1}, \dots, v_j$. Its degree is therefore Δ_Y^{j-i} . In the reverse direction, every $u \in V_j$ has degree Δ_Z^{j-i} . Hence $G_{i,j}$ is biregular.

Equality of the two sampling distributions. Every full chain is uniquely specified by

$$z_1 \in Z, \quad (y_p, z_p) \in E \quad (2 \leq p \leq \ell),$$

through

$$x_i = (z_1, \dots, z_i, y_{i+1}, \dots, y_\ell).$$

Thus independent uniform choices of these variables give a uniform full chain. A uniform edge of the biregular base graph has uniform marginals on Y and Z . For each fixed $i < j$, the shared prefix, intervening base edges, and shared suffix of (x_i, x_j) are therefore independent and uniform on

$$Z^i, \quad E^{j-i}, \quad Y^{\ell-j},$$

respectively. Using the fact that the constraints between these layers are in bijection with

$$Z^i \times E^{j-i} \times Y^{\ell-j},$$

for every $(v, u) \in \Pi_{i,j}$ we obtain

$$\Pr[(x_i, x_j) = (v, u)] = \frac{1}{|Z|^i |E|^{j-i} |Y|^{\ell-j}} = \frac{1}{|\Pi_{i,j}|}.$$

An independent uniform layer pair therefore gives each constraint in $\Pi_{i,j}$ probability $\binom{\ell}{2}^{-1} |\Pi_{i,j}|^{-1}$ under both procedures.

Alphabet sizes and reduction size. For every layer,

$$R_i = 2^{ti} 6^{t(\ell-i)} \leq 6^{t\ell} = 2^{O(\ell r)}.$$

All operations are deterministic and have polynomial size for fixed ℓ, r : exact degree regularization has linear size, repetition uses $t = O(r)$ coordinates, and layering uses ℓ coordinates. The integer $d = 3^t$ depends only on the fixed soundness parameter r .

3. Since

$$R_i = |B|^i |A|^{\ell-i}, \quad |A| = d|B|,$$

we have $R_{i-1} = dR_i$. For any target label $b = (b_1, \dots, b_\ell) \in \mathcal{A}_j$, the unchanged coordinates of a preimage under π_{vu} are forced, and each changed coordinate has exactly d choices. Thus

$$|\pi_{vu}^{-1}(b)| = \prod_{p=i+1}^j |\rho_{v_p u_p}^{-1}(b_p)| = d^{j-i}.$$

This proves the theorem and the sampling assertion. □