HARDNESS OF SOLVING SPARSE OVERDETERMINED LINEAR SYSTEMS: A 3-QUERY PCP OVER INTEGERS

VENKATESAN GURUSWAMI AND PRASAD RAGHAVENDRA

Abstract. A classic result due to Håstad established that for every constant $\varepsilon > 0$, given an overdetermined system of linear equations over a finite field $\mathbb{F}_q$ where each equation depends on exactly 3 variables and at least a fraction $(1-\varepsilon)$ of the equations can be satisfied, it is NP-hard to satisfy even a fraction $\left(\frac{1}{q}+\varepsilon\right)$ of the equations.

In this work, we prove the analog of Håstad's result for equations over the integers (as well as the reals). Formally, we prove that for every $\varepsilon, \delta > 0$, given a system of linear equations with integer coefficients where each equation is on 3 variables, it is NP-hard to distinguish between the following two cases: (i) There is an assignment of integer values to the variables that satisfies at least a fraction $(1-\varepsilon)$ of the equations, and (ii) No assignment even of real values to the variables satisfies more than a fraction δ of the equations.

1. Introduction

Solving a system of linear equations over the rationals or reals is a fundamental algorithmic task arising in numerous applications. It is possible to tell in polynomial time, by Gaussian elimination, if a given system admits a solution, and if so to find one. However, Gaussian elimination is not robust against noise, and given an overdetermined system of equations, of which say only 99% of the equations are simultaneously satisfiable, no efficient algorithm for finding a good solution satisfying a good fraction (say 50%) of equations is known. Indeed, it was recently shown that, for any constant $\varepsilon > 0$, given a $(1-\varepsilon)$-satisfiable linear system over the rationals, it is NP-hard to find an assignment to the variables that satisfies even a fraction ε of the equations [9, 8]. A similar hardness result over large finite fields was established in a classic paper by Håstad [11].

This work is motivated by the complexity of solving *sparse* overdetermined linear systems, where each equation is on a small constant number of variables. (The result in [9] applies to linear systems where each equation has a constant $c(\varepsilon)$ number of variables where $c(\varepsilon) \to \infty$ as $\varepsilon \to 0$, and we are interested in the case when each equation has at most an absolute constant, say 3, variables.)

The theory of probabilistically checkable proofs (PCP) has led to immense progress in understanding the approximability of constraint systems where each constraint is local and depends only on a fixed constant number of variables. A celebrated hardness result due to Håstad [11] shows that for every constant $\varepsilon > 0$, given a $(1-\varepsilon)$-satisfiable system of linear equations over a finite field $\mathbb{F}_q$ where each equation depends on at most 3 variables, it is NP-hard to satisfy

Preliminary version appeared in *Proc. of the 39th ACM Symposium on Theory of Computing*, June 2007. Research supported in part by NSF CCF-0343672 and a fellowship from the David and Lucile Packard Foundation.

more than a fraction $\left(\frac{1}{q}+\varepsilon\right)$ of the equations. Underlying this result is a 3-query PCP verifier that queries 3 symbols from purported codewords of the "long code" (a code first defined and considered in [4]) and checks a linear constraint on them, and a tight estimate of the soundness of such a verifier using Fourier analysis. The method of designing long-code based PCP verifiers with tests that closely parallel the underlying constraint in the optimization problem of interest (3-variable linear equations in the above case), and analyzing their performance using Fourier analysis has been highly influential since (for instance, see Khot's survey [12]).

In this work, we prove the analog of Håstad's 3-variable linear equations result for equations over the integers (as well as the reals). Formally, we prove that for every $\varepsilon, \delta > 0$, given a system of linear equations with integer coefficients where each equation is on 3 variables, it is NP-hard to distinguish between the following two cases: (i) There is an assignment of integer values to the variables that satisfies at least a fraction $(1-\varepsilon)$ of the equations, and (ii) No assignment even of real values to the variables satisfies more than a fraction δ of the equations.

We stress that there seems to be no easy reduction from the problem of solving linear equations over finite fields to solving equations over the real numbers. It is straightforward to obtain a hardness result over integers from the hardness result of Håstad [11] over finite fields. Specifically, for every $\mod p$ equation of the form $x+y-z = c \mod p$, introduce an auxiliary variable w and an equation $x + y - z - pw = c$ over integers. However this reduction yields hardness of linear systems with 4 variables per equation instead of 3. More importantly, this reduction does not yield any hardness for linear systems over real numbers.

Obtaining a hardness of approximation result for linear systems with very few variables per constraint was mentioned as an open question in [8]. The result for general linear equations was obtained via a simple reduction from Label Cover in [9], and via a natural tensoring based approach to amplify the gap in [8]. Obtaining a result for 3-variable equations seems harder, and our proof is based on Fourier analysis of a long code based PCP over integers (hence our title for the paper). In Section 2, we present an overview of our proof technique highlighting some of the key challenges in the integers case, our technical contributions to address them, and connections to derandomized linearity testing.

1.1. **Previous related results.** For sparse linear equations over integers, in fact with at most 2 variables per equation, it is shown in [2] (via a reduction from vertex cover on bounded degree graphs) that for some absolute constants $\rho_2 < \rho_1 < 1$, it is NP-hard to tell if such a system is at least ρ_1-satisfiable or at most ρ_2-satisfiable. By boosting this gap using a natural "product" construction, strong hardness results have been shown for the problem (called MAX-SATISFY in the literature) of approximating the number of satisfied equations in an overdetermined system of (not necessarily sparse) linear equations over the rationals [2, 7]. In [7], it is shown that unless NP $\subset$ BPP, for every $\varepsilon > 0$, MAX-SATISFY cannot be approximated within a ratio of $n^{1-\varepsilon}$ where n is the number of equations in the system. (On the algorithmic side, the best approximation algorithm for the problem, due to Halldorsson [10], achieves ratio $O(n/\log n)$.)

However, the product construction destroys the sparsity of the original system, and also reduces the completeness to about ρ_1^k for a k-fold product. Consequently, even without the sparsity requirement, these results do not yield any hardness for near-satisfiable instances where an assignment satisfying a $(1-\varepsilon)$ fraction of the equations is promised to exist (for an arbitrarily small parameter $\varepsilon > 0$). For such near-satisfiable instances, a result showing

NP-hardness of satisfying even an ε fraction of the equations was obtained only recently in [9, 8].

For the complementary objective of minimizing the number of unsatisfied equations, a problem called MIN-UNSATISFY, hardness of approximation within ratio $2^{\log^{0.99} n}$ is shown in [2] (see also [1]).

2. Proof Overview

Our proof follows along the lines of Håstad's result for 3-variable linear equations over prime fields $\mathbb{F}_p$. We give a 3-query PCP verifier that reads 3 appropriately chosen locations of the proof (each of whose entry holds an integer in some finite range) and checks a linear equation on them. The starting point is an instance of Label Cover over a fixed alphabet Σ consisting of a bipartite graph and projection constraints $\pi_e : \Sigma \to \Sigma$ on the edges e; the projection constraint on edge (u, v) imposes the condition $\pi_{(u,v)}(\ell(v)) = \ell(u)$ where $\ell(w)$ is the label assigned to vertex w. The verifier checks satisfiability of the Label cover instance by picking a random edge (u, v) of the Label Cover graph and then checking that the labels assigned to the endpoints of that edge satisfy the projection constraint. To aid the verifier to perform the latter check in a query-efficient way, the prover is expected to write down the *integer long code* encodings (in some large finite range) of all the vertex labels. The verifier picks one location x, with probability $P(x)$ for some distribution P, from the supposed long code A of u's label, and two locations y, y' with probability $P'(y)$ according to distribution P', from the suppose long code B of v's label. (Here $y' = y + x$ is determined once x, y are picked — in the actual test, as in Håstad's test [11], a small noise according to some distribution is added to $y + x$ to get y', and this is crucial. However, for the following description let us pretend that y' is determined once x, y are picked.) The verifier then checks that the values $A(x), B(y)$ and $B(y')$ obey a linear constraint.

Let M be a large enough integer such that the total mass of distributions P and P' outside a cube of dimension M is tiny. Now any test of the above form that works for integers must also work modulo all large enough primes (that are much bigger than the range in which we allow the long code values to lie). In particular, picking p large enough compared to M, we will have a 3-query long code test modulo p that only queries a negligible fraction of the domain $\mathbb{F}_p^\Sigma$ of the long code. Therefore, our results imply a highly derandomized version of Håstad's test (though our target soundness ε is necessarily much larger than $1/p$). In particular, we obtain a test whose total randomness used depends only on the soundness and the dimension, and is independent of the domain size.

Technically, the difficulty imposed by this manifests itself in trying to extend the "decoding" procedure where the tables A and B are used to produce a small list of candidate labels for u and v. Håstad's decoding procedure uses the large Fourier coefficients of A to decode a small list of labels for u. The Fourier transform $\hat{A}_P$ of A with respect to the distribution P can have many large coefficients since P is very far from uniform. In fact, the sum of squares of the Fourier coefficients grows exponentially in the dimension (size of the alphabet). A key technical lemma we show (Lemma 4.3) implies that the Fourier spectrum $\hat{A}_P$ cannot have many large coefficients that are "far-off" from each other. Here the notion of two Fourier coefficients being "far-off" refers to the natural l_∞ metric between the corresponding linear

functions being large. We then show how this can be exploited to decode a small set of labels for u from A (Claim 6.1). A "folding" property of the long code ensures that the set of decoded labels is in fact nonempty (Lemma 5.3). The property of the distribution P needed to show that A has few large pairwise far-off coefficients is an (ε, δ)-*concentration* property, namely $\sum_x P(x)e^{-i\omega\cdot x} \leqslant \varepsilon$ for all $\|\omega\|_\infty \geqslant 2\pi\delta$. Essentially for an (ε, δ)-concentrated distribution P, most of its weight is concentrated around the origin in the Fourier domain.

We are certainly not the first to attempt a derandomization of PCP tests. In particular, we want to point out the work of Ben-Sasson, Sudan, Vadhan, and Wigderson [5] who studied derandomized versions of the BLR linearity test [6] and the low-degree tests underlying PCP constructions. Their derandomized BLR test (for the field $\mathbb{F}_2$) picks a triple $(x, y, y' = y + x)$ of locations to query where y is uniformly distributed on the whole domain $\mathbb{F}_2^{|\Sigma|}$, but x is distributed uniformly on a much smaller subset S of the domain — the only requirement is that S is ε-biased, which means that for all nonzero $\omega \in \{0, \pi\}^{|\Sigma|}$, the Fourier coefficient $\frac{1}{|S|} \cdot \sum_{x\in S} e^{-i\omega\cdot x} \leqslant \varepsilon$. In our terminology, this means that the distribution on x is $(\varepsilon, 1/2)$-concentrated. However this derandomization is inadequate for our case, since y ranges over the entire domain.

It is our hope that ideas from this work might perhaps be useful to reduce the size of long code based PCPs. This could enable giving such PCP constructions for much larger values of parameters, and in turn lead to some improved hardness of approximation results.

3. Our Results

We begin with formal definitions of the problems for which we obtain hardness results. The problem $\mathsf{MAX3LIN}_{\mathbb{Z}}$ consists of finding an assignment that satisfies maximum number of a set of linear equations over integers, each of which has 3 variables. Formally

Definition 3.1. *For constants c, s satisfying $0 \leqslant s < c \leqslant 1$, define $\mathsf{MAX3LIN}_{\mathbb{Z}}(c, s)$ to be the following Promise problem : The input consists of a multiset of linear equations over variables $\{x_1, x_2 \ldots, x_n\}$ with each equation consisting of at most 3 variables. The problem is to distinguish between the following two cases:*

- *There is an integer assignment that satisfies at least a fraction c of the equations.*
- *Every integer assignment satisfies less than a fraction s of the equations.*

$\mathsf{MAX3LIN}_{\mathbb{R}}(c, s)$ is the corresponding problem over real numbers instead of integers.

The hardness results in this paper are obtained by reductions from the Label Cover problem defined below.

Definition 3.2. *An instance of $\mathsf{LABELCOVER}(c, s)$ represented as $\Gamma = (U, V, E, \Sigma, \Pi)$, consists of a bipartite graph over node sets U,V with the edges E between them, such that all nodes in U are of the same degree. Also part of the instance is a set of labels Σ, and a set of mappings $\pi_e : \Sigma \to \Sigma$ for each edge $e \in E$. An assignment Γ of labels to vertices is said to satisfy an edge $e = (u, v)$ where $u \in U$ and $v \in V$, if $\pi_e(A(v)) = A(u)$. The problem is to distinguish between the following two cases:*

- *There exists an assignment* Γ *that satisfies at least a fraction* c *of the edge constraints* Π
- *Every assignment satisfies less than a fraction* s *of the constraints in* Π.

The following strong hardness result for the label cover problem is the starting point for our reductions.

Proposition 3.3. [13, 3] *There exists an absolute constant* $\gamma > 0$ *such that for all large enough* R*, the gap problem* $\mathsf{LABELCOVER}(1, \frac{1}{R^\gamma})$ *is NP-hard, where* $R = |\Sigma|$ *is the size of the alphabet.*

In this paper, we prove the following hardness result for $\mathsf{MAX3LIN}_{\mathbb{Z}}$,

Theorem 3.4 (Main)**.** *For all constants* $\varepsilon, \delta > 0$ *the problem* $\mathsf{MAX3LIN}_{\mathbb{Z}}(1-\varepsilon, \delta)$ *is NP-hard. Further it is NP-hard even when all the equations are of the form* $x_i + x_j = x_k + c$ *for some integer constants* c*.*

It is easy to see that the above result implies a similar hardness result for $\mathsf{MAX3LIN}_{\mathbb{R}}$. The details of the reduction from $\mathsf{MAX3LIN}_{\mathbb{Z}}$ are as follows:

Theorem 3.5. *For all constants* $\varepsilon, \delta > 0$*, the problem* $\mathsf{MAX3LIN}_{\mathbb{R}}(1-\varepsilon, \delta)$ *is NP-hard.*

Proof. Let $\mathcal{I}$ be an instance of $\mathsf{MAX3LIN}_Z(1-\varepsilon, \frac{\delta}{8})$ with the additional restriction that all equations are of the form $x_i + x_j = x_k + c$ for some integer constants c. View this system of equations, as equations over $\mathbb{R}$ to obtain a $\mathsf{MAX3LIN}_{\mathbb{R}}(1-\varepsilon, \delta)$ instance.

In the completeness case, there is an integer assignment that satisfies at least $(1-\varepsilon)$ fraction of the equations. Clearly the same assignment is also a real assignment that satisfies at least $(1-\varepsilon)$ fraction of the equations.

Suppose there is a real assignment $A_{\mathbb{R}}$ that satisfies more than δ fraction of the equations. Obtain an integer assignment $A_{\mathbb{Z}}$ as follows: For each variable x_i, $A_{\mathbb{Z}}(x_i)$ is randomly assigned either $\lceil A_{\mathbb{R}}(x_i) \rceil$ or $\lfloor A_{\mathbb{R}}(x_i) \rfloor$. For every equation $x_i + x_j = x_k + c$ that is satisfied by $A_{\mathbb{R}}$ we have

$$A_{\mathbb{R}}(x_i) + A_{\mathbb{R}}(x_j) - A_{\mathbb{R}}(x_k) = c$$

Since c is an integer, there exists at least one rounding (either ceiling or floor) of $A_{\mathbb{R}}(x_i), A_{\mathbb{R}}(x_j), A_{\mathbb{R}}(x_k)$ such that the above equation continues to hold after rounding. With two choices for each $A_{\mathbb{R}}(x_i)$, there are 8 possible ways to round the 3 variables. Hence with probability at least $\frac{1}{8}$ the equation still holds after rounding. So the expected number of equations satisfied by the rounded solution $A_{\mathbb{Z}}$ is at least $\frac{\delta}{8}$. □

4. Analytic Machinery

4.1. **Fourier Preliminaries.** Let $\mathbb{F}_p$ denote the prime field with p elements. Here we recall the definition of Fourier transform and a few useful identities. For a function $A : \mathbb{F}_p^R \to \mathbb{C}$, define the function $\hat{A}(\omega)$ as follows:

$$\hat{A}(\omega) = \frac{1}{p^R} \sum_{x \in \mathbb{F}_p^R} A(x) e^{-i\omega \cdot x}$$

Hence $\hat{A}(\omega)$ is a function defined over $[0,2\pi]^R$. Let $S_p = \{0, \frac{2\pi}{p}, \ldots, \frac{2\pi j}{p}, \ldots, \frac{2\pi(p-1)}{p}\}$. The values of $\hat{A}(\omega)$ on the finite set S_p^R is the Fourier transform of the function A on $\mathbb{F}_p^R$. Throughout the analysis, we will only be using these Fourier coefficients, i.e., the values of $\hat{A}(\omega)$ on S_p^R. The Fourier coefficients satisfy the following identities:

Inverse Transform:

$$A(x) \quad = \quad \sum_{\omega \in S_p^R} \hat{A}(\omega) e^{i\omega \cdot x}$$

Parseval's identity:

$$\frac{1}{p^R} \sum_{x \in \mathbb{F}_p^R} |A(x)|^2 \quad = \quad \sum_{\omega \in S_p^R} |\hat{A}(\omega)|^2$$

Although we will be applying Fourier Transform over a large prime field $\mathbb{F}_p$, it is instructive to think of the Fourier transform $\hat{A}(\omega)$ as a function over the continuous domain $[0,2\pi]^R$. Operations like addition, subtraction, multiplication by scalars, of elements in $[0,2\pi]^R$ are all done modulo 2π. For instance, if $\omega' = 3\omega$ then the i^{th} coordinate of ω' is given by $\omega'_i = 3\omega_i \mod 2\pi$. For $\theta \in [0,2\pi]$ we will use $\|\theta\|_{2\pi}$ to denote $\min(\theta, 2\pi - \theta)$. For any $\omega \in [0,2\pi]^R$ define $\|\omega\|_\infty = \max_{i \in \{1,\ldots,R\}} \|\omega_i\|_{2\pi}$ This defines a metric on $[0,2\pi]^R$ given by $d(\omega,\omega') = \|\omega - \omega'\|_\infty$ for any two ω, ω'.

We shall denote by $\mathbb{Z}_+$ the set of non negative integers. For a general probability distribution P on $\mathbb{Z}_+^R$, and a function $A : \mathbb{Z}_+^R \to \mathbb{C}$, we define

$$\hat{A}_P(\omega) = \mathop{\mathbf{E}}_{x \in P}[A(x) e^{-i\omega \cdot x}] \tag{1}$$

The numbers $\hat{A}_P(\omega)$ can be thought of as the Fourier coefficients of A with respect to the distribution P. Notice that, if A were a function on $\mathbb{F}_p^R$, and P was the uniform distribution over $\mathbb{F}_p^R$, $A_P(\omega)$ would reduce to the traditional definition of Fourier coefficient of A.

4.2. **(ε,δ)-concentrated distributions.** Let $\hat{\mathbf{1}}$ denote the constant function on $\mathbb{Z}_+^R$ which is always equal to 1. The notion of an (ε,δ)-concentrated distribution is defined as follows:

Definition 4.1. *For $\varepsilon, \delta > 0$, a probability distribution P on $\mathbb{Z}_+^R$ is said to be (ε,δ)-concentrated if $|\hat{\mathbf{1}}_P(\omega)| \leqslant \varepsilon$ for all $\|\omega\|_\infty \geqslant 2\pi\delta$.*

Intuitively a probability distribution is (ε,δ)-concentrated if its Fourier transform is concentrated around the origin. In what follows, we will derive some results on the distribution of large Fourier coefficients $\hat{A}_P(\omega)$ in $[0,2\pi]^R$ if an arbitrary function A, and an (ε,δ)-concentrated distributions P. Let $\ell_2(\mathbb{Z}_+^R)$ denote the vector space of all functions from $\mathbb{Z}_+^R$ to $\mathbb{C}$ such that $\sum_{x \in \mathbb{Z}_+^R} |A(x)|^2 < \infty$. Let $v_1 \cdot v_2 = \sum_{x \in \mathbb{Z}_+^R} v_1(x)\overline{v_2(x)}$ denote the natural inner product for two functions v_1, v_2 in $\ell_2(\mathbb{Z}_+^R)$.

Lemma 4.2. *Let P be a (ε,δ)-concentrated probability distribution. For any $\omega_1, \omega_2 \in [0,2\pi]^R$ such that $\|\omega_1 - \omega_2\|_\infty \geqslant 2\pi\delta$ the functions $v_1(x) = \sqrt{P(x)} e^{i\omega_1 \cdot x}$ and $v_2(x) = \sqrt{P(x)} e^{i\omega_2 \cdot x}$ are nearly orthogonal, i.e., $v_1 \cdot v_2 \leqslant \varepsilon$.*

Proof. We have

$$\begin{aligned} v_1 \cdot v_2 = & \sum_{x \in \mathbb{Z}_+^R} \sqrt{P(x)} e^{i\omega_1 \cdot x} \overline{\sqrt{P(x)} e^{i\omega_2 \cdot x}} \\ = & \hat{\mathbf{1}}_P(\omega_2 - \omega_1) \leqslant \varepsilon \end{aligned}$$

where the last inequality follows from $\|\omega_1 - \omega_2\|_\infty \geqslant 2\pi\delta$ and the fact that P is (ε, δ)-concentrated. □

Let $A : \mathbb{F}_p^R \to \mathbb{C}$ be a function that is bounded, say $|A(x)| = 1$ for all x. By Parseval's identity, the sum of squares of Fourier coefficients $\hat{A}(\omega)$ is 1. In particular, this implies that not more than $\frac{1}{\varepsilon^2}$ of the Fourier coefficients can be more than ε. Now, consider a function $A : \mathbb{Z}_+^R \to \mathbb{C}$ satisfying $|A(x)| = 1$ for all x. The Fourier coefficients $\hat{A}_P(\omega)$ do not satisfy the Parseval's identity. In fact, the sum of the squares of Fourier coefficients could be exponentially large in R, thus giving us no bound on the number of large Fourier coefficients.

However, the following lemma asserts that there cannot be many large Fourier coefficients that are all far from each other. Specifically, although there could be exponentially many ω for which $\hat{A}_P(\omega)$ is large, they are all clustered together in to very few clusters.

Lemma 4.3 (Few far-off Fourier coefficients)**.** *For $0 \leqslant \varepsilon < \frac{1}{9}, \delta > 0$, let P be a (ε^5, δ)-concentrated probability distribution. Let $A : \mathbb{Z}_+^R \to \mathbb{C}$ be a function such that $|A(x)| = 1$ for all $x \in \mathbb{Z}_+^R$. Let $\Omega = \{\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(k)}\} \subset [0, 2\pi]^R$ be a set such that $\|\omega^{(j)} - \omega^{(j')}\|_\infty \geqslant 2\pi\delta$ and $|\hat{A}_P(\omega^{(j)})| \geqslant \varepsilon$ for all j, j'. Then $|\Omega| < \frac{3}{\varepsilon^2}$.*

Proof. On the contrary, let us say there exists a set Ω such that $|\Omega| \geqslant \frac{3}{\varepsilon^2}$. By deleting some elements from the set, we can assume $k = |\Omega| = \frac{3}{\varepsilon^2}$. Consider functions $v(x) = \sqrt{P(x)} A(x)$, $v_j(x) = \sqrt{P(x)} e^{i\omega^{(j)} \cdot x}$ for all $1 \leqslant j \leqslant |\Omega|$. Observe that all of them are unit vectors in $\ell_2(\mathbb{Z}_+^R)$. Since $v \cdot v_j = \hat{A}_P(\omega^{(j)})$ we have $|v \cdot v_j| \geqslant \varepsilon$. Further using Lemma 4.2, we know $|v_j \cdot v_{j'}| \leqslant \varepsilon^5$. Now consider

$$\begin{aligned} |v - \sum_{i=j}^{k} (v \cdot v_j) v_j|^2 &= |v|^2 + \sum_{j=1}^{k} (v \cdot v_j)^2 |v_j|^2 - 2 \sum_{j=1}^{k} (v \cdot v_j)^2 \\ &\quad + 2 \sum_{1 \leqslant j' < j \leqslant k} (v \cdot v_{j'})(v \cdot v_j)(v_{j'} \cdot v_j) \\ &\leqslant 1 - k\varepsilon^2 + 2\binom{k}{2}\varepsilon^5 \end{aligned}$$

Substituting $k = \frac{3}{\varepsilon^2}$, $|v - \sum_{j=1}^{k} (v \cdot v_j) v_j|^2 < 1 - 3 + 18\varepsilon < 0$, a contradiction. Hence we must have $|\Omega| < \frac{3}{\varepsilon^2}$. □

4.3. **An explicit (ε, δ)-concentrated distribution.** It can be shown that the uniform distribution over the cube $[M]^R$ is (ε, δ)-concentrated for a sufficiently large integer M. However we will use the exponential probability distribution to simplify some of the calculations. Formally, define a probability distribution P on $\mathbb{Z}_+^R$ as :

$$P(x) = \gamma e^{-c \sum_{i=1}^{R} x_i} \text{ for some } c > 0 \text{ and } \gamma = (1 - e^{-c})^R \tag{2}$$

The constant γ in the above definition is the correct normalization constant to ensure that P is a distribution. In showing that P has the desired properties, we will use the following fact:

Fact 4.4. *For $c > 0$ and $\omega \in [0, 2\pi]$ the following inequality holds $|1 - e^{-c-i\omega}| \geqslant \frac{2e^{-c}}{\pi}\|\omega\|_{2\pi}$.*

Proof. We have $|1 - e^{-c-i\omega}| \geqslant |e^{-c} - e^{-c-i\omega}| \geqslant e^{-c}|1 - e^{-i\omega}| \geqslant e^{-c}|2\sin\frac{\omega}{2}|$. Using the fact that $|\sin\theta| \geqslant \frac{2\theta}{\pi}$ for $\theta \in [0, \frac{\pi}{2}]$, we conclude

$$|1 - e^{-c-i\omega}| \geqslant \frac{2e^{-c}}{\pi}|\min(\omega, 2\pi - \omega)| = \frac{2e^{-c}}{\pi}\|\omega\|_{2\pi}$$

□

Lemma 4.5. *For all constants $\varepsilon, \delta > 0$ and $0 < c < \ln(1 + 4\delta\varepsilon)$, the distribution P defined in Equation (2) is (ε, δ)-concentrated.*

Proof. Let $\omega \in [0, 2\pi]^R$ be such that $\|\omega\|_\infty \geqslant 2\pi\delta$. In particular, let j_0 be an index such that $\min(\omega_{j_0}, 2\pi - \omega_{j_0}) > 2\pi\delta$.

$$\begin{aligned}
\hat{\mathbf{1}}_P(\omega) &= \sum_{x \in \mathbb{Z}_+^R} P(x) e^{-i\omega\cdot x} \\
&= (1 - e^{-c})^R \prod_{j=1}^{R} \sum_{x_j=0}^{\infty} e^{-cx_j} e^{-i\omega_j x_j} \\
&= \prod_{j=1}^{R} \frac{(1 - e^{-c})}{(1 - e^{-c-i\omega_j})}
\end{aligned}$$

However from Fact 4.4 we know

$$\begin{aligned}
|1 - e^{-c-i\omega_{j_0}}| &\geqslant \frac{2e^{-c}}{\pi}\|\omega_{j_0}\|_{2\pi} \\
&\geqslant 4e^{-c}\delta
\end{aligned}$$

Substituting in the expression for $\hat{\mathbf{1}}_P(\omega)$ we get

$$|\hat{\mathbf{1}}_P(\omega)| = \left(\prod_{j=1, j\neq j_0}^{R} \frac{|(1 - e^{-c})|}{|(1 - e^{-c-i\omega_j})|}\right) \frac{|(1 - e^{-c})|}{|(1 - e^{-c-i\omega_{j_0}})|} \leqslant \frac{|(1 - e^{-c})|}{4e^{-c}\delta} = \frac{e^c - 1}{4\delta}$$

which is less than ε for $c < \ln(1 + 4\delta\varepsilon)$. □

5. Label Cover Test

The reduction from label cover proceeds along the lines of [11]. We will present the reduction as a PCP system for label cover which makes linear tests on three proof locations. The connection to $\mathsf{MAX3LIN}_{\mathbb{Z}}$ will be immediate. Towards this, we define a long code over integers as follows:

Definition 5.1. *The long code for label $i \in \{1, \ldots, R\}$ consists of the function $\mathbf{F}_i : \mathbb{Z}_+^R \to \mathbb{Z}$ defined by $\mathbf{F}_i(x) = x_i$ for all $x \in \mathbb{Z}_+^R$.*

5.1. **Folding.** Denote by $\mathbb{Z}_0^R \subset \mathbb{Z}_+^R$ the set of all points in $\mathbb{Z}_+^R$ with the one of its coordinates equal to zero.

Definition 5.2. *A $\vec{1}$-folded long code is a function $\mathbf{F'}_i : \mathbb{Z}_0^R \to \mathbb{Z}$ defined by $\mathbf{F}'_i(x) = x_i$. More generally, a function $a : \mathbb{Z}_+^R \to \mathbb{Z}$ is a $\vec{1}$-folded function if $a(x + \vec{1}) = a(x) + 1$.*

Given a $\vec{1}$-folded long code $\mathbf{F'}_i$, it is possible to retrieve the value of the full long code at any location x. This is achieved by expressing $x \in \mathbb{Z}_+^R$ as $x = x_0 + t\vec{1}$ where $x_0 \in \mathbb{Z}_0^R$, and then using $\mathbf{F}(x) = \mathbf{F'}(x_0) + t$. By using $\vec{1}$-folded long codes, the reduction ensures that all functions under consideration are $\vec{1}$-folded functions.

If a $\vec{1}$-folded function a is linear, then clearly it must be of the form $a(x) = \sum_{i=1}^R a_i x_i$ where $\sum_{i=1}^R a_i = 1$. The following lemma asserts that the significant Fourier coefficients ω corresponding to an arbitrary $\vec{1}$-folded function a also approximately satisfy $\sum_{i=1}^R \omega_i = 1$.

Lemma 5.3 (Folding lemma)**.** *Let $a : \mathbb{Z}_+^R \to \mathbb{Z}$ be a function such that $a(x + \vec{1}) = a(x) + 1$ for all $x \in \mathbb{Z}_+^R$. Let $A(x) = e^{i\frac{2\pi k a(x)}{p}}$. For all $\delta > 0$ and $c < \frac{1}{R}\ln\left(1 + 4\delta^2\right)$ the following holds : for all $\omega \in [0, 2\pi]^R$ with $\|\omega \cdot \vec{1} - \frac{2\pi k}{p}\|_{2\pi} \geqslant 2\pi\delta$:*

$$|\hat{A}_P(\omega)| \leqslant \delta$$

Proof. Recall that $\mathbb{Z}_0^R \subset \mathbb{Z}_+^R$ denotes the set of all points in $\mathbb{Z}_+^R$ with the one of its coordinates equal to zero. For every $x \in \mathbb{Z}_+^R$, there exists unique $x_0 \in \mathbb{Z}_0^R, t \in \mathbb{Z}$ such that $x = x_0 + t\vec{1}$. By definition of P we have $P(x) = P(x_0)e^{-cRt}$. Hence picking x with probability $P(x)$ is the same as:

- Pick $x_0 \in \mathbb{Z}_0^R$ with probability $\tilde{P}(x_0) = \sum_{t=0}^{\infty} P(x_0 + t\vec{1})$
- Pick t with probability $p(t) = (1 - e^{-cR})e^{-cRt}$

Decompose the expression for $\hat{A}_P(\omega)$ as follows:

$$\begin{aligned}\hat{A}_P(\omega) &= \mathop{\mathbf{E}}_{x\in P}[A(x)e^{-i\omega\cdot x}] \\ &= \mathop{\mathbf{E}}_{x_0\in\tilde{P}}\mathop{\mathbf{E}}_{t\in p}[A(x_0 + t\vec{1})e^{-i\omega\cdot(x_0+t\vec{1})}]\end{aligned}$$

However since $a(x_0 + \vec{1}) = a(x) + 1$, we know $A(x_0 + t\vec{1}) = A(x_0)e^{\frac{2\pi kt}{p}}$. Substituting we get

$$\hat{A}_P(\omega) = \mathop{\mathbf{E}}_{x_0\in\tilde{P}}[A(x_0)e^{-i\omega\cdot x_0}]\mathop{\mathbf{E}}_{t\in p}[e^{\frac{2\pi kt}{p}}e^{-i\omega\cdot t\vec{1}}]$$

Now to compute

$$\begin{aligned}\left|\mathop{\mathbf{E}}_{t\in p}[e^{\frac{2\pi kt}{p}}e^{-i\omega\cdot t\vec{1}}]\right| &= \left|(1 - e^{-cR})\sum_{t=0}^{\infty} e^{-cRt}e^{it(\frac{2\pi k}{p} - \omega\cdot\vec{1})}\right| \\ &= \frac{|(1 - e^{-cR})|}{|1 - e^{-cR+i\Delta}|}\end{aligned}$$

where $\Delta = \frac{2\pi k}{p} - \omega \cdot \vec{1}$. By our assumption $\|\Delta\|_{2\pi} \geqslant 2\pi\delta$, hence using fact 4.4, we get

$$|\mathop{\mathbf{E}}_{t\in p}[e^{\frac{2\pi kt}{p}} e^{-i\omega\cdot t\vec{1}}]| \quad \leqslant \quad \frac{|(1-e^{-cR})|}{4e^{-cR}|\delta|} \leqslant \delta$$

for all $c < \frac{1}{R}\ln\left(1+4\delta^2\right)$. Since $|A(x)| = 1$ for all x, we know $|\mathbf{E}_{x_0\in\tilde{P}}[A(x_0)e^{-i\omega\cdot x_0}]| \leqslant 1$. Together with the bound on $|\mathbf{E}_{t\in p}[e^{\frac{2\pi kt}{p}} e^{-i\omega\cdot t\vec{1}}]|$, this implies the required result. □

5.2. **Verifier.** As defined above, long codes are infinite objects that cannot be written down. Throughout this article, we will be dealing with long codes that are truncated by restricting the domain from $\mathbb{Z}_+^R$ to $[M]^R$ for some large M. However for the purposes of analysis, it is convenient to ignore the truncation and assume that the entire long code is available. As we shall see later, this truncation can be carried out since the verifier queries the values outside a sufficiently large box $[M]^R$ with very low probability.

Let $\Gamma = (U, V, E, \Sigma, \Pi)$ be an instance of Label Cover with $|\Sigma| = R$. Let us assume that labels are indexed by $\{1, \ldots, R\}$. Given an assignment A to the instance Γ, the corresponding PCP proof consists of the $\vec{1}$-folded long codes of the labels assigned to each of the vertices in $U \cup V$. For instance if A is an assignment then for every vertex $u \in U \cup V$ the proof contains the $\vec{1}$-folded long code $\mathbf{F}'_{A(u)}$.

Recall that given a $\vec{1}$-folded long code $\mathbf{F}'_i$, it is possible to retrieve the value of the full long code at any location x. Henceforth, we shall describe the verifier as having access to the full long code of the labels. Clearly the linear tests of the verifier on the full long code can be converted to linear tests on the $\vec{1}$-folded long code.

Given a function $\pi : [R] \to [R]$ and a vector $x \in \mathbb{Z}_+^R$ define $x \circ \pi \in \mathbb{Z}_+^R$ as $(x \circ \pi)_i = x_{\pi(i)}$. Let P and P' be exponential decay probability distributions over $\mathbb{Z}_+^R$ whose parameters will be chosen later. Intuitively, the distribution P will be chosen to be a sufficiently slowly decaying exponential distribution, while the distribution P' decays at a much slower rate than P. The verifier is described below:

3-Query PCP Verifier

(1) Pick a random edge $e = (u, v) \in E$. Let $a : \mathbb{Z}_+^R \to Z$,$b : \mathbb{Z}_+^R \to \mathbb{Z}$ be the long codes corresponding to vertices u, v respectively.
(2) Pick a random $x \in \mathbb{Z}_+^R$ with the distribution P, a random $y \in \mathbb{Z}_+^R$ with the distribution P'
(3) Generate a *noise* vector $\mu \in \mathbb{Z}_+^R$ from the following distribution : Each coordinate μ_i is chosen
   - 0 with probability $(1 - \varepsilon)$.
   - Chosen uniformly at random from $\{1, \ldots, m\}$ with probability ε.
(4) Accept if the following equation holds

$$a(x) = b(x \circ \pi + y + \mu) - b(y)$$

For technical reasons, we will need the following simple lemmas in the soundness analysis.

Lemma 5.4. *The total weight of the distribution P outside the set $[N]^R$ (on the set $\mathbb{Z}_+^R - [N]^R$) is less than δ for $N \geqslant \frac{1}{c} \ln \frac{R}{\delta(1-e^{-c})^R}$*

Proof. We have

$$\begin{aligned} \sum_{x \in \mathbb{Z}_+^R - [N]^R} P(x) &\leqslant \sum_{i=1}^{R} \sum_{x_i \geqslant N} P(x) \\ &\leqslant \frac{Re^{-cN}}{(1-e^{-c})^R} \end{aligned}$$

which is less than δ for $N \geqslant \frac{1}{c} \ln \frac{R}{\delta(1-e^{-c})^R}$. □

Lemma 5.5. *For all $M > 0$, $c \leqslant \frac{\ln 4}{RM}$, for all $x \in [M]^R, y \in \mathbb{Z}_+^R$ the following is true : $P(x+y) \geqslant P(y)/4$*

Proof. Clearly we have

$$\frac{P(x+y)}{P(y)} = e^{-c\sum_i x_i} \geqslant e^{-cRM} \geqslant \frac{1}{4}$$

□

5.3. **Noise Stability.** Notice that in Step (3), the 3-query PCP verifier generates a *noise* vector μ. Finally, instead of querying the location $b(x \circ \pi + y)$, the verifier queries the value of a nearby location $b(x \circ \pi + y + \mu)$.

Introducing *noise* into the locations queried by the verifier is a powerful recurring theme in dictatorship (long code) tests and PCP constructions ever since its use in Håstad [11]. Roughly speaking, using this technique, the verifier can ensure that the function being queried does not depend on too many coordinates. Specifically, if the function b was a long code then $b(x \circ \pi + y + \mu) = b(x \circ \pi + y)$ with high probability over the choice of the noise vector μ. On the other hand, if b is a linear function depending on too many coordinates, then the noise μ would affect the value, thus reducing the probability of success.

Denote by Q the distribution on $\mathbb{Z}_+^R$ of the *noise* vector μ. That is each coordinate of μ is chosen independently to be 0 with probability $(1-\varepsilon)$ and a uniformly random element in $\{1, \ldots, m\}$ with probability ε. Along the lines of Håstad [11], we need to bound the contribution of the Fourier coefficients of b corresponding to linear forms depending on many coordinates. However, in our setting, the coefficients of the linear forms are not discrete. Thus, we say a linear function depends on many coordinates if it has more than C (defined below in Lemma 5.6) *large enough* coefficients. The following lemma will be used in the soundness analysis to bound the contribution of the Fourier coefficients corresponding to these linear functions:

Lemma 5.6. *For all $\varepsilon_1 > 0, 0 < \delta_1 \leqslant \frac{1}{4}$ and constants $m = \lceil \frac{1}{\delta_1} \rceil$, $C = \lceil \log_{1-\frac{\varepsilon}{2}} \varepsilon_1 \rceil$ the following is true: For all $\omega \in [0, 2\pi]^R$ with more than C coordinates ω_i satisfying $\|\omega_i\|_{2\pi} \geqslant 2\pi\delta_1$,*

$$|\hat{\mathbf{1}}_Q(\omega)| \leqslant \varepsilon_1$$

Proof. Let S denote the set of indices $j \in \{1, \ldots, R\}$ such that $\|\omega_j\|_{2\pi} \geqslant 2\pi\delta_1$. Then by definition $|S| \geqslant C$

$$\begin{aligned}
|\hat{\mathbf{1}}_Q(\omega)| &= \left| \sum_{x \in \mathbb{Z}_+^R} Q(x) e^{-i\omega \cdot x} \right| \\
&= \left| \prod_{j=1}^{R} \left[(1-\varepsilon) e^{i\omega_j \cdot 0} + \frac{\varepsilon}{m} \sum_{t=1}^{m} e^{i\omega_j t} \right] \right| \\
&\leqslant \prod_{j=1}^{R} \left[1 - \varepsilon + \frac{\varepsilon}{m} \left| \frac{e^{i\omega_j(m+1)} - e^{i\omega_j}}{e^{i\omega_j} - 1} \right| \right] \\
&\leqslant \prod_{j \in S} \left[1 - \varepsilon + \frac{\varepsilon}{m} \frac{2}{|e^{i\omega_j} - 1|} \right]
\end{aligned}$$

By definition of S, $\|\omega_j\|_{2\pi} > 2\pi\delta_1$ for $j \in S$. Hence using Fact 4.4 with $c = 0$ we get

$$|\hat{\mathbf{1}}_Q(\omega)| \leqslant \prod_{j \in S} \Big(1 - \varepsilon + \frac{\varepsilon}{2m\delta_1} \Big)$$

which for $m \geqslant 1/\delta_1$ and $C \geqslant \log_{1-\varepsilon/2} \varepsilon_1$ is at most $\prod_{j \in S} \left(1 - \varepsilon + \frac{\varepsilon}{2}\right) = \left(1 - \frac{\varepsilon}{2}\right)^C \leqslant \varepsilon_1$. □

6. Proof of main Theorem 3.4

In this section, we will present the proof of Theorem 3.4. Towards this, we first describe the parameters for the verifier in section 5.

Choose an integer $m > \frac{R^2}{\delta}$. Choose a c less than both $\frac{1}{R}\ln(1+4\delta^2)$ and $\ln(1+4(\frac{\delta}{4})^5 \frac{\delta}{2R})$. Denote by P the exponential decay probability distribution with parameter c. In particular, P is $((\frac{\delta}{4})^5, \frac{\delta}{2R})$-concentrated. Let N be the integer obtained from Lemma 5.4, such that weight of P outside $[N]^R$ is less than δ. Let c' be a real number less than $\frac{\ln 4}{R(N+m)}$. Let P' denote the exponential probability distribution with parameter c'.

Completeness : Suppose $\mathcal{A}$ is an assignment that satisfies all the edge constraints Π. The corresponding long code assignment is accepted by the verifier with probability at least $1 - \varepsilon$. For an edge $e = (u, v) \in E$, the verifier rejects the long code assignment only if $\mu_{\mathcal{A}(v)} \neq 0$. It is clear from the choice of μ that this happens with probability exactly $1 - \varepsilon$.

Soundness : Suppose the verifier accepts with probability greater than 19δ. Let $\chi^{uv}(x, y, \mu)$ be the indicator variable that is 1 if the test on edge $e = (u, v)$ succeeds with random choices x, y, μ. Then we can write the probability of acceptance of the test as follows:

$$\mathbf{Pr}[\text{test accepts}] = \mathop{\mathbf{E}}_{u,v} \left[\sum_{\substack{x,y, \\ \mu \in \mathbb{Z}_+^R}} P(x) P'(y) Q(\mu) \chi^{uv}(x, y, \mu) \right] \geqslant 19\delta$$

Notice that the support of the distribution μ is $\{0, 1, \ldots m\}^R$. Further from Lemma 5.4 the total weight of the distribution P outside $[N]^R$ is less than δ. Hence we can truncate the

summation over x and conclude

$$\mathop{\mathbf{E}}_{u,v}\left[\sum_{\substack{x\in[N]^R,\ \mu\in[m]^R\\ y\in\mathbb{Z}_+^R}} P(x)P'(y)Q(\mu)\chi^{uv}(x,y,\mu)\right] \geqslant 18\delta$$

where $[N]^R \subset \mathbb{Z}_+^R$ defined as $[N]^R = \{0,1,\ldots,N\}^R$. Clearly for $x \in [N]^R, \mu \in [m]^R$ the vector $x \circ \pi + \mu \in [N+m]^R$. Recall that the distribution P' is chosen to be sufficiently slowly decaying in comparison to $P(x)$ and $Q(\mu)$. That is by Lemma 5.5 for all $y \in \mathbb{Z}_+^R, z \in [N+m]^R$ we have $P'(y+z) \geqslant \frac{P'(y)}{4}$. In particular, $P'(y+x\circ\pi_{uv}+\mu) \geqslant \frac{P'(y)}{4}$, or equivalently $2\sqrt{P'(y+x\circ\pi_{uv}+\mu)P'(y)} \geqslant P'(y)$. Henceforth we will use y' to denote $y+x\circ\pi_{uv}+\mu$.

Using this inequality in the expression for probability of acceptance we get:

$$\mathop{\mathbf{E}}_{u,v}\left[\sum_{\substack{x\in[N]^R,\mu\in[m]^R\\ y\in\mathbb{Z}_+^R}} P(x)\sqrt{P'(y)P'(y')}Q(\mu)\chi^{uv}(x,y,\mu)\right] \geqslant 9\delta$$

For a prime p define $\chi_p^{uv}(x,y,\mu)$ to be 1 if $a(x)+b(y)-b(x\circ\pi+y+\mu) = 0 \mod p$ and zero otherwise. Clearly $\chi_p^{uv}(x,y,\mu) \geqslant \chi^{uv}(x,y,\mu)$ for all integers x,y,μ. Replacing χ^{uv} by χ_p^{uv} we get:

$$\mathop{\mathbf{E}}_{u,v}\left[\sum_{\substack{x\in[N]^R,\mu\in[m]^R\\ y\in\mathbb{Z}_+^R}} P(x)\sqrt{P'(y)P'(y')}Q(\mu)\chi_p^{uv}(x,y,\mu)\right] \geqslant 9\delta$$

The prime p can be chosen to be sufficiently large so that truncating the summation over y to $[p]^R$ does not alter the probability value significantly. Further, by picking p sufficiently large, it is possible to ensure that the total weight of the distributions P, P', Q outside $[\frac{p}{3}]^R$ is less than δ. Hence computing $y' = y + x\circ\pi_{uv}+\mu$ modulo p is same as computing y' over integers for all but a δ fraction of (x,y,μ). In particular, we can conclude

$$\mathop{\mathbf{E}}_{u,v}\left[\sum_{\substack{x,\mu,\\ y\in[p]^R}} P(x)\sqrt{P'(y)P'(y')}Q(\mu)\chi_p^{uv}(x,y,\mu)\right] \geqslant 8\delta \tag{3}$$

where $y' = y + x\circ\pi_{uv}+\mu$ is computed modulo p. Notice that the parameter p is an artifact in the analysis, and is chosen to be sufficiently large compared to all other parameters. It is instructive to think of p as tending to infinity while all other parameters are fixed.

Now we fix an edge $e = (u,v)$ and analyze the probability that the test succeeds. Let π denote the projection constraint on the edge e. The following is an arithmetization for χ_p^{uv} :

$$\chi_p^{uv}(x,y,\mu) = \frac{1}{p}\sum_{k=0}^{p-1}\beta^{k[a(x)+b(y)-b(x\circ\pi+y+\mu)]}$$

where $\beta = e^{\frac{2\pi i}{p}}$. Now we define the following notation:

$$A(x) = \beta^{a(x)} \qquad B(x) = \beta^{b(x)}$$
$$\mathbf{A}^k(x) = P(x)\beta^{ka(x)} \qquad \mathbf{B}(x) = \sqrt{P'(x)}\beta^{kb(x)}$$

Substituting the above expressions in (3) we get:

$$\mathop{\mathbf{E}}_{u,v}\left[\frac{1}{p}\sum_{k=0}^{p-1}\sum_{\substack{x,\mu,\\ y\in[p]^R}} Q(\mu)\mathbf{A}^k(x)\mathbf{B}^k(y)\overline{\mathbf{B}^k(y')}\right] \geqslant 8\delta\ . \tag{4}$$

Given an $\omega \in [0,2\pi]^R$ and a function $\pi : [R] \to [R]$, the vector $\pi(\omega) \in [0,2\pi]^R$ is defined by $(\pi(\omega))_i = \sum_{j\in\pi^{-1}(i)} \omega_j$. The expression inside the expectation in (4) is similar to the one obtained in [11], and using a standard computation over $\mathbb{F}_p$ it can be written in terms of the Fourier coefficients. For the sake of completeness, we include the details below. The expression within the expectation in (4) is equal to

$$\frac{1}{p}\sum_{k=0}^{p-1}\sum_{\substack{x,\mu,\\ y\in[p]^R}} Q(\mu)\sum_{\omega_1\in S_p^R}\hat{\mathbf{A}^k}(\omega_1)e^{i\omega_1\cdot x}\sum_{\omega_2\in S_p^R}\hat{\mathbf{B}^k}(\omega_2)e^{i\omega_2\cdot y}\overline{\sum_{\omega_3\in S_p^R}\hat{\mathbf{B}^k}(\omega_3)e^{i\omega_3\cdot(x\circ\pi+y+\mu)}}$$

$$= \frac{1}{p}\sum_{k=0}^{p-1}\sum_{\omega_1,\omega_2,\omega_3\in S_p^R}\hat{\mathbf{A}^k}(\omega_1)\hat{\mathbf{B}^k}(\omega_2)\overline{\hat{\mathbf{B}^k}}(\omega_3)\sum_{\mu\in[p]^R}Q(\mu)e^{-i\omega_3\cdot\mu}\sum_{x\in[p]^R}e^{i(\omega_1-\pi(\omega_3))\cdot x}\sum_{y\in[p]^R}e^{i(\omega_2-\omega_3)\cdot y}$$

Since $\omega_1,\omega_2,\omega_3 \in S_p^R$, we have

$$\begin{aligned}\sum_{x\in[p]^R} e^{i(\omega_1-\pi(\omega_3))\cdot x} &= 0 \text{ unless } \omega_1 = \pi(\omega_3)\\ \sum_{y\in[p]^R} e^{i(\omega_2-\omega_3)\cdot y} &= 0 \text{ unless } \omega_2 = \omega_3\end{aligned}$$

Using these relations in the expression, and renaming ω_3 to be ω we get

$$\frac{1}{p}\sum_{k=0}^{p-1}\sum_{\omega\in S_p^R}\left(p^R\hat{\mathbf{A}^k}(\pi(\omega))\right)\left(p^R\left|\hat{\mathbf{B}^k}(\omega)\right|^2\right)\left(\sum_{\mu\in[p]^R}Q(\mu)e^{-i\omega\cdot\mu}\right)$$

Recall that for $Q(\mu) = 0$ for all $\mu \notin [m]^R$, hence for $p > m$ we have $\sum_{\mu\in[p]^R} Q(\mu)e^{-i\omega\cdot\mu} = \hat{\mathbf{1}}_Q(\omega)$. Therefore we have

$$\frac{1}{p}\sum_{k=0}^{p-1}\mathop{\mathbf{E}}_{u,v}\left[\sum_{\omega\in S_p^R}\left(p^R|\hat{\mathbf{B}^k}(\omega)|^2\right)\left|p^R\hat{\mathbf{A}^k}(\pi(\omega))\hat{\mathbf{1}}_Q(\omega)\right|\right] \geqslant 8\delta \tag{5}$$

From Parseval's identity we have,

$$p^R\sum_{\omega\in S_p^R}|\hat{\mathbf{B}^k}(\omega)|^2 = \sum_{x\in[p]^R}|\sqrt{P'(x)}\beta^{kb(y)}|^2 \leqslant 1 \tag{6}$$

Further we have $|p^R\hat{\mathbf{A}^k}(\omega)|, |\hat{\mathbf{1}}_Q(\omega)| \leqslant 1$ for all ω. Hence for all k

$$\sum_{\omega \in S_p^R} \left(p^R |\hat{\mathbf{B}^k}(\omega)|^2\right) \left|p^R \hat{\mathbf{A}^k}(\pi(\omega))\hat{\mathbf{1}}_Q(\omega)\right| \leqslant 1$$

The inequality (5) asserts that the average of p such terms is larger than 8δ. By an averaging argument, there exists $2\delta p \leqslant k \leqslant p(1-2\delta)$ such that

$$\mathop{\mathbf{E}}_{u,v}\left[\sum_{\omega \in S_p^R} \left(p^R |\hat{\mathbf{B}^k}(\omega)|^2\right) \left|p^R \hat{\mathbf{A}^k}(\pi(\omega))\hat{\mathbf{1}}_Q(\omega)\right|\right] \geqslant 4\delta$$

Fix some such k for the rest of the argument. Observe that

$$p^R \hat{\mathbf{A}^k}(\pi(\omega)) = \sum_{x \in [p]^R} P(x)A^k(x)e^{-i\pi(\omega)\cdot x}$$

By Definition 1, the Fourier coefficient $\hat{A}_P^k(\pi(\omega))$ with respect to distribution P is given by

$$\hat{A}_P^k(\pi(\omega)) = \sum_{x \in \mathbb{Z}_+^R} P(x)A^k(x)e^{-i\pi(\omega)\cdot x} \ .$$

For sufficiently large choice of the prime p, we have

$$|p^R \hat{\mathbf{A}^k}(\pi(\omega)) - \hat{A}_P^k(\pi(\omega))| \leqslant \delta \ .$$

Substituting $p^R \hat{\mathbf{A}^k}(\pi(\omega))$ by $\hat{A_P^k}(\pi(\omega))$ and using equation 6 we get

$$\mathop{\mathbf{E}}_{u,v}\left[\sum_{\omega \in S_p^R} \left(p^R |\hat{\mathbf{B}^k}(\omega)|^2\right) \left|\hat{A_P^k}(\pi(\omega))\hat{\mathbf{1}}_Q(\omega)\right|\right] \geqslant 3\delta \tag{7}$$

6.1. **Restricting to "sparse" Fourier coefficients.** The expectation (7) above looks similar to the expression that is used to derive labels in Håstad's work on 3-variable linear equations modulo 2. This latter expression is of the form $\sum_\beta |\hat{B}(\beta)|^2 |\hat{A}_{\pi_2(\beta)}|$ summed over all β of small size — see [11] for details. Along the lines of [11], we will use the Fourier coefficients in the above expression to obtain a decoding of labels to the vertices u, v. Roughly speaking, Håstad's decoding proceeds as follows:

> For each vertex $v \in U \cup V$, sample a sparse Fourier coefficient ω from an appropriate distribution, and sample uniformly random non-zero coordinate of ω. Assign to vertex v the label corresponding to the coordinate.

The Fourier coefficients ω in our case do not take discrete values. Although for the purposes of analysis we have used ω in a discrete set S_p^R, recall that p is chosen to be sufficiently large compared to every other parameter including R. In fact, it is instructive to think of $p \to \infty$ while all other parameters stay fixed.

In the continous setting, the notion of a *sparse* Fourier coefficient ω needs to be redefined. Specifically, a sparse Fourier coefficient ω would have a few large coordinates ω_i, while the remaining coordinates are small in absolute value. To this end, we define two subsets $\Omega_1, \Omega_2 \subset S_p^R$ as follows:

- Ω_1 : set of ω such that $\|\omega \cdot \vec{1}\|_{2\pi} \geqslant 2\pi\delta$. In other words, for every $\omega \in \Omega_1$ there is at least one *large* coordinate, i.e, a coordinate ω_i with $\|\omega_i\|_{2\pi} \geqslant \frac{2\pi\delta}{R}$.
- Ω_2 : subset of ω which have very few *large* coordinates. In particular, for all $\omega \in \Omega_2$ at most C of its coordinates satisfy $\|\omega_i\|_{2\pi} \geqslant \frac{2\pi\delta}{R^2}$. (Here C is the constant from Lemma 5.6.)

Here $\Omega_1 \cap \Omega_2$ would be the set of *sparse* Fourier coefficients for our purpose.

Firstly, we will bound the contribution of Fourier coefficients with no large coordinate using Lemma 5.3. This corresponds to bounding the contribution of trivial Fourier coefficient in [11]. Notice that $\omega \cdot \vec{1} = \pi(\omega) \cdot \vec{1}$. Hence for $\omega \notin \Omega_1$, $\|\pi(\omega) \cdot \vec{1} - \frac{2\pi k}{p}\|_{2\pi} \geqslant \|2\pi\delta - \frac{2\pi(2\delta p)}{p}\|_{2\pi} \geqslant 2\pi\delta$. From Lemma 5.3 and choice of distribution P, $|\hat{A^k_P}(\pi(\omega))| < \delta$ when $\|\pi(\omega) \cdot \vec{1} - \frac{2\pi k}{p}\|_{2\pi} \geqslant 2\pi\delta$. This implies that $|\hat{A^k_P}(\pi(\omega))| < \delta$ for all $\omega \notin \Omega_1$.

$$\mathop{\mathbf{E}}_{u,v}\left[\sum_{\omega\in\Omega_1}\left(p^R|\hat{\mathbf{B}^k}(\omega)|^2\right)\left|\hat{A^k_P}(\pi(\omega))\hat{\mathbf{1}}_Q(\omega)\right|\right] \geqslant 2\delta$$

To bound the contribution of Fourier coefficients with too many large coordinates, we will use the noise μ introduced by the verifier. More precisely, we have $|\hat{\mathbf{1}}_Q(\omega)| \leqslant \delta$ for all $\omega \notin \Omega_2$ from Lemma 5.6. Therefore,

$$\mathop{\mathbf{E}}_{u,v}\left[\sum_{\omega\in\Omega_1\cap\Omega_2}\left(p^R|\hat{\mathbf{B}^k}(\omega)|^2\right)\left|\hat{A^k_P}(\pi(\omega))\right|\right] \geqslant \delta \tag{8}$$

We will next see how one can decode labels satisfying many Label Cover constraints based on (8).

6.2. **Decoding Label Sets.** For $\omega \in [0, 2\pi]^R$ and $\delta > 0$, let $L_\delta(\omega) \subseteq [R]$ denote the subset of indices ω_i such that $\|\omega_i\|_{2\pi} \geqslant 2\pi\delta$.

For every vertex $v \in V$ with the corresponding Fourier transform $\hat{\mathbf{B}^k}$, define P_v to be the distribution obtained by normalizing $p^R|\hat{\mathbf{B}^k}(\omega)|^2$. Since $\sum_{\omega\in S^R_p} p^R|\hat{\mathbf{B}^k}(\omega)|^2 \leqslant 1$, $P_v = \gamma p^R|\hat{\mathbf{B}^k}(\omega)|^2$ for some $\gamma \geqslant 1$. For a vertex $u \in U$ with the corresponding Fourier transform $\hat{A^k_P}$, define the set Ω_A of significant frequencies as follows:

$$\Omega_A = \left\{\omega \in \Omega_1 \cap \Omega_2 \ : \ |\hat{A^k_P}(\omega)| \geqslant \frac{\delta}{4}\right\} . \tag{9}$$

Define the set $L(u)$ as follows:

$$L(u) = \bigcup_{\omega\in\Omega_A} L_{\frac{\delta}{R}}(\omega) \ . \tag{10}$$

Intuitively $L(u)$ is the set of all *large* coordinates of those ω for which the Fourier coefficient $|\hat{A^k_P}(\omega)|$ is *large*. The decoding algorithm proceeds as follows:

- For $v \in V$, pick a $\omega \in S^R_p$ with probability P_v. Assign a label uniformly at random from $L_{\frac{\delta}{R^2}}(\omega)$ if it is nonempty, else assign a random label.

- For every vertex $u \in U$, assign a label uniformly at random from $L(u)$ if it is nonempty, else assign a random label.

Every Fourier coefficient $\omega \in \Omega_A$ is *sparse* in that it at most C large coordinates. A trivial bound on the size of $L(u)$ is given by $C \cdot |\Omega_A|$. In Håstad's work [11], this bound suffices since the size of Ω_A is bounded using Parseval's identity. The main technical challenge in our setting is that $\sum_\alpha |\hat{A^k_P}(\alpha)|^2$, the sum of squared Fourier coefficients with respect to the distribution P, could be very large. In particular, bounding this sum by Parseval's we get

$$\sum_\alpha \left(p^R |\hat{\mathbf{A}^k}(\alpha))|\right)^2 \leqslant p^R \sum_x P(x)^2 \ .$$

When $P(x)$ is uniform, i.e., $P(x) = 1/p^R$ for every x, this bound equals 1, but the bound could be exponentially larger in R for distributions P that are very non-uniform (as in our case). Thus the obvious extension of Håstad's argument will lead to a list size bound that is too large to be useful as a decoding strategy.

Although the size of Ω_A could be exponentially large, Lemma 4.3 shows that the large Fourier coefficients are all clustered in to a few clusters. Using this property, we obtain the following bound on the size of $L(u)$.

Claim 6.1. *For every vertex $u \in U$, the cardinality of the set $L(u)$ is at most $\frac{48C}{\delta^2}$.*

Proof. Recall that by definition, every $\omega \in \Omega_2$ has at most C coordinates ω_i satisfying $\|\omega_i\|_{2\pi} \geqslant \frac{2\pi\delta}{R^2}$. Hence for all $\omega \in \Omega_1 \cap \Omega_2$ each of the sets $L_{\frac{\delta}{R}}(\omega)$ and $L_{\frac{\delta}{2R}}(\omega)$ have a cardinality of at most C.

Suppose the assertion of the claim is false. We will inductively produce a *large* set of distant ω, for all of which $\hat{A}^k_P(\omega)$ is *large*. This will contradict the Lemma 4.3 since the distribution P is concentrated.

Construct the set $\Omega' \subset \Omega_A$ iteratively as follows: To start with pick an $\omega^{(1)} \in \Omega_A$. After $t \geqslant 1$ steps, let $L_t = \cup_{i=1}^t L_{\frac{\delta}{2R}}(\omega^{(i)})$. Since each $L_{\frac{\delta}{2R}}$ has at most C elements, the cardinality of L_t is at most $C \cdot t$. Since $L(u) > \frac{48C}{\delta^2}$, when $t \leqslant \frac{48}{\delta^2}$ we have $|L(u)| > |L_t|$. In particular, there exists some $\omega^{(t+1)} \in \Omega_A$ such that the set $L_{\frac{\delta}{R}}(\omega^{(t+1)}) - L_t$ is nonempty. Let us assume $j \in L_{\frac{\delta}{R}}(\omega^{(t+1)}) - L_t$. For any $1 \leqslant i \leqslant t$, the distance $\|\omega^{(i)} - \omega^{(t+1)}\|_\infty \geqslant \|\omega_j^{(t+1)} - \omega_j^{(i)}\|_{2\pi}$. Since $j \in L_{\frac{\delta}{R}}(\omega^{(t+1)}) - L_t$, we have $\|\omega_j^{(t+1)}\|_{2\pi} \geqslant \frac{2\pi\delta}{R}$ and $\|\omega_j^{(i)}\|_{2\pi} \leqslant \frac{2\pi\delta}{2R}$. Hence the distance $\|\omega^{(i)} - \omega^{(t+1)}\|_\infty$ is at least $\frac{2\pi\delta}{2R}$.

By iterating the above process, it is possible to construct a set $\Omega' \subseteq \Omega_A$ with cardinality at least $\frac{48}{\delta^2}$ such that for all $\omega^{(i)}, \omega^{(j)} \in \Omega'$, $\|\omega^{(i)} - \omega^{(j)}\|_\infty \geqslant \frac{2\pi\delta}{2R}$. This will contradict Lemma 4.3, since P is a $(\left(\frac{\delta}{4}\right)^5, \frac{\delta}{2R})$-concentrated. □

6.3. **Soundness analysis wrap-up using the label sets.** By an averaging argument applied to (8), at least for a fraction $\frac{\delta}{2}$ of the edges the following inequality holds:

$$\sum_{\omega \in \Omega_1 \cap \Omega_2} \left(p^R |\hat{\mathbf{B}^k}(\omega)|^2 \right) \left| \hat{A_P^k}(\pi(\omega)) \right| \geqslant \frac{\delta}{2}$$

We refer to these edges (u, v) as *good* edges. Consider a good edge $e = (u, v)$. On choosing ω over the probability distribution $P_v(\omega)$ with probability at least $\frac{\delta}{4}$ we have $|\hat{A_P^k}(\pi_{uv}(\omega))| \geqslant \frac{\delta}{4}$ and $\omega \in \Omega_1 \cap \Omega_2$. Since $\omega \in \Omega_1$ we have $\|\pi(\omega) \cdot 1\|_{2\pi} \geqslant 2\pi\delta$. Consequently, there have to be large coordinates of $\pi(\omega)$, i.e., there must exist $i \in [R]$ such that $\|[\pi(\omega)]_i\|_{2\pi} \geqslant \frac{2\pi\delta}{R}$. Suppose $i \in L_{\frac{\delta}{R}}(\pi(\omega))$ is a large coordinate of $\pi(\omega)$ then there must be a large coordinate of ω in $\pi^{-1}(i)$, i.e., a $j \in \pi^{-1}(i)$ such that $\|\omega_j\|_{2\pi} \geqslant \frac{2\pi\delta}{R^2}$. Recall that $\omega \in \Omega_2$ has at most C large coordinates. Therefore with probability at least $\frac{1}{C}$, the vertex v is assigned label j. Further using Claim 6.1, we conclude that vertex u is assigned label i with probability at least $\frac{\delta^2}{48C}$. The edge (u, v) is satisfied when u is assigned i and v is assigned j. Hence the edge e is satisfied with probability at least $\frac{\delta}{4} \cdot \frac{1}{C} \cdot \frac{\delta^2}{48C} = \frac{\delta^3}{192C^2}$. As there are at least a fraction $\frac{\delta}{2}$ of *good* edges, the expected fraction of edges satisfied is at least $\frac{\delta^4}{384C^2}$ which is greater than $\frac{1}{R^\gamma}$ for large enough R.

We have thus shown that the 3-query PCP has completeness $(1 - \varepsilon)$ and soundness at most 19δ. The tests it makes are linear equations. Therefore, we immediately get that the promise problem $\mathsf{MAX3LIN}_{1-\varepsilon,19\delta}$ is NP-hard. Since $\varepsilon, \delta > 0$ are arbitrary, the proof of Theorem 3.4 is complete.

Acknowledgments

We thank Johan Håstad and the anonymous reviewers for very useful feedback which helped us improve the presentation of the paper.

References

[1] E. Amaldi and V. Kann. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, 109:237–260, 1998.

[2] S. Arora, L. Babai, J. Stern, and Z. Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. *Journal of Computer System Sciences*, 54(2):317–331, 1997.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[4] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCP and non-approximability - towards tight results. *SIAM Journal on Computing*, 27(3):804–915, June 1998.

[5] E. Ben-Sasson, M. Sudan, S. P. Vadhan, and A. Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 612–621, 2003.

[6] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, 47(3):549–595, 1993.

[7] U. Feige and D. Reichman. On the hardness of approximating Max-Satisfy. *Information Processing Letters*, 97(1):31–35, 2006.

[8] V. Feldman, P. Gopalan, S. Khot, and A. K. Ponnuswami. New results for learning noisy parities and halfspaces. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 563–572, 2006.
[9] V. Guruswami and P. Raghavendra. Hardness of learning halfspaces with noise. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 543–552, 2006.
[10] M. Halldorsson. Approximations of weighted indpendent set and hereditary subset problems. *J. Gaph Algorithms Appl.*, 4(1), 2000.
[11] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.
[12] S. Khot. Inapproximability results via long code based PCPs. *SIGACT News*, 36(2), June 2005.
[13] R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195. Email: {venkat,prasad}@cs.washington.edu