Electronic Colloquium on Computational Complexity, Report No. 109 (2012)

Towards An Optimal Query Efficient PCP?

Subhash Khot [*]
NYU & U.Chicago

Muli Safra
Tel Aviv University

Madhur Tulsiani
TTI-Chicago

Abstract

We construct a PCP based on the hyper-graph linearity test with 3 free queries. It has near-perfect completeness and soundness strictly less than $\frac{1}{8}$. Such a PCP was known before only assuming the Unique Games Conjecture, albeit with soundness arbitrarily close to $\frac{1}{16}$.

At a technical level, our main contribution is constructing a new outer PCP which is "robust" against bounded degree polynomials, and showing that it can be composed with the hyper-graph linearity test with 3 free queries. We believe this outer PCP may be useful in obtaining the optimal query vs. soundness tradeoff for PCPs.

[*] Research supported by NSF CAREER grant CCF-0833228, NSF Expeditions grant CCF-0832795, NSF Waterman Award and BSF grant 2008059.

ISSN 1433-8092

1 Introduction

Testing procedures for certain classes of boolean functions have been extremely useful towards constructing Probabilistically Checkable Proofs (PCPs) for NP. Two classes of boolean functions that are most relevant from the PCP viewpoint are the class of linear functions and the class of dictator functions.

A function $f : \{0,1\}^n \mapsto \{0,1\}$ is called linear if there is some $\alpha \in \{0,1\}^n$ such that $f(x) = \alpha \cdot x = \sum_{i=1}^n \alpha_i x_i$ (all computations are over $GF(2)$). Blum, Luby and Rubinfeld [5] suggested the following test: pick two inputs $x, y \in \{0,1\}^n$ uniformly at random and accept iff $f(x) + f(y) = f(x+y)$. They analyzed the test combinatorially and subsequently a Fourier analytic approach led to the following theorem (see for instance [11]).

Theorem 1.1 *The BLR Test*

- *has perfect completeness, meaning if f is linear, the test passes with probability* 1.
- *has soundness $\frac{1}{2}$, in the sense that if the test passes with probability $\frac{1}{2} + \varepsilon$, then f has a non-negligible Fourier coefficient.*

We note that the soundness characterization in terms of existence of a non-negligible Fourier coefficient is the most relevant from PCP viewpoint. Here non-negligible means a constant that is independent of the dimension n. In the above theorem, there is in fact a Fourier coefficient with magnitude $\Omega(\varepsilon)$, but we will ignore a specific dependence since it typically is less important in PCP applications. The BLR test and its generalizations have found many applications, e.g. in the (original) proof of the PCP Theorem [9, 4, 3], Håstad's 3-bit PCP [11], and the "query efficient" PCPs [21] that are the focus of this paper.

A generalization of the BLR test is the *hyper-graph linearity test* (HLT) with k free queries[1], $k \geq 2$: pick k inputs $x_1, \ldots, x_k \in \{0,1\}^n$ uniformly at random and accept iff

$$\forall S \subseteq \{1, \ldots, k\}, |S| \geq 2, \quad f\left(\sum_{i \in S} x_i\right) = \sum_{i \in S} f(x_i).$$

A variant of the HLT is the *graph linearity test* (GLT) where the above test is carried out only for the $\binom{k}{2}$ sets S with $|S| = 2$. Clearly both GLT and HLT have perfect completeness, meaning a linear function passes the test with probability 1. Samorodnitsky and Trevisan [21] showed (see also [12]) that the GLT has soundness $2^{-\binom{k}{2}}$ (in the sense of Theorem 1.1) and the soundness does not improve with the HLT. Both these are remarkable facts! In the GLT, each of the $\binom{k}{2}$ queries in addition to the k free queries *cuts down* the soundness by a factor half, and afterwards, all the other queries in HLT put together (all those corresponding to sets with $|S| > 2$) do not help at all. Specifically, [21] exhibits a function f that passes the HLT with probability $2^{-\binom{k}{2}}$ and has no non-negligible Fourier coefficient. On the other hand, if the acceptance probability exceeds $2^{-\binom{k}{2}}$, even in the GLT, the function must have a non-negligible Fourier coefficient.

After one designs a linearity test (or a dictatorship test), the next question from the PCP viewpoint is whether the test can translated to a full-fledged PCP. One expects the PCP to have near-perfect

[1] A set of queries is called "free" if the answer to every other query is determined by the answers to these "free" queries.

completeness, the same acceptance predicate as the test and (ideally) the same soundness as the test. Indeed, the BLR test can be translated to a PCP, namely the celebrated 3-bit PCP of Håstad [11], and so can be the GLT test [21, 8] (the first paper gets down to soundness $2^{-k^2/4}$ for even k and the latter to the "correct" bound of $2^{-\binom{k}{2}}$). We would like to emphasize that this translation is often highly non-trivial and as we will see later, there are tests, especially for the class of dictatorship functions, that are currently known to translate to PCPs only assuming the Unique Games Conjecture [14]. When a PCP construction is viewed as an inapproximability result for a *constraint satisfaction problem* (CSP), these results imply that is it hard to distinguish whether the CSP on the "GLT predicate" has optimum $1-o(1)$ or at most $2^{-\binom{k}{2}}+o(1)$. The GLT predicate is a predicate on $k+\binom{k}{2}$ variables where the last $\binom{k}{2}$ variables are pairwise sums of the first k variables and negated variables are allowed in the CSP. Another application to inapproximability is a much simpler proof of Håstad's $n^{1-\varepsilon}$ inapproximability result for clique [10]. For the clique application one desires a PCP whose *amortized free bit complexity* is arbitrarily small. Indeed, for the [21, 8] PCP, this parameter, defined as the ratio of the number of free queries to the logarithm of the inverse soundness, is $\approx \frac{k}{\binom{k}{2}}$ which becomes arbitrarily small as k gets larger[2].

After successfully translating the GLT to a PCP, it is natural (and fruitful) to further investigate the HLT. As we already mentioned, the HLT, when viewed as a linearity test, gets stuck at the soundness threshold of $2^{-\binom{k}{2}}$, i.e. there are functions with no non-negligible Fourier coefficient but pass the HLT with probability $2^{-\binom{k}{2}}$. One approach to get around this barrier is to use the HLT as a dictatorship test instead (and this is how the BLR Test and the GLT are actually used in [11, 21, 8], but the way these tests are used, the distinction between a linearity test and a dictatorship test is hazy). A function $f : \{0,1\}^n \mapsto \{0,1\}$ is called a dictatorship if $f(x) = x_i$ for some $i \in [n]$. In other words, the i^{th} co-ordinate has *influence* 1 where influence of a co-ordinate is the probability that flipping that co-ordinate flips the value of the function on a random input. The HLT (and similarly the BLR Test and the GLT) can be adapted into a dictatorship test, called the *hyper-graph dictatorship test* (HDT), where the test is the same except that a small noise is added to each query.[3] Viewed now as a dictatorship test, the soundness barrier of $2^{-\binom{k}{2}}$ is broken, as shown in [22]! Specifically,

Theorem 1.2 *The HDT*

- *has near-perfect completeness, meaning if f is a dictatorship, it passes the test with probability close to* 1*.*
- *has soundness $2^{-(2^k-k-1)}$, in the sense that if the test passes with probability slightly above this threshold, then f has a co-ordinate with non-negligible influence*[4]

Note that the threshold $2^{-(2^k-k-1)}$ indicates that each of the 2^k-k-1 queries (corresponding to sets $S \subseteq \{1,\ldots,k\}, |S| \geq 2$) in addition to the k free queries *cuts down* the soundness by a factor half. HDT is a remarkable test and it would be even more remarkable if it were to translate into a PCP. In addition to being an interesting PCP in its own right, it would have (at least)

[2] The results in [21] are stated in terms of the *amortized query complexity* which is the ratio of the number of queries to the logarithm of the inverse soundness, and is $1+O(\frac{1}{k})$ for the PCPs in [21, 8]

[3] Dictatorship functions form a restricted sub-class of linear functions. Thus a dictatorship test is required to "kill" not just the non-linear functions but also the linear functions with a "large support". The purpose of introducing noise is precisely to "kill" the latter.

[4] There is an issue of influence versus the *low-degree influence* which we omit from this discussion.

two applications to inapproximability: firstly, for the the Max-K-CSP problem, it would be NP-hard to distinguish whether a CSP on $K = 2^k - 1$ variables has optimum $1 - o(1)$ or at most $2^{-(2^k-k-1)} + o(1)$, which would essentially match the best known algorithm for the problem [7]. Secondly, it would be NP-hard to approximate the independent set problem on degree d graphs within factor $d/\text{polylog}(d)$ [22].

However, one knows how to translate HDT to a PCP only assuming the Unique Games Conjecture [22]. This is the case for many other dictatorship tests, perhaps the *Max-Cut Test* being the most prominent example. Therein, we know a two-query test that "works" and analyzed via the Majority Is Stablest Theorem [17], but one knows how to translate the test to a PCP and obtain optimal inapproximability result for Max-Cut only assuming the UGC [15].

Since the UGC, if true, remains out of reach of current techniques, it is worth-while to attempt proving (at least some of) the implications of UGC without relying on it. This paper may be viewed as a modest step in this direction for the specific problem of Max-K-CSP and the related HLT test.

Fix the number of free queries to $k = 3$. Let us call the corresponding GLT and HLT predicates as $\mathsf{3GLT}$ and $\mathsf{3HLT}$. Note that $\mathsf{3GLT}$ is a predicate on 6 variables (x, y, z, a, b, c) which is True iff $a = x + y, b = y + z, c = x + z$. Similarly, $\mathsf{3HLT}$ is a predicate on 7 variables (x, y, z, a, b, c, d) which is True iff in addition $d = x + y + z$. As mentioned before, $\mathsf{3GLT}$ is known to be NP-hard with gap $(1 - o(1), \frac{1}{8} + o(1))$ [8] and $\mathsf{3HLT}$ is known to be Unique Games hard with gap $(1 - o(1), \frac{1}{16} + o(1))$ [22]. The predicate $\mathsf{3HLT}$ has one additional "query" and the UGC tells us that this query helps bring the soundness down to $\frac{1}{16}$ from $\frac{1}{8}$. However, without relying on UGC, we didn't know whether this additional query helps at all. Our main result is to show that it does:

Theorem 1.3 *There exists a constant $s < \frac{1}{8}$ such that for an arbitrarily small constant $\varepsilon > 0$, it is NP-hard to distinguish whether the CSP on $\mathsf{3HLT}$ predicate has optimum at least $1 - \varepsilon$ or at most $s + \varepsilon$.*

The constant s for which we can prove the above theorem is $\frac{1}{8} - \frac{1}{320}$.

Techniques

At a high level, our approach is to revisit the analysis of HLT and not get *discouraged* by lack of a non-negligible Fourier coefficient. A non-negligible Fourier coefficient would allow us to conclude that the boolean function f is correlated with a linear function. We instead conclude that the function is close (in Hamming distance) to a unique quadratic polynomial and hence the function may be *decoded uniquely* to a quadratic polynomial. This constitutes the so-called *Inner PCP.* To integrate it inside a full PCP however requires an *Outer PCP* that is *robust against quadratic polynomials.* We construct such an Outer PCP from scratch (it is robust against degree d polynomials for any fixed d). In the following, we give a more detailed overview of the techniques.

Consider the HLT with three free queries. Given a function $f : \{0,1\}^n \mapsto \{0,1\}$:

- Pick inputs $x, y, z \in \{0,1\}^n$ uniformly at random.
- Accept iff:

$$f(x+y) = f(x)+f(y), f(y+z) = f(y)+f(z), f(x+z) = f(x)+f(z), f(x+y+z) = f(x)+f(y)+f(z).$$

A (by-now) standard way to analyze such a test is to think of the function f as $\{1,-1\}$ valued and *arithmetize* the acceptance predicate:

$$\frac{1+f(x)f(y)f(x+y)}{2}\cdot\frac{1+f(y)f(z)f(y+z)}{2}\cdot\frac{1+f(x)f(z)f(x+z)}{2}\cdot\frac{1+f(x)f(y)f(z)f(x+y+z)}{2}.$$

The above expression equals 1 if the test accepts and 0 otherwise and thus the acceptance probability equals its expectation over the choice of x, y, z. The expression can be written as a sum of 16 terms, with the constant term equal to $\frac{1}{16}$. Fourteen of the terms, after taking expectation over x, y, z, can be written in terms of Fourier coefficients either as $\sum_\alpha \widehat{f}(\alpha)^3$ or as $\sum_\alpha \widehat{f}(\alpha)^4$. This step is standard and allows us to conclude that if either of these fourteen terms is non-negligible, then f must have a non-negligible Fourier coefficient. Moreover f may be *decoded* into a small list of non-negligible Fourier coefficients. In this sense, these fourteen terms may be "handled well" in the PCP setting (if and when we get to the PCP) and hence may be assumed to be arbitrarily small. The only "troublesome term" is the last one:

$$\mathop{\mathbb{E}}_{x,y,z}\left[f(x)f(y)f(z)f(x+y)f(y+z)f(x+z)f(x+y+z)\right].$$

Indeed, if f is a quadratic polynomial (or to be precise, a quadratic phase $(-1)^{g(x)}$ where g is a quadratic polynomial, since we are now thinking of f as being $\{1,-1\}$ valued), the above expectation equals 1 whereas there are quadratic phase functions that have no non-negligible Fourier coefficient. Still, for a small enough constant $\theta \in (0, \frac{1}{20})$, it was shown by Alon et al. [1] that if the above expectation is at least $1-\theta$, then f is $(1-\theta)$ close in Hamming distance to a quadratic phase function, which then may be taken as a unique decoding of f. Thus we may upper bound the "troublesome term" by $1-\theta$ and get an overall soundness of $\frac{1}{16}+\frac{1}{16}(1-\theta) = s < \frac{1}{8}$.

The next issue is whether the decoding to a quadratic polynomial can be integrated with an appropriate Outer PCP. The known Outer PCPs are obtained by parallel repetition of a basic "equation versus variable" game and this basic game is not robust enough towards this end. We design a new game that is robust against low degree polynomials. Its description is a bit cumbersome and non-intuitive, so we elaborate it in some detail.

Let us first describe the standard "equation versus variable" game. One starts with a hard instance of 3LIN problem: a set of variables $x_1, \ldots, x_n$ and equations $E_1, \ldots, E_m$ over $GF(2)$ with each equation containing three variables. It is known [11] that distinguishing between a gap $(1-o(1), \frac{1}{2}+o(1))$ is NP-hard. For the ease of exposition, assume that all right hand sides of the equations are zero (this would actually make the instance trivial, but the right hand sides serve as "offsets" that can easily be incorporated via "folding"). The basic game consists of two provers and a verifier. The verifier picks a random equation, say $E_\ell : x_i + x_j + x_k = 0$. The first prover is asked to furnish values of two variables in the equation, say x_i and x_j (and given the equation). The second prover is asked to furnish the value of one of the three variables chosen at random (without revealing the equation). The verifier accepts iff the two provers' answers are consistent on variables x_i, x_j and on the variable x_k, the sum of the values of variables x_i, x_j furnished by the first prover equals the value of variable x_k furnished by the second prover. The value of the game is the maximum probability with which the provers can make the verifier accept. It is easily checked that the hardness of 3LIN translates to hardness of distinguishing whether the value of the game is $1-o(1)$ or bounded away from 1.

Let us now reformulate the above game so as to incorporate the notion of robustness against low degree polynomials. Let $r = 2$. We first describe the reformulated game in abstract and then show how it indeed is a reformulation of the above game.

- The question to the first prover is an index $\ell \in [m]$. His answer is a linear function $f_\ell : \{0,1\}^r \mapsto \{0,1\}$.
- The verifier chooses a random non-zero input $z \in \{0,1\}^r$. Depending on the index ℓ and input z, the verifier calculates an index $t \in [n]$.
- The question to the second prover is the index $t \in [n]$. His answer is a single bit, say y_t.
- The verifier accepts iff $f_\ell(z) = y_t$.

It is intended that $y_t = x_t$ in a supposedly almost-satisfying assignment to the 3LIN instance. The function f_ℓ is intended to be the linear function

$$f_\ell(z_1, z_2) = x_i \cdot z_1 + x_j \cdot z_2,$$

where the index ℓ identifies the equation $E_\ell : x_i + x_j + x_k = 0$. After choosing ℓ and non-zero $z \in \{0,1\}^r$, verifier calculates the index t as: if $z = \text{“10”}, t = i$, if $z = \text{“01”}, t = j$, and if $z = \text{“11”}, t = k$. Note that the input z is not revealed to either of the provers. It is clear that the game is a simple reformulation of the game before and hence the hardness of its gap-version holds.

With the reformulated version, it is now straightforward to introduce the notion of robustness. We prove that it is NP-hard to distinguish whether the value of the game is $1 - o(1)$ (and the first prover furnishes linear functions as required) or whether the value of the game is bounded away from 1 even if the first prover is allowed to furnish functions $f_\ell : \{0,1\}^r \mapsto \{0,1\}$ that are degree d polynomials. I.e. the soundness is robust even against degree d polynomials. Our result holds with $r = d + 1$ and the soundness is $1 - 2^{-O(d)}$. A formal statement of the result appears as Theorem 3.1 and proved in Section 3. The proof is a rather significant and non-intuitive variant (in authors' opinion) of Håstad's 3-bit PCP and instead of attempting to give some intuition here, the reader is referred directly to Section 3.

The Outer PCP is now obtained by a *smooth* parallel repetition of the basic robust game. The term *smooth* refers to a variant of the parallel repetition where the actual game is played only on a small fraction of co-ordinates and on the remaining co-ordinates, the same question is sent to both provers. As demonstrated in [16], when Hadamard Code (i.e. table of values of a linear function) is “plugged in” at the Inner PCP level, the smoothness of the parallel repetition game ensures the so-called *subcode covering* property. In short, the property says that the code (or rather the encoding space) corresponding to a fixed question to the “larger” prover is almost uniformly covered by the codes (or rather the encoding spaces) of the questions to the “smaller” prover. This property is very convenient (and essential towards the result in [16]) and we make use of it as well.

Related and Future Work

The question of obtaining a PCP with $2^k - 1$ queries and the optimal soundness of $2^{-(2^k-k-1)}$ (without relying on the UGC) which was the motivating application for our construction, was subsequently resolved by a beautiful result of Chan [6]. Implicit in his construction, is also a PCP similar to ours, which is robust against degree-d polynomials. However, our PCP construction still leads to several other questions which might be interesting for future work.

We construct a PCP based on the 3HLT predicate with soundness strictly below $\frac{1}{8}$. A natural (and very interesting in authors's opinion) question is whether the soundness of this particular PCP can be pushed arbitrarily close to $\frac{1}{16}$, matching the UGC-based result. At the Outer PCP level, one may use the same Outer PCP as herein: a smooth parallel repetition of the basic robust game. At the Inner PCP level however we can no longer appeal to unique decoding to a quadratic polynomial and a naive appeal to list decoding does not work because the list would be too large. If a function has a non-negligible correlation with a a linear polynomial, i.e. has a non-negligible Fourier coefficient, then the list decoding consists of making a list of all non-negligible Fourier coefficients, and since the sum of squares of the Fourier coefficients is 1, the list is bounded (by $O(1/\varepsilon^2)$ if the correlation is ε). For the class of quadratic polynomials, we (fortunately) do have a theorem [20] stating that if the expectation

$$\mathop{\mathbb{E}}_{x,y,z}\left[f(x)f(y)f(z)f(x+y)f(y+z)f(x+z)f(x+y+z)\right]$$

is non-negligible, then $f : \{0,1\}^n \mapsto \{1,-1\}$ must be correlated with a quadratic polynomial (or rather a quadratic phase). Unfortunately however, the list of all quadratic phases with which f may have a non-negligible correlation could be as large as $2^{O(n)}$. This seems like an insurmountable difficulty from the PCP viewpoint, a priori at least. The authors however do have some indications that there might be a way to bypass this difficulty.

The nature of our PCP also leads to a natural question about a variant of parallel repetition. We formulate is as a concrete problem in Section 7. The resolution of this problem would provide an alternate approach to the construction of optimal query efficient PCPs, bypassing the above difficulty.

2 Preliminaries and Definitions

2.1 Fourier Analysis

The vector space of all functions $f : \{0,1\}^n \mapsto \mathbb{R}$ has an orthonormal basis $\{\chi_\alpha \mid \alpha \in \{0,1\}^n\}$ where the inner product between two functions $f, g : \{0,1\}^n \to \mathbb{R}$ is defined as $\langle f, g\rangle := \mathbb{E}_x\left[f(x)g(x)\right]$. Hence, every $f : \{0,1\}^n \mapsto \mathbb{R}$ can be expressed uniquely as $f = \sum_{\alpha\in\{0,1\}^n} \widehat{f}(\alpha)\ \chi_\alpha$. The coefficients $\widehat{f}(\alpha) \in \mathbb{R}$ are called Fourier coefficients and are defined by: $\widehat{f}(\alpha) = \langle f, \chi_\alpha\rangle = \mathbb{E}_x\left[f(x)\overline{\chi_\alpha(x)}\right]$. By Parseval's identity, $\sum_\alpha |\widehat{f}(\alpha)|^2 = \|f\|_2^2 = \mathbb{E}_x\left[|f(x)|^2\right]$. In particular, for a function taking values in $\{-1,1\}$, the sum of squared absolute values of all its Fourier coefficients equals 1.

Definition 2.1 *Let $f : \{0,1\}^n \mapsto \{-1,1\}$ be a Boolean function. We define the* influence *of the i^{th} variable in f as the probability over all inputs that changing the i^{th} bit changes the value of f.*

$$\mathsf{Inf}_i(f) \ := \ \mathop{\mathbb{P}}_{x\in\{0,1\}^n}\left[f(x) \neq f(x+e_i)\right] \ = \ \sum_{\substack{\alpha\in\{0,1\}^n \\ \alpha_i=1}} |\widehat{f}(\alpha)|^2 \,.$$

Here e_i represents the input with 1 only in the i^{th} position. The influence for $f : \{0,1\}^n \to \{0,1\}$ is also defined as the same probability as above.

For $\eta > 0$, let $\mu \sim_\eta \{0,1\}^n$ denote sampling $\mu \in \{0,1\}^n$ from a distribution with each bit independently set to 1 with probability η and 0 with probability $1-\eta$. For a function $f : \{0,1\}^n \mapsto \{-1,1\}$, it's noise stability at noise η is given by

$$\mathop{\mathbb{E}}_{\substack{x\in\{0,1\}^n \\ \mu\sim_\eta\{0,1\}^n}}\left[f(x)f(x+\mu)\right] \ = \ \sum_\alpha (1-2\eta)^{|\alpha|}\cdot|\widehat{f}(\alpha)|^2 \,,$$

where $|\alpha|$ denotes the number of 1s in α.

2.2 Hellinger and Statistical Distance

The squared Hellinger distance between distributions D_1 and D_2 over a discrete probability space $\mathcal{A}$ is

$$H^2(D_1, D_2) \ := \ \frac{1}{2}\sum_{a\in\mathcal{A}}\left(\sqrt{D_1(a)} - \sqrt{D_2(a)}\right)^2 \ = \ 1 - \sum_{a\in\mathcal{A}}\sqrt{D_1(a)D_2(a)}.$$

It is clear that $1 - H^2(\cdot,\cdot)$ is multiplicative for product distributions D_1^k, D_2^k on space $\mathcal{A}^k$, i.e.

$$1 - H^2(D_1^k, D_2^k) = (1 - H^2(D_1, D_2))^k.$$

The statistical distance between D_1 and D_2 is:

$$\Delta(D_1, D_2) := \frac{1}{2}\sum_{a\in\mathcal{A}}|D_1(a) - D_2(a)|\,.$$

We have the standard inequality:

Lemma 2.2

$$H^2(D_1, D_2) \le \Delta(D_1, D_2) \le \sqrt{2}\cdot H(D_1, D_2).$$

2.3 Label Cover and 2 Prover 1 Round Games

Definition 2.3 *An instance $\mathcal{L}$ of* Label Cover *with projection property is defined by a tuple $(U, V, E, \Sigma_1, \Sigma_2, \Pi)$. Here, (U, V, E) is a bipartite graph, $R_1 \ge R_2$ are natural numbers (called alphabet sizes) and $\Pi = \{\pi_{uv}\}_{(u,v)\in E}$ is a collection of projection functions $\pi_{uv} : [R_1] \mapsto [R_2]$. A labeling is pair of maps $A : U \mapsto [R_1]$, $B : V \mapsto [R_2]$. We say that an edge $e = (u, v)$ is* satisfied *by the labeling if $\pi_{uv}(A(u)) = B(v)$. We define:*

$$\mathsf{Opt}(\mathcal{L}) := \max_{\substack{A:U\to[R_1]\\ B:V\to[R_2]}} \ \mathop{\mathbb{P}}_{(u,v)\in E}\left[\pi_{uv}(A(u)) = B(v)\right]\,.$$

It is known that for all $\delta > 0$, there exist R_1, R_2 such that it is NP-hard to decide if for a given label cover instance with alphabet sizes R_1, R_2, $\mathsf{Opt}(\mathcal{L}) = 1$ or $\mathsf{Opt}(\mathcal{L}) < \delta$. This result also holds for the special case of instances in which the bipartite graph (U, V, E) is regular on both sides.

Definition 2.4 *A* 2P1R Game *$\mathcal{G}(U, V, \mu, \mathcal{R}, \mathcal{S}, \{\pi_{uv}\})$ consists of sets of questions U, V and sets of answers $\mathcal{R}, \mathcal{S}$ for the two provers respectively, a distribution μ on the set of question pairs $U \times V$ and for every question pair (u, v) in the support of μ, a predicate $\pi_{uv} : \mathcal{R} \times \mathcal{S} \mapsto \{0, 1\}$ that defines the pairs of accepting answers. A strategy of provers is a map $\varphi : V \mapsto \mathcal{R}, \varphi : U \mapsto \mathcal{S}$. The value of the strategy φ is:*

$$\mathsf{Val}(\varphi, \mathcal{G}) := \mathop{\mathbb{P}}_{(u,v)\sim\mu}\left[\pi_{uv}(\varphi(u), \varphi(v)) = 1\right]\,.$$

The value of the game $\mathsf{Val}(\mathcal{G})$ *is the maximum value of any prover strategy. A Projection Game is one where for every answer of the first prover, there is exactly one accepting answer of the second prover. For a projection game, the predicate π_{uv} can be thought of as a map $\pi_{uv} : \mathcal{R} \mapsto \mathcal{S}$ and the accepting answers are of the form $(r, \pi_{uv}(r))$ for $r \in \mathcal{R}$. For a projection game, $|\mathcal{S}| \le |\mathcal{R}|$.*

A 2P1R Game is best viewed as a game between the two provers and a verifier. The verifier picks a random question pair (u, v) from the distribution μ, asks one question each to the two prover respectively, and accepts if and only if the provers' answers satisfy the predicate π_{uv}. The probability of acceptance of the verifier is same as the value of a provers' strategy. Each bipartite label cover instance can be viewed as a 2P1R Game where the verifier picks a random edge $(u, v) \in E$, and the questions sent to the two provers are the vertices u and v. The acceptance predicate is the same as the projection π_{uv} for the edge.

Definition 2.5 *Given a 2P1R Game $\mathcal{G}(U, V, \mu, \mathcal{R}, \mathcal{S}, \{\pi_{uv}\})$, the k-wise repeated game is*

$$\mathcal{G}^{\otimes k}(U^k, V^k, \mu^k, \mathcal{R}^k, \mathcal{S}^k, \{\pi^k_{\mathbf{uv}}\}),$$

where for $\mathbf{u} = (u_1, \ldots, u_k)$ and $\mathbf{v} = (v_1, \ldots, v_k)$, $\pi^k_{\mathbf{uv}} := \wedge_{i=1}^k \pi_{u_i v_i}$.

We state below Raz's Parallel Repetition Theorem along with the recent improvements (and simplifications) by Holenstein and Rao.

Theorem 2.6 ([19, 13, 18]) *There exists an absolute constant $c > 0$ such that for a 2P1R Game $\mathcal{G}$ with answer sets $\mathcal{R}, \mathcal{S}$ and $\mathrm{val}(\mathcal{G}) = 1 - \varepsilon$,*

$$\mathsf{Val}(\mathcal{G}^k) \leq (1 - \varepsilon^3)^{ck/\log(|\mathcal{R}||\mathcal{S}|)}.$$

For a Projection Game, the bound of $(1 - \varepsilon^2)^{ck}$ holds.

2.4 Linear vs. Degree-d Labeling

We will, in fact, consider label cover instances in which the labels on one side are functions. These will have the property that in the YES case one can satisfy almost all constraints with labels corresponding to linear functions, while in the NO case it not possible to satisfy many constraints even with labels corresponding to degree-d polynomials (for technical reasons, we only argue about polynomials with no constant term). We define the problem below.

Definition 2.7 *An instance of* Linear vs. Degree-d Label-Cover *with parameters $\eta, \gamma > 0$ is a graph (U, V, E) and a set of (possibly weighted) constraints. A label for each vertex $u \in U$ is a function $F_u : \{0,1\}^{d+1} \to \{0,1\}$ and a label for each $v \in V$ is a bit $A(v) \in \{0,1\}$. Each constraint is indexed by a tuple (u, v, x) for $x \in \{0,1\}^{d+1} \setminus \{0^{d+1}\}$ and is of the form*

$$F_u(x) \;=\; A(v) + c_{uv,x} \mod 2\,,$$

for $c_{uv,x} \in \mathbb{F}_2$. The goal is to distinguish between the cases:

- YES*: $\exists A : V \to \{0,1\}$ and homogeneous linear functions $\{F_u : \{0,1\}^{d+1} \to \{0,1\}\}_{u \in U}$ satisfying $1 - \eta$ fraction of the constraints.*

- NO*: $\forall A : V \to \{0,1\}$, any collection of degree-d polynomials $\{F_u : \{0,1\}^{d+1} \to \{0,1\}\}_{u \in U}$ (with $F_u(0) = 0\ \forall u$) satisfies at most $1 - \gamma$ fraction of the constraints.*

Remark 2.8 *The instances of* Linear vs. Degree-d Label-Cover *we will consider will have the additional property that for all $u \in U$ and all $x, x' \in \{0,1\}^{d+1} \setminus \{0^{d+1}\}$, the total weight of the constraints involving the pair (u, x) will be the same as the weight of the constraints involving the pair (u, x'). This assumption is not crucial, but it makes the PCP construction a little simpler.*

It is perhaps also useful to view Linear vs. Degree-d Label-Cover as a gap version of a constraint satisfaction problem, where the hardness result involves two kinds of constraints. Let $L : \{0,1\}^{2^{d+1}} \to \{0,1\}$ be a predicate on 2^{d+1} variables which is 1 iff the input corresponds to the evaluation table of a linear function in $d+1$ variables. Let $D : \{0,1\}^{2^{d+1}} \to \{0,1\}$ be a similar predicate which is 1 iff the input corresponds to the table of a polynomial of degree at most D.

Then an instance of Linear vs. Degree-d Label-Cover as above can be considered as a constraint satisfaction problem, where each constraint is imposed on the 2^{d+1} values of the form $A(v)$ corresponding to all the inputs of a single function F_u (after shifting the values $A(v)$ by appropriate constants $c_{uv,x}$). The goal is to distinguish between the cases when $1-\eta$ fraction of the constraints can be satisfied even when we consider each constraint to be the predicate L, and when not even $1-\gamma$ fraction of the constraints can be satisfied even when we take the constraints to be predicate D (the accepting assignments for which are a superset of those for L).

3 Hardness of Linear vs. Degree-d Labeling

We now give a reduction from Label Cover to the Linear vs. Degree-d Label-Cover problem as defined in Definition 2.7. We will prove the following theorem:

Theorem 3.1 *For any $d \geq 1$, there exists $\gamma = 2^{-O(d)}$ such that for any $\eta \in (0,\gamma)$, the* Linear vs. Degree-d Label-Cover *problem with parameters η and γ is NP-hard.*

Let (U, V, R_1, R_2, Π) be an instance of Label Cover where $\Pi = \{\pi_{uv}\}_{(u,v)\in E}$ is a collection of projections $\pi_{uv} : [R_1] \to [R_2]$. Assume the constraint graph is regular on both sides. To prove the hardness of Linear vs. Degree-d Label-Cover with parameters η, γ, we will need δ for the Label Cover instance to be less than $\eta \cdot 2^{-\Omega(d)}/\sqrt{\gamma}$.

Reduction from Label-Cover. Identify $\{0,1\}^{d+1}$ with $2^{[d+1]}$ in the canonical way and for $y_1, y_2, \ldots, y_t \in [d+1]$, let $\mathbb{1}_{\{y_1,\ldots,y_t\}} \in \{0,1\}^{d+1}$ denote the string which has 1s only in the positions corresponding to $y_1, \ldots, y_t$.

For $u \in U, v \in V$ and $S_1 \subseteq \{0,1\}^{R_1}, S_2 \subseteq \{0,1\}^{R_2}, |S_1| = |S_2| = d+1$, we will impose constraints on functions $F_{u,S_1} : \{0,1\}^{d+1} \to \{0,1\}$, $F_{v,S_2} : \{0,1\}^{d+1} \to \{0,1\}$, $A_u : \{0,1\}^{R_1} \to \{0,1\}$ and $A_v : \{0,1\}^{R_2} \to \{0,1\}$. Thus, for the resulting instance of Linear vs. Degree-d Label-Cover, the left vertices will be of the form (u, S_1) and (v, S_2) for $u \in U, v \in V$ and $S_1 \subseteq \{0,1\}^{R_1}, S_2, \subseteq \{0,1\}^{R_2}, |S_1| = |S_2| = d+1$. The labels for these vertices are functions from $\{0,1\}^{d+1}$ to $\{0,1\}$. The right vertices are of the form (u, x) and (v, y) for $u \in U, v \in V, x \in \{0,1\}^{R_1}$ and $y \in \{0,1\}^{R_2}$. The labels for these vertices are bits. The collection of labels for the vertices (u, x) for all x (and similarly (v, y) for all y) can be viewed as a function $A_u : \{0,1\}^{R_1} \to \{0,1\}$ and we write the constraints for the values of such functions.

Remark 3.2 *We will write our constraints assuming that the functions A_u and A_v are* folded *for all u, v i.e., they satisfy $A_u(x + \mathbf{1}) = A_u(x) + 1$ (and similarly for A_v) where $\mathbf{1}$ denotes the all-1s string. This is implemented by requiring the labeling A_u to be specified only for half the inputs in $\{0,1\}^{R_1}$ (say with the first bit 0) and constructing the value for the remaining inputs using the above relation. This is what results in the constants $c_{uv,x}$ in the constraints.*

We also assume $A_u(0) = A_v(0) = 0$ for all u, v. This will be true for the labeling we give in the YES *case. Also, any labeling can be modified to satisfy this without significantly affecting the number of satisfies constraints.*

We include the following types of constraints with equal probability:

1. For each $u \in U$, select a random set $S \subseteq \{0,1\}^{R_1}$ with $|S| = d+1$ and a random $T \subseteq S$, $T \neq \emptyset$, and include the constraint:

$$F_{u,S}(\mathbb{1}_T) \;=\; A_u\left(\sum_{y \in T} y\right).$$

   Similarly, for each $v \in V$, $S \subseteq \{0,1\}^{R_2}$ with $|S| = d+1$ and $T \subseteq S, T \neq \emptyset$, include the constraint:

$$F_{v,S}(\mathbb{1}_T) \;=\; A_v\left(\sum_{z \in T} z\right).$$

2. For each $u \in U$, select a random $S \subseteq \{0,1\}^{R_1}$ with $|S| = d+1$, $T \subseteq S, T \neq \emptyset$ and $\mu \sim_\eta \{0,1\}^{R_1}$. Include the constraint [5]:

$$F_{u,S}(\mathbb{1}_T) \;=\; A_u\left(\sum_{y \in T} y + \mu\right).$$

   Similarly, for each $v \in V$, $S \subseteq \{0,1\}^{R_2}$ with $|S| = d+1$, $T \subseteq S, T \neq \emptyset$ and $\mu \sim_\eta \{0,1\}^{R_2}$, include:

$$F_{v,S}(\mathbb{1}_T) \;=\; A_v\left(\sum_{z \in T} z + \mu\right).$$

3. For each edge $(u,v) \in E$, pick $y_1, \ldots, y_d \in \{0,1\}^{R_1}$ and $z \in \{0,1\}^{R_2}$. Let $\pi_{uv}^{-1}(z) \in \{0,1\}^{R_1}$ denote the string obtained by copying appropriate bits of z. Let $S = \{y_1, \ldots, y_d, y_1 + \ldots + y_d + \pi_{uv}^{-1}(z)\}$. Include the constraint

$$F_{u,S}(\mathbb{1}_S) \;=\; A_v(z).$$

   Also, for each $T \subseteq S$, $1 \le |T| \le d$, include the constraint

$$F_{u,S}(\mathbb{1}_T) \;=\; A_u\left(\sum_{y \in T} y\right).$$

Note that some of the constraints of type (3) are the same as constraints of type (1). However, type (1) constraints would have small total weight for special sets S of the form $\{y_1, \ldots, y_d, y_1 + \ldots + y_d + \pi_{uv}^{-1}(z)\}$, whereas constraints of type (3) are exclusively for sets of this form, and thus have large total weight.

[5] We will need to use these constraints only for singletons sets T. However, we include them for all non-empty T to maintain the property that all non-zero inputs to the functions $F_{u,S}$ are equally likely to appear in a constraint, as discussed in Remark 2.8.

3.1 Completeness

Let $\mathcal{L} : U \cup V :\to \Sigma_1 \cup \Sigma_2$ be a labeling which satisfies all the constraints in the Label-Cover instance. Choose the functions A_u, A_v to be dictator functions according to $\mathcal{L}$. For each u and $S \subseteq \{0,1\}^{R_1}$, let let $F_{u,S}$ be the linear function defined as $F_{u,S}(x) = \sum_{y \in S} x_y \cdot A_u(y) \mod 2$, where we think of x being in $\{0,1\}^S$. We define $F_{v,S}$ similarly. We now verify the completeness for each type of constraints.

1. Since $F_{u,S}$ and A_u are both linear, for any set $T \subseteq S$ we have

$$F_{u,S}(\mathbb{1}_T) \;=\; \sum_{y \in S} \mathbb{1}_T(y) \cdot A_u(y) \;=\; \sum_{y \in T} A_u(y) \;=\; A_u\left(\sum_{y \in S} y\right).$$

   A similar argument holds for any $v \in V$.

2. Since A_u (similarly A_v) is a dictator function, with probability $1-\eta$ over the choice of μ, we have

$$F_{u,S}(\mathbb{1}_T) \;=\; \sum_{y \in S} \mathbb{1}_T(y) \cdot A_u(y) \;=\; \sum_{y \in T} A_u(y) \;=\; A_u\left(\sum_{y \in T} y\right) \;=\; A_u\left(\sum_{y \in T} y + \mu\right)$$

   Hence, a $1-\eta$ fraction of constraints of this type are satisfied.

3. By linearity of $F_{u,S}$ and A_u, we get

$$F_{u,S}\left(\mathbb{1}_{y_1,\ldots,y_d,y_1+\ldots+y_d+\pi_{uv}^{-1}(z)}\right) \;=\; \sum_{i=1}^{d} A_u(y_i) + A_u(y_1 + \ldots + y_d + \pi_{uv}^{-1}(z)) \;=\; A_u(\pi_{uv}^{-1}(z)),$$

   which is equal to $A_v(z)$ since A_u and A_v are dictator functions according to a labeling satisfying the constraint on the edge (u,v). The rest of the type (3) constraints are satisfied simply as type (1) constraints.

Thus, if the starting instance of Label-Cover is satisfiable, there exist linear functions $\{F_{w,S}\}_{w \in U \cup V, |S|=d+1}$ satisfying at least $1-\eta$ fraction of the constraints.

3.2 Soundness

Suppose there exist functions $\{F_w, A_w\}_{w \in U \cup V}$ satisfying $1-\gamma$ fraction of the constraints. Also, the functions $\{F_{w,S}\}_{w \in U \cup V, |S|=d+1}$ are degree-d polynomials.

For at least $1 - 3\sqrt{\gamma}$ fraction of the edges (u,v) we have that at least $1-\sqrt{\gamma}$ fraction of the constraints of type (1) and (2) are satisfied for u and v, and that at least $1-\sqrt{\gamma}$ fraction of constraints of type (3) are satisfied. For the rest of the argument, we fix such an edge (u,v) and drop the subscript from the projection π_{uv},

Using the fact that $1-\sqrt{\gamma}$ fraction of the constraints of type (1) are satisfied and that each $F_{u,S}$ is a degree-d polynomial over $\{0,1\}^{d+1}$, the following claim gives that A_u must also be close to a degree-d polynomial over $\{0,1\}^{R_1}$.

Claim 3.3 *With probability $1 - O(2^d\sqrt{\gamma})$ over the choice of the tuples $(y_1, \ldots, y_{d+1})$, the function A_u must satisfy*

$$\sum_{I \subseteq [d+1]} A_u\left(\sum_{i \in I} y_i\right) \;=\; 0\,.$$

Proof: Note that if the tuple $(y_1, \ldots, y_{d+1})$ has $y_i = y_j$ for any $i \neq j$ then the expression on the left is identically 0. Hence, we only need to prove the claim when $y_1, \ldots, y_{d+1}$ are all distinct.

Since the functions $F_{u,S}$ and A_u satisfy $1 - \sqrt{\gamma}$ fraction of the constraints of type (1), we have

$$\mathop{\mathbb{P}}_{|S|=d+1, T\subseteq S}\left[F_{u,S}(\mathbb{1}_T) \;=\; A_u\left(\sum_{y\in T} y\right)\right] \;\geq\; 1 - \sqrt{\gamma}\,,$$

Where we have also included $T = \emptyset$ since $F_{u,S}(0) = 0$ by assumption and $A_u(0) = 0$ by assumption. By averaging, we get

$$\mathop{\mathbb{P}}_{|S|=d+1}\left[\mathop{\mathbb{P}}_{T\subseteq S}\left[F_{u,S}(\mathbb{1}_T) \;=\; A_u\left(\sum_{y\in T} y\right)\right] \;\geq\; 1 - 2^{-(d+2)}\right] \;\geq\; 1 - \sqrt{\gamma}\cdot 2^{d+2}\,.$$

This means that for a large fraction of the subsets S, the constraint is satisfied for all $T \subseteq S$, since the number of subsets is only 2^{d+1}. Since picking a set $S \subseteq \{0,1\}^{R_1}$ with $|S| = d+1$ is the same as picking $d+1$ *distinct* elements $y_1, \ldots, y_{d+1}$, we have

$$\mathop{\mathbb{P}}_{y_1,\ldots,y_{d+1}}\left[\forall T \subseteq \{y_1, \ldots, y_{d+1}\} \quad F_{u,S}(\mathbb{1}_T) \;=\; A_u\left(\sum_{y\in T} y\right)\right] \;\geq\; 1 - \sqrt{\gamma}\cdot 2^{d+2}\,.$$

Also, since each $F_{u,S}$ is a degree-d polynomial, $\sum_{T\subseteq S} F_{u,S}(\mathbb{1}_T) = 0$ for *every* $S = \{y_1, \ldots, y_{d+1}\}$. Combining this with the above, we get that

$$\mathop{\mathbb{P}}_{y_1,\ldots,y_{d+1}}\left[\sum_{T\subseteq\{y_1,\ldots,y_{d+1}\}} A_u\left(\sum_{y\in T} y\right) = 0\right] \;\geq\; 1 - \sqrt{\gamma}\cdot 2^{d+2}\,,$$

which proves the claim. ■

By [1], there exits a degree-d polynomial $\tilde{A}_u$ which is at distance at most $2^{O(d)}\sqrt{\gamma}$ from A_u. Similarly, A_v is close to a degree-d polynomial $\tilde{A}_v$. We will need to show that these polynomials can have only few relevant variables (variables with non-zero influence). We need the following lower bound on the influence of a relevant variable.

Claim 3.4 *Let $P : \{0,1\}^R \to \{0,1\}$ be a degree-d polynomial over $\mathbb{F}_2$ and let the $\mathsf{Inf}_i(P) > 0$. Then $\mathsf{Inf}_i(P) \geq 2^{-(d-1)}$.*

Proof: This simply follows by noting that $\mathsf{Inf}_i(P)$ is the probability that the degree $d-1$ polynomial $P(y + e_i) - P(y)$ is non-zero. Since $\mathsf{Inf}_i(P) > 0$, the polynomial is non-trivial and hence must be non-zero with probability at least $2^{-(d-1)}$. ■

With probability $1-\sqrt{\gamma}\cdot 2^{d+1}$ over the choice of y, $S \ni y$ and μ, we have that $F_{u,S}(\mathbb{1}_y) = A_u(y+\mu)$ (since they form at least a $2^{-(d+1)}$ fraction of the type (2) constraints). Also with probability $1-\sqrt{\gamma}\cdot 2^{d+1}$ we have $F_{u,S}(\mathbb{1}_y) = A_u(y)$ (since these form a $2^{-(d+1)}$ fraction of the type (1) constraints). Thus, with probability at least $1-2^{d+2}\sqrt{\gamma}$, $A_u(y) = A_u(y+\mu)$. We then have that for $\tilde{A}_u$,

$$\mathop{\mathbb{P}}_{y,\mu}\left[\tilde{A}_u(y) \;=\; \tilde{A}_u(y+\mu)\right] \;\geq\; 1-2^{O(d)}\cdot\sqrt{\gamma}\,,$$

and similarly for $\tilde{A}_v$. However, since $\tilde{A}_u$ and $\tilde{A}_v$ are degree-d polynomials and highly resistant to noise by the above, the following claim gives that they cannot have too many variables with non-zero influence.

Claim 3.5 *Let P be a degree-d polynomial over $\mathbb{F}_2^R$ with at least t variables having non-zero influence. Let y be uniform in $\{0,1\}^R$ and $\mu \sim_\eta \{0,1\}^R$. Then*

$$\mathop{\mathbb{P}}_{y,\mu}\left[P(y) = P(y+\mu)\right] \;\leq\; 1-\frac{1}{2^{d+1}}+\frac{(1-2\eta)^{t/2^d}}{2^{d+1}}\,.$$

Proof: Let the polynomial P have *exactly* t variables with non-zero influence. Let f denote the function $f(y) = (-1)^{P(y)}$. Note that

$$\mathop{\mathbb{P}}_{y,\mu}\left[P(y) = P(y+\mu)\right] \;=\; \frac{1}{2}+\frac{1}{2}\cdot\mathop{\mathbb{E}}_{y,\mu}\left[f(y)f(y+\mu)\right] \;=\; \frac{1}{2}+\frac{1}{2}\cdot\sum_{\alpha\in\{0,1\}^R}(1-2\eta)^{|\alpha|}(\hat{f}(\alpha))^2\,.$$

Since $\sum_\alpha(\hat{f}(\alpha))^2 = \mathbb{E}_y\left[f^2(y)\right] = 1$, we can think of these values as giving a distribution, say $\mathcal{D}$, on the vectors α. We want to upper bound the quantity $\mathbb{E}_{\alpha\sim\mathcal{D}}\left[(1-2\eta)^{|\alpha|}\right]$. We have by Claim 3.4 that

$$\sum_{i\in[R]}\mathsf{Inf}_i(f) \;=\; \sum_{\alpha\in\{0,1\}^R}|\alpha|\cdot(\hat{f}(\alpha))^2 \;=\; \mathop{\mathbb{E}}_{\alpha\sim\mathcal{D}}\left[|\alpha|\right] \;\geq\; \frac{t}{2^{d-1}}\,.$$

On the other hand, we also know that since P (and hence f) has exactly t relevant variables, $\mathcal{D}$ is supported only on vectors α of size at most t. Thus, we have that $\mathbb{P}_{\alpha\sim\mathcal{D}}\left[|\alpha|\geq\frac{t}{2^d}\right]\geq\frac{1}{2^d}$. This gives the required bound as

$$\mathop{\mathbb{E}}_{\alpha\sim\mathcal{D}}\left[(1-2\eta)^{|\alpha|}\right] \;\leq\; \left(1-\frac{1}{2^d}\right)\cdot 1+\frac{1}{2^d}\cdot(1-2\eta)^{t/2^d}\,.$$

■

Thus, both $\tilde{A}_u$ and $\tilde{A}_v$ can have at most $t \leq \sqrt{\gamma}\cdot\frac{2^{O(d)}}{\eta}$ variables with non-zero influence.

We now consider constraints of type (3) which are satisfied with probability $1-\sqrt{\gamma}$.

Claim 3.6 *With probability $1-2^{O(d)}\sqrt{\gamma}$ over the choice of $y_1,\ldots,y_d\in\{0,1\}^{R_1}$ and $z\in\{0,1\}^{R_2}$, we have that*

$$\sum_{\substack{T\subseteq[d+1]\\1\leq|T|\leq d}} A_u\left(\sum_{i\in T}y_i\right) \;=\; A_v(z)\,,$$

where we take $y_{d+1} = y_1+\ldots+y_d+\pi^{-1}(z)$.

Proof: The proof is virtually identical to that of Claim 3.3. Let S be a random set chosen by picking $y_1, \ldots, y_d \in \{0,1\}^{R_1}$ and $z \in \{0,1\}^{R_2}$ at random and taking $S = \{y_1, \ldots, y_d, y_{d+1}\}$ with $y_{d+1} = y_1 + \ldots + y_d + \pi^{-1}(z)$. As in Claim 3.3, we have that

$$\mathop{\mathbb{P}}_{S}\left[\forall T \subseteq S, T \neq S \ F_{u,S}(\mathbb{1}_T) \ = \ A_u\left(\sum_{y \in T} y\right) \quad \text{and} \quad F_{u,S}(\mathbb{1}_S) \ = \ A_v(z)\right] \ \geq \ 1 - \sqrt{\gamma} \cdot 2^{d+2} \, .$$

However, since $F_{u,S}$ is a degree d polynomial with $F_{u,S}(0) = 0$, we have that for any $S = \{y_1, \ldots, y_{d+1}\}$, $\sum_{T \subseteq S, T \neq \emptyset} F_{u,S}(\mathbb{1}_T) = 0$. Thus, with probability $1 - \sqrt{\gamma} \cdot 2^{d+2}$ over the choice of S as above, we have that

$$\sum_{\substack{T \subseteq [d+1] \\ 1 \leq |T| \leq d}} A_u\left(\sum_{i \in T} y_i\right) + A_v(z) \ = \ 0 \, ,$$

which proves the claim. ■

Note that we cannot simplify the left hand side in the above equation using Claim 3.3 since the $d+1$ elements $y_1, \ldots, y_{d+1}$ are not independent elements of $\{0,1\}^{R_1}$. However, we can replace A_u by $\tilde{A}_u$ and then simplify the LHS as it holds for *every* $y_1, \ldots, y_{d+1}$ that

$$\sum_{\substack{T \subseteq [d+1] \\ 1 \leq |T| \leq d}} \tilde{A}_u\left(\sum_{i \in T} y_i\right) \ = \ \tilde{A}_u\left(\sum_{i=1}^{d+1} y_i\right) \, .$$

However, $y_1 + \ldots + y_{d+1} = \pi^{-1}(z)$. We then have with probability $1 - 2^{O(d)}\sqrt{\gamma}$ over the choice of $z \in \{0,1\}^{R_2}$ that

$$\tilde{A}_u(\pi^{-1}(z)) \ = \ \tilde{A}_v(z) \, .$$

But $\tilde{A}_u(\pi^{-1}(z)) - \tilde{A}_v(z)$ is a degree-d polynomial in z. If $2^{O(d)}\sqrt{\gamma} \leq 2^{-d}$, then the above implies that the polynomial must be *always* 0 i.e.

$$\tilde{A}_u(\pi^{-1}(z)) \ = \ \tilde{A}_v(z) \ \ \forall z \in \{0,1\}^{R_2} \, .$$

Thus, if $\mathsf{Inf}_i(\tilde{A}_v) > 0$ then $\exists j \in \pi^{-1}(i)$ such that $\mathsf{Inf}_j(\tilde{A}_u) > 0$. If we construct the unique polynomials $\tilde{A}_u$ and $\tilde{A}_v$ close to A_u and A_v and randomly pick the index of a variable with non-zero influence it gives a labeling satisfying the edge with probability $1/t$, where t is the maximum number of relevant variables in either of these polynomials. By Claim 3.5, t is at most $2^{O(d)}\sqrt{\gamma}/\eta$. This gives a labeling which in expectation satisfies $(1 - 3\sqrt{\gamma}) \cdot \frac{1}{t} = 2^{-O(d)}\eta/\sqrt{\gamma}$ fraction of the constraints of the label cover instance. By choosing the parameter δ for the Label-Cover instance to be smaller than $\eta/2^{O(d)}$, we can conclude that if the Label-Cover instance we start from is such that at most δ-fraction of the constraints are satisfiable, then one cannot have degree-d polynomials satisfying $1 - \gamma$ fraction of the constraints for Linear vs. Degree-d labeling, with $\gamma = 2^{-O(d)}$.

4 The Outer PCP

We will construct an outer PCP starting with an instance of Linear vs. Degree-d Label-Cover and performing (smooth) parallel repetition. Let $\Phi = (U, V, E)$ be an instance of Linear vs. Degree-d

Label-Cover with parameters $\eta, \gamma > 0$. This can be equivalently thought of as a two-prover game in which the verifier picks a constraint indexed by (u, v, x) and sends u to the first prover and v to the second prover. We define the two-prover game corresponding to the outer PCP according to the following procedure:

- The verifier picks constraints $(u_1, v_1, x'_1), \ldots, (u_k, v_k, x'_k) \in \Phi$.
- The question sent to the first prover is $\mathbf{u} = (u_1, \ldots, u_k)$.
- The question sent to the second prover is $\mathbf{w} = (w_1, \ldots, w_k)$ where each $w_i = u_i$ with probability $1 - \beta$ and v_i with probability β. Let $R_\mathbf{w} \stackrel{\text{def}}{=} \{i \in [k] : w_i = u_i\}$ and let $|\mathbf{w}|$ denote $|R_\mathbf{w}|$.
- The provers are required to respond with functions $F_\mathbf{u} : \{0,1\}^{(d+1)k} \to \{0,1\}$ and $G_\mathbf{w} : \{0,1\}^{(d+1)|\mathbf{w}|+(k-|\mathbf{w}|)} \to \{0,1\}$. We think of each input x to $F_\mathbf{u}$ as $x = (x_1, \ldots, x_k)$ where each $x_i \in \{0,1\}^{d+1}$. Similarly, we think of an input y to $G_\mathbf{w}$ as $y = (y_1, \ldots, y_k)$ where $y_i \in \{0,1\}^{d+1}$ if $i \in R_\mathbf{w}$ and $y_i \in \{0,1\}$ otherwise.
- Let $F_\mathbf{u}|_i$ denote the function on $d+1$ bits obtained by setting all inputs to $F_\mathbf{u}$ except the i^{th} one to 0 i.e., $F_\mathbf{u}|_i(z) = F_\mathbf{u}(x_1, \ldots x_k)$ where $x_j = z$ if $j = i$ and $x_j = 0$ otherwise. Define $G_\mathbf{w}|_i$ similarly. The verifier accepts if the following conditions are satisfied:

$$\begin{aligned} F_\mathbf{u}|_i(z) &= G_\mathbf{w}|_i(z) && \forall i \in R_\mathbf{w}, \forall z \in \{0,1\}^{d+1} \\ F_\mathbf{u}|_i(x'_i) &= G_\mathbf{w}|_i(1) + c_{u_i v_i, x'_i} && \forall i \notin R_\mathbf{w} \end{aligned}$$

  where the inputs x'_i for $i \notin R_\mathbf{w}$ are picked initially as part of the constraints.

Let $\mathcal{G}(\Phi, k, \beta)$ denote the above two-prover game with parameters k, β, starting from an instance Φ of Linear vs. Degree-d Label-Cover. Let $\mathsf{Val}^{(l)}(\mathcal{G}(\Phi, k, \beta))$ denote the maximum acceptance probability of the verifier over the prover strategies where all the functions are degree-l polynomials.

Completeness

Lemma 4.1 *Let the instance Φ of* Linear vs. Degree-d Label-Cover *be such that there exists a labeling $A : V \to \{0,1\}$ and linear functions $\{F_u\}_{u \in U}$ satisfying $1 - \eta$ fraction of the constraints. Then $\mathsf{Val}^{(1)}(\mathcal{G}(\Phi, k, \beta)) \geq 1 - k\eta$.*

Proof: The provers define the functions $F_\mathbf{u}$ and $G_\mathbf{w}$ as

$$\begin{aligned} F_\mathbf{u}(x_1, \ldots, x_k) &\stackrel{\text{def}}{=} F_{u_1}(x_1) + \ldots + F_{u_k}(x_k) && \forall x_1, \ldots, x_k \\ G_\mathbf{w}(y_1, \ldots, y_k) &\stackrel{\text{def}}{=} \sum_{i \in R_\mathbf{w}} F_{u_i}(y_i) + \sum_{i \notin R_\mathbf{w}} y_i \cdot A(v_i) && \forall y_1, \ldots, y_k\,. \end{aligned}$$

By definition of the functions, for any pair of questions $(\mathbf{u}, \mathbf{w})$ generated by the verifier, the answers of the provers always satisfy $F_\mathbf{u}|_i(z) = G_\mathbf{w}|_i(z)$ for all $i \in R_\mathbf{w}, z \in \{0,1\}^{d+1}$. Moreover, with probability at least $1 - k\eta$ over the choice of the initial constraints $(u_1, v_1, x'_1), \ldots, (u_k, v_k, x'_k) \in \Phi$, it is true that for all $i \in [k]$, $F_{u_i}(x'_i) = A(v_i) + c_{u_i v_i, x'_i}$. Then the answers of the provers also satisfy $F_\mathbf{u}|_i(x'_i) = G_\mathbf{w}|_i(1) + c_{u_i v_i, x'_i}$ for all $i \notin R_\mathbf{w}$. ■

Soundness

Lemma 4.2 *Let the instance Φ of* Linear vs. Degree-d Label-Cover *be such any labeling $A : V \to \{0,1\}$ and degree-d polynomials $\{F_u\}_{u \in U}$ satisfy at most $1-\gamma$ fraction of the constraints. Then* $\mathsf{Val}^{(d)}(\mathcal{G}(\Phi,k,\beta)) \le (1-\gamma^2)^{\Omega(\beta k)}$.

Proof: Let the prover strategy make the verifier accept with probability s. Note that constraints checked by the verifier are only on the polynomials obtained by setting all inputs except one (x_i for some $i \in [k]$) to zero. Thus, if $F_{\mathbf{u}}$ and $G_{\mathbf{w}}$ are a pair of degree-d polynomials which make the verifier accept, then so do the polynomials $F'_{\mathbf{u}}$ and $G'_{\mathbf{w}}$ obtained by dropping all the monomials involving bits from two different inputs x_i and x_j for $i \neq j$. Also, the constant terms in both $F_{\mathbf{u}}$ and $G_{\mathbf{w}}$ must be the same for the verifier to accept and we can drop these in $F'_{\mathbf{u}}$ and $G'_{\mathbf{w}}$.

We can then write $F'_{\mathbf{u}}$ and $G'_{\mathbf{w}}$ as

$$\begin{aligned} F'_{\mathbf{u}}(x_1,\ldots,x_k) &= F_{u_1}(x_1)+\ldots+F_{u_k}(x_k) \qquad \forall x_1,\ldots,x_k \\ G'_{\mathbf{w}}(y_1,\ldots,y_k) &= G_{w_1}(y_1)+\ldots+G_{w_k}(y_k) \qquad \forall y_1,\ldots,y_k\,, \end{aligned}$$

where $F_{u_1},\ldots,F_{u_k}$ and $G_{w_1},\ldots,G_{w_k}$ are degree-d polynomials with $F_{u_i}(0)=G_{w_i}(0)=0\ \forall i$. Since the above pair is accepted by the verifier, it must be true that $F_{u_i}=G_{w_i}$ for $i \in R_{\mathbf{w}}$. Also, for $i \notin R_{\mathbf{w}}$, since G_{w_i} is a homogeneous function on a 1-bit input, it must be of the form $A(v_i)\cdot y_i$. Moreover, if v_i was chosen according to a constraint indexed by (u_i,v_i,x'_i), then $A(v_i)$ must satisfy $F_{u_i}(x'_i)=A(v_i)+c_{u_i v_i,x'_i}$.

Thus, the prover strategy also gives a strategy for game obtained by parallel repetition of Φ, when viewed as a two-prover game, which makes the verifier accept with probability s. However, as argued in [16], then we must have $s \le (1-\gamma^2)^{\Omega(\beta k)}$, which completes the proof. As in [16], we consider the set of *useful* coordinates, which are the coordinates $i \notin R_{\mathbf{w}}$. With probability at least $1-2^{-\Omega(\beta k)}$, the number of useful coordinates is at least $\beta k/2$. The repeated game restricted to questions with at least $\beta k/2$ useful coordinates is a convex combination of sub-games, in each of which the basic game corresponding to Φ is repeated at least $\beta k/2$ times. By Theorem 2.6, each such sub-game has value at most $(1-\gamma^2)^{\Omega(\beta k)}$ and hence so does the game $\mathcal{G}(\Phi,k,\beta)$. ■

5 The Inner PCP

The inner PCP we will simply use the Hypergraph Test with 7 queries as described by Samorodnitsky and Trevisan [22]. Given a function $f : \{0,1\}^m \to \{-1,1\}$, the test $\mathsf{HypergraphTest}(f)$ picks $x,y,z \in \{0,1\}^m$ at random and accepts if and only if

$$\begin{aligned} f(x+y) &= f(x)\cdot f(y)\,, \\ f(y+z) &= f(y)\cdot f(z)\,, \\ f(z+x) &= f(z)\cdot f(x)\,, \\ \text{and}\quad f(x+y+z) &= f(x)\cdot f(y)\cdot f(z)\,. \end{aligned}$$

Let $s(f)$ denote the probability that $\mathsf{HypergraphTest}(f)$ accepts. We can write $s(f)$ as

$$s(f) = \mathop{\mathbb{E}}_{x,y,z}\left[\begin{array}{c} \left(\frac{1+f(x)f(y)f(x+y)}{2}\right)\left(\frac{1+f(y)f(z)f(y+z)}{2}\right) \\ \times\left(\frac{1+f(z)f(x)f(z+x)}{2}\right)\left(\frac{1+f(x)f(y)f(z)f(x+y+z)}{2}\right) \end{array}\right]$$

We can also write $s(f)$ as a multilinear polynomial in the values of f since $f(x)^2 = 1$ for $f : \{0,1\}^m \to \{-1,1\}$. We define $s(f)$ for a function f taking values in $[-1,1]$ to be the expectation of the corresponding multilinear polynomial. We will also need the following (well-known) concentration result for Lipschitz functions, which follows from Azuma's inequality:

Lemma 5.1 (Theorem 7.4.2 in [2]) *Let $\mathbf{F} = \mathbf{F}(X_1, \ldots, X_N)$ be a function in random variables $X_1, \ldots, X_N$ taking value ± 1, which is C-Lipschitz in each of the variables i.e., for any $i \in [N]$, $x_1, \ldots, x_N \in \{-1,1\}$ and $x_i' \in \{-1,1\}$, $|\mathbf{F}(x_1, \ldots, x_i, \ldots, x_N) - \mathbf{F}(x_1, \ldots, x_i', \ldots, x_N)| \leq C$. Then*

$$\mathbb{P}\left[|\mathbf{F} - \mathbb{E}[\mathbf{F}]| \ > \ \gamma\right] \ \leq \ \exp\left(-\gamma^2/(2C^2N)\right) .$$

The following is easy to prove using the results from [1].

Lemma 5.2 *Let $f : \{0,1\}^m \to [-1,1]$ be such that $s(f) > \frac{1}{8} - \frac{1}{320} + \varepsilon$. Then,*

- *there exists $\alpha \in \{0,1\}^m$ such that $\left|\hat{f}(\alpha)\right| \geq \frac{\varepsilon}{2}$, or*
- *there exists a quadratic polynomial $q : \{0,1\}^m \to \{0,1\}$ such that $\mathbb{E}_x\left[f(x)(-1)^{q(x)}\right] \geq \frac{9}{10} - \varepsilon$.*

Proof: We first argue that we can take f to be Boolean. Given an f as above, consider a random function $\tilde{f}$ such that for each $x \in \{0,1\}^m$, $\tilde{f}(x) = 1$ with probability $(1 + f(x))/2$ and -1 with probability $(1 - f(x))/2$. Thus, for each x, $\mathbb{E}\left[\tilde{f}(x)\right] = f(x)$.

The quantity $s(\tilde{f})$ is a degree-7 polynomial in the 2^m random variables $\left\{\tilde{f}(x)\right\}_{x \in \{0,1\}^m}$ with expectation $s(f)$ (since we wrote s as a multilinear polynomial in the values of f). Also, it is C-Lipschitz with $C = 7/2^n$. Thus, by Lemma 5.1, the probability that $\left|s(f) - s(\tilde{f})\right| > \varepsilon/4$ is at most $\exp(-\Omega_\varepsilon(2^m))$. Also, the correlation of $\tilde{f}$ with any quadratic polynomial q is $\mathbb{E}_x\left[\tilde{f}(x)(-1)^{q(x)}\right]$, which is a linear polynomial in the values of $\tilde{f}$, with expectation $\mathbb{E}_x\left[f(x)(-1)^{q(x)}\right]$. The probability that the correlation differs from it's expected value by more than $\varepsilon/2$ for *any* quadratic polynomial is at most $2^{m^2} \cdot \exp -\Omega_\varepsilon(2^m)$. Since the set of quadratic polynomials also includes all linear functions, this also gives that with high probability, all Fourier coefficients of $\tilde{f}$ are within an additive $\varepsilon/4$ of the corresponding Fourier coefficients of f. Thus, up to a change of an additive $\varepsilon/4$ in the correlations and the quantity $s(f)$, we can replace the function f by a "good" $\tilde{f}$. We this assume f takes values in $\{-1,1\}$ in the argument below.

Expanding the terms and changing variables to simply, we can re-write $s(f)$ as

$$\begin{aligned} s(f) \ &= \ \frac{1}{16} \ + \ \frac{7}{16} \cdot \mathop{\mathbb{E}}_{x,y}\left[f(x)f(y)f(x+y)\right] \ + \ \frac{7}{16} \cdot \mathop{\mathbb{E}}_{x,y,z}\left[f(x)f(y)f(x+z)f(y+z)\right] + \\ &\quad \frac{1}{16} \cdot \mathop{\mathbb{E}}_{x,y,z}\left[f(x)f(y)f(z)f(x+y)f(y+z)f(z+x)f(x+y+z)\right] \\ &= \ \frac{1}{16} \ + \ \frac{7}{16}\sum_\alpha (\hat{f}(\alpha))^3 \ + \ \frac{7}{16}\sum_\alpha (\hat{f}(\alpha))^4 \ + \\ &\quad \frac{1}{16} \cdot \mathop{\mathbb{E}}_{x,y,z}\left[f(x)f(y)f(z)f(x+y)f(y+z)f(z+x)f(x+y+z)\right] \end{aligned}$$

Let $s^*(f)$ denote the expectation in the last term. If $s(f) \geq \frac{1}{8} - \frac{1}{320} + \varepsilon$, then we must have $\sum_\alpha (\hat{f}(\alpha))^3 \geq \varepsilon$, $\sum_\alpha (\hat{f}(\alpha))^4 \geq \varepsilon$ or $s^*(f) > 1 - \frac{1}{20}$. The first two cases both imply

$$\max_\alpha \left|\hat{f}(\alpha)\right| \ = \ \max_\alpha \left|\hat{f}(\alpha)\right| \cdot \sum_\alpha (\hat{f}(\alpha))^2 \ \geq \ \sum_\alpha \left|\hat{f}(\alpha)\right| \cdot (\hat{f}(\alpha))^2 \ \geq \ \varepsilon .$$

On the other hand Alon et al. [1] show that for $f : \{0,1\}^m \to \{-1,1\}$, if $s^*(f) > 1 - \nu$ for $\nu \leq \frac{1}{20}$, then there exists a quadratic function $q : \{0,1\}^m \to \{0,1\}$ such that $\mathbb{E}_x\left[f(x)(-1)^{q(x)}\right] \geq 1 - 2\nu$. Using $\nu = 1/20$ and adjusting all correlations for the loss of $\varepsilon/4$ in passing from f to $\tilde{f}$ gives the required result. ■

6 The Composed PCP

We instantiate the outer PCP with $d = 2$. The final proof comprises of the tables of functions $\{G_{\mathbf{w}}\}_{\mathbf{w}\in(U\cup V)^k}$. The verifier chooses a random $\mathbf{u} \in U^k$ and performs $\mathsf{HypergraphTest}(\tilde{F}_{\mathbf{u}})$, where $\tilde{F}_{\mathbf{u}}$ is a *virtual* function described below. To query $\tilde{F}_{\mathbf{u}}$ on an input $x = (x_1, \ldots, x_k) \in \{0,1\}^{3k}$, the verifier generates a random $\mathbf{w} \in (U \cup V)^k$ and $y = (y_1, \ldots, y_k) \in \{0,1\}^{3|\mathbf{w}|+(k-|\mathbf{w}|)}$ and queries $G_{\mathbf{w}}(y)$. The tuple $\mathbf{w}$ is generated by taking $w_i = u_i$ for each $i \in R_{\mathbf{w}}$, generating a string $x'_i \in \{0,1\}^3$ for each $i \notin R_{\mathbf{w}}$ and taking w_i as $v_i \sim (u_i, x'_i)$, where this notation denotes taking a random constraint (u_i, v_i, x'_i) (with probability proportional to weights) involving the pair (u_i, x'_i) and choosing the corresponding v_i. Below we describe the distribution of $(\mathbf{w}, y, x')$ given an input x.

Suppose that we have already (randomly) chosen the set $R_{\mathbf{w}}$. Recall that the intended solution is

$$\begin{aligned} F_{\mathbf{u}}(x_1, \ldots, x_k) &\overset{\text{def}}{=} F_{u_1}(x_1) + \ldots + F_{u_k}(x_k) && \forall x_1, \ldots, x_k \\ G_{\mathbf{w}}(y_1, \ldots, y_k) &\overset{\text{def}}{=} \sum_{i\in R_{\mathbf{w}}} F_{u_i}(y_i) + \sum_{i\notin R_{\mathbf{w}}} y_i \cdot A(v_i) && \forall y_1, \ldots, y_k\,, \end{aligned}$$

where $F_{u_1}, \ldots, F_{u_k}$ are polynomials with $F_{u_i}(0) = 0\ \forall i$. Thus, to obtain $F_{\mathbf{u}}$ from the function $G_{\mathbf{w}}$, we simply want to take $y_i = x_i$ if $i \in R_{\mathbf{w}}$. Also, for $i \notin R_{\mathbf{w}}$, $x_i \neq 000$ and $v_i \sim (u_i, x_i)$, $F_{u_i}(x_i)$ and $A(v_i)$ must be related as $F_{u_i}(x_i) = A(v_i) + c_{u_i v_i, x_i}$. Thus, if $i \notin R_{\mathbf{w}}$ and $x_i \neq 000$, the verifier must choose $x'_i = x_i$, $v_i \sim (u_i, x'_i)$ and $y_i = 1$. The contribution of the intended solution in the i^{th} coordinate is then $A(v_i)$ and the verifier can add $c_{u_i v_i, x'_i}$ for this coordinate to obtain the (intended) contribution from the i^{th} coordinate of $F_{\mathbf{u}}$.

However, when $i \notin R_{\mathbf{w}}$ and $x_i = 000$, the verifier has more freedom in choosing v_i. Suppose we choose *any* $x'_i \neq 000$ and $v_i \sim (u_i, x'_i)$ (recall that constraints (u_i, v_i, x'_i) are only defined for $x'_i \neq 000$). If y_i is chosen to be 0, then the contribution from the i^{th} coordinate of $G_{\mathbf{w}}(y)$ is $y_i \cdot A(v_i) = 0$, and so is $F_{u_i}(0)$ for the intended solution. Thus, when $i \notin R_{\mathbf{w}}$ and $x_i = 000$, we take $y_i = 0$ and choose x'_i randomly from $\{0,1\}^3 \setminus \{000\}$. Note that this does *not* complete the description of the sampling process since $\mathbb{P}\left[i \in R_{\mathbf{w}} \mid x_i\right]$ might (and will) depend on the value of x_i.

From the above description, given the tuple $(\mathbf{w}, y, x')$, we can uniquely identify the input x i.e. there exists a map φ such that $\varphi(\mathbf{w}, y, x') = x$ given for the i^{th} coordinate by

$$x_i \;=\; \big(\varphi(\mathbf{w}, y, x')\big)_i \;=\; \begin{cases} y_i & \text{if } i \in R_{\mathbf{w}} \\ x'_i & \text{if } i \notin R_{\mathbf{w}} \text{ and } y_i = 1 \\ 000 & \text{otherwise} \end{cases}$$

We will now define a distribution $\mathcal{D}_{\mathbf{u}}$ on the tuples $(\mathbf{w}, y, x')$ (for a fixed $\mathbf{u}$) such that y is uniformly distributed in $\{0,1\}^{3|\mathbf{w}|+(k-|\mathbf{w}|)}$ and $\varphi(\mathbf{w}, y, x')$ is close to the uniform distribution on $\{0,1\}^{3k}$. Given an input x, the verifier will simply sample a tuple $(\mathbf{w}, y, x')$ from $\mathcal{D}_{\mathbf{u}}$ conditioned on $\varphi(\mathbf{w}, y, x') = x$. Given the tuple, the function $\tilde{F}_{\mathbf{u}}$ is taken to be

$$\tilde{F}_{\mathbf{u}}(x) \;=\; G_{\mathbf{w}}(y) + \sum_{i\notin R_{\mathbf{w}}} y_i \cdot c_{u_i v_i, x'_i}\,.$$

We now describe the distribution $\mathcal{D}_{\mathbf{u}}$. For a fixed $\mathbf{u} \in U^k$, we sample a tuple $(\mathbf{w}, y, x')$ according to $\mathcal{D}_{\mathbf{u}}$ by performing the following process for each $i \in [k]$:

- with probability $1-\beta$, set $w_i = u_i$ and sample $y_i \in_R \{0,1\}^3$.
- with probability $\beta/2$, take $y_i = 1$, $x'_i \in_R \{0,1\}^3 \setminus \{000\}$ and $w_i = v_i \sim (u_i, x'_i)$.
- with probability $\beta/2$, take $y_i = 0$, $x'_i \in_R \{0,1\}^3 \setminus \{000\}$ and $w_i = v_i \sim (u_i, x'_i)$.

Let $\varphi(\mathcal{D}_{\mathbf{u}})$ denote the distribution on $\{0,1\}^{3k}$ obtained by sampling $(\mathbf{w}, y, x')$ according to $\mathcal{D}_{\mathbf{u}}$ and then computing $\varphi(\mathbf{w}, y, x')$. We show that it is very close to the uniform distribution on $\{0,1\}^{3k}$, denoted by $\mathcal{U}_{3k}$.

Claim 6.1 $\Delta\left(\varphi(\mathcal{D}_{\mathbf{u}}), \mathcal{U}_{3k}\right) \;=\; O(\beta\sqrt{k})\,.$

Proof: It is easy to see that $\varphi(\mathcal{D}_{\mathbf{u}})$ is also a product distribution. For each $i \in [k]$,

$$\begin{aligned}
\mathop{\mathbb{P}}_{x \sim \mathcal{D}_{\mathbf{u}}}[x_i = 000] &= \frac{1-\beta}{8} + \frac{\beta}{2} = \frac{1}{8} + \frac{3\beta}{8} \\
\mathop{\mathbb{P}}_{x \sim \mathcal{D}_{\mathbf{u}}}[x_i = b] &= \frac{1-\beta}{8} + \frac{\beta}{2} \cdot \frac{1}{7} = \frac{1}{8} - \frac{3\beta}{56} \qquad \forall b \neq 000\,.
\end{aligned}$$

Let $\mathcal{D}_i$ denote the distribution of x_i when x is sampled according to $\varphi(\mathcal{D}_{\mathbf{u}})$. By the above, we can bound the squared Hellinger distance of $\mathcal{D}_i$ and $\mathcal{U}_3$, the uniform distribution on 3 bits as

$$H^2\left(\mathcal{D}_i, \mathcal{U}_3\right) = 1 - \sum_{b \in \{0,1\}^3} \sqrt{\frac{1}{8} \cdot \mathcal{D}_i(b)} = 1 - \sqrt{\frac{1}{8} \cdot \left(\frac{1}{8} + \frac{3\beta}{8}\right)} - 7\sqrt{\frac{1}{8} \cdot \left(\frac{1}{8} - \frac{3\beta}{56}\right)} = O(\beta^2)\,.$$

By multiplicativity of $1 - H^2(\cdot,\cdot)$ for product distributions, we get

$$1 - H^2\left(\varphi(\mathcal{D}_{\mathbf{u}}), \mathcal{U}_{3k}\right) \;\geq\; (1 - O(\beta^2))^k \;\geq\; 1 - O(\beta^2 k)\,.$$

Finally, using the relationship between statistical and Hellinger distance gives

$$\Delta\left(\varphi(\mathcal{D}_{\mathbf{u}}), \mathcal{U}_{3k}\right) \;\leq\; \sqrt{2}H\left(\varphi(\mathcal{D}_{\mathbf{u}}), \mathcal{U}_{3k}\right) \;=\; O(\beta\sqrt{k})$$

as claimed. ■

It is also easy to observe that if $\mathbf{u}$ is chosen with probability proportional to the weight of constraints incident on it, then the distribution of the pair $(\mathbf{u}, \mathbf{w})$ is the same as in the outer PCP.

Claim 6.2 *Let $\mathbf{u} \in U^k$ be sampled with probability proportional to the weight of constraints (in the outer PCP) incident on it and let $(\mathbf{w}, y, x')$ be sampled according to $\mathcal{D}_{\mathbf{u}}$. Then the distribution of the pair $(\mathbf{u}, \mathbf{w})$ is the same as the questions of the verifier in $\mathcal{G}(\Phi, k, \beta)$.*

Proof: By definition of $\mathcal{D}_{\mathbf{u}}$, the distribution is a product distribution over the coordinates. It only remains to analyze the distribution of a pair of questions (u_i, w_i) in a single coordinate. With probability $1-\beta$, $w_i = u_i$ and with probability β, $w_i = v_i \sim (u_i, x'_i)$ for a random x'_i i.e. it is a random constraint incident on u_i. This is exactly the same as the distribution of questions for the game $\mathcal{G}(\Phi, k, \beta)$. ■

We can now analyze the composed PCP.

6.1 Completeness

Lemma 6.3 *Let* $\mathsf{Val}^{(1)}(\mathcal{G}(\Phi,k,\beta)) \geq 1-\eta$*. Then there exists a proof that the verifier accepts with probability at least* $1-7\eta$*.*

Proof: Let $\{F_{\mathbf{u}}\}_{\mathbf{u}\in U^k}$ and $\{G_{\mathbf{w}}\}_{\mathbf{w}\in(U\cup V)^k}$ be a family of linear functions which make the verifier in $\mathcal{G}(\Phi,k,\beta)$ accept with probability $1-\eta$. Since any pair of functions $F_{\mathbf{u}}, G_{\mathbf{w}}$ satisfying the constraint corresponding to the pair of questions $(\mathbf{u},\mathbf{w})$ must have the same constant term, we can assume that all the polynomials have constant term 0. We can then write the functions for each $\mathbf{u}$ and $\mathbf{w}$ as a sum of linear functions

$$\begin{aligned} F_{\mathbf{u}}(x_1,\ldots,x_k) &= F_{u_1}(x_1)+\ldots+F_{u_k}(x_k) \qquad \forall x_1,\ldots,x_k \\ G_{\mathbf{w}}(y_1,\ldots,y_k) &= G_{w_1}(y_1)+\ldots+G_{w_k}(y_k) \qquad \forall y_1,\ldots,y_k \end{aligned}$$

Let the proof for the final verifier consist of the tables $G_{\mathbf{w}}$ as above. By Claim 6.2, the distribution of the pairs $(\mathbf{u},\mathbf{w})$ generated by the final verifier is the same as that in the game $\mathcal{G}(\Phi,k,\beta)$ and thus, for any x, $\mathbb{P}_{\mathbf{u},r}\left[F_{\mathbf{u}}(x)=\tilde{F_{\mathbf{u}}}(x)\right] \geq 1-\eta$, where r is the randomness used in sampling a random $G_{\mathbf{w}}$ to read the value of $\tilde{F_{\mathbf{u}}}$. Recall that the verifier performs $\mathsf{HypergraphTest}(\tilde{F_{\mathbf{u}}})$ which reads the value of $\tilde{F_{\mathbf{u}}}$ on 7 inputs. By the above, we have that

$$\mathop{\mathbb{P}}_{\mathbf{u}}\left[\text{All 7 values read by } \mathsf{HypergraphTest} \text{ agree with the function } F_{\mathbf{u}}\right] \;\geq\; 1-7\eta\,.$$

However, if all 7 values read by $\mathsf{HypergraphTest}$ agree with $F_{\mathbf{u}}$, then the verifier will accept since $F_{\mathbf{u}}$ is linear. ■

6.2 Soundness

Lemma 6.4 *Let* $\{G_{\mathbf{w}}\}_{\mathbf{w}\in(U\cup V)^k}$ *be a proof that makes the verifier accept with probability at least* $1/8-1/320+\varepsilon$*. Let* $\mathcal{G}(\Phi,k,\beta)$ *denote the game constructed by the outer verifier as before with* β,k *chosen so that* $\beta\sqrt{k}=o(\varepsilon)$*. Then* $\mathsf{Val}^{(2)}(\mathcal{G}(\Phi,k,\beta)) \geq \varepsilon^6/2^{18}$*.*

Proof: Let $s(f)$ be the expression for the probability of acceptance of $\mathsf{HypergraphTest}(\mathrm{f})$ as defined in Section 5. Let $\tilde{F_{\mathbf{u}}}:\{0,1\}^{3k}\to\{0,1\}$ be the (random) function queried by the verifier as above. Then,

$$\mathbb{P}\left[\text{Verifier accepts}\right] \;=\; \mathop{\mathbb{E}}_{\mathbf{u}}\left[s\left((-1)^{\tilde{F_{\mathbf{u}}}}\right)\right]\,.$$

If r is the randomness used by the verifier in sampling $\tilde{F_{\mathbf{u}}}$, let $H_{\mathbf{u}}$ denote the real-valued function

$$H_{\mathbf{u}}(x)=\mathop{\mathbb{E}}_{r}\left[(-1)^{\tilde{F_{\mathbf{u}}}(x)}\right]=\mathop{\mathbb{E}}_{(\mathbf{w},y,x')|x}\left[(-1)^{G_{\mathbf{w}}(y)+\sum_{i\notin R_{\mathbf{w}}} y_i\cdot c_{u_i v_i,x'_i}}\right]\,.$$

Since the verifier uses fresh randomness for sampling $\tilde{F_{\mathbf{u}}}(x)$ for each input x, we can re-write the acceptance probability of the verifier as

$$\mathbb{P}\left[\text{Verifier accepts}\right] \;=\; \mathop{\mathbb{E}}_{\mathbf{u}}\left[s\left(H_{\mathbf{u}}\right)\right]\,.$$

Since the verifier accepts with probability $\frac{1}{8}-\frac{1}{320}+\varepsilon$, we have

$$\mathop{\mathbb{E}}_{\mathbf{u}}\left[s(H_{\mathbf{u}})\right] \geq \frac{1}{8}-\frac{1}{320}+\varepsilon \;\Rightarrow\; \mathop{\mathbb{P}}_{u}\left[s(H_{\mathbf{u}}) \geq \frac{1}{8}-\frac{1}{320}+\frac{\varepsilon}{2}\right] \geq \frac{\varepsilon}{2}$$

By Lemma 5.2, for each $\mathbf{u}$ with $s(H_{\mathbf{u}}) \geq \frac{1}{8} - \frac{1}{320} + \frac{\varepsilon}{2}$, either there exists an $\alpha_{\mathbf{u}} \in \{0,1\}^{3k}$ such that $\left|\hat{H}_{\mathbf{u}}(\alpha_{\mathbf{u}})\right| \geq \frac{\varepsilon}{4}$ or there exists a quadratic function $q_{\mathbf{u}}$ such that $\mathbb{E}_x\left[H_{\mathbf{u}}(x)(-1)^{q_{\mathbf{u}}(x)}\right] \geq \frac{9}{10} - \frac{\varepsilon}{2}$. Thus we have

$$\mathop{\mathbb{P}}_{\mathbf{u}}\left[\max_{\alpha}\left\{\left|\hat{H}_{\mathbf{u}}(\alpha)\right|\right\} \geq \frac{\varepsilon}{4}\right] \geq \frac{\varepsilon}{4} \quad \text{or} \quad \mathop{\mathbb{P}}_{\mathbf{u}}\left[\exists q_{\mathbf{u}} \text{ such that } \mathop{\mathbb{E}}_{x}\left[H_{\mathbf{u}}(x)(-1)^{q_{\mathbf{u}}(x)}\right] \geq \frac{9}{10} - \frac{\varepsilon}{2}\right] \geq \frac{\varepsilon}{4}\,.$$

We first consider the case that a significant fraction of the functions $H_{\mathbf{u}}$ have a large Fourier coefficient.

Claim 6.5 *Let* $\mathbb{P}_{\mathbf{u}}\left[\max_{\alpha}\left\{\left|\hat{H}_{\mathbf{u}}(\alpha)\right|\right\} \geq \frac{\varepsilon}{4}\right] \geq \frac{\varepsilon}{4}$. *Then* $\mathsf{Val}^{(1)}(\mathcal{G}(\Phi, k, \beta)) \geq \varepsilon^6/2^{18}$.

Proof: Let $\mathbf{u}$ be such that $\left|\hat{H}_{\mathbf{u}}(\alpha)\right| \geq \frac{\varepsilon}{4}$ for some $\alpha \in \{0,1\}^{3k}$. We will show that there exists a prover strategy which satisfies a significant fraction of constraints incident on all such $\mathbf{u}$. We give the argument for the case when $\hat{H}_{\mathbf{u}}(\alpha) \geq \frac{\varepsilon}{4}$. The argument for the case when the value is negative is identical. By assumption, we have

$$\mathop{\mathbb{E}}_{x \sim \mathcal{U}_{3k}}\left[H_{\mathbf{u}}(x)(-1)^{\alpha \cdot x}\right] \;\geq\; \frac{\varepsilon}{4}\,.$$

By Claim 6.1, the distribution $\mathcal{U}_{3k}$ is very close to the distribution $\varphi(\mathcal{D}_{\mathbf{u}})$. Hence, we also have

$$\mathop{\mathbb{E}}_{x \sim \varphi(\mathcal{D}_{\mathbf{u}})}\left[H_{\mathbf{u}}(x)(-1)^{\alpha \cdot x}\right] \;\geq\; \frac{\varepsilon}{4} - O(\beta\sqrt{k}) \;\geq\; \frac{\varepsilon}{8}$$

since $\beta\sqrt{k} = o(\varepsilon)$. Using the definition of the function $H_{\mathbf{u}}$ we have

$$\begin{aligned} \mathop{\mathbb{E}}_{x \sim \varphi(\mathcal{D}_{\mathbf{u}})}\left[\mathop{\mathbb{E}}_{(\mathbf{w},y,x')|x}\left[(-1)^{G_{\mathbf{w}}(y) + \sum_{i \notin R_{\mathbf{w}}} y_i \cdot c_{u_i v_i, x'_i}}\right] \cdot (-1)^{\alpha \cdot x}\right] &\;\geq\; \frac{\varepsilon}{8} \\ \Rightarrow \quad \mathop{\mathbb{E}}_{(\mathbf{w},y,x') \sim \mathcal{D}_{\mathbf{u}}}\left[(-1)^{G_{\mathbf{w}}(y) + \sum_{i \notin R_{\mathbf{w}}} y_i \cdot c_{u_i v_i, x'_i} + \alpha \cdot \varphi(\mathbf{w},y,x')}\right] &\;\geq\; \frac{\varepsilon}{8} \end{aligned}$$

since sampling $x \sim \varphi(\mathcal{D}_{\mathbf{u}})$ and $(\mathbf{w}, y, x')$ from $\mathcal{D}_{\mathbf{u}}$ conditioned on $\varphi(\mathbf{w}, y, x') = x$ is the same as sampling the tuple $(\mathbf{w}, y, x')$ from $\mathcal{D}_{\mathbf{u}}$ and taking $x = \varphi(\mathbf{w}, y, x')$. The above gives that

$$\mathop{\mathbb{P}}_{(\mathbf{w},x')}\left[\mathop{\mathbb{E}}_{y|(\mathbf{w},x')}\left[(-1)^{G_{\mathbf{w}}(y) + \sum_{i \notin R_{\mathbf{w}}} y_i \cdot c_{u_i v_i, x'_i} + \alpha \cdot \varphi(\mathbf{w},y,x')}\right] \geq \frac{\varepsilon}{16}\right] \;\geq\; \frac{\varepsilon}{16}\,.$$

Note that according to the distribution $\mathcal{D}_{\mathbf{u}}$, given the tuple $(\mathbf{w}, x')$, y is distributed uniformly in $\{0,1\}^{3|\mathbf{w}| + (k - |\mathbf{w}|)}$. We can now analyze the inner expression. We can write $\alpha \cdot \varphi(\mathbf{w}, y, x')$ as $\sum_{i \in R_{\mathbf{w}}} \alpha_i \cdot (\varphi(\mathbf{w}, y, x'))_i + \sum_{i \notin R_{\mathbf{w}}} \alpha_i \cdot (\varphi(\mathbf{w}, y, x'))_i$ where α_i denotes the 3 bits of α in the positions $(3i-2, 3i-1, 3i)$ and $(\varphi(\mathbf{w}, y, x'))_i$ denotes the corresponding 3 bits of $x = \varphi(\mathbf{w}, y, x')$. For $i \in R_{\mathbf{w}}$, we have $x_i = y_i$. Also, for $i \notin R_{\mathbf{w}}$, $x_i = 000$ if $y_i = 0$ and x'_i if $y_i = 1$. Thus, we can write $\alpha \cdot \varphi(\mathbf{w}, y, x')$ as

$$\alpha \cdot \varphi(\mathbf{w}, y, x') \;=\; \sum_{i \in R_{\mathbf{w}}} \alpha_i \cdot y_i + \sum_{i \notin R_{\mathbf{w}}} y_i \cdot (\alpha_i \cdot x'_i)\,.$$

Consider a pair $(\mathbf{w}, x')$ for which the inner expectation is large. Grouping the terms in the expectation, we have

$$\mathop{\mathbb{E}}_{y|(\mathbf{w},x')}\left[(-1)^{G_{\mathbf{w}}(y) + \sum_{i \notin R_{\mathbf{w}}} y_i \cdot (\alpha_i \cdot x'_i + c_{u_i v_i, x'_i}) + \sum_{i \in R_{\mathbf{w}}} y_i \cdot \alpha_i}\right] \geq \frac{\varepsilon}{16}$$

Since y is uniformly distributed, the above expression gives that $\mathbb{E}_y\left[(-1)^{G_{\mathbf{w}}(y)}(-1)^{\overline{\alpha}\cdot y}\right] \geq \frac{\varepsilon}{16}$ where $\overline{\alpha} \in \{0,1\}^{3|\mathbf{w}|+(k-|\mathbf{w}|)}$ is defined as

$$\overline{\alpha}_i \;=\; \begin{cases} \alpha_i & \text{for } i \in R_{\mathbf{w}} \\ \alpha_i \cdot x'_i + c_{u_i v_i, x'_i} & \text{for } i \notin R_{\mathbf{w}} \end{cases}$$

This implies that for a question pair $(\mathbf{u},\mathbf{w})$ if the first prover answers with the linear function $L_{\mathbf{u}}(x) = \alpha \cdot x$ and the second prover answers with the linear function $L_{\mathbf{w}}(x) = \overline{\alpha}\cdot y$, then their answers satisfy

$$\begin{aligned} L_{\mathbf{u}}|_i(z) \;&=\; \alpha_i \cdot z \;=\; \overline{\alpha}_i \cdot z \;=\; L_{\mathbf{w}}|_i(z) & \forall i \in R_{\mathbf{w}}, \forall z \in \{0,1\}^3 \\ L_{\mathbf{u}}|_i(x'_i) \;&=\; \alpha_i \cdot x'_i \;=\; \overline{\alpha}_i + c_{u_i v_i, x'_i} \;=\; L_{\mathbf{w}}|_i(1) + c_{u_i v_i, x'_i} & \forall i \notin R_{\mathbf{w}} \end{aligned}$$

which are all the constraints checked by the outer verifier. Thus, the strategy for the provers is simply as follows. Given a $\mathbf{u}$, the first prover gives a linear function $\alpha \cdot x$ for a random α such that $\left|\hat{H}_{\mathbf{u}}\right| \geq \frac{\varepsilon}{4}$. Similarly, given a $\mathbf{w}$, the second prover answers with the function $\overline{\alpha}\cdot y$ for a random $\overline{\alpha}$ such that $\mathbb{E}_y\left[(-1)^{G_{\mathbf{w}}(y)}(-1)^{\overline{\alpha}\cdot y}\right] \geq \frac{\varepsilon}{16}$. The probability that they choose a consistent pair $(\alpha, \overline{\alpha})$ is at least $\frac{\varepsilon^2}{16}\cdot\frac{\varepsilon^2}{256} = \frac{\varepsilon^4}{2^{12}}$. Thus, the probability that they make the outer verifier accept over the choice of the pair $(\mathbf{u},\mathbf{w})$ is at least $\frac{\varepsilon}{4}\cdot\frac{\varepsilon}{16}\cdot\frac{\varepsilon^4}{2^{12}} = \frac{\varepsilon^6}{2^{18}}$. ■

The argument for the second case is almost identical.

Claim 6.6 *Let* $\mathbb{P}_{\mathbf{u}}\left[\exists q_{\mathbf{u}} \text{ such that } \mathbb{E}_x\left[H_{\mathbf{u}}(x)(-1)^{q_{\mathbf{u}}(x)}\right] \geq \frac{9}{10} - \frac{\varepsilon}{2}\right] \;\geq\; \frac{\varepsilon}{4}$. *Then* $\mathsf{Val}^{(2)}(\mathcal{G}, k, \beta) \geq \varepsilon^2/4$.

Proof: Let $\mathbf{u}$ be such that $\mathbb{E}_{x\sim\mathcal{U}_{3k}}\left[H_{\mathbf{u}}(x)(-1)^{q(x)}\right] \geq \frac{9}{10} - \frac{\varepsilon}{2}$ for some quadratic function q. As before, this gives

$$\begin{aligned} & \underset{x\sim\varphi(\mathcal{D}_{\mathbf{u}})}{\mathbb{E}}\left[H_{\mathbf{u}}(x)(-1)^{q(x)}\right] \;\geq\; \frac{9}{10} - \varepsilon \\ \Rightarrow\; & \underset{(\mathbf{w},y,x')\sim\mathcal{D}_{\mathbf{u}}}{\mathbb{E}}\left[(-1)^{G_{\mathbf{w}}(y)+\sum_{i\notin R_{\mathbf{w}}} y_i\cdot c_{u_i v_i, x'_i} + q(\varphi(\mathbf{w},y,x'))}\right] \;\geq\; \frac{9}{10} - \varepsilon \\ \Rightarrow\; & \underset{(\mathbf{w},x')}{\mathbb{P}}\left[\underset{y}{\mathbb{E}}\left[(-1)^{G_{\mathbf{w}}(y)+\sum_{i\notin R_{\mathbf{w}}} y_i\cdot c_{u_i v_i, x'_i} + q(\varphi(\mathbf{w},y,x'))}\right] \geq \frac{9}{10} - 2\varepsilon\right] \;\geq\; \varepsilon\,. \end{aligned}$$

Now, given x', we can interpret $q(\varphi(\mathbf{w},y,x'))$ as a polynomial in y. Let $q' : \{0,1\}^{3|\mathbf{w}|+(k-|\mathbf{w}|)} \to \{0,1\}$ be the quadratic polynomial in y obtained as follows. Replace the 3 bits corresponding to the input x_i in q by the 3 bits corresponding to y_i for $i \in R_{\mathbf{w}}$. Also, for $i \notin R_{\mathbf{w}}$, let x_{i1}, x_{i2}, x_{i3} be the 3 bits corresponding to x_i. Replace them by $y_i x'_{i1}, y_i x'_{i2}, y_i x'_{i3}$ where $x'_{i1}, x'_{i2}, x'_{i3}$ are the 3 bits of x'_i.

Let $\overline{q}$ denote the polynomial $q' + \sum_{i\notin R_{\mathbf{w}}} y_i \cdot c_{u_i v_i, x'_i}$. It is easy to check that

$$\begin{aligned} q|_i(z) \;&=\; \overline{q}|_i(z) & \forall i \in R_{\mathbf{w}}, \forall z \in \{0,1\}^3 \\ q|_i(x'_i) \;&=\; q'|_i(1) \;=\; \overline{q}|_i(1) + c_{u_i v_i, x'_i} & \forall i \notin R_{\mathbf{w}}\,. \end{aligned}$$

Thus, in this case, given $\mathbf{u}$, the first prover answers with the *unique* quadratic polynomial q such that $\mathbb{E}_x\left[H_{\mathbf{u}}(x)(-1)^{q(x)}\right] \geq \frac{9}{10} - \varepsilon$, and the second prover, given $\mathbf{w}$, answers with the *unique* quadratic polynomial $\overline{q}$ such that $\mathbb{E}_y\left[(-1)^{G_{\mathbf{w}}(y)(-1)^{\overline{q}(y)}}\right] \geq \frac{9}{10} - 2\varepsilon$. If $\mathbf{u},\mathbf{w}$ are as above, then q and $\overline{q}$ satisfy the required conditions and the verifier accepts. Hence, this strategy makes the verifier accept with probability at least $(\varepsilon/4)\cdot\varepsilon = \varepsilon^2/4$. ■

Since $\mathsf{Val}^{(2)}(\mathcal{G}(\Phi,k,\beta)) \geq \mathsf{Val}^{(1)}(\mathcal{G}(\Phi,k,\beta))$, we have $\mathsf{Val}^{(2)}(\mathcal{G}(\Phi,k,\beta)) \geq \varepsilon^6/2^{18}$ in both cases. ■

6.3 Choice of Parameters

We require that $\beta\sqrt{k} = o(\varepsilon)$ and the soundness of the outer PCP, which is $(1-\gamma^2)^{\Omega(\beta k)}$ is less than $\varepsilon^6/2^{18}$. Since we are working with $d = 2$, we have $\gamma = \Omega(1)$ and so this means that we require $\beta k = \Omega(\log(1/\varepsilon))$. Choosing $k = \Omega((1/\varepsilon^3) \cdot \log^2(1/\varepsilon))$ and $\beta = O(\varepsilon^3/\log(1/\varepsilon))$ meets both these conditions. Finally, we choose the parameter η for the Linear vs. Degree-d Label-Cover problem to be $O(\varepsilon/k)$ so that the completeness of the composed PCP is $1-\varepsilon$.

7 Parallel Repetition of Linear vs. Degree-d PCPs

Let Φ be an instance of Linear vs. Degree-d Label-Cover as defined in Definition 2.7. Consider a new instance of Linear vs. Degree-d Label-Cover defined by the following two-prover game:

- The verifier picks constraints $(u_1, v_1, x'_1), \ldots, (u_k, v_k, x'_k) \in \Phi$.
- The question sent to the first prover is $\mathbf{u} = (u_1, \ldots, u_k)$.
- The question sent to the second prover is $\mathbf{v} = (v_1, \ldots, v_k)$.
- The first prover is required to respond with a degree-d polynomial $F_{\mathbf{u}} : \{0,1\}^{(d+1)k} \to \{0,1\}$ and the second prover responds with a bit $A(\mathbf{v}) \in \{0,1\}$. The verifier accepts if

$$F_{\mathbf{u}}(x'_1, \ldots, x'_k) = A(\mathbf{v}) + c_{u_1 v_1, x'_1} + \ldots + c_{u_k v_k, x'_k}\,.$$

Note that the label for $\mathbf{u} = (u_1, \ldots, u_k)$ is now a single polynomial, which is supposed to be the XOR of the polynomials $F_{u_1}, \ldots, F_{u_k}$. Similarly, the label for $\mathbf{v}$ is supposed to be the XOR of $A(v_1), \ldots, A(v_k)$.

It is natural to ask that if any degree-d labeling can satisfy (say) at most $s = 1 - 2^{-O(d)}$ fraction of constraints in Φ, then does the fraction of constraints satisfiable in the new instance obtained above decrease to $1/2 + \exp(-\Omega_s(r))$. Note that since the first prover is required to prove a single polynomial $F_{\mathbf{u}}$ in $(d+1)k$ variables, which might involve cross-terms between the inputs $x_1, \ldots, x_k$, the above does not follow from known results on parallel repetition.

If it is indeed possible to reduce the soundness of the Linear vs. Degree-d Label-Cover label cover to a value arbitrarily close to $1/2$ as above, then one can in fact avoid the complication of going through the smoothed parallel repetition approach in the outer PCP. In our current proof, both provers are required to decode a degree-2 polynomial in a way that the polynomial of the second prover is a projection of the polynomial of the first prover. We are unable to do this unless there is only a unique choice of the quadratic polynomial for each prover.

However, using the above instance of Linear vs. Degree-d Label-Cover, one can instead apply the HLT directly the functions $F_{\mathbf{u}}$ by only reading the corresponding values for $A(\mathbf{v})$. Then, if the table of values $A(\mathbf{v})$ satisfies the verifier in the composed PCP with probability more than $1/16$, the first prover can find functions $F_{\mathbf{u}}$ which are polynomials of degree at most 2 (by using the table $A(\mathbf{v})$), which satisfy significantly more than half of the constraints in the instance above. We omit the details of the analysis.

We believe that if the parallel repetition as above can indeed be used to reduce the soundness even for degree-d labeling then an appropriate modification of our PCP might even give an alternate proof of the soundness guarantee of $2^{-(2^{d+1}-(d+1)-1)}$ with $2^{d+1}-1$ queries.

References

[1] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron. Testing Reed-Muller codes. *IEEE Transactions on Information Theory*, 51(11):4032–4039, 2005.

[2] N. Alon and J. Spencer. *The Probabilistic Method.* Wiley, 2008.

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

[4] S. Arora and S. Safra. Probabilistic checking of proofs : A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.

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

[6] S. O. Chan. Personal Communication, 2012.

[7] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for unique games. In *Proc. ACM Symposium on the Theory of Computing*, pages 205–214, 2006.

[8] L. Engebretsen and J. Holmerin. More efficient queries in pcps for np and improved approximation hardness of maximum csp. *Random Struct. Algorithms*, 33(4):497–514, 2008.

[9] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.

[10] J. Hastad. Clique is hard to approximate within $n^{1-\varepsilon}$. *Acta Mathematica*, 182:105–142, 1999.

[11] J. Hastad. Some optimal inapproximability results. *Journal of ACM*, 48:798–859, 2001.

[12] J. Hastad and A. Wigderson. Simple analysis of graph tests for linearity and PCP. In *Proc. 16th IEEE Conference on Computational Complexity*, 2001.

[13] T. Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proc. ACM Symposium on the Theory of Computing*, pages 411–419, 2007.

[14] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symposium on Theory of Computing*, 2002.

[15] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.

[16] S. Khot and M. Safra. A two prover one round game with strong soundness. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science*, pages 648–657, 2011.

[17] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Proc. 46th IEEE Symposium on Foundations of Computer Science*, pages 21–30, 2005.

[18] A. Rao. Parallel repetition in projection games and a concentration bound. In *Proc. ACM Symposium on the Theory of Computing*, pages 1–10, 2008.

[19] R. Raz. A parallel repetition theorem. *SIAM J. of Computing*, 27(3):763–803, 1998.

[20] A. Samorodnitsky. Low-degree tests at large distances. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, pages 506–515, 2007.

[21] A. Samorodnitsky and L. Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proc. 32nd ACM Symposium on Theory of Computing*, pages 191–199, 2000.

[22] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influence of variables, and PCPs. In *Proc. 38th ACM Symposium on Theory of Computing*, 2006.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de