The Biased Homogeneous r-Lin Problem

Suprovat Ghoshal ✉
University of Michigan, Ann Arbor, MI, USA

Abstract

The p-biased Homogeneous r-Lin problem ($\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$) is the following: given a *homogeneous* system of r-variable equations over $\mathbb{F}_2$, the goal is to find an assignment of relative weight p that satisfies the maximum number of equations. In a celebrated work, Håstad (JACM 2001) showed that the unconstrained variant of this i.e., Max-3-Lin, is hard to approximate beyond a factor of 1/2. This is also tight due to the naive random guessing algorithm which sets every variable uniformly from $\{0,1\}$. Subsequently, Holmerin and Khot (STOC 2004) showed that the same holds for the *balanced* $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}$ problem as well. In this work, we explore the approximability of the $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$ problem beyond the balanced setting (i.e., $p \neq 1/2$), and investigate whether the (p-biased) random guessing algorithm is optimal for every p. Our results include the following:

- The $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$ problem has no efficient $\frac{1}{2} + \frac{1}{2}(1-2p)^{r-2} + \varepsilon$-approximation algorithm for every p if r is even, and for $p \in (0, 1/2]$ if r is odd, unless $\mathsf{NP} \subset \cup_{\varepsilon>0}\mathsf{DTIME}(2^{n^{\varepsilon}})$.
- For any r and any p, there exists an efficient $\frac{1}{2}(1-e^{-2})$-approximation algorithm for $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$. We show that this is also tight for odd values of r (up to $o_r(1)$-additive factors) assuming the Unique Games Conjecture.

Our results imply that when r is even, then for large values of r, random guessing is near optimal for every p. On the other hand, when r is odd, our results illustrate an interesting contrast between the regimes $p \in (0, 1/2)$ (where random guessing is near optimal) and $p \to 1$ (where random guessing is far from optimal). A key technical contribution of our work is a generalization of Håstad's 3-query dictatorship test to the p-biased setting.

2012 ACM Subject Classification Theory of computation → Problems, reductions and completeness

Keywords and phrases Biased Approximation Resistance, Constraint Satisfaction Problems

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2022.47

Category APPROX

Acknowledgements Most of this work was carried out when the author was visiting KTH, Stockholm. The author thanks Johan Håstad for inviting him to visit KTH, for the many insightful discussions and ideas that have played a key role in this work, and for his comments on a previous draft of the manuscript. The author thanks Per Austrin for several useful discussions on related topics. The author also thanks the anonymous reviewers for several helpful suggestions on the manuscript, and for fixing an issue in a previous version of the work.

1 Introduction

The problem of finding solutions to systems of linear equations is one of fundamental importance. While in theory, the exact running time complexity of even efficiently solvable instances has profound implications in the theory of algorithms [26], the question of approximability of infeasible systems is also fundamental and has been studied widely [19, 12, 24, 14, 8]. A particularly useful instantiation of this is the $\mathsf{Max}\text{-}r\text{-}\mathsf{Lin}$ problem[1] where given a (possibly infeasible) system of r-variables equations over $\mathbb{F}_2$, the objective is to find an assignment

[1] In the literature, the $\mathsf{Max}\text{-}r\text{-}\mathsf{Lin}$ problem is typically referred to as $\mathsf{Max}\text{-}r\text{-}\mathsf{Lin}_q$, where the indexing by q indicates that the equations are over $\mathbb{F}_q$. We drop the indexing by q since the current work deals only with the setting $q = 2$.

to the variables that satisfies the maximum fraction of equations. The Max-r-Lin problem manifests as various computational problems in error correcting codes, combinatorial optimization, probabilistically checkable proofs among many others. In particular, its study played a seminal role in the development Proabablistically Checkable Proofs based reductions [7, 19, 24], and techniques introduced for studying its hardness have now become staple tools in hardness of approximation.

A standout result among these is the celebrated work of Håstad [19] who showed that for $r \geq 3$, the Max-r-Lin problem is NP-hard to approximate beyond a factor of $1/2$. This is clearly tight since the naive algorithm which outputs a uniformly random assignment also satisfies at least $1/2$-fraction of constraints[2]. This property of the "random guessing algorithm being optimal" also happens to hold for a much broader class of combinatorial optimization problems, and is formally studied under the notion of "approximation resistance". The ubiquity of this notion has lead to several works which systematically study such problems [5, 4, 10], including landmark results such as which give complete conditional and unconditional characterizations of approximation resistant predicates [3, 25].

A key problem studied in this context is the *Balanced Homogeneous* Max-3-Lin problem, where given a *homogeneous* system of linear equations, the goal is to find a *balanced* assignment that satisfies the maximum fraction of constraints. Clearly, naive random guessing is still a candidate algorithm for this setting as well, since it produces balanced[3] assignments that satisfy at least half of the constraints. Naturally, this leads one to ask if random guessing is still optimal in this setting as well? This was answered in the affirmative by Holmerin and Khot [22] who ruled out efficiently approximability beyond $1/2$ assuming SAT does not admit sub-exponential time algorithms. Subsequently, Håstad and Manokaran [21] strengthened the above hardness result to rule out quasi-polynomial time algorithms which give better than $1/2$ approximation assuming $\mathsf{NP} \not\subset \mathsf{DTIME}(\exp(\log n)^{O(1)})$.

In this work, we study a natural generalization of the above and investigate the approximability of the homogeneous Max-3-Lin problem beyond the balanced setting. Formally, we study the p-biased version of the above problem, which we refer to as the Hom-r-Lin$_p$ problem. We define it formally below:

▶ **Definition 1** (Hom-r-Lin$_p$). *Given $p \in (0,1)$, an instance $\psi([n], E)$ of the p-biased* Hom-r-Lin *problem is given by a set of homogeneous equations over $\mathbb{F}_2$ defined by a set of r-arity hyperedges $E := \{e_1, \ldots, e_m\}$ over variables $\{x_1, \ldots, x_n\}$, where the i^{th} hyperedge e_i implies the constraint $\oplus_{j \in e_i} x_j = 0$. Here, the objective is to find a labeling of relative weight p which satisfies the maximum fraction of hyperedges (constraints).*

Clearly, for $p = 1/2$, the above recovers the balanced setting, for which the aforementioned works show that the uniformly random guessing algorithm is optimal. On the other hand, for any $p \in (0,1)$, one can naturally consider the following extension of random guessing: set each variable to 1 independently with probability p – we refer to this as *p-biased* random guessing. Clearly, with high probability, p-biased random guessing will return an assignment

---

[2] In fact, a simpler deterministic $1/2$-approximation algorithm is known for Max-r-Lin: given any assignment to the set of the variables, that or its negation will always satisfy at least $1/2$ of the constraints – hence, outputting the best of any assignment and its negation is gives a trivial $1/2$-approximation. However, since negating the assignment can also change its relative weight, this approach doesn't yield an algorithm for the weight constrained setting considered in this paper.

[3] Strictly speaking, it produces almost balanced assignments, which can be converted to exactly balanced assignments by changing $o_n(1)$-variables. This only affects the approximation factor in lower order $o(1)$ terms.

with relative weight $\approx p$, i.e., it is again a feasible candidate algorithm. And keeping with the above trend, one may ask if the p-biased random guessing algorithm is still optimal for the $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$ problem, for every p. In other words, we ask the following:

Is the $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$ problem approximation resistant for every $p \in (0,1)$?

The above is the main motivating question which we seek to address in the current work. At a finer level, our goal is to understand the approximability of the $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$ problem as a function of the parameter p and the arity r. This formulation of the problem brings in several additional dimensions to the existing literature on approximation resistance which typically deals with uniformly random guessing, as opposed to the more general p-biased guessing studied in the current work.

1.1 Our Results

In this work, we study the bias dependent approximability of $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$, and make substantial progress towards understanding the above question. In the interest of keeping the presentation concise, we will first state our results for the setting when r is odd, since this setting exhibits a more interesting dependence on the parameter p. We will then point out how the results change when r is even.

The $p \leq 1/2$ setting. Our first result is the following theorem which shows that $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$ predicate is close to being approximation resistant for large values of r.

▶ **Theorem 2.** *Fix $p, \eta \in (0, 1/2)$, and $r \geq 3$. Then assuming $\mathsf{NP} \not\subseteq \cup_{\varepsilon>0}\mathsf{DTIME}(2^{n^\varepsilon})$ the following holds. Given an instance ψ of $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}$, there is no polynomial time algorithm that can distinguish between the following cases:*

- **YES Case**. *There exists an assignment of relative hamming weight p, which satisfies at least $1-\eta$ fraction of constraints.*
- **NO Case**. *No assignment of relative hamming weight p satisfies more than $\frac{1}{2} + \frac{1}{2}(1-2p)^{r-2} + \eta$ fraction of constraints.*

The above theorem implies that there are no efficient algorithms which give a better than $\frac{1}{2}(1 + (1-2p)^{r-2})$-approximation. On the other hand, it is easy to verify that the p-biased random guessing is a $\frac{1}{2}(1 + (1-2p)^{r})$-approximation algorithm for $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$. Hence, the above theorem implies that for large r, biased random guessing is almost optimal.

The $p > 1/2$-setting. Our second result shows that the almost approximation resistant behavior of the $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$ predicate breaks down in the $p > 1/2$ setting. This is implied by the following theorem which gives an efficient randomized p-independent approximation algorithm for $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$.

▶ **Theorem 3.** *For every $r \geq 3$ and every $p \in (0,1)$, there exists an efficient randomized $\beta_r/2$-approximation algorithm for $\mathsf{Hom}\text{-}r\text{-}\mathsf{Lin}_p$. Here $\beta_r := 1 - (1-1/r)^{2r}$ is a decreasing function of r satisfying $\lim_{r\to\infty} \beta_r = (1-e^{-2})$.*

The algorithm for the above theorem is based on a linear programming + rounding approach inspired by algorithms for hitting set and is markedly different from random guessing. The above theorem implies that even for $r = 3$, the random guessing algorithm is *strictly sub-optimal* for all $p > 1/2(1 + e^{-2/3})$. We also show that the above approximation guarantee is tight (up to $o_r(1)$-factors) assuming the Unique Games Conjecture [23].

▶ **Theorem 4.** *Assuming the Unique Games Conjecture, the following holds for every* **odd** $r \geq 3$ *and* $\eta \in (0,1)$*. Let* $p := 1 - 1/r$*. Given a* Hom-r-Lin *instance* ψ *it is* NP*-hard to distinguish between the following cases:*

- **YES Case**. *There exists an assignment of relative weight* p *which satisfies at least* $(1-\eta)$*-fraction of constraints.*
- **NO Case**. *No assignment of relative weight* p *satisfies more than* $\left(\frac{\beta_r}{2} + O(1/r)\right)$*-fraction of constraints.*

The even r-setting . It is easy to see that when r is even, the $p \geq 1/2$ and $p \leq 1/2$ regimes are symmetric. To see this, given a system of equations $\psi([n], E)$, and an assignment $(x_1, \ldots, x_n)$ of relative weight p, consider the negated assignment $x'_i = 1 \oplus x_i$. Since r is even and the constraints are homogeneous, the negated assignment will satisfy a constraint if and only if the original assignment satisfied the constraint. Furthermore, the negated assignment will have a relative weight of $1-p$. This observation implies that the Hom-r-Lin$_p$ problem behaves identically under the bias constraints p and $1-p$ for every p. This observation along with Theorem 2 results in the following corollary.

▶ **Corollary 5.** *For every* $p, \varepsilon \in (0,1)$ *and even* $r \geq 4$*, there is no polynomial time* $\frac{1}{2}(1 + (1-2p)^{r-2}) + \varepsilon$*-approximation algorithm for* Hom-r-Lin$_p$ *unless* NP $\subset$ DTIME$(2^{n^{\varepsilon}})$ *for any* $\varepsilon > 0$*.*

Threshold Phenomena. The above results show that when $p < 1/2$, then the random guessing algorithm is almost optimal, whereas this behavior breaks down for the $p > 1/2$ regime when r is odd. In particular, in the setting of large p's the Hom-r-Lin predicate exhibits a *hitting set* like behavior – our algorithm and hardness results (Theorem 4 and 3) are based on this connection. In fact, we believe that when $p \leq 1/2$, the p-biased random guessing algorithm is indeed optimal, and the current gap in the hardness result is due to technical bottlenecks[4] that arise more generally in the context of hardness reductions involving problems with global constraints.

Furthermore, our results hint at the possibility of the existence of a threshold p_r beyond which the approximation resistance of Hom-r-Lin breaks. In particular, the hard distribution for Theorem 4 seems to indicate that this threshold is $1 - 1/r$. Lastly, we point out that the even and odd r settings contrast nicely against each other as while Hom-r-Lin$_p$ can be approximation resistant only for a certain range of p when r is odd, it is possibly approximation resistant for every p when r is even.

1.2 Related Works

The Max-r-Lin Problem. The Max-r-Lin problem has been studied extensively in the literature. In particular, when $r = 2$, it expresses the affine UNIQUEGAMES problem which is central to Khot's Unique Games Conjecture (UGC) [23], and has been extensively studied by several works [12, 11, 24]. In particular, the algorithmic results from [11] show that the

[4] In particular, the gap of $(1-2p)^2$ in the second term is primarily due to the following reason. As is standard in Label Cover based reductions, out of the r-queries made by out dictatorship test, 2 of its queries are made to the large side table for the consistency check. However, the outer verifier (i.e., Mixing Label Cover) can only guarantee mixing w.r.t. vertices on the smaller side (due to which we are able to recover the $(1-2p)^{r-2}$)-factor, and doesn't guarantee mixing on the larger side (due to which we lose out by a factor $(1-2p)^2$ in the p-dependent term). The question of constructing hard outer verifiers with mixing guarantees with respect all vertices is a fundamental technical challenge in itself in this area.

$r = 2$ setting is not approximation resistant. For $r \geq 3$, Håstad [19] showed that Max-r-Lin is hard to approximate beyond a factor of $1/2 + \varepsilon$. In fact, Håstad's result actually shows that the Max-r-Lin problem is approximation resistant over any finite abelian group, which was later strengthened to the setting of infeasible instances over non-abelian groups [14]. More recently, Bhangale and Khot [8] give tight hardness results for satisfiable Hom-r-Lin instances over non-abelian finite groups. Specifically, given a non-abelian group G, they showed that it is hard to approximate the satisfiable Hom-r-Lin problem beyond a factor of $1/|[G,G]| + \varepsilon$, where $[G,G]$ is the commutator sub-group of G – this is matched by a folklore algorithm for the same, we refer interested readers to [8] for more details on this.

Approximation Resistance. Starting with the work of Håstad [19], the question of understanding the conditions under which random guessing is optimal has generated great interest, and has been studied by several works. In particular, the work of Håstad [20] showed that 2-arity CSPs can never be approximation resistant. For when the arity r is 3, it is known that a CSP on a predicate can be approximation resistant if and only if the predicate is implied by 3-XOR [19, 32]. The work of Hast [18], gives an almost complete characterization for 4-arity CSPs. On the other hand, stronger results are known assuming stronger hypotheses. For e.g., assuming UGC, Håstad and Austrin [3] showed that a uniformly random predicate is approximation resistant with high probability. Austrin and Khot [4] gave a complete characterization of approximation resistance for k-partite CSPs, which was later strengthened to the setting of all CSPs by Khot, Tulsiani and Worah [25].

Globally Constrained CSPs. The Hom-r-Lin$_p$ problem also falls within the framework of Max-CSPs with global cardinality constraints i.e., CSPs where the objective is to find a labeling that satisfies the maximum number of edge constraints while "strictly" respecting global constraints. Such CSPs express extensively studied problems such as Max-Bisection [28, 2], Densest-k-Subgraph [15, 9], Small Set Expansion [29, 30]. There have been several works which also systematically study such CSPs under a more general framework. For e.g., the works of Guruswami and Sinop [17] and Bansal et al. [1] propose general purpose algorithmic frameworks for solving globally constrained CSPs. A closely related work is that of Ghoshal and Lee [16], who study the bias parameter dependent approximation curve for globally constrained Boolean CSPs, and give upper and lower bounds which are matching up to constant factors for constant arity, assuming the Small Set Expansion Hypothesis. In particular, their results imply that there exists a $\Omega(2^{-r})$-approximation algorithm, which again is p-independent but deteriorates with increasing r.

2 Warm-up : The $p \leq 1/2$ vs. $p > 1/2$ settings

In this section, we first provide some intuition on why the behavior of the approximation curve of the Hom-r-Lin predicate in the $p \geq 1/2$ regime is different from that of $p < 1/2$ when r is odd. Consider the random guessing algorithm for the Hom-r-Lin problem which sets every variable to 1 with probability p. Then, elementary computation shows that this assignment satisfies $\frac{1}{2} + \frac{1}{2}(1-2p)^r$-fraction of constraints in expectation and w.h.p. the relative weight of the assignment returned by the algorithm is $\approx p$. Clearly, the approximation guarantee of this algorithm improves as p decreases and is trivially optimal in the limit $p = 0$, and hence it might not be a stretch to posit that the random guessing algorithm is indeed optimal in this regime.

On the other hand, as p increases, the approximation guarantee of the random guessing algorithm worsens, eventually reaching 0 as $p \to 1$. However, while random guessing might be almost optimal when $p \in (0, 1/2)$, there is an inherent reason as to why it should be sub-optimal in the almost all-ones regime. The key insight here is that the random guessing can be wasteful in term of choosing the zeros when the budget of zeros is small (i.e, when $p \to 1$). For instance, consider a Hom-r-Lin$_p$ instance $\psi([n], E)$ whose optimal assignment satisfies nearly all constraints. Then one can show that there exists a small set of vertices in V which intersects (hits) nearly all hyperedges in ψ – this is witnessed by the zero set of the optimal assignment. In contrast, the solutions output by the random guessing algorithm will have the set of zeros spread uniformly throughout the constraint hypergraph, and hence such assignments are likely to miss out on the potential exploits guaranteed by the combinatorial structure of the instance, thus ending up satisfying far fewer than the optimal fraction of constraints. On the other hand, the existence of such a hitting set opens up other possible approach which can use this. For instance, the surrogate problem of finding a small size set which hits the maximum number of hyperedges itself is known to admit efficient constant factor approximation algorithms using simple greedy/linear programming based approaches. The above observations indicate that one might be able to strictly better than random guessing by exploiting the combinatorial structure of the instance.

3 Hardness for $p \leq 1/2$

Our result for $p \leq 1/2$ (Theorem 2) is based on a careful generalization of the standard 3-query dictatorship test of Håstad to the setting of biased long codes. In order to highlight the challenges towards establishing the hardness result, it will be useful to go over the reduction for the hardness of balanced setting (i.e., $p = 1/2$) from [22]. The reduction in [22] for Hom-r-Lin consisted of two key components:

(i) The 3-query dictatorship test of Håstad for 3-Lin.

(ii) A variant of Label Cover[5] with one-sided *mixing* properties.

While the 3-query test from (i) was established much before the work of Holmerin and Khot [22], the key contribution of [22] was a variant of Label Cover with the property that the larger side of the Label Cover instance $\mathcal{L}$ has good expansion – this is needed to ensure that globally balanced assignments also translate to locally balanced assignments in the reduction. As we shall see, while we can use the outer verifier from (ii) as is in our reduction, most of the work will go into modifying and analyzing the inner verifier (i.e, the dictatorship test). Now we shall first briefly describe how (i) and (ii) can be put together to prove the hardness for the balanced setting, and then we will discuss the challenges and the techniques used for going beyond the balanced case.

The 3-query test. In order to understand the challenges towards designing a r-query test that works in our setting, it is instructive to recall the well known 3-query test of [19]. The design of the 3-query test is based on the principle that linear functions with small number of influential coordinates must be close to being dictators, which readily yields test in Figure 1.

[5] A Label Cover instance $\mathcal{L}(U, V, E, [s], [l], \{\pi_e\}_{e \in E})$ is a 2-CSP on the bipartite constraint graph with left and right vertex sets U and V respectively, with the edge constraint set E. Every edge $(u, v) \in E$ is identified with a projection constraint $\pi_{u,v} : [l] \to [s]$. The objective is to find a global labeling of U and V using $[s]$ and $[l]$ respectively, that satisfies the maximum fraction of edge constraints in E.

Input. Long code table $f : \{0,1\}^k \to \{\pm 1\}$ satisfying $\mathsf{E}_{x\sim\{0,1\}^k}[f(x)] = 0$.
Test.
1. Sample $x, y \sim \{0,1\}^k_{1/2}$ and $\rho \sim \{0,1\}^k_\eta$. Set $z := x \oplus y \oplus \rho$.
2. Accept if and only if

$$f(x) \cdot f(y) \cdot f(z) = 1.$$

Figure 1 3-query test.

The analysis of this test proceeds through the following well-traversed path. Firstly, it is easy to see that the dictator assignment $f = \chi_i$ is balanced and passes the test with probability $(1-\eta)$. On the other hand, the soundness direction proceeds as follows: given a balanced assignment $f : \{0,1\}^k \to \{\pm 1\}$, we can arithmetize the acceptance probability of the test in term of f and express it as:

$$\Pr[\text{ Test Accepts }] = \frac{1}{2} + \frac{1}{2}\mathsf{E}_{x,y,z}[f(x)f(y)f(z)].$$

Furthermore, by standard Fourier analytic arguments, we can manipulate the RHS of the above and further write the RHS in terms of the Fourier coefficients of f as

$$\Pr[\text{ Test Accepts }] = \frac{1}{2} + \frac{1}{2}\sum_{\beta \neq \mathbf{0}_k} \widehat{f}(\beta)^3(1-\eta)^{|\beta|}, \tag{1}$$

where note that the summation term does not feature the term $\beta = \mathbf{0}_k$ since the Fourier coefficient corresponding to the all zeros term vanishes due to the *balancedness* of the long code table. Hence, if the test passes with probability strictly bounded away from $1/2$, then the summation term is strictly positive, which allows one to show that there exists $\beta \in \{0,1\}^k$ such that $\widehat{f}(\beta) \geq \Omega(1)$ and $|\beta| \leq O(1/\eta)$ i.e., f has non-trivial correlation with a low degree term. In particular, note that since the summation omits the $\beta = \mathbf{0}_k$ term, the low degree term is guaranteed to be non-trivial – this property is used crucially in the composition step which we describe next.

Composing with Mixing Label Cover. Given the above dictatorship test, the next step is to compose it with the outer verifier (i.e., Label Cover). This is a standard step in dictatorship test based reductions, and roughly goes as follows. Informally, given a Label Cover instance $\mathcal{L}(U, V, E, \Sigma, \{\pi_e\}_{e\in E})$, the reduction introduces a long code table $f_w : \{0,1\}^{|\Sigma_w|} \to \{\pm 1\}$ for every vertex $w \in U \cup V$ – the entries of the long code tables correspond to the variable set of the reduction. The constraint set of the reduction corresponds to the distribution over checks generated by the following process:

- Sample a random edge $(u, v) \sim \mathcal{L}$.
- Run the 3-query test by querying the positions from the long code tables $f_u(x)$ $f_v(v)$ and $f_v(\pi_e(x) \oplus y \oplus z)$.

The above seeks to simultaneously test the following (i) the tables f_u and f_v are correlated with non-trivial low degree terms, and (ii) the sets corresponding to the low degree terms have non-trivial intersection under the projection map π_e. These checks are indeed useful due to the following principle: if (i) and (ii) hold simultaneously for a non-negligible faction of fraction of edges incident on a vertex $v \in V$, then the assignment to the long code tables can be used to decode a labeling that satisfies a non-trivial fraction of constraints incident on v in $\mathcal{L}$. However, lifting the soundness guarantee of the test to ensure (i) and (ii) from above requires the empty set Fourier coefficient term to vanish in expectation for a random choice of $u \sim N_{\mathcal{L}}(v)$ (recall that this is critical in ensuring that the low degree term in (i) is non-trivial). This is precisely where the mixing property is useful – it ensures that if the global set of U-tables are balanced, then for most choices of $v \in V$, the long code tables in the neighborhood $N_{\mathcal{L}}(v)$ of v are also balanced (in expectation). In terms of the soundness analysis of the full reduction, this translates to the guarantee that for most choices of v, empty set Fourier coefficient term has negligible contribution. However, the mixing property of the Label Cover instance comes at the cost of super-polynomial sized constructions, due to which the inapproximability only holds under stronger assumptions such as $\mathsf{NP} \not\subset \cup_{\varepsilon>0}\mathsf{DTIME}(2^{n^{\varepsilon}})$.

3.1 The Biased r-query test

Towards generalizing the above to the setting of any $p \leq 1/2$, a clear first obstacle is to design a dictatorship test that has the right completeness and soundness guarantees as functions of p and r. It turns out that this is the key issue that we need to address, as once we are equipped with the right test, the composition step and its analysis follows almost as is using the techniques from the balanced case. Formally, our goal is to design a r-query test with the following property:

- **Completeness**: If $f : \{0,1\}^k \to \{\pm 1\}$ is a dictator, then it passes the test with probability $1-\eta$, and induces an assignment of relative weight p i.e., $\mathsf{E}_{x \sim \{0,1\}^k_p}[f(x)] = 1 - 2p$.
- **Soundness**: If f is an assignment of relative weight p and passes the probability at least $\frac{1}{2}(1 + (1-2p)^{r-2}) + \Omega(1)$, then f is correlated with a non-trivial Fourier character of low-degree .

It is easy to see that the 3-query test from Figure 1 does not imply the above conditions, and in particular, fails the completeness guarantee. This is due to the observation that since the marginal distribution of the points queried by the test is uniform over $\{0,1\}^k$, any dictator assignment will be balanced under the distribution. This leads us to consider analogous p-"biased" dictatorship tests over the p-biased hypercube[6].

Towards designing such a "biased" dictatorship test, a natural first approach (while ignoring the noise component ρ) would be to consider the distribution over triples over choices of (x, y, z) such that $x \oplus y \oplus z = \mathbf{0}_k$ such that x, y and z are marginally distributed as $\{0,1\}^k_p$ – such a test was explored (for slices of the hypercube) in the context of *direct sum testing* in [13] for the regime where the soundness of the test approaches 1. However, to the best of our knowledge, the techniques used in [13] seem to rely on the soundness parameter being close to 1, whereas our application would require us to establish guarantees in settings where the soundness parameter is bounded away from 1. Furthermore, even in the regime where the soundness approaches 1, it is not easy to see how the techniques of [13] generalize to the setting of higher values of r. Finally, the techniques used in the analysis of the above test (Figure 1) do not generalize well to this setting, since establishing (1) heavily relies on the fact that the Fourier characters for the unbiased distribution are linear operators over $\mathbb{F}_2^k$ i.e., $\chi_\alpha(x \oplus y) = \chi_\alpha(x) \oplus \chi_\alpha(y)$, a property that does not hold for the general p-biased Fourier characters (i.e., when $p \neq 1/2$).

[6] The p-biased Boolean k-hypercube is the k-hypercube $\{0,1\}^k$ equipped with the following measure: $\mu(x) := p^{|x|}(1-p)^{k-|x|}$.

The above observations instead motivate us to consider a test with an asymmetric test distribution, where the first $(r-2)$-query positions are independent p-biased strings and the remaining two strings are uniformly distributed but correlated. Formally, we consider the distribution in the following figure.

Input. Long code table $f : \{0,1\}^k \to \{\pm 1\}$ satisfying $\mathsf{E}_{x \sim \{0,1\}^k_p}[f(x)] = 1 - 2p$.
Test.
1. Sample $x_1, \ldots, x_{r-2} \sim \{0,1\}^k_p$.
2. Sample $y \sim \{0,1\}^k_{1/2}$ and $\rho \sim \{0,1\}^k_\eta$. Set $z := \left(\oplus_{i \in [r-2]} x_i\right) \oplus y \oplus \rho$.
3. Accept if and only if

$$\left(\prod_{i \in [r-2]} f(x_i)\right) \cdot f(y) \cdot f(z) = 1.$$

Figure 2 The p-biased r-query test.

The above distribution allows us to have *the best of both worlds*: the $(r-2)$ independent p-biased strings ensure that the completeness and soundness parameters have the desired dependence on the parameters p and r, where as the uniformity of y and z ensures that the analysis can exploit the linearity of Fourier characters in the necessary steps and recover the quadratic term required to show correlation with a low degree term. We conclude our discussion by giving a brief sketch of the completeness and soundness analysis of the above test. As in the 3-query test, it is straightforward to establish that a dictator function will again pass the test with probability $1-\eta$ and induce an assignment of weight $1-2p$. On the other hand, analyzing the soundness direction is requires additional care. In particular, note that since the marginal distributions of the first $(r-2)$-distribution are p-biased, where as the remaining two strings are uniformly distributed, we have to expand the arithmetization in a hybrid bases consisting of p-biased Fourier expansion for the first $(r-2)$-strings and the unbiased Fourier expansion for the others. This approach introduces additional complications since the p-biased Fourier characters are non-linear operators over $\mathbb{F}_2$ and hence the arguments from the $p = 1/2$ setting don't apply as is. Nevertheless, using the properties of the Fourier characters we can still establish the following analogue of (1).

$$\Pr[\text{ Test Accepts }] \le \frac{1}{2} + \frac{1}{2}(1-2p)^{r-2} + 2^r \cdot \sum_{\beta \neq \mathbf{0}_k} \sum_{\alpha \subseteq \beta} \widehat{f}_p(\alpha) \widehat{f}(\beta)^2 (1-\eta)^{|\beta|}$$

where $\{\widehat{f}_p(\alpha)\}_\alpha$ and $\{\widehat{f}(\beta)\}_\beta$ are the Fourier coefficients of f with respect to the p-biased and unbiased Fourier expansion respectively. Establishing the above involves the most of the work in the soundness analysis and requires a careful application of properties of the p-biased Fourier characters. Note that the above immediately implies the soundness guarantee of the test: if the test passes with probability bounded away from $\frac{1}{2}(1 + (1-2p)^{r-2})$, the summation term of the above equation is strictly positive, which with some additional work can be used to show that f has a low degree term of significant magnitude, thus implying non-trivial correlation with a non-trivial low degree Fourier character.

4 $\frac{1}{2}(1 - e^{-2})$-approximation for all p

As observed earlier in Section 2, the key to doing better than the random guessing algorithm lies in carefully identifying the zero set of the assignment – in particular, we use the fact that the zero set of the optimal assignment must hit a large fraction of hyperedges. Formally we observe the following. Let $\psi([n], E)$ be a $\textsf{Hom-}r\textsf{-Lin}_p$ instance whose optimal value is α. Then we claim that there exists a set of size $(1 - p)n$ which hits at least α-fraction of hyperedges (constraints) in ψ. This is due to the observation that since r is odd, the all-ones string is not a satisfying assignment to the $\textsf{Hom-}r\textsf{-Lin}_p$ predicate, and hence if an assignment satisfies a constraint $e \in E$, then at least one of the variables in e must be set to 0. Furthermore, the problem of hitting the largest number of hyperedges with a cardinality constraint is a *coverage type* problem, which readily admits a linear programming based $(1 - 1/e)$-approximation algorithm.

The above observations immediately suggests the following approach which leads to a constant factor but sub-optimal approximation guarantee. (i) Find the set S of size $(1 - p)n$ which is a $(1-1/e)$-approximation to the coverage problem. (ii) Set all variables in $V \setminus S$ to 1 and all variables variables in S uniformly. This results in a $(1-(1-p)/2)$-weight assignment which satisfies $\frac{1}{2}(1-1/e) \cdot \textsf{Opt}(\psi)$-fraction of constraints[7]. Our actual algorithm uses a slight modification of the above approach, and is based on the observation that simply using a single hitting set of size $(1 - p)n$ to round off the final assignment is wasteful, since that only sets $(1 - p)/2$-fraction of variables to 0, where as the final solution allows for $(1 - p)$-fraction of variables to be set to 0. Instead, our algorithm actually first independently rounds off two hitting sets of size $(1 - p)n$ (from the same fractional solution), and then uses the union of these two sets to construct an assignment of weight p. We outline the steps of our algorithm below.

- Solve the following linear programming relaxation for finding a set which hits maximum fraction of hyperedges in ψ.

$$\begin{aligned}
\text{Maximize } & \sum_{e\in E} x_e & \\
\text{Subject to } & \sum_{i\in e} z_i \geq x_e & \forall\ e \in E \\
& \sum_{i\in V} z_i \leq 2(1-p)n & \\
& 0 \leq x_e, z_i \leq 1 & \forall e \in E, i \in V.
\end{aligned}$$

- Independently round off two sets S_1, S_2 of size $\approx (1-p)n$ by independent rounding using $\{z_i\}_{i\in V}$.
- Set all variables in $[n] \setminus (S_1 \cup S_2)$ to 1 and every variable in $S_1 \cup S_2$ is set to $\{0, 1\}$ uniformly.

We give a brief sketch of the correctness of the above algorithm. Firstly, using a slight modification of the standard analysis of LP rounding for coverage type problems (for e.g., see Section 16.3 [31]), we can show that $S_1 \cup S_2$ will hit at least $(1 - e^{-2}) \cdot \textsf{Opt}(\psi)$-fraction of hyperedges – the negative exponent of e in the approximation factor is 2 instead of 1 due to the fact that we are using the union of two independently rounded sets to hit hyperedges.

[7] Here Opt(ψ) denotes the optimal number of constraints that can be satisfied in ψ.

Furthermore, we observe that any constraint hit by $S_1 \cup S_2$ is going to be satisfied with probability $1/2$ under the rounding, and hence the expected fraction of constraints satisfied by the assignment returned is $\frac{1}{2}(1-e^{-2}) \cdot \mathsf{Opt}(\psi)$. Finally, observe that using the constraint $\sum_i z_i = (1-p)n$, we have $\mathsf{E}[|S_1|] = \mathsf{E}[|S_2|] = (1-p)n$ and hence using Chernoff bound, we can argue that w.h.p. we have $|S_1 \cup S_2| \le 2(1-p)n$, and the weight of the assignment returned by the algorithm will be at least $p(1-o(1))n$ with high probability..

5 $\frac{1}{2}(1-e^{-2})$-hardness assuming UGC

As is standard, our UGC based matching hardness from Theorem 3 is again a dictatorship test based reduction. However, unlike the reduction for Theorem 2, here our test isn't a generalization of Håstad's 3-query test and instead uses the existence of pairwise independent distributions with certain biases that are supported on the set of accepting strings of the Hom-r-Lin predicate. Formally, we show the following.

▶ **Lemma 6** (Informal). *For every large odd $r \in \mathbb{N}$, there exists a pairwise independent permutation invariant distribution supported on the set of accepting strings of* Hom-r-Lin *predictate, such that marginally each bit is $(1-1/(r-1))$-biased.*

As in [5], the above distribution can be immediately used to construct a dictatorship test, as described in Figure 3.

Input. Long code $f : \{0,1\}^k \to \{\pm 1\}$ satisfying $\mathsf{E}_{x\sim\{0,1\}^k_p}[f(x)] = 1-2p$ where $p := 1-2/r$.
Test Let μ be the distribution from Lemma 6.
1. Sample row vectors $x^{(1)}, \dots, x^{(k)} \sim \mu$ independently.
2. Sample $(1-\eta)$-correlated copies $x_i' \underset{1-\eta}{\sim} x_i$ for every $i \in [r]$.
3. Accept if and only if

$$\prod_{i\in[r]} f(x_i') = 1.$$

Figure 3 UGC Test.

We now briefly summarize the completeness and soundness guarantees of the test. For the completeness direction, observe that if $f = \chi_\ell$ is a dictator function, then $\mathsf{E}_{x\sim\{0,1\}^k_p}[f(x)] = 1-2p$ i.e., f is feasible for the test. Furthermore, with probability at least $1-r\eta$ we have

$$\prod_{i\in[r]} f(x_i') = \prod_{i\in[r]} \chi_\ell(x_i') = \prod_{i\in[r]} x_i(\ell) = 1,$$

where in the last step we used the fact that $x^{(\ell)} = (x_i(\ell))_{i\in[k]}$ is always an accepting string for Hom-r-Lin due to our choice of μ from Lemma 6. On the other hand, for the soundness direction, suppose $f : \{0,1\}^k \to \{\pm 1\}$ is a long code satisfying the weight constraint having no influential coordinates[8]. Then as a first step, we again proceed by arithmetizing the probability of the test accepting as

[8] In the context of this dictatorship test, we use a function having no influential coordinates as the notion of being non-correlated with low-degree terms – this is quite standard in Unique Games based reductions [24].

$$\Pr[\text{ Test Accepts }] = \frac{1}{2} + \frac{1}{2}\mathsf{E}_{x\sim\mu^{\otimes k}}\left[\prod_{i\in[r]} f(x_i')\right].$$

Now note that the distribution of $(x_i)_{i\in[r]}$ is pairwise independent and hence its covariance structure matches that of the fully independent p-biased distribution $\{0,1\}_p^{k\otimes r}$ i.e, the distribution where $x_1,\ldots,x_r$ are independent p-biased k-length strings. This along with the fact that f has no influential coordinates allows us to pass on from μ to be the fully independent p-biased distribution using the Invariance principle [27] i.e.,

$$\frac{1}{2} + \frac{1}{2}\mathsf{E}_{x\sim\mu^{\otimes k}}\left[\prod_{i\in[r]} f(x_i')\right] \approx \frac{1}{2} + \frac{1}{2}\mathsf{E}_{x\sim(\{0,1\}_p^r)^{\otimes k}}\left[\prod_{i\in[r]} f(x_i')\right].$$

Finally, using the fact that the expected average weight of f is $1-2p$ and using the independence of x_i''s in the new distribution we have

$$\frac{1}{2} + \frac{1}{2}\mathsf{E}_{x\sim(\{0,1\}_p^r)^{\otimes k}}\left[\prod_{i\in[r]} f(x_i')\right] = \frac{1}{2} + \frac{1}{2}\left(1 - 2\left(1-2/r\right)^r\right) \approx \frac{1}{2}(1-e^{-2}),$$

which gives us the desired soundness. Our final reduction composes the above test with UniqueGames as the outer verifier. While the completeness of the reduction follows as is, additional care has to taken in establishing the soundness of the full reduction owing to the following issue. In the setting of the full reduction from UniqueGames instance $\mathcal{G}(V_{\mathcal{G}}, E_{\mathcal{G}}, [k], \{\pi_e\}_{e\in E})$, the set of variables is defined by a collection of long codes $\{f_v\}_{v\in V_{\mathcal{G}}}$ satisfying the global bias constraint $\mathsf{E}_v\mathsf{E}_x f_v(x) = 1-2p$. Now, by combining standard techniques for analyzing UniqueGames based reductions with the soundness analysis of the test from above, we can show that for NO instances, the soundness of the reduction can be expressed as:

$$\frac{1}{2} + \frac{1}{2}\mathsf{E}_{v\sim V}\left[(1-2p_v)^r\right]$$

where $p_v = \mathsf{E}_{w\sim N_G(v)}\mathsf{E}_{x\sim\{0,1\}_p^k}\left[f_v(x)\right]$ is the local average weight of long codes around v. As before, if we could show that $p_v \approx p$ for most choices of $v \in V_{\mathcal{G}}$, then we would be done. However, unlike in the setting of Theorem 2, the outer UniqueGames instance cannot have strong mixing properties[9], and hence here we cannot hope to show that $p_v \approx v$ for most choices of v. Instead, by combining the fact that r is odd with a careful argument we can show that the mapping $p_v \mapsto (1-2p_v)^{r-2}$ is *concave* for most points of the distribution over p_v, and hence using Jensen's inequality, we can push the expectation operator inside to show that

$$\frac{1}{2} + \frac{1}{2}\mathsf{E}_{v\sim V}\left[(1-2p_v)^r\right] \le \frac{1}{2} + \frac{1}{2}\left(1 - 2\mathsf{E}_{v\sim V_{\mathcal{G}}}\left[p_v\right]\right)^r + o_r(1) = \frac{1}{2}(1-e^{-2}) + o_r(1),$$

where the $o_r(1)$-additive factor is due to the fact that the function is concave in all but $o_r(1)$-mass of the distribution.

[9] In fact, even mildly expanding UniqueGames instances are known to admit polynomial time algorithms [6]

References

1 Sepehr Abbasi-Zadeh, Nikhil Bansal, Guru Guruganesh, Aleksandar Nikolov, Roy Schwartz, and Mohit Singh. Sticky brownian rounding and its applications to constraint satisfaction problems. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 854–873. SIAM, 2020.

2 Per Austrin, Siavosh Benabbas, and Konstantinos Georgiou. Better balance by being biased: A 0.8776-approximation for max bisection. *ACM Transactions on Algorithms (TALG)*, 13(1):1–27, 2016.

3 Per Austrin and Johan Håstad. Randomly supported independence and resistance. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 483–492. ACM, 2009.

4 Per Austrin and Subhash Khot. A characterization of approximation resistance for even k-partite csps. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 187–196, 2013.

5 Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009.

6 Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *2011 ieee 52nd annual symposium on foundations of computer science*, pages 472–481. IEEE, 2011.

7 Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, pcps, and nonapproximability—towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998.

8 Amey Bhangale and Subhash Khot. Optimal inapproximability of satisfiable k-lin over non-abelian groups. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1615–1628, 2021.

9 Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an o (n 1/4) approximation for densest k-subgraph. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 201–210, 2010.

10 Siu On Chan. Approximation resistance from pairwise-independent subgroups. *Journal of the ACM (JACM)*, 63(3):1–32, 2016.

11 Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 205–214. ACM, 2006.

12 Eden Chlamtac, Konstantin Makarychev, and Yury Makarychev. How to play unique games using embeddings. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 687–696. IEEE, 2006.

13 Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct sum testing. *SIAM Journal on Computing*, 46(4):1336–1369, 2017.

14 Lars Engebretsen, Jonas Holmerin, and Alexander Russell. Inapproximability results for equations over finite groups. *Theoretical Computer Science*, 312(1):17–45, 2004.

15 U Feige and M Seltser. On the densest k-subgraph problems, 1997.

16 Suprovat Ghoshal and Euiwoong Lee. A characterization of approximability for biased csps. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 989–997. ACM, 2022. doi:10.1145/3519935.3520072.

17 Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with PSD objectives. In Rafail Ostrovsky, editor, *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 482–491. IEEE Computer Society, 2011.

18 Gustav Hast. *Beating a random assignment: Approximating constraint satisfaction problems*. PhD thesis, KTH, 2005.

19 Johan Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001.

20 Johan Håstad. Every 2-csp allows nontrivial approximation. *Comput. Complex.*, 17(4):549–566, 2008. doi:10.1007/s00037-008-0256-y.

21 Johan Håstad and Rajsekar Manokaran. On the hardness of approximating balanced homogenous 3-lin. *Theory Of Computing*, 13(1):1–19, 2017.

22 Jonas Holmerin and Subhash Khot. A new pcp outer verifier with applications to homogeneous linear equations and max-bisection. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 11–20, 2004.

23 Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 767–775, 2002.

24 Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM Journal on Computing*, 37(1):319–357, 2007.

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

26 Rasmus Kyng and Peng Zhang. Hardness results for structured linear systems. *SIAM J. Comput.*, 49(4), 2020.

27 Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010.

28 Prasad Raghavendra and Tselil Schramm. Gap amplification for small-set expansion via random walks. *arXiv preprint*, 2013. arXiv:1310.1493.

29 Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 755–764. ACM, 2010.

30 Prasad Raghavendra, David Steurer, and Prasad Tetali. Approximations for the isoperimetric and spectral profile of graphs and related parameters. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 631–640. ACM, 2010.

31 Vijay V Vazirani. *Approximation algorithms*, volume 1. Springer, 2001.

32 Uri Zwick. Computer assisted proof of optimal approximability results. In *SODA*, pages 496–505, 2002.