Hardness of Bipartite Expansion

Subhash Khot[1] and Rishi Saket[2]

1 **Department of Computer Science, New York University, NY, USA**
`khot@cs.nyu.edu`
2 **IBM Research, Bangalore, India**
`rissaket@in.ibm.com`

Abstract

We study the natural problem of estimating the expansion of subsets of vertices on one side of a bipartite graph. More precisely, given a bipartite graph $G(U, V, E)$ and a parameter β, the goal is to find a subset $V' \subseteq V$ containing β fraction of the vertices of V which minimizes the size of $N(V')$, the neighborhood of V'. This problem, which we call *Bipartite Expansion*, is a special case of submodular minimization subject to a cardinality constraint, and is also related to other problems in graph partitioning and expansion. Previous to this work, there was no hardness of approximation known for Bipartite Expansion.

In this paper we show the following strong inapproximability for Bipartite Expansion: for any constants $\tau, \gamma > 0$ there is no algorithm which, given a constant $\beta > 0$ and a bipartite graph $G(U, V, E)$, runs in polynomial time and decides whether

- (YES case) There is a subset $S^* \subseteq V$ s.t. $|S^*| \geq \beta|V|$ satisfying $|N(S^*)| \leq \gamma|U|$, or
- (NO case) Any subset $S \subseteq V$ s.t. $|S| \geq \tau\beta|V|$ satisfies $|N(S)| \geq (1-\gamma)|U|$,

unless NP $\subseteq \cap_{\varepsilon>0}$DTIME $\left(2^{n^{\varepsilon}}\right)$ i.e. NP has subexponential time algorithms.

We note that our hardness result stated above is a vertex expansion analogue of the Small Set (Edge) Expansion Conjecture of Raghavendra and Steurer [23].

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases inapproximability, bipartite expansion, PCP, submodular minimization

Digital Object Identifier 10.4230/LIPIcs.ESA.2016.55

1 Introduction

Graph partitioning and graph expansion are very well studied topics in graph theory, combinatorics and theoretical computer science. A central goal in this line of research is to decide how well a given graph can be partitioned into smaller parts. Generally speaking, a partitioning is considered good if the graph is decomposed into reasonably sized components while removing only a small number of vertices or edges. Specific variants of the graph partitioning question are addressed by a number of well known problems – such as Vertex Separator, Sparsest Cut and Balanced Separator – which have been studied extensively in several previous works [18, 16, 17, 5, 15].

Related to the above is the *Bipartite Expansion* problem which measures the vertex expansion of subsets on one of the sides of a bipartite graph. Specifically, given a bipartite graph $G(U, V, E)$, the goal is to find a subset $V' \subseteq V$ of size at least $\beta|V|$ to minimize $|N(V')|$ for some parameter $\beta \in (0, 1)$, where $N(V')$ is the neighborhood of V' in U. The absence of V' with a small neighborhood implies that there is an edge between any two large enough subsets – one each of U and V, presenting a bottleneck to a good partitioning of the graph. Such graphs are known as *bipartite expanders*. They have been studied in several applications such as parallel sorting [22, 2], constructing good codes [26], randomness extractors [12, 25]

and more recently for constructing secure public key cryptographic systems [4]. Since vertex expansion is a submodular function, Bipartite Expansion is a special case of minimizing a submodular function subject to a cardinality constraint, for which Svitkina and Fleischer [27] gave an $O\left(\sqrt{n/\ln n}\right)$ pseudo-approximation. In spite of this being a natural problem, we are not aware of any hardness of approximation results for Bipartite Expansion.

In this work we establish the following strong hardness of approximation result for Bipartite Expansion under the assumption that NP does not have subexponential time algorithms.

▶ **Theorem 1.1** (Main). *For any constants $\gamma, \tau > 0$, there is no algorithm which, given a constant $\beta > 0$ and an n-vertex bipartite graph $G(U, V, E \subseteq U \times V)$, runs in $O(n^c)$ time where $c = c(\tau, \gamma, \beta)$ and decides between the following cases,*

- (YES case) *There is a subset $S^* \subseteq V$ s.t. $|S^*| \geq \beta|V|$ satisfying $|N(S^*)| \leq \gamma|U|$, or*
- (NO case) *Any subset $S \subseteq V$ s.t. $|S| \geq \tau\beta|V|$ satisfies $|N(S)| \geq (1-\gamma)|U|$,*

unless NP $\subseteq \cap_{\varepsilon>0}$DTIME$\left(2^{n^\varepsilon}\right)$.

More concretely, the above shows (for example) that even if there exists a good subset V^* of at least β fraction of the vertices of V such that its neighborhood is 1% of U, it is hard to find V' of $\beta/100$ fraction of the vertices of V, such that the neighborhood of V' is at most 99% of U.

We note that Bipartite Expansion problem seems to bear resemblance to the Small Set Expansion problem which has recently received attention due to its connection to Khot's [13] Unique Games Conjecture. This connection was established by Raghavendra and Steurer [23] who proved that the Small Set Expansion Conjecture (see [23] for the statement) implies the Unique Games Conjecture. Theorem 1.1 is, in some sense, a vertex expansion analogue of the statement of the Small Set Expansion Conjecture which deals with edge expansion. Louis, Raghavendra, and Vempala [19] have shown that the Small Set Expansion Conjecture implies hardness of approximation for a variant of vertex expansion on general graphs. While a similar reverse reduction from vertex expansion to edge expansion of small sets is not known, this raises the intriguing possibility that techniques similar to those of this work could throw some light on the Small Set Expansion and Unique Games conjectures.

Our result also serves as a complexity theoretic lower bound for the $O\left(\sqrt{n/\ln n}\right)$ pseudo-approximation of [27] for minimizing submodular functions subject to a cardinality constraint, even when the function is monotone like bipartite vertex expansion.

1.1 Related Work

Bipartite Expansion is related to graph partitioning problems including Vertex Separator and Balanced Separator, and is known to be NP-hard via a reduction from the Balanced Vertex Separator problem [21].

Leighton and Rao [16] gave an $O(\log n)$ (pseudo-)approximation for Balanced Separator and Vertex Separator problems. For Balanced Separator, the factor was improved to the currently best known $O(\sqrt{\log n})$ in a seminal work of Arora, Rao, and Vazirani [5]. Subsequent work by Feige *et al.* [11] and Agarwal *et al.* [1] also proved $O(\sqrt{\log n})$ approximation for Vertex Separator. For both these problems PTAS was ruled out under standard complexity assumptions by Ambuhl, Mastrolilli, and Svensson [3] using the quasi-random Probabilistically Checkable Proof (PCP) of Khot [14]. Subsequently work of Raghavendra, Steurer, and Tulsiani [24] has ruled out a constant factor approximation for Balanced Separator based on the Small Set Expansion Conjecture [23] which implies the Unique Games Conjecture

of Khot [13]. The latter conjecture was used in previous works [15, 7] to prove similar inapproximability for a non-uniform version of Balanced Separator. As mentioned above, the work of Svitkina and Fleischer [27] shows that Bipartite Expansion admits a $O\left(\sqrt{n/\ln n}\right)$ pseudo-approximation: given the existence of a subset V^* of size $\beta|V|$ and $|N(V^*)| \leq \gamma|U|$, the algorithm outputs a subset V' of size at least $\sigma\beta|V|$ and $|N(V')| \leq \rho\gamma|U|$, with $\rho/\sigma \leq O\left(\sqrt{n/\ln n}\right)$.

While not much is known about the inapproximability of Bipartite Expansion, the problem of explicitly constructing bipartite expanders has been fairly well studied [20, 22, 2]. These constructions and their variants have applications in sorting networks [22, 2], error-correcting codes [26] and randomness extractors [12, 25]. We end with a brief mention of a result by Applebaum, Barak, and Wigderson [4] who construct public-key encryption schemes based on the (assumed) average case hardness of detecting a random unbalanced bipartite graph from one which has a randomly planted "shrinking" set S of $O(\log n)$ vertices on the larger side such that $|N(S)| \leq |S|/3$. While the parameters they consider are different from our setting, their work has, in part, motivated this study of bipartite expansion.

1.2 Our Techniques

The starting point of our reduction is the quasi-random PCP constructed by Khot [14]. Unlike previously constructed PCPs, Khot's construction essentially showed that the YES and NO cases differ in how randomly the queries of the verifier's tests are distributed over the locations of the proof. This crucial quasi-randomness property – which we describe below – was used by Khot [14] to rule out PTAS for Min-Bisection, Dense k-Subgraph and Bipartite Clique, results which were only known earlier assuming the average case hardness of Random-3SAT [10].

The construction in [14] proceeded by (i) proving the inapproximability of an Homogeneous Algebraic CSP over a large field, (ii) transforming the latter into an *Outer Verifier* based on an algebraic test, and (iii) composing the Outer Verifier with a d-query *Inner Verifier* based on the Hadamard Encoding over $\mathbb{F}[2]$. The quasi-randomness finally obtained by this series of reductions can be roughly summarized as follows: in the YES case there is a subset of half the locations of the proof which contains all the d queries of $\approx 1/2^{d-1}$ fraction of tests of the Inner Verifier, while in the NO case any such subset of the proof locations completely contains $\approx 1/2^d$ fraction of the tests. Taking the locations of the proof on the LHS, the tests of the verifier on the RHS, and connecting a location with a test if it queries the former, this already gives us an instance of Bipartite Expansion with a small hardness factor.

However, the inapproximability obtained above is far too weak for us to amplify the hardness gap using (say) graph powering. For this we modify the construction of [14] to encode the proof of the Inner Verifier using a Hadamard Code over a larger field $\mathbb{F}[q]$ where $q \gg 2$. A similar abstraction of the Inner Verifier as a bipartite graph $G(U, V, E)$ yields a $\eta\delta$ versus η gap in the expansion of similar sized subsets of V, for arbitrarily small $\eta, \delta > 0$. Taking the bipartite kth graph power using OR-product of the edges, where $k \approx C/\eta$, amplifies the above to a $C\delta$ versus $(1 - \exp(-C))$ gap in the expansion. The modified quasi-random PCP also yields a gap in the sizes of the relevant subsets of V which is preserved by the powering operation. This, along with the expansion gap, is sufficient to prove Theorem 1.1.

The construction and the analysis of the modified quasi-random PCP proceed along the same lines as in [14]. The parameters of the construction are set appropriately so that the subsequent OR-product graph powering amplifies the gap as desired.

Organization of the paper. The next section formally defines Bipartite Expansion as a decision problem, and restates our hardness result as Theorem 2.2. Section 3 provides the description of the Homogeneous Algebraic CSP and the quasi-random PCP of Khot [14], along with the statement (Theorem 3.4) of the aforementioned modified quasi-random PCP. Section 4 is devoted to proving Theorem 2.2 starting from Theorem 3.4 and the construction of the modified quasi-random PCP, i.e. the proof of Theorem 3.4, is given in Section 5.

2 Our Results

Bipartite Expansion is defined as the following decision problem.

▶ **Definition 2.1.** For parameters $\tau, \gamma, \beta > 0$, the BipartiteExpansion (τ, γ, β) problem is: given a bipartite graph $G(U, V, E \subseteq U \times V)$ distinguish between the following cases.

- (YES Case) There is a subset $S^* \subseteq V$, $|S^*| \geq \beta|V|$ s.t. $|N(S^*)| \leq \gamma|U|$.
- (NO Case) For any subset $S \subseteq V$ s.t. $|S| \geq \tau\beta|V|$, $|N(S)| \geq (1-\gamma)|U|$.

We prove the following hardness of BipartiteExpansion which implies Theorem 1.1.

▶ **Theorem 2.2.** *For any choice of constants $\varepsilon, \tau, \gamma > 0$, there exists $\beta > 0$, such that there is a* DTIME (2^{n^ε}) *reduction from* SAT *to* BipartiteExpansion (τ, γ, β).

3 Preliminaries

3.1 The HomAlgCSP Problem

▶ **Definition 3.1.** Let an HomAlgCSP instance $\mathcal{A}(k, d, m, \mathbb{F}, \mathcal{C})$ be the following problem:

1. $\mathcal{C}$ is a system of constraints on functions $f : \mathbb{F}^m \mapsto \mathbb{F}$ where every constraint is on values of f on k different points and is given by a conjunction of homogeneous linear constraints on those k values. A constraint $C \in \mathcal{C}$ on $f(\overline{p}_1), \ldots, f(\overline{p}_k)$ is is given as

$$\sum_{i=1}^{k} \gamma_{ij} f(\overline{p}_i) = 0 \quad \text{for } j = 1, 2, \ldots \qquad \text{where} \overline{p}_i \in \mathbb{F}^m \text{ and } \gamma_{ij} \in \mathbb{F}.$$

   We denote a constraint C by the set of points $\{\overline{p}_i\}_{i=1}^k$, while the γ_{ij}'s will be implicit.
2. $\mathcal{C}$ has $|\mathbb{F}|^{O(m)}$ constraints.

The goal is to find a m-variate polynomial f of total degree at most d, not identically zero, so as to maximize the fraction of constraints satisfied.

The following inapproximability of HomAlgCSP was shown in [14] following from Theorems 1.5 and 3.4 proved therein.

▶ **Theorem 3.2.** *There is a universal constant $\Delta > 0$, such that for any constant $K > 0$ and any constant $\tilde{d} > 0$ possibly depending on K, there is a reduction from a* SAT *formula of size n to an instance $\mathcal{A}(k, d^* = 10\tilde{d}, m = O(\tilde{m}^3\tilde{d}), \mathbb{F}, \mathcal{C})$ of* HomAlgCSP *with $k = 21$, $N \leq |\mathbb{F}| \leq N^2$ where $N = n^{\Delta K}$, and any choice of $\tilde{m}$ satisfying $\binom{\tilde{m}}{\tilde{d}} \geq N$. The size of the instance $\mathcal{A}$ is $|\mathbb{F}|^{O(m)}$, where the field $\mathbb{F}$ is any suitably sized extension of $\mathbb{F}[2]$. The reduction is a* DTIME $(|\mathbb{F}|^{O(m)})$ *procedure*[1] *such that,*

[1] The work of Khot [14] was based on a randomized hardness reduction for *Minimum Distance of Codeword* [9], which can be made deterministic using subsequent results of Cheng and Wan [8] and Austrin and Khot [6].

1. (YES Case) *If the* SAT *formula is satisfiable then there is a degree d^* multivariate polynomial f, not identically zero, which satisfies $1 - 1/2^K$ fraction of constraints of $\mathcal{A}$.*
2. (NO Case) *If the* SAT *formula is unsatisfiable then no degree $1000d^*$ multivariate polynomial, which is not identically zero, satisfies more than $1/2^{2^K}$ fraction of constraints of $\mathcal{A}$.*

3.2 Quasi-Random PCP of Khot [14]

The following is the statement of Khot's quasi-random PCP.

▶ **Theorem 3.3** (Theorem 1.9 of [14]). *For every $\varepsilon > 0$, there exists an integer $d = O(1/\varepsilon \log(1/\varepsilon))$ such that the following holds : there is a PCP verifier for a* SAT *instance of size n satisfying:*

1. *The proof for the verifier is of size $2^{O(n^\varepsilon)}$.*
2. *The verifier uses $O(n^\varepsilon)$ random bits, runs in time $2^{O(n^\varepsilon)}$, and reads d locations from the proof. Let Q be the d locations queried by the verifier in a random test.*
3. *Every query location is uniformly distributed over the proof, though different query locations within Q are correlated.*
4. (YES Case) *Suppose that the* SAT *instance is satisfiable. Then there exists a subset Π^* of half the locations of the proof such that,*

$$\Pr_Q [Q \subseteq \Pi^*] \geq \frac{1}{2^{d-1}} \left(1 - O\left(\frac{1}{d}\right)\right),$$

where the probability is taken over a random test of the verifier.

5. (NO Case) *Suppose that the* SAT *instance is unsatisfiable, and let Π' be any set of half the locations in the proof. Then,*

$$\left|\Pr_Q [Q \subseteq \Pi'] - \frac{1}{2^d}\right| \leq \frac{1}{2^{20d}}.$$

3.3 Modified Quasi-Random PCP

As discussed in Section 1.2, for our hardness result we construct the quasi-random PCP with an Inner Verifier encoding over a large field $\mathbb{F}[q]$. While the details of the construction and its analysis are given in Section 5, here we abstract out the bounds on the distribution of the PCP queries required for our purposes.

▶ **Theorem 3.4.** *For every positive integer (power of two) $R > 2$, and arbitrarily small $\varepsilon > 0$, there exists an integer $d = \Theta((1/\varepsilon)\log((\log R)/\varepsilon))$ along with the setting $q := R^{4d}$, such that the following holds : there is a PCP verifier for a* SAT *instance of size n satisfying properties (1)-(3) of Theorem 3.3 along with,*

4. (YES Case) *Suppose that the* SAT *instance is satisfiable. Then there exists a subset Π^* of $1/q$ fraction of the locations of the proof, such that*

$$\Pr_Q [Q \subseteq \Pi^*] \geq \frac{1}{q^{d-1}} \left(1 - O\left(\frac{1}{d^2}\right)\right), \tag{1}$$

where the probability is taken over a random test of the verifier.

5. (NO Case) *Suppose that the* SAT *instance is unsatisfiable, and let Π' be any set of $\zeta \in [0,1]$ fraction of the locations of the proof. Then,*

$$\left|\Pr_Q [Q \subseteq \Pi'] - \zeta^d\right| \leq \frac{1}{q^{2d^2}}. \tag{2}$$

4 Reducing the Modified Quasi-Random PCP to BipartiteExpansion

For convenience we first abstract out the Modified Quasi-Random PCP as a bipartite graph and translate its YES and NO cases into a gap in expansion of small subsets on one side of the bipartition. This gap in expansion is then strengthened using an appropriate powering of the initial bipartite graph to yield the desired hardness for BipartiteExpansion. We assume for the rest of this section that that the parameters R, d and q in Theorem 3.4 are large enough constants.

4.1 Modified Quasi-Random PCP as a Bipartite Graph

Starting from an instance of the Modified Quasi-Random PCP in Theorem 3.4 define the bipartite graph $G(U, V, E \subseteq U \times V)$ where U is the set of proof locations, V is the set of d-query tests of the verifier and $(u, v) \in E$ iff the test v contains the query location u. Restating the YES and NO cases in terms of expansion of subsets of V we have the following lemmas.

▶ **Lemma 4.1.** *If G is a* YES *instance then there is a subset $S^* \subseteq V$ of size at least $0.99|V|/q^{d-1}$ such that $|N(S^*)| \leq |U|/q$.*

Proof. Take S^* to be the set of tests completely contained in the $1/q$ fraction of the proof locations given by the YES case. The lemma follows from (1) and large enough d. ◀

▶ **Lemma 4.2.** *If G is a* NO *instance then for any subset $S \subseteq V$ s.t. $|S| = a|V|$ where $a \in [1/q^{2d^2}, 1]$,*

$$\frac{|N(S)|}{|U|} \geq a^{1/d} - \frac{1}{aq^{2d^2}}.$$

Proof. In the NO case we let $\zeta = \frac{|N(S)|}{|U|}$, and thus from (2) we have

$$a \leq \zeta^d + \frac{1}{q^{2d^2}} \quad \Rightarrow \quad \zeta^d \geq a - \frac{1}{q^{2d^2}} \geq 0,$$

since $a \geq 1/q^{2d^2}$. This implies that,

$$\zeta \geq \left(a - \frac{1}{q^{2d^2}}\right)^{\frac{1}{d}} \geq a^{1/d}\left(1 - \frac{1}{aq^{2d^2}}\right)^{\frac{1}{d}} \geq a^{1/d}\left(1 - \frac{1}{aq^{2d^2}}\right) \geq a^{1/d} - \frac{1}{aq^{2d^2}},$$

since $a^{1/d} \leq 1$. ◀

4.2 Graph powering using OR-product

Fix a parameter $k := R^{4d-1}$. From $G(U, V, E)$ obtained above we construct the bipartite graph $\overline{G}(\overline{U}, \overline{V}, \overline{E} \subseteq \overline{U} \times \overline{V})$ as follows.

- $\overline{U} = U^k$ and $\overline{V} = V^k$. For any $\overline{u} \in \overline{U}$, $j \in [k]$, $\overline{u}_j \in U$ denotes the jth coordinate of $\overline{u}$. Similarly for $\overline{v} \in \overline{V}$.
- $(\overline{u}, \overline{v}) \in \overline{E}$ iff $\exists j \in [k]$ s.t. $(\overline{u}_j, \overline{v}_j) \in E$.

The rest of this section is devoted to proving the desired YES and NO cases completing the proof of Theorem 2.2.

4.2.1 YES Case

We prove the following lemma.

▶ **Lemma 4.3.** *If G is a* YES *instance then there exists a subset $T^* \subseteq \overline{V}$ such that,*

$$|T^*| \geq \left(\frac{0.99}{q^{d-1}}\right)^k |\overline{V}|,$$

and,

$$|N(T^*)| \leq |\overline{U}|/R.$$

Proof. Let $T^* = (S^*)^k$ where S^* is as given in Lemma 4.1. The first condition above is directly satisfied by the bound on the size of S^* in Lemma 4.1. Further, by union bound over all the k coordinates,

$$\frac{|N(T^*)|}{|\overline{U}|} \leq k \cdot \frac{|N(S^*)|}{|U|} \leq \frac{k}{q} = \frac{R^{4d-1}}{R^{4d}} = R^{-1},$$

where $|N(S^*)|/|U| \leq 1/q$ as given in Lemma 4.1. ◀

4.2.2 NO Case

For convenience let $h := 1/q^{d-1/2}$. The NO case is given by the following lemma.

▶ **Lemma 4.4.** *If G is a* NO *instance then for any subset $T \subseteq \overline{V}$ s.t. $|T| \geq h^k|\overline{V}|$,*

$$|N(T)| \geq \left(1 - e^{-R/2}\right)|\overline{U}|.$$

Proof. Let us first define the projections $T_1, \ldots, T_k \subseteq V$ of T as: $T_j = \{v \in V \mid \exists \overline{v} \in T \text{ s.t. } \overline{v}_j = v\}$. By construction, $|T| \leq \prod_{j=1}^k |T_j|$. Let $a_j := |T_j|/|V|$. Thus, we have,

$$\prod_{j=1}^k (a_j|V|) \geq h^k|\overline{V}| = h^k|V|^k,$$

which implies

$$\prod_{j=1}^k a_j \geq h^k. \tag{3}$$

By the AM-GM inequality we have,

$$\mathbb{E}_{j\in[k]}\left[a_j^{1/d}\right] \geq \left(\prod_{j=1}^k a_j^{1/d}\right)^{1/k} = \left(\prod_{j=1}^k a_j\right)^{1/kd} \geq h^{1/d}. \tag{4}$$

We also have the following simple lemma.

▶ **Lemma 4.5.** *For at most k/d values $j \in \{1, \ldots, k\}$, $a_j < h^d$.*

Proof. Assuming that for $t > k/d$ values $j \in \{1, \ldots, k\}$ $a_j < h^d$, we obtain that $\prod_{j=1}^k a_j \leq h^{td} < h^k$ (since $h < 1$), which contradicts (3). ◀

Let us define $b_j := |N(T_j)|/|U|$ for $j \in [k]$. Since, by our setting, $h^d \geq q^{-d^2} \geq q^{-2d^2}$, Lemma 4.2 yields

$$\left\{a_j \geq h^d\right\} \Rightarrow \left\{b_j \geq a_j^{1/d} - \frac{1}{a_j q^{2d^2}}\right\} \Rightarrow \left\{b_j \geq a_j^{1/d} - \frac{1}{q^{d^2}}\right\}. \tag{5}$$

Therefore,

$$\begin{aligned}
\sum_{j=1}^{k} b_j \geq \sum_{\substack{j\in[k]\\ a_j \geq h^d}} \left(a_j^{1/d} - \frac{1}{q^{d^2}}\right) &= \sum_{j=1}^{k}\left(a_j^{1/d} - \frac{1}{q^{d^2}}\right) - \sum_{\substack{j\in[k]\\ a_j < h^d}} \left(a_j^{1/d} - \frac{1}{q^{d^2}}\right) \\
&\geq \sum_{j=1}^{k}\left(a_j^{1/d} - \frac{1}{q^{d^2}}\right) - \sum_{\substack{j\in[k]\\ a_j < h^d}} a_j^{1/d} \\
&\geq \sum_{j=1}^{k}\left(a_j^{1/d} - \frac{1}{q^{d^2}}\right) - \sum_{\substack{j\in[k]\\ a_j < h^d}} h \\
&\geq \sum_{j=1}^{k}\left(a_j^{1/d} - \frac{1}{q^{d^2}}\right) - \left(\frac{k}{d}\right) h,
\end{aligned}$$

where the last inequality uses Lemma 4.5. Taking an expectation we obtain,

$$\mathbb{E}_{j\in[k]}\left[b_j\right] \;\geq\; \mathbb{E}_{j\in[k]}\left[a_j^{1/d}\right] - \frac{1}{q^{d^2}} - \frac{h}{d} \;\geq\; h^{1/d} - \frac{1}{q^{d^2}} - \frac{h}{d} \;\geq\; h^{1/d}/2, \tag{6}$$

where second last inequality follows from (4) and the last inequality is due to the large enough setting of the parameters. Observe that,

$$h^{1/d} = \left(q^{-(d-1/2)}\right)^{1/d} = \left(R^{-4d(d-1/2)}\right)^{1/d} = R^{-4d+2}.$$

Using the above along with the construction of $\overline{G}$ we have,

$$\begin{aligned}
1 - \frac{|N(T)|}{|\overline{U}|} &= \prod_{j=1}^{k}(1-b_j) \\
&\leq \left(\mathbb{E}_{j\in[k]}[1-b_j]\right)^k && \text{(By the AM-GM inequality)} \\
&\leq \left(1 - h^{1/d}/2\right)^k && \text{(Using (6))} \\
&\leq \left(1 - \frac{1}{2R^{4d-2}}\right)^{R^{4d-1}} && \leq\; e^{-R/2},
\end{aligned}$$

which completes the proof of Lemma 4.4. ◀

4.2.3 Gap in the domain subset sizes

In the YES case there is a subset of $\overline{V}$ of fractional size at least

$$\beta := \left(\frac{0.99}{q^{d-1}}\right)^k \tag{7}$$

with neighborhood size at most $R^{-1}|\overline{U}|$, while in the NO case every subset of $\overline{V}$ of fractional size at least

$$h^k = \left(\frac{1}{q^{d-1/2}}\right)^k \tag{8}$$

has neighborhood of size at least $(1 - e^{-R/2})|\overline{U}|$. The subset size threshold in the NO case is much smaller than in the YES case with the gap being,

$$\left(\frac{\beta}{h^k}\right) = \left(\frac{0.99q^{d-1/2}}{q^{d-1}}\right)^k \geq q^{k/3} \geq e^R,$$

for large enough R, q, d which we may assume.

4.2.4 Setting the parameters and proof of Theorem 2.2

Given ε, τ and $\gamma > 0$, choose R large enough so that $\gamma \geq \max\{R^{-1}, e^{-R/2}\}$, and $\tau \geq e^{-R}$. Setting $d = \Theta((1/\varepsilon)\log((\log R)/\varepsilon))$ as per Theorem 3.4 along with Lemmas 4.3, Lemma 4.4, and Section 4.2.3 yields the proof of Theorem 2.2 with β given by (7).

5 Construction of the Quasi-Random PCP

The PCP given in Theorem 3.4 is a composition of an Outer Verifier which is an algebraic test on an instance of HomAlgCSP, with a Hadamard code based encoding (Inner Verifier). This is almost the same as the construction of [14], except that the Inner Verifier's encoding is over a larger field rather than $\mathbb{F}[2]$. We refer the reader to [14] for motivation behind this construction and its nuances, and instead give a concise description of the PCP and its analysis.

Let the HomAlgCSP instance be $\mathcal{A}(k = 21, d^*, m, \mathbb{F}, \mathcal{C})$. The Outer Verifier is given the polynomial f as a table of values at each point in $\mathbb{F}^m$, and it samples a constraint from $\mathcal{C}$ uniformly at random and attempts to verify whether it is satisfied by f, and whether the table of f is a polynomial of degree $\approx d^*$. We need the following definition of a *curve*.

▶ **Definition 5.1.** A *curve* L in $\mathbb{F}^m$ is a function $L : \mathbb{F} \mapsto \mathbb{F}^m$, where $L(t) = (a_1(t), \ldots, a_m(t))$. It is of degree d if each of the coordinate functions a_i is degree d (univariate) polynomial. A *line* is a curve of degree 1.

Let $t_1, t_2, \ldots, t_{k+3}$ be distinct field elements in $\mathbb{F}$ which we fix for the rest of the construction. Suppose the verifier chooses the constraint $C(\{\overline{p}_i\}_{i=1}^k) \in \mathcal{C}$ uniformly at random. For $\overline{a}, \overline{b}, \overline{c} \in \mathbb{F}^m$, define $L = L_{\overline{a},\overline{b},\overline{c}}$ be the unique degree $(k+2)$ curve that satisfies

$$L(t_i) = \overline{p}_i, \quad 1 \leq i \leq k, \quad L(t_{k+1}) = \overline{a}, \quad L(t_{k+2}) = \overline{b}, \quad L(t_{k+3}) = \overline{c}.$$

If f is a degree d^* multivariate polynomial over the vector space $\mathbb{F}^m$ then its restriction to the curve $L(t) = L_{\overline{a},\overline{b},\overline{c}}(t)$, denoted by $f|_L$, is a degree $d - 1 := (k+2)d^*$ univariate polynomial in t. This polynomial can be interpolated from any d values of f on the curve, which is then used to test its consistency at an additional random point. Similarly, given a line ℓ, the restriction of f, denoted by $f|_\ell$ is a degree d^* univariate polynomial. Allowing it degree up to $(d-2)$ it is interpolated using the values of f at $(d-1)$ random points on ℓ, which is used to run the Low Degree test. The following is the description of the Outer Verifier.

5.1 Outer Verifier

Steps of the Outer Verifier

1. Pick a constraint $C = \{\overline{p}_i\}_{i=1}^{k} \in \mathcal{C}$ at random.
2. Pick a random line ℓ in $\mathbb{F}^m$ and pick random points $\overline{v}_1, \ldots, \overline{v}_{d-1}, \overline{v}_d$ on the line.
3. Pick $t \in \mathbb{F} \setminus \{t_1, \ldots, t_{k+3}\}$ at random, points $\overline{a}, \overline{b}$ at random from $\mathbb{F}^m$ and let L be the unique degree $k+2$ curve $L = L_{\overline{a},\overline{b},\overline{c}}$ such that,

$$L(t_i) = \overline{p}_i, \quad 1 \leq i \leq k, \quad L(t_{k+1}) = \overline{a}, \quad L(t_{k+2}) = \overline{b}, \quad L(t) = \overline{v}_d,$$

   and $\overline{c}$ is implicitly defined as $L(t_{k+3})$.
4. Pick random points $\overline{v}_{d+1}, \ldots, \overline{v}_{2d}$ on L.
5. Let $f|_\ell$ be the unique degree $d-2$ polynomial interpolated using the values $\{f(\overline{v}_i)\}_{i=1}^{d-1}$.
6. Let $f|_L$ be the unique degree $d-1$ polynomial interpolated using the values $\{f(\overline{v}_i)\}_{i=d+1}^{2d}$.
7. Check if,

$$f|_L(\overline{v}_d) = f(\overline{v}_d) = f|_\ell(\overline{v}_d).$$

8. Check if the values of $f|_L$ at points $\{\overline{p}_i\}_{i=1}^{k}$ satisfy the constraint C.
9. Check that the values $f(\overline{v}_i),\ 1 \leq i \leq 2d$ are not all zero.

As in [14], the Outer Verifier can be replaced by the following *Modified Outer Verifier* which reads more values from the proof and makes additional tests, and additionally abstracts out: (i) interpolation into multiplication by an invertible matrix, and (ii) checking the homogeneous constraints of the Outer Verifier into checking orthogonality with a certain subspace. Our construction is the same, except that instead of $\mathbb{F}[2]$ we shall use an extension field $\mathbb{F}[q]$ as the underlying field of representation, where q is as given in Theorem 3.4 and $\mathbb{F}$ in Theorem 3.2 is chosen to be an extension of $\mathbb{F}[q]$.

5.2 Modified Outer Verifier

Since $\mathbb{F}$ is an extension of $\mathbb{F}[q]$ the elements of $\mathbb{F}$ are represented as $\mathbb{F}[q]$-vectors of a length $l = (\log|\mathbb{F}|)/(\log q)$. Moreover, the representation can be chosen such that addition over $\mathbb{F}$ and multiplication by a constant in $\mathbb{F}$ are homogeneous linear operations on these vectors. The Modified Outer Verifier is given a table of values $f(\overline{v})$ (in the form of l length $\mathbb{F}[q]$-vectors) for every point $\overline{v} \in \mathbb{F}^m$ and it executes the following steps:

Steps of the Modified Outer Verifier

1. Pick a constraint $C = \{\overline{p}_i\}_{i=1}^{k} \in \mathcal{C}$ at random.
2. Pick a random line ℓ in $\mathbb{F}^m$ and pick random points $\overline{v}_1, \ldots, \overline{v}_{d-1}, \overline{v}_d$ on the line.
3. Pick $t \in \mathbb{F} \setminus \{t_1, \ldots, t_{k+3}\}$ at random, points $\overline{a}, \overline{b}$ at random from $\mathbb{F}^m$ and let L be the unique degree $k+2$ curve $L = L_{\overline{a},\overline{b},\overline{c}}$ such that,

$$L(t_i) = \overline{p}_i, \quad 1 \leq i \leq k, \quad L(t_{k+1}) = \overline{a}, \quad L(t_{k+2}) = \overline{b}, \quad L(t) = \overline{v}_d.$$

   and $\overline{c}$ is implicitly defined to be $L(t_{k+3})$.
4. Pick random points $\overline{v}_{d+1}, \ldots, \overline{v}_{2d}$ on the curve L.
5. Pick additional random points $\overline{u}_1 \ldots \overline{u}_d$ on the line ℓ and $\overline{u}_{d+1}, \ldots, \overline{u}_{2d}$ from the curve L.
6. Let $T_{2ld \times 2ld}$ be an appropriate invertible matrix over $\mathbb{F}[q]$ and H be an appropriate subspace of $\mathbb{F}[q]^{2ld}$. Both depend only on the choice of the points $\{\overline{v}_i\}_{i=1}^{2d}$ and $\{\overline{u}_j\}_{j=1}^{2d}$. Remark 5.2 explains how T and H are chosen.

7. Read the values of the function f from the table at the points $\overline{v}_1, \ldots, \overline{v}_{2d}$ and $\overline{u}_1, \ldots, \overline{u}_{2d}$. Since the values are represented by length l $\mathbb{F}[q]$-vectors, let
$$x = f(\overline{v}_1) \circ f(\overline{v}_2) \circ \cdots \circ f(\overline{v}_{2d}) \tag{9}$$
$$y = f(\overline{u}_1) \circ f(\overline{u}_2) \circ \cdots \circ f(\overline{u}_{2d}) \tag{10}$$
where $\circ$ represents concatenation of vectors.
8. Accept iff,
$$x \neq 0,\ x = Ty \qquad \text{and} \qquad h \cdot x = 0 \quad \forall\ h \in H \quad (\text{i.e. } x \perp H). \tag{11}$$

▶ Remark 5.2. The choice of H is such that $h \cdot x = 0\ \forall\ h \in H$ abstracts out the conditions: (i) the values at the field elements $\{t_i\}_{i=1}^k$ of the degree $d-1$ univariate polynomial interpolated from $f(\overline{v}_{d+1}) \ldots f(\overline{v}_{2d})$ satisfy the homogeneous linear constraints of C, and (ii) the polynomial interpolated from the values $f(\overline{v}_1) \ldots f(\overline{v}_{d-1})$ agrees with the degree $d-1$ polynomial interpolated from $f(\overline{v}_{d+1}) \ldots f(\overline{v}_{2d})$ at the point $\overline{v}_d$, where both evaluate to $f(\overline{v}_d)$.

The invertible matrix T is chosen such that the constraint $x = Ty$ abstracts out the conditions: (i) the degree $d-1$ polynomial interpolated from the values $f(\overline{v}_1) \ldots f(\overline{v}_d)$ is the same as the polynomial interpolated from the values $f(\overline{u}_1) \ldots f(\overline{u}_d)$, and (ii) the degree $d-1$ polynomial interpolated from $f(\overline{v}_{d+1}) \ldots f(\overline{v}_{2d})$ is the same as the polynomial interpolated from the values $f(\overline{u}_{d+1}) \ldots f(\overline{u}_{2d})$.

The condition $x \neq 0$ essentially ensures that f is not a zero polynomial.

The following theorem, regarding the acceptance probability of the Outer Verifier, was proved in [14].

▶ **Theorem 5.3.** *There are constants c_1, c_2 such that the following holds. If, after picking a constraint $C(\{\overline{p}_i\}_{i=1}^k) \in \mathcal{C}$, the Outer Verifier (or the Modified Outer Verifier) accepts with probability δ, then for $1 \leq t \leq 2c_2/(\delta/2)^{c_1}$, $P_1, P_2, \ldots, P_t$ are all the degree d polynomials that have agreement at least $(\delta/2)^{c_1}/c_2$ with f and for some $1 \leq j \leq t$, P_j is a non-zero polynomial whose values at the points $\{\overline{p}_i\}_{i=1}^k$ satisfies the constraint C.*

5.3 Inner Verifier

The Inner Verifier expects, for every point $\overline{v} \in \mathbb{F}^m$, the Hadamard Code of $f(\overline{v}) \in \mathbb{F}[q]^l$. (See Section 7 for a description of the Hadamard Code).

Steps of the Inner Verifier

1. Pick a constraint $C \in \mathcal{C}$ and the points $\overline{v}_1, \ldots, \overline{v}_{2d}$ and $\overline{u}_1, \ldots, \overline{u}_{2d}$ as in steps $1-5$ of the Modified Outer Verifier.
2. Let $T_{2ld \times 2ld}$ and H be the matrix and subspace respectively chosen as in step 7 of the Modified Outer Verifier.
3. Pick a random string $z \in (\mathbb{F}[q]^l)^{2d}$ and a random $h \in H$. Write,
$$z = z_1 \circ z_2 \circ \cdots \circ z_{2d}$$
$$h = h_1 \circ h_2 \circ \cdots \circ h_{2d}$$
$$zT = w_1 \circ w_2 \circ \cdots \circ w_{2d}.$$
4. Let $A_1, \ldots, A_{2d}$ and $B_1, \ldots, B_{2d}$ be the tables giving the supposed Hadamard Codes of $f(\overline{v}_1), \ldots, f(\overline{v}_{2d})$ and $f(\overline{u}_1), \ldots, f(\overline{u}_{2d})$ respectively.
5. Accept iff $\sum_{i=1}^{2d} A_i(z_i + h_i) + \sum_{j=1}^{2d} B_j(w_j) = 0$.

5.4 Analysis

Let Π_* be a subset of locations of the proof Π given to the Inner Verifier. Setting the locations of Π_* to be 1 and the rest of the locations to zero, we obtain the tables A_i and B_i $(1 \leq j \leq 2d)$ which are queried in the description of the Inner Verifier. We wish to analyze the probability over a random test Q of the Inner Verifier that the locations queried by it are contained inside $\Pi*$, i.e.

$$\Pr_Q \left[Q \subseteq \Pi_* \right].$$

This is first arithmetized to,

$$\mathbb{E}_Q \left[\prod_{i=1}^{2d} A_i(z_i + h_i) \prod_{j=1}^{2d} B_j(w_j) \right]. \tag{12}$$

Here Q depends on the choice of the constraint C, the line ℓ and curve L, the points $\overline{v}_1, \ldots, \overline{v}_{2d}, \overline{u}_1, \ldots, \overline{u}_{2d}$, and the choice of z and h. Plugging in the Fourier expansion (see Section 6) of the A_i and B_i we obtain,

$$\begin{aligned} & \mathbb{E}_Q \left[\sum_{\alpha_1,\ldots,\alpha_{2d},\beta_1,\ldots,\beta_{2d}} \left[\prod_{i=1}^{2d} \widehat{A}_{i,\alpha_i} \prod_{j=1}^{2d} \widehat{B}_{j,\beta_j} \prod_{i=1}^{2d} \chi_{\alpha_i}(z_i + h_i) \prod_{j=1}^{2d} \chi_{\beta_j}(w_j) \right] \right] \\ = \quad & \mathbb{E}_Q \left[\sum_{\substack{\alpha=\alpha_1 \circ \cdots \circ \alpha_{2d}, \\ \beta=\beta_1 \circ \cdots \circ \beta_{2d}}} \left[\prod_{i=1}^{2d} \widehat{A}_{i,\alpha_i} \prod_{j=1}^{2d} \widehat{B}_{j,\beta_j} \cdot \phi(\alpha \cdot z + \beta \cdot w) \cdot \phi(\alpha \cdot h) \right] \right], \end{aligned} \tag{13}$$

where $\phi : \mathbb{F}[q] \to \{-1, 1\}$ is defined in Section 6. Now, since h is randomly chosen from H, the above expectation is zero unless $\alpha \perp H$. Also,

$$\begin{aligned} z \cdot \alpha + w \cdot \beta &= z \cdot \alpha + zT \cdot \beta \\ &= z \cdot (\alpha + T\beta) \end{aligned}$$

which implies that the expectation in (13) is zero unless $\alpha = T\beta$, since z is chosen randomly from $\mathbb{F}[q]^{2ld}$. Therefore we obtain the following expression,

$$\mathbb{E}_{\substack{C,\ell,L,\overline{v}_1,\ldots,\overline{v}_{2d} \\ \overline{u}_1,\ldots,\overline{u}_{2d}}} \left[\sum_{\substack{\alpha=\alpha_1 \circ \cdots \circ \alpha_{2d}, \\ \beta=\alpha_1 \circ \cdots \circ \beta_{2d}, \\ \alpha \perp H,\ \beta=T^{-1}\alpha}} \left[\prod_{i=1}^{2d} \widehat{A}_{i,\alpha_i} \prod_{j=1}^{2d} \widehat{B}_{j,\beta_j} \right] \right]. \tag{14}$$

5.4.1 YES Case

We prove the following lemma.

▶ **Lemma 5.4.** *If the instance $\mathcal{A}$ of* HomAlgCSP *is a* YES *instance then there exists a subset Π_* of $1/q$ fraction of the proof locations such that*

$$\Pr_Q \left[Q \subseteq \Pi_* \right] \geq \frac{1}{q^{4d-1}} \left(1 - \frac{1}{2^K} \right).$$

Proof. Let f be the polynomial given by the YES case. Since f is of degree at most d^*, it is nonzero at all points except for a negligible fraction ($O(d^*/|\mathbb{F}|)$) which we ignore. Construct the proof Π as follows: For each point $\overline{v} \in \mathbb{F}^m$ let the corresponding table $A_{\overline{v}}$ be defined as:

$$A_{\overline{v}}(x) = \begin{cases} 1 & \text{if the Hadamard Code of } f(\overline{v}) \text{ at location } x \text{ is } 0 \in \mathbb{F}[q], \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Now, let Π_* be the subset of locations where Π is 1. Since we are dealing with Hadamard Codes of nonzero values, Π_* is exactly $(1/q)$ fraction of the locations. Also, from Lemma 7.1,

$$A_{\overline{v}} = (1/q) \sum_{t \in \mathbb{F}[q]} \chi_{tf(\overline{v})},$$

where $f(\overline{v})$ in the is represented as an element of $\mathbb{F}[q]^l$. Using this, we see that when the constraint C is satisfied by f, for each $t \in \mathbb{F}[q]$ setting $\alpha_i = tf(\overline{v}_i)$ and $\beta_i = tf(\overline{u}_i)$ $(1 \leq i \leq 2d$ where the f values are represented as elements of $\mathbb{F}[q]^l)$ contributes $1/q^{4d}$ in (14). Since $1 - 1/2^K$ fraction of the constraints are satisfied by f in the YES case of Theorem 3.2 the lemma follows. ◀

5.4.2 NO Case

The NO case soundness is given by the following lemma.

▶ **Lemma 5.5.** *Let Π_* be any subset of $\zeta \in [0,1]$ fraction of locations of the proof Π. Then, if $\mathcal{A}$ is a* NO *instance then,*

$$\left| \Pr_Q [Q \subseteq \Pi_*] - \zeta^{4d} \right| \leq \frac{C_0}{2^{2^K/C_1}}, \tag{16}$$

for some universal constants $C_0, C_1 > 0$.

Proof. Suppose that,

$$\left| \Pr_Q [Q \subseteq \Pi_*] - \zeta^{4d} \right| = \delta. \tag{17}$$

Let the proof Π evaluate to 1 at the locations in Π^* and zero otherwise. Thus,

$$\mathbb{E}_{\overline{v}} \left[\widehat{A}_{\overline{v}} \right] = \zeta.$$

Using the mixing property of curves and lines (refer to Appendix A.4 of [14]) we obtain that for the random line ℓ and curve L chosen by the Outer Verifier, except with probability $O(1/|\mathbb{F}|^{1/3})$ (which we shall ignore),

$$\forall\ i = 1, \ldots, 2d \quad \mathbb{E}_{\overline{v}_i \in \ell}[\widehat{A}_{i,0}] \approx \zeta \quad , \quad \mathbb{E}_{\overline{u}_j \in \ell}[\widehat{B}_{j,0}] \approx \zeta$$

where again the error in the above approximations is bounded by $O(1/|\mathbb{F}|^{1/3})$ which we shall ignore. Thus, the contribution of $\alpha = 0$ in (14) is (up to negligible error) ζ^{4d}. From (17), an analysis identical to that in Section 10.5 of [14] yields a table of values f such that the Outer Verifier accepts with probability δ^2 over a randomly chosen constraint $C \in \mathcal{C}$ of the HomAlgCSP instance $\mathcal{A}$. Thus, for at least $\delta^2/2$ of the constraints C, the Outer Verifier accepts with probability $\delta^2/2$. Using Theorem 5.3, this implies that there is a degree $d \leq (21+3)d^* \leq 100d^*$ polynomial which satisfies at least $(\delta/C_0)^{C_1}$ fraction of the constraints of $\mathcal{A}$ for some universal constants $C_0, C_1 > 0$. This contradicts the NO case of Theorem 3.2 unless δ is at most the RHS of (16), thus completing the proof of the lemma. We omit further details and refer the reader to [14]. ◀

5.4.3 Setting the parameters

The various parameters of the PCP reduction from HomAlgCSP are set so that Lemmas 5.4 and5.5 along with Theorem 3.2 yield Theorem 3.4. First we change $4d$ to d in Lemmas 5.4 and 5.5. As in Theorem 3.4, $q := R^{4d}$. We set $d := \Theta(2^{K/3}/(\log R))$ so that,

$$\frac{1}{2^K} = O\left(\frac{1}{d^3}\right) \quad \text{and} \quad \frac{C_0}{2^{2^K/C_1}} \leq \frac{1}{q^{2d^2}},$$

appropriately bounding the errors in Lemmas 5.4 and 5.5. Note that $d \leq (21+3)\cdot 10 \cdot 4 \cdot \tilde{d} \leq 1000\tilde{d}$ (where $\tilde{d}$ is as in Theorem 3.2). Thus, choosing $m = n^{1000\Delta K/d}$ yields $\binom{m}{d} \geq N$ as required, and that the entire reduction runs in time 2^{n^ε} where $\varepsilon = \Theta(K/d)$ can be made arbitrarily small by choosing K large enough. Rearranging, $d = \Theta((1/\varepsilon)\log((\log R)/\varepsilon))$.

6 Fourier Analysis

We will be working over the field $\mathbb{F}[q] := \mathbb{F}[2^r]$ for $r > 0$, which is a field extension of $\mathbb{F}[2]$. Let φ be the isomorphism from the additive group $(\mathbb{F}[2^r], +)$ to $(\mathbb{F}[2]^r, +)$. Define the following homomorphism ϕ from $(\mathbb{F}[2^r], +)$ to the multiplicative group $(\{-1, 1\}, .)$.

$$\phi(a) = \begin{cases} 1 & \text{if } \varphi(a) \text{ contains even number of 1s} \\ -1 & \text{otherwise} \end{cases}$$

for any $a \in \mathbb{F}[2^r]$. Note that $\phi(a+b) = \phi(a)\phi(b),\ \forall a, b \in \mathbb{F}[2^r]$. We now define the 'characters' $\psi_a : \mathbb{F}[2^r] \mapsto \{-1, 1\}$ for $a \in \mathbb{F}[2^r]$ as follows.

$$\psi_a(b) := \phi(ab)$$

The characters ψ_a satisfy the following properties.

$$\begin{aligned} \psi_0(b) &= 1 \qquad && \forall b \in \mathbb{F}[2^r] \\ \psi_a(0) &= 1 && \forall a \in \mathbb{F}[2^r] \\ \psi_{a+b}(c) &= \psi_a(c)\psi_b(c) \end{aligned}$$

and,

$$\sum_{a \in \mathbb{F}[2^r]} \psi_a(b) = \begin{cases} |\mathbb{F}[2^r]| & \text{if } b = 0 \\ 0 & \text{otherwise} \end{cases}$$

We note that the 'character' functions form an orthonormal basis for the space $L^2(\mathbb{F}[2^r])$. We have that,

$$\langle \psi_a, \psi_b \rangle = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases}$$

where,

$$\langle \psi_a, \psi_b \rangle := \mathbb{E}_{c \in \mathbb{F}[2^r]} \left[\psi_a(c)\psi_b(c)\right].$$

We now consider the vector space $\mathbb{F}[2^r]^m$ for some positive integer m. We define the 'characters' $\chi_\alpha : \mathbb{F}[2^r]^m \mapsto \{-1, 1\}$ for every $\alpha \in \mathbb{F}[2^r]^m$ as,

$$\chi_\alpha(f) := \phi(\alpha \cdot f), \qquad f \in \mathbb{F}[2^r]^m$$

where '$\cdot$' is the inner product in the vector space $\mathbb{F}[2^r]^m$. From the way we defined the characters ψ_a, we have,

$$\chi_\alpha(f) = \prod_{i=1}^{m} \psi_{\alpha_i}(f_i),$$

where α_i and f_i are the i^{th} coordinates of α and f respectively. The characters χ_α satisfy the following properties,

$$\begin{aligned} \chi_0(f) &= 1 \qquad && \forall f \in \mathbb{F}[2^r]^m \\ \chi_\alpha(0) &= 1 && \forall \alpha \in \mathbb{F}[2^r]^m \\ \chi_{\alpha+\beta}(f) &= \chi_\alpha(f)\chi_\beta(f) \\ \chi_\alpha(f+g) &= \chi_\alpha(f)\chi_\alpha(g) \end{aligned}$$

and,

$$\mathbb{E}_{f\in\mathbb{F}[2^r]^m}\left[\chi_\alpha(f)\right] = \begin{cases} 1 & \text{if } \alpha = 0 \\ 0 & \text{otherwise} \end{cases}$$

The characters χ_α form an orthonormal basis for $L^2(\mathbb{F}[2^r]^m)$. We have,

$$\langle \chi_\alpha, \chi_\beta \rangle = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{otherwise} \end{cases}$$

where,

$$\langle \chi_\alpha, \chi_\beta \rangle := \mathbb{E}_{f\in\mathbb{F}[2^r]^m}\left[\chi_\alpha(f)\chi_\beta(f)\right].$$

Let $A : \mathbb{F}[2^r]^m \mapsto \mathbb{R}$ be any real valued function. Then the Fourier expansion of A is given by,

$$A(x) = \sum_{\alpha\in\mathbb{F}[2^r]^m} \widehat{A}_\alpha \chi_\alpha(x),$$

where,

$$\widehat{A}_\alpha = \mathbb{E}_{x\in\mathbb{F}[2^r]^m}[A(x)\chi_\alpha(x)].$$

A useful equality is:

$$\widehat{A}_0 = \mathbb{E}_{x\in\mathbb{F}[2^r]^m}[A(x)].$$

7 Hadamard Codes

Let l be a positive integer and $\mathbb{F}[q]$ be an extension of $\mathbb{F}[2]$. Then, for any $a \in \mathbb{F}[q]^l$, its Hadamard Code $H_a : \mathbb{F}[q]^l \to \mathbb{F}[q]$ is given by $H_a(x) = a \cdot x = \sum_{i=1}^{l} a_i x_i$. We have the following simple lemma.

▶ **Lemma 7.1.** *For any $a \in \mathbb{F}[q]^l$, let $A : \mathbb{F}[q]^l \to \{0,1\}$ be defined as $A(x) := \mathbb{1}\{H_a(x) = 0\}$. Then,*

$$A = (1/q) \sum_{t\in\mathbb{F}[q]} \chi_{ta}.$$

Proof. If $a = 0$, then A is identically 1, and thus $A = \chi_0 = (1/q)\sum_{t\in\mathbb{F}[q]} \chi_{ta}$. If $a \neq 0$, then $\mathbb{E}_x[A(x)] = \Pr_x[a \cdot x = 0] = 1/q$. Further,

$$\{A(x) = 1\} \Leftrightarrow \{a \cdot x = 0\} \Leftrightarrow \{ta \cdot x = 0,\ \forall t \in \mathbb{F}[q]\} \Rightarrow \{\chi_{ta}(x) = 1,\ \forall t \in \mathbb{F}[q]\}.$$

Thus, $\widehat{A}_{ta} = \mathbb{E}_x[A(x)\chi_{ta}(x)] = \mathbb{E}_x[A(x)] = 1/q$, for all $t \in \mathbb{F}[q]$. By Parseval's identity these are the only non-zero Fourier coefficients. ◀

References

1. A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. O(sqrt(log n)) approximation algorithms for min UnCut, min 2CNF deletion, and directed cut problems. In *Proceedings of the ACM Symposium on the Theory of Computing*, pages 573–581, 2005.
2. N. Alon. Eigenvalues, geometric expanders, sorting in rounds, and Ramsey theory. *Combinatorica*, 6(3):207–219, 1986.
3. C. Ambühl, M. Mastrolilli, and O. Svensson. Inapproximability results for maximum edge biclique, minimum linear arrangement, and sparsest cut. *SIAM Journal of Computing*, 40(2):567–596, 2011.
4. B. Applebaum, B. Barak, and A. Wigderson. Public-key cryptography from different assumptions. In *Proceedings of the ACM Symposium on the Theory of Computing*, pages 171–180, 2010.
5. S. Arora, S. Rao, and U. V. Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM*, 56(2):1–37, 2009.
6. P. Austrin and S. Khot. A simple deterministic reduction for the gap minimum distance of code problem. In *Proceedings of ICALP*, pages 474–485, 2011.
7. S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. *Computational Complexity*, 15(2):94–114, 2006.
8. Q. Cheng and D. Wan. A deterministic reduction for the gap minimum distance problem: [extended abstract]. In *Proceedings of the ACM Symposium on the Theory of Computing*, pages 33–38, 2009.
9. I. Dumer, D. Micciancio, and M. Sudan. Hardness of approximating the minimum distance of a linear code. *IEEE Trans. Information Theory*, 49(1):22–37, 2003.
10. U. Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the ACM Symposium on the Theory of Computing*, pages 534–543, 2002.
11. U. Feige, M. Hajiaghayi, and J. R. Lee. Improved approximation algorithms for minimum weight vertex separators. *SIAM Journal of Computing*, 38(2):629–657, 2008.
12. V. Guruswami, C. Umans, and S. P. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *Journal of the ACM*, 56(4), 2009.
13. S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the ACM Symposium on the Theory of Computing*, pages 767–775, 2002.
14. S. Khot. Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM Journal of Computing*, 36(4):1025–1071, 2006.
15. S. Khot and N. K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proceedings of the Annual Symposium on Foundations of Computer Science*, pages 53–62, 2005.
16. F. T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM*, 46(6):787–832, 1999.
17. N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.
18. R. J. Lipton and R. E. Tarjan. Applications of a planar separator theorem. *SIAM Journal of Computing*, 9(3):615–627, 1980.

19 A. Louis, P. Raghavendra, and S. Vempala. The complexity of approximating vertex expansion. In *Proceedings of the Annual Symposium on Foundations of Computer Science*, pages 360–369, 2013.

20 A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–227, 1988.

21 R. Müller and D. Wagner. α-vertex separator is NP-hard even for 3-regular graphs. *Computing*, 46:343–353, 1991.

22 N. Pippenger. Sorting and selecting in rounds. *SIAM Journal of Computing*, 16(6):1032–1038, 1987.

23 P. Raghavendra and D. Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the ACM Symposium on the Theory of Computing*, pages 755–764, 2010.

24 P. Raghavendra, D. Steurer, and M. Tulsiani. Reductions between expansion problems. In *Proceedings of the Annual IEEE Conference on Computational Complexity*, pages 64–73, 2012.

25 A. Rao. *Randomness Extractors for Independent Sources and Applications*. PhD thesis, University of Texas at Austin, 2007.

26 M. Sipser and D. A. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42(6):1710–1722, 1996.

27 Z. Svitkina and L. Fleischer. Submodular approximation: Sampling-based algorithms and lower bounds. *SIAM Journal of Computing*, 40(6):1715–1737, 2011.