Electronic Colloquium on Computational Complexity, Report No. 129 (2024)

On Approximability of Satisfiable k-CSPs: V

Amey Bhangale* Subhash Khot† Dor Minzer‡

Abstract

We propose a framework of algorithm vs. hardness for all Max-CSPs and demonstrate it for a large class of predicates. This framework extends the work of Raghavendra [STOC, 2008], who showed a similar result for almost satisfiable Max-CSPs.

Our framework is based on a new *hybrid approximation algorithm*, which uses a combination of the Gaussian elimination technique (i.e., solving a system of linear equations over an Abelian group) and the semidefinite programming relaxation. We complement our algorithm with a matching dictator vs. quasirandom test that has perfect completeness.

The analysis of our dictator vs. quasirandom test is based on a novel invariance principle, which we call the *mixed invariance principle*. Our mixed invariance principle is an extension of the invariance principle of Mossel, O'Donnell and Oleszkiewicz [Annals of Mathematics, 2010] which plays a crucial role in Raghavendra's work. The mixed invariance principle allows one to relate 3-wise correlations over discrete probability spaces with expectations over spaces that are a mixture of Guassian spaces and Abelian groups, and may be of independent interest.

1 Introduction

1.1 Constraint Satisfaction Problems

The class of constraint satisfaction problems (CSPs in short) consists of some of the most studied computational problems in artificial intelligence, database theory, logic, graph theory, and computational complexity. Given a predicate $P : \Sigma^k \to \{0,1\}$ for some finite alphabet Σ, a P-CSP instance consists of a set of variables $X = \{x_1, x_2, \ldots, x_n\}$ and a collection of *local* constraints $C_1, C_2, \ldots, C_m$, each one of the form $P(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = 1$. Here and throughout, we refer to the parameter k as the arity of the CSP. For a class of predicates $\mathcal{P} \subseteq \{P \colon \Sigma^k \to \{0,1\}\}$, an instance of $\mathcal{P}$-CSP consists of a set of variables X and a colleciton of constraints $C_1, \ldots, C_m$ each one of the form $P(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) = 1$ for some $P \in \mathcal{P}$. The value of an instance Υ, denoted by $\mathbf{val}(\Upsilon)$, is the maximum fraction of the constraints that can be satisfied by an assignment to the variables.

The most natural decision problem associated with instances of $\mathcal{P}$-CSP is the satisfiability problem: given an instance Υ of $\mathcal{P}$-CSP, determine if it is satisfiable, i.e. , if there exists an assignment $A \colon X \to \Sigma$ satisfying all of the constraints of Υ. In a relaxation of this problem called the Max-$\mathcal{P}$-CSP problem (which is most relevant to this paper), one is given an instance Υ of $\mathcal{P}$-CSP, and the task is to efficiently find an assignment to the variables that satisfies as many of the constraints as possible. An α-approximation

*Department of Computer Science and Engineering, University of California, Riverside. Supported by the Hellman Fellowship award.

†Department of Computer Science, Courant Institute of Mathematical Sciences, New York University. Supported by the NSF Award CCF-1422159, NSF CCF award 2130816, and the Simons Investigator Award.

‡Department of Mathematics, Massachusetts Institute of Technology. Supported by NSF CCF award 2227876 and NSF CAREER award 2239160.

ISSN 1433-8092

algorithm is a polynomial-time algorithm that, given an instance Υ, finds an assignment satisfying at least $\alpha \cdot \textsc{Opt}(\Upsilon)$ fraction of the constraints, where $\textsc{Opt}(\Upsilon)$ is the value of the optimal assignment.

The study of CSPs has driven some of the most influential developments in theoretical computer science, including the theory of NP-completeness [21, 40], the PCP theorem [1, 2, 24], the development of semidefinite programming algorithms [28, 35, 53, 54, 46, 18], the Unique Games Conjecture (UGC in short) and its consequences [36, 46], the dichotomy theorem [23, 17, 52] and more. Below, we elaborate on two of these topics.

The Dichotomy Theorem: a systematic study of the complexity of solving CSPs was started by Schaefer in 1978 [48], who showed that for every predicate P over a Boolean alphabet, the problem of checking satisfiability of a P-CSP instance is either in P or is NP-complete. Note that this is not a trivial statement: by Lander's theorem [39], if P$\neq$NP, then there are languages that are not in P nor are NP-hard; these are often called NP-intermediate problems. Thus, another way of stating Shaefer's theorem is that CSPs over Boolean alphabets cannot be NP-intermediate. Feder and Vardi [23] conjectured that Shaefer's theorem holds for all finite alphabets, and this statement is often referred to as the Dichotomy Conjecture. Much effort has gone into studying the dichotomy conjecture, mainly using the tools of abstract algebra. The conjecture was recently resolved by Bulatov and Zhuk (independently) [17, 52], who proved that, indeed, for any family of prediactes $\mathcal{P}$, the decision problem $\mathcal{P}$-CSP is either in P or is NP-complete.

Approximating Almost Satisfiable Instances: the complexity of approximating *almost satisfiable* instances is rather well understood by now. Here and throughout, we say that an instance Υ is almost satisfiable if $\textsc{Opt}(\Upsilon) \geqslant 1 - \varepsilon$ where $\varepsilon > 0$ is a small constant.

Some of the theory here is based on the PCP theorem [24, 2, 1]. As an example, an important result of Håstad [33] states that for all $\varepsilon > 0$, given an instance Υ of the Max-3-Lin problem promised to have $\textsc{Opt}(\Upsilon) \geqslant 1 - \varepsilon$, it is NP-hard to find an assignment satisfying at least $1/2 + \varepsilon$ fraction of the constraints. Here, the Max-3LIN problem is the problem Max-$\mathcal{P}$-CSP where $\mathcal{P} = \{P_0, P_1\}$ and $P_a \colon \mathbb{F}_2^3 \to \{0, 1\}$ is defined by $P_a(x, y, z) = 1_{x+y+z=a}$. In fact, Håstad's hardness result [33] also applies to the *decision version* of the problem: for every $\varepsilon > 0$, it is NP-hard to distinguish between the cases $\mathbf{val}(\Upsilon) \geqslant 1 - \varepsilon$ and $\mathbf{val}(\Upsilon) \leqslant 1/2 + \varepsilon$.

Getting a more comprehensive understanding of the approximability of almost satisfiable CSPs requires a stronger PCP characterization of NP, in the form of the Unique Games Conjecture (UGC) [36]. Assuming UGC, Raghavendra [46] showed that for every collection of predicates $\mathcal{P}$ and $\varepsilon > 0$, there is a constant $\beta_{\mathcal{P}}$ such that:

1. Algorithm: there exists a polynomial-time algorithm that given an instance Υ of $\mathcal{P}$-CSP promised to have $\textsc{Opt}(\Upsilon) \geqslant 1-\varepsilon$, outputs an assignment satisfying at least $\beta_{\mathcal{P}}$ fraction of the constraints. Clearly, this also means that there exists a polynomial-time algorithm that distinguishes instances with a value at least $1 - \varepsilon$ from instances with a value at most almost $\beta_{\mathcal{P}}$.

2. Hardness: for all $\delta > 0$, given an instance Υ, it is NP-hard to distinguish between the case that $\mathbf{val}(\Upsilon) \geqslant 1 - \varepsilon - \delta$, and the case that $\mathbf{val}(\Upsilon) \leqslant \beta_{\mathcal{P}} + \delta$.

In words, Raghavendra's result asserts that for every collection of predicates, the approximability of almost satisfiable instances exhibits a dichotomy between approximation ratios that can be achieved in polynomial time, and those that are NP-hard (assuming UGC).

1.2 Approxmiating Satisfiable Instances

The complexity of approximating satisfiable CSPs is much more complicated and remains mostly unsolved as of now (even under reasonable conjectures in the style of UGC [36]). The proof of the dichotomy theorem implies that for $\mathcal{P}$ for which $\mathcal{P}$-CSP is NP-complete, there exists a constant $0 < \delta_{\mathcal{P}} < 1$ such that it is NP-hard to distinguish satisfiable instances from instances with value at most $\delta_{\mathcal{P}}$. However, unlike the almost-satisfiable case, we do not know tight inapproximability results for satisfiable instances for every $\mathcal{P}$. There are only a few $\mathcal{P}$'s for which we know the existence of $\alpha_{\mathcal{P}}$ for which efficient $\alpha_{\mathcal{P}}$ approximation of Max-$\mathcal{P}$-CSP is possible, while $\alpha_{\mathcal{P}} + \varepsilon$ approximation is NP-hard, and even fewer $\mathcal{P}$'s for which we know the value of $\alpha_{\mathcal{P}}$. An example for such problem is the 3-SAT problem, for which Håstad [33] proved that $\alpha_{3SAT} = 7/8$ works (and more generally $\alpha_{kSAT} = 1 - 1/2^k$ for the k-SAT problem). Another example is the NTW predicate[1] that received much attention [43, 44, 37], for which the optimal threshold of $\alpha_{NTW} = 5/8$ was shown by Håstad [34]. The problem of determining the existence of the ratio α and pinning it down gets very difficult very quickly though; see [15] for the case of the NAE predicate.

1.3 The Dichotomy Approximation Conjecture

Motivated by the dichotomy theorem and Raghavendra's theorem discussed above, in [7] the authors suggested the following statement, referred to as the *Approximation Dichotomy Conjecture*:

Conjecture 1.1 (Approximation Dichotomy Conjecture)**.** *For all $k \in \mathbb{N}$ and for all collections of k-ary predicates $\mathcal{P}$ for which $\mathcal{P}$-CSP is NP-hard, there exists a constant $\alpha_{\mathcal{P}}$ such that:*

1. *Algorithm: there is a polynomial-time algorithm that distinguishes satisfiable instances of Max-$\mathcal{P}$-CSP from instances with value at most $\alpha_{\mathcal{P}}$.*

2. *Hardness: for all $\varepsilon > 0$, it is NP-hard to distinguish satisfiable instances of Max-$\mathcal{P}$-CSP from instances with value at most $\alpha_{\mathcal{P}} + \varepsilon$.*

In words, the approximation dichotomy conjecture states that the complexity of the approximating Max-$\mathcal{P}$-CSP exhibits a rapid phase transition between approximation ratios that can be achieved by polynomial-time algorithms, and approximation ratios that are NP-hard to achieve. In other words, the approximation problem is never NP-intermediate.

A sequence of works [7, 9, 10, 11] made progress on the case $k = 3$ of Conjecture 1.1, focusing on the "hardness" part; we elaborate on these works below. The goal of the current paper is to make further progress on the case that 3-ary predicates and address the "algorithmic" part of Conjecture 1.1. In particular, we propose an approximation algorithm for a wide class of CSPs, and develops tools to bridge between its performance and the "hardness" side of the conjecture.

1.3.1 Why 3-ary Predicates?

In full generality, Conjecture 1.1 is likely to be very difficult to settle. Settling it, even for certain classes of CSPs, requires one to study associated, very general analytical problems. These problems include within them (as subcases) the inverse theorems for Gowers' uniformity norms over finite fields [5, 50, 51]. In fact, the resulting analytical problem for k-ary predicates implies a generalization of the inverse theorem for Gowers' U^{k-1}-norms. For instance, the main result of [11] is a solution to this analytical problem for

[1] The accepting assignments of the predicate NTW are $\{(0,0,0),(0,0,1),(0,1,0),(1,0,0),(1,1,1)\}$.

$k = 3$, and it can be used to study the density of restricted 3-AP free sets [8], a result which was previously unknown. As the difficulty in the study of Gowers' uniformity norms sharply increases when k is large, it stands to reason the associated analytical problems we study also become more challenging as k increases. Hence we focus on the simplest case that is not understood, $k = 3$, which is already highly non-trivial. It is also reasonable to expect that the resolution of the analytical problem for general k will ultimately proceed by induction on k, in which scenario the case $k = 3$ will be the base case of this induction.[2]

We remark that prior works, and more specifically Raghavendra's theorem [46] as well as the dichotomy theorem [17, 52], have side-stepped these issues. First, due to the imperfect completeness Raghavendra requires only fairly simple inverse theorems (corresponding to distributions with full support), and the argument is essentially the same for all k. Second, in the context of the dichotomy theorem one does not have to worry about preserving approximation ratios, and the difference between this case and our case is analogous to the difference between the Hales-Jewett theorem [32] and the density Hales-Jewett theorem [25, 27, 45].

1.3.2 Approximation Algorithms vs Hardness Reductions

To characterize the right approximation threshold for Max-$\mathcal{P}$-CSP, one has to work in two fronts and make them meet: the algorithmic front and the hardness front. In this section we discuss these two fronts, the way it is manifested in almost satisfiable CSPs, and the difference between that and the case of satisfiable CSPs.

Dictatorship tests as evidence for hardness: dictatorship tests are one of the most important components in hardness of approximation results. A function $f : \Sigma^n \to \Sigma$ is called a dictatorship function if it depends only on one variable. A dictatorship test is a procedure that queries f at a few (correlated) locations randomly, and based on the function values at these locations it decides if f is a dictator function or *far* from any dictator function. For brevity, we often refer to the latter type of functions as *quasirandom* functions.

We briefly describe the notion of being *far* from dictator functions here. The influence of a coordinate i on a function $f\colon (\Sigma^n, \nu^n) \to \mathbb{C}$ is the amount it affects the value of f at a random input, i.e.,

$$I_i[f] = \mathop{\mathbb{E}}_{\substack{(x_1,\ldots,x_n)\sim\nu^n \\ x_i'\sim\nu}} \left[\left|f(x_1,\ldots,x_i,\ldots,x_n) - f(x_1,\ldots,x_i',\ldots,x_n)\right|^2\right].$$

Note that dictatorship functions have one coordinate whose influence is 1. A function is called far from dictatorship functions if, for every coordinate i, the influence of the coordinate i in f is small. There are three important properties of a dictatorship test that are useful in getting the hardness of approximation result for Max-$\mathcal{P}$-CSP:

1. The completeness c, which is the probability that test accepts any dictatorship function.
2. The soundness s, which is the probability that the test accepts any function which is far from being a dictatorship.
3. The check that the test makes: if the dictatorship test makes k queries, say to the points $x(1), \ldots, x(k)$, and performs a check of the form $P(f(x(1)), \ldots, f(x(k))) = 1$, for $P \in \mathcal{P}$, then it can be used to get a hardness result for Max-$\mathcal{P}$-CSP.

Typically, a dictatorship test with parameters $0 < s < c \leqslant 1$ that uses a collection of predicates $\mathcal{P}$, can be converted to an NP-hardness result for the following promise problem: given an instance Υ of CSP-$\mathcal{P}$

[2] Many of the proofs of the inverse theorem for Gowers' uniformity norms indeed proceed in this fashion.

promised to be at least c-satisfiable, find an assignment satisfying at least s fraction of the constraints. While this transformation is far from being automatic in general,[3] dictatorship tests are often thought of as strong evidence towards this form of hardness of approximation result. Thus, the hardness part of Conjecture 1.1 requires dictatorship tests with $c = 1$, which we refer to as perfect completeness.

Dictatorship tests with perfect completeness: the papers [7, 9, 10, 11] develop tools to analyze dictatorship tests with perfect completeness. This case turns out to be considerably more complicated than the case of imperfect completeness. Whereas in the latter case, one can change the dictatorship test slightly (to gain desirable features from it) at only a mild cost in the completeness parameter, this is unaffordable once we take $c = 1$. In particular, the paper [11] gives a "quasirandom" versus "dictatorship" result for a wide class of 3-ary predicates.

Matching the dictatorship test via an algorithm: to match the performance of the dictatorship test, an approximation algorithm has to gain insights from the soundness analysis of the test. At a high level, the algorithm has to be able to utilize any quasirandom function in an algorithmic way.

In the context of Raghavendra's theorem [46], the class of quasirandom functions consists of low-degree functions with no influential coordinates [41]. To utilize these functions algorithmically, the powerful invariance principle of [42] is used, asserting that these type of functions essentially come from Gaussian space. As Gaussian samples can be produced algorithmically by solving the semi-definite programming relaxation and multiplying by vector-valued Gaussian random variables, this gives rise to an algorithm.

In the context of satisfiable CSPs, the class of quasirandom functions is more rich in general. In fact, this class may even depend on the specific predicate in question. In many cases of interest this class includes low-degree functions as well as Fourier characters, and it is not immediately clear how to use these types of functions algorithmically. The invariance principle of [42] fails for these types of functions, and more information (besides the one gained from the semi-definite programming relaxation) seems necessary for an algorithm. The main contribution of the current work is to propose such an algorithm and prove a generalization of the invariance principle that facilitates the use of this type of functions algorithmically.

1.4 Our Main Result

We now state the main result of this paper, and towards this end we require a few definitions. We begin with the notion of Abelain embeddings, also referred to as linear embeddings.

Definition 1.2. *Let $\Sigma_1, \dots, \Sigma_k$ be finite alphabets and let $A \subseteq \prod_{i=1}^{k} \Sigma_i$. For an Abelian group $(G, +)$, we say maps $\sigma_i \colon \Sigma_i \to G$ for $i = 1, \dots, k$ form an Abelian embedding of A if*

$$(a_1, \dots, a_k) \in A \implies \sum_{i=1}^{k} \sigma_i(a_i) = 0_G.$$

We say A is Abelianly embeddable if there are maps σ_i that are not all constant that form an Abelian embedding of A. We say that a distribution μ over $\prod_{i=1}^{k} \Sigma_i$ is Abelialy embeddable if $\mathsf{supp}(\mu)$ is.

[3] Often times one requires a plausible complexity-theoretic assumption, such as the Unique-Games Conjecture [36] or the Rich-2-to-1 Games Conjecture [16].

The notion of Abelian embeddings is central to the study of satisfiable CSPs, and the existence of embeddings into a group G should be thought of as hinting towards a subspace-type structure inside A. For technical reasons though, we are only able to handle Abelian groups G that are finite. Thus, we say that A admits a $(\mathbb{Z}, +)$ embedding if it has an Abelian embedding into the group $(\mathbb{Z}, +)$. With this in mind, we wish to define the class of predicates we handle in the current paper, and we begin by giving a few examples.

1. **Vector valued punctured 3-Lin:** let p be a prime number, let $m \geqslant 2$ and take $\Sigma = \mathbb{F}_p^m \setminus \{\vec{0}\}$. Take the collection of predicates $\mathcal{P} = \{P\}$, where the set of satisfying assignments of P consists of tuples (x, y, z) such that $x, y, z \in \mathbb{F}_p^m \setminus \{\vec{0}\}$, each pair of x, y, z is linearly independent and $x + y + z = \vec{0}$. It is easy to observe that P admits an Abelian embedding into $\mathbb{F}_p^m$. There are other Abelian embeddings that are induced by it, such as $\sigma_1(x) = \langle \alpha, x\rangle$, $\sigma_2(y) = \langle \alpha, y\rangle$, $\sigma_3(z) = \langle \alpha, z\rangle$ for any $\alpha \in \mathbb{F}_p^m$.

   The predicate P does not admit a $(\mathbb{Z}, +)$ embedding, and furthermore there is a group action on Σ that preserves satisfying assignments of P. Namely, for each invertible $M \in \mathbb{F}_p^{m\times m}$ we can take $\tau_M\colon \mathbb{F}_p^m \to \mathbb{F}_p^m$ defined as $\tau_M u = Mu$. It is easy to see that for each such M, if $P(x, y, z) = 1$, then $P(\tau_M(x), \tau_M(y), \tau_M(z)) = 1$. Using this, one can show that the collection $\mathcal{P}$ is MILDLY-SYMMETRIC as defined below, and our results therefore apply to $\mathcal{P}$.

2. **3-Uniform hypergraph p-strong coloring:** let $p \geqslant 5$ be a prime, take $\Sigma = \mathbb{F}_p$ and consider the predicate $P\colon \Sigma \to \{0, 1\}$ where $P(x, y, z) = 1_{x,y,z \text{ are distinct}}$ and $\mathcal{P} = \{P\}$. The problem Max-$\mathcal{P}$-CSP is also known [12] as the p-strong coloring problem: an instance can be viewed as a hypergraph (whose hyperedges are the triplets that are involved in some constraint), and a solution can be viewed as coloring the graph maximizing the fraction of hyperedges with all distinct colors.

   The predicate P does not admit Abelian embeddings, however its support does contain subsets that do admit Abelian embeddings. The collection $\mathcal{P}$ can be seen to be MILDLY-SYMMETRIC (see Section A for a proof), and so our result applies to $\mathcal{P}$.

3. **3-Uniform hypergraph 3-rainbow coloring:** consider the previous example except that we take $p = 3$. This problem is also known [31] as the rainbow hypergraph coloring problem, wherein an hyperedge is properly colored if *all* the colors are present. This time the predicate P has Abelian embeddings, such as $\sigma, \gamma, \phi\colon \Sigma \to \mathbb{F}_3$ defined as $\sigma(x) = x$, $\gamma(y) = y - 1$, $\phi(z) = z - 2$. In fact, this is even a $(\mathbb{Z}, +)$ embedding. We also note that there is a group action $\{\tau_{a,b}\}_{a,b\in\mathbb{F}_3, a\neq 0}$ on Σ defined as $\tau_{a,b}(u) = au + b$ that preserves the satisfying assignment of P.

   Because P admits a $(\mathbb{Z}, +)$ embedding, the collection $\mathcal{P}$ is not MILDLY-SYMMETRIC as defined below, and our result does not apply to it. As discussed in Sections 1.7.3 and 1.9 we view the $(\mathbb{Z}, +)$-embedding obstruction as an issue of a technical nature (as opposed to the issue of pairwise connectivity, which appears much more fundamental). Therefore, we suspect that a variant of our result should hold for this predicate as well.

Having looked at a few examples and non-examples, we now formally define the class of MILDLY-SYMMETRIC predicates as follows:

Definition 1.3 (MILDLY-SYMMETRIC predicates)**.** *A family of predicates $\mathcal{P} \subseteq \{P : \Sigma^3 \to \{0, 1\}\}$ is called* MILDLY-SYMMETRIC *if there are actions $\tau_1, \tau_2, \ldots, \tau_\ell : \Sigma \to \Sigma$ such that:*

1. *For every $P \in \mathcal{P}$, every $i \in [\ell]$ and every satisfying assignment $\sigma \in \Sigma^3$ of P, the assignment $(\tau_i(\sigma_1), \tau_i(\sigma_2), \tau_i(\sigma_3))$ is a satisfying assignment of P.*

2. *For every $P \in \mathcal{P}$ and for every satisfying assignment $\sigma \in \Sigma^3$ of P, the set $\{(\tau_i(\sigma_1), \tau_i(\sigma_2), \tau_i(\sigma_3)) \mid i \in [\ell]\} \subseteq \Sigma^3$ does not have a $(\mathbb{Z}, +)$-embedding.*

In words, a collection of predicates $\mathcal{P}$ is called MILDLY-SYMMETRIC if there are maps on the alphabet Σ that both (1) preserve the satisfying assignments of all predicates in $\mathcal{P}$, and (2) the orbit of each satisfying assignment under the maps is rich enough that it doesn't admit any $(\mathbb{Z}, +)$ embedding.

Our main result with regard to Conjecture 1.1 is the following statement.

Theorem 1.1. *Let $\mathcal{P}$ be a collection of* MILDLY-SYMMETRIC 3*-ary predicates. Then there exists $\alpha_{\mathcal{P}}$ such that for all $\varepsilon > 0$:*

1. *Hardness: there is a dictatorship vs quasirandom test using $\mathcal{P}$ with perfect completeness and soundnesss $\alpha_{\mathcal{P}} + \varepsilon$.*
2. *Algorithm: there exists a polynomial-time algorithm that distinguishes between satisfiable instances of $\mathcal{P}$-CSP from instances of $\mathcal{P}$-CSP with value at most $\alpha_{\mathcal{P}}$.*

Organization: the rest of this introductory section is organized as follows. In Section 1.5 we discuss the algorithmic approach for CSPs, and in Section 1.6 we present our *hybrid algorithm*. In Section 1.7 we discuss the analysis of the hybrid algorithm and in Section 1.8 we discuss our main technical contribution, the *mixed invariance principle*. In Section 1.9 we discuss other related works.

1.5 Approximation Algorithms for CSPs

In this section we discuss two algorithmic techniques that are vital towards our hybrid algorithm, and are used in Raghavendra's theorem and in the dichotomy theorem.

1.5.1 Raghavendra's Algorithm: Semi-definite Programming Relaxations

For almost satisfiable CSPs, Raghavendra [46] showed that **semi-definite programming** (SDP) based algorithms give the optimal approximation algorithms (assuming the UGC). More specifically, we can write down a basic SDP relaxation of a given instance of Max-$\mathcal{P}$-CSP as shown in Figure 1. Here, $\mathcal{V}$ is the set of variables of the instance and $\mathcal{C}$ is a distribution over constraints of the instance (representing a weighted instance of CSP-$\mathcal{P}$), and for each $c \in \mathcal{C}$, $\mathcal{V}(c)$ is the set of variables appearing in c. For each variable $i \in \mathcal{V}$ and an alphabet symbol $a \in \Sigma$ the program has a vector $\boldsymbol{b}_{i,a}$ and additionally there is a global vector $\boldsymbol{b}_0$. We think of these vectors as describing a distribution over good assignments to the instance, and write down conditions corresponding to that.

The SDP-solution is then rounded, via a non-trivial rounding procedure, to an assignment to the variables. More precisely, given a solution to the SDP program that lives in dimension m, Raghavendra samples Gaussians $g^{(1)}, \ldots, g^{(R)} \in \mathbb{R}^m$ in which each coordinate is an independent standard Gaussian random variable, and produces jointly distributed Gaussians $z_{\ell,(i,a)} = \langle g^{(\ell)}, \boldsymbol{b}_{i,a} \rangle$. It is easy to see that the z's pairwise correlations match the inner products of the SDP solution vectors. Using these samples (as inputs), Raghavendra shows that any quasirandom function that performs well in the dictatorship tests yields a rounding function that satisfies (in expectation) at least s fraction of the constraints, where s is the soundness of the dictatorship test. We stress that in this context, quasirandom functions refer to low-degree functions in which all variables have small influence.

It is, therefore, natural to ask if semi-definite programming relaxation also gives the best approximation algorithm in the setting of Theorem 1.1. This turns out to be false as can be seen from the following CSP. Let $(G, \cdot)$ be a non-Abelian group and consider the problem 3-Lin$_G$. In this problem, we have variables $x_1, \dots, x_n$ that are supposed to be assigned values from G, and the constraints are of the form $x_i \cdot x_j \cdot x_k = c$ where $c \in G$ are constants. In [4] it is shown that the basic SDP for 3-LIN$_G$ has an integrality gap of 1 vs. $1/|G|$, meaning that an algorithm that only uses SDP rounding cannot achieve a better ratio than $1/|G|$ on satisfiable instances. However, there is an algorithm that achieves $1/|[G, G]|$ factor approximation on satisfiable instances, where $[G, G]$ is a commutator subgroup of G; see [6] for example.[4]

1.5.2 Gaussian Elimination

Another important algorithmic technique is **Gaussian elimination**, i.e., solving a system of linear equations over an Abelian group [22]. Gaussian elimination can at times be more powerful than semi-definite programming relaxations: using it one can decide whether a given system of linear equations over $\mathbb{F}_p$ is perfectly satisfiable or not, but SDPs fail to do so [30, 49]. Still, by itself it is rather weak, and it is not clear how to use it to obtain non-trivial approximation algorithms for problems such as Max-Cut. We remark that Gaussian elimination is not enough to check the satisfiability of bounded width P-CSPs [3], which otherwise are tractable using local-propagation algorithms [3].

1.5.3 Our Hybrid Algorithm

Our hybrid algorithm blends the two aforementioned algorithmic tools in a nontrivial way. We note that these two techniques have recently been used together to produce nontrivial algorithms in the area of *promise CSPs*. Indeed, therein a combination of semidefinite program/linear program (SDP/LP) and affine integer program (AIP) was used [13, 14, 19] to solve a few tractable cases. Similarly, SDP+AIP was shown [20] not to be enough to solve the approximate graph coloring problem. However, these prior works consider solving the SDP and the system of linear equations once and using these solutions to output the final decision. In contrast, our hybrid algorithm iteratively modifies the SDP program and the system of linear equations before coming up with the final decision.

1.6 A New Approximation Algorithm for Satisfiable CSPs

In this section we present our hybrid algorithm that will be used in the analysis of our dictatorship test.

Let $\mathcal{P}$ be a collection of 3-ary predicates, all of which are embeddable in a finite Abelian group G via the map $\sigma \colon \Sigma \to G$.[5] Fix an instance $\Upsilon = (\mathcal{V}, \mathcal{C})$ of Max-$\mathcal{P}$-CSP, where $\mathcal{V}$ is identified with the set $\{1, 2, \dots, n\}$, and $\mathcal{C}$ is a set of constraints on $\mathcal{V}$. The basic semidefinite programming relaxation of the instance is given on the left side of Figure 1. It consists of vectors $\{\boldsymbol{b}_{i,a}\}_{i \in \mathcal{V}, a \in \Sigma}$, distributions $\{\mu_c\}_{c \in \mathcal{C}}$ over the local assignments (i.e., on $\Sigma^{\mathcal{V}(c)}$, where $\mathcal{V}(c)$ denotes the tuple of variables in c) and a unit vector $\boldsymbol{b}_0$. The notation $\blacktriangle(Z)$ refers to the collection of all probability distributions over the set Z. Observe that this is indeed a relaxation: given a satisfying assignment $\alpha : \mathcal{V} \to \Sigma$, choose $\boldsymbol{b}_0$ to be some unit vector, and

[4] For many non-Abelian groups G, such as for the group $G = S_m$ for example, the size of the commutator subgroup $[G, G]$ is strictly smaller than the size of G. In particular, a $1/|[G, G]|$-approximation is strictly better than a $1/|G|$-approximation.

[5] For simplicity, we assume here that this is essentially the only group in which the predicates are embeddable and all the embedding maps are identical, i.e., $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$.

consider the vector assignment

$$\boldsymbol{b}_{i,a} = \begin{cases} \boldsymbol{b}_0 & \text{if } \alpha(i) = a, \\ \vec{0} & \text{otherwise.} \end{cases} \tag{1}$$

For every $c \in \mathcal{C}$, the distribution μ_c is set to be supported on the assignment $\alpha|_{\mathcal{V}(c)}$, which by definition satisfies the constraint c. It can be easily observed that all the constraints (1)–(4) are satisfied, and furthermore, the objective value of this vector solution is 1, as $c(x) = 1$ if $x \sim \mu_c$ for every $c \in \mathcal{C}$.

Semidefinite Program	System of linear equations over G
maximize $\mathop{\mathbb{E}}_{c\in\mathcal{C}} \mathop{\mathbb{E}}_{x\sim\mu_c} [c(x)]$ subject to (1) $\langle \boldsymbol{b}_{i,a}, \boldsymbol{b}_{j,b} \rangle = \Pr_{x\sim\mu_c}[x_i = a, x_j = b]$ $c \in \mathcal{C}, \quad i,j \in \mathcal{V}(c), \quad a,b \in \Sigma,$ (2) $\langle \boldsymbol{b}_{i,a}, \boldsymbol{b}_0 \rangle = \\|\boldsymbol{b}_{i,a}\\|_2^2, \forall i \in \mathcal{V}, a \in \Sigma,$ (3) $\\|\boldsymbol{b}_0\\|_2^2 = 1,$ (4) $\mu_c \in \blacktriangle(\Sigma^{\mathcal{V}(c)}), c \in \mathcal{C}.$	Find $\vartheta : \{y_1, y_2, \ldots, y_n\} \to (G, +)$ such that $\forall c \in \mathcal{C}$ with $\mathcal{V}(c) = (i_1, i_2, i_3),$ we have $\vartheta(y_{i_1}) + \vartheta(y_{i_2}) + \vartheta(y_{i_3}) = 0_G.$

Figure 1: A semidefinite programming formulation and a system of linear equations for a given Max-P-CSP instance $\Upsilon = (\mathcal{V}, \mathcal{C})$.

We also write down a system of linear equations over an Abelian group G, given on the right-side of Figure 1. Note that if the instance Υ is satisfiable, then the system of linear equations over G also has solutions. This follows from the definition of Abelian embeddability (Definition 1.2): take any satisfying assignment $\alpha \in \Sigma^n$ and assign $\vartheta(y_i) = \sigma(\alpha_i)$ for all $i \in [n]$.

We modify the semi-definite program and the system of linear equations iteratively as follows. At every step, we work with an SDP solution where for every satisfying assignment $\alpha : \mathcal{V} \to \Sigma$ to the instance Υ, $\mu_c(\alpha|_{\mathcal{V}(c)}) > 0$.[6] In other words, every satisfying assignment to the instance 'survives' in the SDP solution.

- For a constraint $c \in \mathcal{C}$ we say that $(g_1, g_2, g_3) \in G^3$ is SDP-unattainable if $(\sigma(a_1), \sigma(a_2), \sigma(a_3)) \neq (g_1, g_2, g_3)$ for all $(a_1, a_2, a_3) \in \mathsf{supp}(\mu_c)$. Consider an SDP-unattainable tuple (g_1, g_2, g_3) not in a subgroup (of G^3) generated by the SDP-attainable tuples. We modify the system of linear equations by eliminating (g_1, g_2, g_3) from being a possible setting to the variables corresponding to c while preserving all the SDP-attainable tuples in a solution.

- For every constraint $c \in \mathcal{C}$ and an assignment $\boldsymbol{a} = (a_1, a_2, a_3) \in \Sigma^3$ if no solution to system of linear equations assigns $(\sigma(a_1), \sigma(a_2), \sigma(a_3))$ to the variables of c, then add a constraint $\mu_c(\boldsymbol{a}) = 0$ to the SDP formulation.

It is easy to see that the process ends in polynomially many steps. The algorithm accepts an instance if the SDP value is 1 at the end of the above procedure. Since we ensured that every satisfying assignment to the instance 'survives' in the SDP solution, for a satisfying assignment α, the vector solution given in (1)) will be a feasible solution to the final SDP. Therefore, the algorithm always accepts satisfiable instances of Max-$\mathcal{P}$-CSP.

[6]This can be achieved in polynomial time, see Lemma 7.9.

As for the soundness of the algorithm, let $\mathcal{S}$ be the collection of all Max-$\mathcal{P}$-CSP instances where the algorithm accepts, and define $s = \inf_{\Upsilon \in \mathcal{S}} \mathbf{val}(\Upsilon)$. To complete the proof of Theorem 1.1 we show an 1 vs. $s + o(1)$ dictatorship test. We explain this in Section 1.7.

1.7 The Dictatorship Test

In this section we explain how to use an integrality gap of the above hybrid algorithm to construct a dictatorship test. Throughout, we have an integrality gap Υ with $\mathbf{val}(\Upsilon) = s + o(1)$. To analyze the dictatorship test, it will be useful for us to consider the following approximation algorithm (we stress that it is specific for the instance Υ):

1. *Solving the programs.* Solve the final SDP program such that the objective value of the solution is 1. As there are $n|\Sigma| + 1$ vectors, we can assume without loss of generality that these vectors live in $\mathbb{R}^m$ for $m \leqslant n|\Sigma| + 1$. Denote by $S \subseteq G^n$ be the set of all the satisfying assignments ϑ for this system of linear equations, and note that S is a subspace.

2. *Setting up rounding functions and sampling.* For a constant $R \geqslant 1$ to be chosen as large enough, we fix a rounding function $f : \mathbb{R}^{R|\Sigma|} \times G^R \to \blacktriangle(\Sigma)$, such that $f(\boldsymbol{z}, \boldsymbol{w})$ gives a probability distribution on Σ. We sample R Gaussian vectors $\boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}, \ldots, \boldsymbol{g}^{(R)} \in \mathbb{R}^m$ where for each $j \in [m]$ and $\ell \in [R]$, $g_j^{(\ell)}$ is distributed according to the standard normal variable. We also sample R uniformly random satisfying assignments over G from the set S, which we denote by $\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)}, \ldots, \boldsymbol{\sigma}^{(R)}$.

3. *SDP rounding component.* We first take the inner product of the Gaussian vectors with the SDP vectors corresponding to the variable i. More formally, let $z_{i,(\ell,a)} = \langle \boldsymbol{b}_{i,a}, \boldsymbol{g}^{(\ell)} \rangle$, for every $i \in \mathcal{V}$, $\ell \in [R]$ and $a \in \Sigma$. This gives a vector $\boldsymbol{z}_i = (z_{i,(\ell,a)})_{\ell \in [R], a \in \Sigma} \in \mathbb{R}^{R|\Sigma|}$ for each variable $i \in \mathcal{V}$.

4. *Gaussian elimination component.* Taking the assignments we sampled form S, we create a string from $\boldsymbol{w}_i \in G^R$ for each variable $i \in \mathcal{V}$, where $w_i = (\boldsymbol{\sigma}^{(\ell)}(y_i))_{\ell \in [R]}$.

5. *Outputting an assignment.* For each $i \in \mathcal{V}$, sample $A(i)$ according to the distribution $f(\boldsymbol{z}_i, \boldsymbol{w}_i)$.

1.7.1 A Dictatorship Test with Imperfect Completeness

Towards the construction of the dictatorship test, it is helpful to first analyze a variant of the above algorithm that only solves the SDP program and applies a rounding scheme (or alternatively, that applies a rounding scheme that ignores its $\boldsymbol{w}$-component). This is the setting in Raghavendra's theorem, and he uses it to construct a dictatorship test with completeness $1 - \varepsilon$ and soundness $s + o(1)$ that uses the collection $\mathcal{P}$, for all $\varepsilon > 0$. Denote by $(\boldsymbol{b}, \boldsymbol{\mu})$ the SDP solution with value 1 and consider the following test to check if a given function $f : \Sigma^R \to \Sigma$ is a dictator function or far from a dictator function:

1. Sample $(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k)$ as follows:
   (a) Sample a random constraint $c \in \mathcal{C}$ and let $P \in \mathcal{P}$ be the predicate it uses.
   (b) For each $i \in [R]$, sample $(y_{1,i}, y_{2,i}, \ldots, y_{k,i})$ according to μ_c independently.
   (c) For each $i \in [R]$, with probability ε resample $(y_{1,i}, y_{2,i}, \ldots, y_{k,i})$ from Σ^k uniformly and independently.

2. Check if $P(f(\boldsymbol{y}_1), f(\boldsymbol{y}_2), \ldots, f(\boldsymbol{y}_k)) = 1$.

Completeness: If f is a dictator function, say $f(\boldsymbol{y}) = y_j$ for some $j \in [R]$, then the probability that the test passes is as follows:

$$\begin{aligned}
\Pr[\text{Test passes}] &= \mathop{\mathbb{E}}_{(c,P)}\left[\mathop{\mathbb{E}}_{(\boldsymbol{y}_1,\boldsymbol{y}_2,\dots,\boldsymbol{y}_k)}[P(f(\boldsymbol{y}_1), f(\boldsymbol{y}_2), \dots, f(\boldsymbol{y}_k))]\right] \\
&\geqslant (1-\varepsilon) \mathop{\mathbb{E}}_{(c,P)}\left[\mathop{\mathbb{E}}_{(y_{1,j},y_{2,j},\dots,y_{k,j})\sim\mu_c}[P(y_{1,j}, y_{2,j}, \dots, y_{k,j})]\right] \\
&= (1-\varepsilon)\cdot 1,
\end{aligned}$$

where the last equality follows from the fact that the SDP value is 1 and hence μ_c is supported on $P^{-1}(1)$ for every $c \in \mathcal{C}$.

Soundness: Let us now analyze the soundness of the test. Towards this, suppose $f\colon \Sigma^R \to \Sigma$ is a quasirandom function.[7] Our goal is to show that the dictatorship test accepts with probability at most $s + o(1)$. First, we express the test passing probability as follows:

$$\Pr[\text{Test passes}] = \mathop{\mathbb{E}}_{(c,P)}\left[\mathop{\mathbb{E}}_{(\boldsymbol{y}_1,\boldsymbol{y}_2,\dots,\boldsymbol{y}_k)}[P(f(\boldsymbol{y}_1), f(\boldsymbol{y}_2), \dots, f(\boldsymbol{y}_k))]\right].$$

Fix c and P and focus on the inner expectation. By expressing P in terms of its multi-linear extension, the expectation above can be written as a linear combination of expectations of the form

$$\mathbb{E}\left[\prod_{i\in S} F_i(\boldsymbol{y}_i)\right], \tag{2}$$

where $S \subseteq [k]$ and $F_i : \Sigma^R \to \{0,1\}$ is an indicator function of the form $F_i(\boldsymbol{y}) = 1_{f(\boldsymbol{y})=a}$ for some $a \in \Sigma$. An important point is that the analysis of the test boils down to analyzing the expectation of product of functions where the output of each function is bounded. From this point onwards, Raghavendra [46] argument proceeds as follows (for the simplicity of presentation we are omitting from the description many important technical details, such as how to keep the functions we work with bounded):

I. For $\varepsilon > 0$, as the distribution on $(y_{1,j}, y_{2,j}, \dots, y_{k,j})$ has full support, a result of [41] asserts the expectation (2) can be approximately computed by only considering the low-degree part $F_i^{\leqslant d}$ of the corresponding F_is, that is,

$$\mathbb{E}\left[\prod_{i\in S} F_i(\boldsymbol{y}_i)\right] \approx \mathbb{E}\left[\prod_{i\in S} F_i^{\leqslant d}(\boldsymbol{y}_i)\right], \tag{3}$$

for some $d = O_\varepsilon(1)$.

II. Using the fact that $F_i^{\leqslant d}$s are low-degree and far from dictator functions, the invariance principle of [42] states that the inputs $F_i^{\leqslant d}$ can be "replaced" by correlated Gaussian random variables that have matching pairwise correlation to the y_i's, i.e.,

$$\mathbb{E}\left[\prod_{i\in S} F_i^{\leqslant d}(\boldsymbol{y}_i)\right] \approx \mathbb{E}\left[\prod_{i\in S} F_i^{\leqslant d}(\boldsymbol{g}_i)\right]. \tag{4}$$

[7] By that, we mean that for all $a \in \Sigma$, the function $1_{f(x)=a}$ is quasirandom.

III. The process of sampling the correlated Gaussian can be simulated by the SDP rounding component as stated above. Thus, using the above observations, an algorithm can generate Gaussian samples such that

$$\left[\mathop{\mathbb{E}}_{(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_k)}\left[P(f(\boldsymbol{y}_1),f(\boldsymbol{y}_2),\ldots,f(\boldsymbol{y}_k))\right]\right]\approx\left[\mathop{\mathbb{E}}_{(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_k)}\left[P(\tilde{f}(\boldsymbol{g}_1),\tilde{f}(\boldsymbol{g}_2),\ldots,\tilde{f}(\boldsymbol{g}_k))\right]\right], \quad (5)$$

for some function $\tilde{f}$. This means that in expectation, the randomized strategy above satisfies at least $\Pr[\text{Test passes}]-o(1)$. On the other hand, $\mathbf{val}(\Upsilon)\leqslant s+o(1)$, so overall we get that $\Pr[\text{Test passes}]\leqslant s+o(1)$ as required.

Note that in the analysis above, even though we started with an integrality gap with perfect completeness, the resulting dictatorship test has completeness $1-\varepsilon$. Tracing this back, the only place in the proof that the fact that $\varepsilon>0$ was used is in (3). In words, that equality asserts that for a k-ary distribution μ, if we want to measure the correlation of $f_1(y_1),\ldots,f_k(y_k)$ where $(y_1,\ldots,y_k)\sim\mu^{\otimes R}$ and $f_i\colon\Sigma^R\to\mathbb{R}$, then this correlation only comes from the low-degree parts of $f_1,\ldots,f_k$. This fact holds if the support of the distribution μ is Σ^k, and more generally when the distribution μ is connected in the sense of [41]. Alas, this fact is false for general distributions.

1.7.2 A Dictatorship Test with Perfect Completeness

To design a dictatorship test with completeness 1 via the above paradigm we are forced to set $\varepsilon=0$, meaning that the distributions arising in (3) are arbitrary. This means that the equality in (3) is no longer true, and it is unclear how to proceed with the analysis of the dictatorship test.

This is the point where the works [7, 9, 10, 11] enter the picture. The goal in these works is to understand what sort of functions may contribute to the left hand side of (3) under only very mild assumptions on the input distribution. This goal was partially achieved in [11], and we now explain how we use that result to proceed with the analysis of the dictatorship test. Throughout the rest of the discussion, we fix $k=3$ and $\varepsilon=0$ in the above dictatorship test.

Decoding correlations and fixing (3)**:** the result of [9] asserts that if μ_c has no Abelian embeddings, then (3) continues to hold, and so the analysis above proceeds in the same way. The result of [11] is a strengthening of it, asserting that if μ_c admits Abelian embeddings but no $(\mathbb{Z},+)$ embeddings, then functions F_1,F_2,F_3 for which the left hand side of (3) is non-negligible must arise from characters of Abelian groups and low-degree functions. More precisely, that result shows that there is a finite Abelian group G, a character $\chi\in\widehat{G}^{\otimes R}$, a map $\sigma\colon\Sigma\to G$ and a low-degree function $L\colon\Sigma^R\to\mathbb{R}$ with 2-norm 1 such that

$$\mathop{\mathbb{E}}_{(\boldsymbol{y}_1,\boldsymbol{y}_2,\boldsymbol{y}_3)\sim\mu_c^{\otimes R}}\left[F_1(\boldsymbol{y}_1)\chi(\sigma(\boldsymbol{y}_1))L(\boldsymbol{y}_1)\right]\geqslant\Omega(1).$$

In this paper, we deduce a list-decoding version of this statement, roughly making the following assertion. We can find functions $G_1,G_2,G_3\colon\Sigma^R\to\mathbb{R}$ that are each a sparse combination of functions of the form $\chi\circ\sigma\cdot L$ for a character χ, $\sigma\colon\Sigma\to G$ and low-degree function L such that a correct version of (3) becomes

$$\mathbb{E}\left[\prod_{i\in S}F_i(\boldsymbol{y}_i)\right]\approx\mathbb{E}\left[\prod_{i\in S}G_i(\boldsymbol{y}_i)\right], \quad (6)$$

for any $S\subseteq\{1,2,3\}$. [8]

[8] We remark that strictly speaking, the description of G_1,G_2,G_3 is incorrect, and they are only close in L_2-distance to functions of this form. Our argument requires that G_1,G_2,G_3 are $O(1)$-bounded, and this results in significant technical complications.

Generalization of the invariance principle and fixing (4)**:** with (6) in hand, the analysis of the dictatorship test would be done provided an algorithm could generate samples that "fool" the functions G_i on the right hand. The invariance principle of [42] asserts that low-degree functions with small influences are "fooled" by Gaussian samples, which we are able to produce using the semi-definite programming relaxation. However, Gaussian samples fail to fool characters. This is where the Gaussian elimination part of the algorithm enters the picture: we argue that characters are fooled by the solutions to the linear system of equations in the hybrid algorithm. Expressing

$$G_1(\boldsymbol{y}_1) = \sum_{j=1}^{W} \chi_j(\sigma(\boldsymbol{y}_1)) L_j(\boldsymbol{y}_1),$$

this leads us to consider the function $\tilde{G}_1 \colon \mathbb{R}^{|\Sigma|R} \times G^R \to \mathbb{R}$ defined by

$$\tilde{G}_1(\boldsymbol{g}_1, \boldsymbol{\sigma}_1) = \sum_{j=1}^{W} \chi_j(\boldsymbol{\sigma}_1) L_j(\boldsymbol{g}_1)$$

A priori, it is not clear what is the relation between G_1 and $\tilde{G}_1$. We have split the input $\boldsymbol{y}_1$ of G_1 into two independent samples from Gaussian space and from the set of solutions to the linear system. We show however, that if all L_j have small influences, then the functions G_1 and $\tilde{G}_1$ are close in a sense that suffices for our purposes (this requires some features from the set of characters χ_j appearing in G_1 that we are able to ensure). This is what we refer to by the "mixed invariance principle", because the discrete probability distribution of $\boldsymbol{y}_1$ is replaced by a mix of a Gaussian distribution and a distribution arising from a solution for a system of linear equations. We defer a more formal statement of the mixed invariance principle to Section 1.8 below.

Algorithmically, we are able to generate inputs to $\tilde{G}_1, \tilde{G}_2, \tilde{G}_3$ by solving the hybrid algorithm, allowing us to replace (4) with

$$\mathop{\mathbb{E}}_{(\boldsymbol{y}_1,\boldsymbol{y}_2,\boldsymbol{y}_3)\sim\mu_c^R}\left[\prod_{i\in S} G_i(\boldsymbol{y}_i)\right] \approx \mathop{\mathbb{E}}_{(\boldsymbol{g},\boldsymbol{\sigma})}\left[\prod_{i\in S} \tilde{G}_i(\boldsymbol{g}_i, \boldsymbol{\sigma}_i)\right] \tag{7}$$

for all $S \subseteq \{1,2,3\}$, and the rest of the arguments proceeds in the same way. In the end we get an analog of (5) wherein the inputs to the function $\tilde{f}$ are generated by the rounding algoirthm presented in the begining of the section. Hence, we showed that the probability the test accepts is at most $\mathbf{val}(\Upsilon) + o(1) \leqslant s + o(1)$.

1.7.3 On the MILDLY-SYMMETRIC Assumption

We finish this section by discussing the MILDLY-SYMMETRIC assumption, and in particular where it is used in the above analysis. To carry out the above argument, our only requirement is that the distributions μ_c arising in the solution of the SDP part of the hybrid algorithm are distribution for which the result of [11] applies. Therein, an inverse theorem holds for all 3-ary distributions that do not admit $(\mathbb{Z},+)$ embeddings, and it stands to reason that result should hold for the more general class of pair-wise connected distributions (in the sense defined therein).

Therefore, if one is interested in a particular family of predicates $\mathcal{P}$ which is not MILDLY-SYMMETRIC , then in principle one could still solve the SDP program, and if the resulting local distributions μ_c do not admit any $(\mathbb{Z},+)$ embeddings, then our argument still goes through and the hybrid algorithm above works.

The conditions stated in MILDLY-SYMMETRIC are a relatively elegant way of ensuring that the local distributions μ_c have no $(\mathbb{Z},+)$ embeddings, and hence Theorem 1.1 is stated in this way. More precisely, we argue that if we have an SDP solution with value 1 for a collection $\mathcal{P}$ that is MILDLY-SYMMETRIC , then can modify it to get an SDP solution with value 1 in which the local distributions μ_c have no $(\mathbb{Z},+)$ embeddings.

1.7.4 On Search vs Decision Algorithm

Our main result, Theorem 1.1, gives a polynomial time algorithm that distinguishes satisfiable instances of Max-$\mathcal{P}$-CSP from instances with value at most $\alpha_{\mathcal{P}}$. The approximation guarantee is matched with the soundness of the corresponding dictatorship test. At this point, we do not know how to convert this decision algorithm into a search algorithm. Namely, an algorithm that given a satisfiable instance as an input, *finds* an assignment with value at least $\alpha_{\mathcal{P}}$ in polynomial time. We believe that modifications of the techniques from Raghavendra-Steurer [47] and Khot, Tulsiani, and Worah [38] could provide the search algorithm, and we leave this as an open problem for the future.

1.8 The Mixed Invariance Principle

In this section we discuss the mixed invariance principle. Let μ be a distribution over Σ^3 that has no $(\mathbb{Z},+)$ embeddings, and let $f_1\colon(\Sigma^n,\mu_1^{\otimes n})\to\mathbb{C}$, $f_2\colon(\Sigma^n,\mu_2^{\otimes n})\to\mathbb{C}$, $f_3\colon(\Sigma^n,\mu_3^{\otimes n})\to\mathbb{C}$ be 1-bounded functions. The goal of the mixed invariance principle is to study expectations of the form

$$\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}}[f_1(x)f_2(y)f_3(z)] \tag{8}$$

and relate them to expectations of related functions over different domains. The above expression should be compared to the left hand side in (6). As explained therein, in this paper we show that we can find functions $\tilde{f}_1,\tilde{f}_2,\tilde{f}_3$ that are 1-bounded, and additionally are close in L_2-distance to sparse sums of functions that are product of characters over some finite Abelian group G and low-degree functions, such that

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}}[f_1(x)f_2(y)f_3(z)]-\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}}\left[\tilde{f}_1(x)\tilde{f}_2(y)\tilde{f}_3(z)\right]\right|=o(1). \tag{9}$$

For the sake of simplicity assume that $\tilde{f}_1,\tilde{f}_2,\tilde{f}_3$ are themselves this sparse combinations (as opposed to just close to them; see Section 2 for a more formal discussion), and write

$$\tilde{f}_1(x)=\sum_{P\in\mathcal{P}_1}P(\sigma(x))L_P(x) \tag{10}$$

and similarly for $\tilde{f}_2$ and $\tilde{f}_3$. Here, $\mathcal{P}_1$ is a set of characters and $\sigma\colon\Sigma\to G$ is some map.[9] Note that in the case that $|\mathcal{P}_1|=1$ and the only element in $\mathcal{P}$ is the constant 1 function, the function $\tilde{f}_1$ is precisely the low-degree part of f_1. Towards the statement of the invaraince principle we decouple the inputs to the characters-part of $\tilde{f}_1$ and the low-degree parts of $\tilde{f}_1$ and define

$$\tilde{f}^1_{\mathsf{decoupled}}(x,x')=\sum_{P\in\mathcal{P}_1}P(\sigma(x))\cdot L_P(x'), \tag{11}$$

and we similarly define $\tilde{f}^2_{\mathsf{decoupled}}$ and $\tilde{f}^3_{\mathsf{decoupled}}$.

[9]We remark that this decomposition is not unique, and often, for our applications, we work with a decomposition with certain extra properties.

Definition 1.4. *We say that a function $\tilde{f}_{\mathsf{decoupled}}$ of the* (11) *has τ-small shifted low-degree influences if for every $j \in [n]$ and every $P \in \mathcal{P}_1$ it holds that the j^{th} influence of L_P is at most τ.*

We are now ready to state our mixed invariance principle.

Theorem 1.2. *(Informal) Suppose μ is a distribution over Σ^3 that has no $\mathbb{Z}$-embedding. There exists an Abelian group G and embeddings $\sigma_i : \Sigma \to G$ such that for 1-bounded functions $f_1, f_2, f_3 : \Sigma^n \to \mathbb{C}$, if the corresponding decoupled functions $\tilde{f}^i_{\mathsf{decoupled}}(x, x')$ have τ-small shifted low-degree influences, then by letting $\tilde{F}_i$ be $\tilde{f}^i_{\mathsf{decoupled}}$ except we replace L_Ps with their corresponding multi-linear expansions, we have*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} [f_1(x)f_2(y)f_3(z)] - \mathop{\mathbb{E}}_{\substack{(x,y,z)\sim\mu^{\otimes n}, \\ (g_x,g_y,g_z)\sim\mathcal{G}^{\otimes n}}} \left[\tilde{F}_1(x, g_x)\tilde{F}_2(y, g_y)\tilde{F}_3(z, g_z)\right] \right| \leqslant \xi(\tau), \tag{12}$$

where $\xi(\tau) \to 0$ as $\tau \to 0$. Here $\mathcal{G}$ has the same pairwise correlations as the distribution μ.

Remark 1.5. *In the above theorem, we use a non-standard notion of the functions that are 'far from the dictator functions.' At this point, we do not know how to use this notion of dictatorship test in the actual (conditional) NP-hardness result (starting with the Rich 2-to-1 Games Conjecture [16]). We believe that new techniques need to be designed for this goal, and we leave it as an open problem for future research.*

1.9 Further Remarks and Other Related Work

As discussed above, we view the condition that the distribution μ has no $(\mathbb{Z}, +)$ embedding as being of technical nature, and expect the above argument to have an analog as long as μ is pairwise connected. Here, a 3-ary distribution over triples $(x, y, z) \in \Sigma^3$ is pairwise connected if the bipartite graph $G_{x,y} = ((\Sigma, \Sigma), E)$ where $E = \{ (x, y) \mid \exists z, (x, y, z) \in \mathsf{supp}(\mu)\}$, is connected, and similarly the analogously defined bipartite graphs $G_{x,z}$ and $G_{y,z}$ are also connected.

We expect that going beyond pairwise connected distribution would be a significant challenge. The class of distributions that are not pairwise connected is very rich, and an inverse theorem for this class is likely to imply effective bounds for the density Hales-Jewett theorem for combinatorial lines of length 3 [25, 27, 45], a prominent open problem in combinatorics. More generally and for larger arities k, this class of distributions captures problems such as multi-dimensional Szemerédi theorems [26], for which only ergodic theoretic proofs are known.

2 Techniques

Is this section we give an outline to the proof of Theorem 1.2. The formal proof appears in Sections 4, 5, 6.

2.1 List Decoding

To begin the discussion, consider an expectation of the form (8), and suppose it is non-negligible in absolute value. We already know, using results of Mossel [41], that if the distribution μ is connected, then each one of the functions f_i must be correlated with a low-degree function. But what happens in the case that μ is not connected? This question was asked in [7, 9, 11], wherein under the condition that μ does not admit $(\mathbb{Z}, +)$ embedding, some conclusion regarding the structure of the functions f_i was made. More precisely, the main result of [11] asserts that there is a constant size Abelian group G_{master} and embeddings $\sigma, \gamma, \phi\colon \Sigma \to G_{\mathsf{master}}$ of μ into G_{master}, such that if f_1, f_2, f_3 are $O(1)$-bounded functions satisfying

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}}\left[f_1(x)f_2(y)f_3(z)\right]\right| \geqslant \varepsilon,$$

then function f_1 (and similarly f_2, f_3) must be correlated with a function of the form $\chi \circ \sigma \cdot L$ where $\chi \in \widehat{G}^n_{\mathsf{master}}$ and L is a function of degree $O_\varepsilon(1)$ and bounded 2-norm. In other words, f_1 is correlated with a single function that appears on the right hand side of (10). Optimistically, one may hope that $\chi \circ \sigma \cdot L$ "explains" all of the magnitude of $\mathbb{E}_{(x,y,z)\sim\mu^{\otimes n}}\left[f_1(x)f_2(y)f_3(z)\right]$ in the sense that f_1 may be replaced by it without changing the value of the expectation by much. Unfortunately, this is incorrect; for once, f_1 could be correlated with numerous functions of this type that are very far from each other. Besides, even if this was true, there is an additional issue: we do not wish to replace the function f_1 by a function that is not bounded (at least not at this point of the argument).

A common strategy for addressing the first issue is by appealing to the notion of list decoding. As a first attempt, one may hope to collect all of the functions of the form $\chi \circ \sigma \cdot L$ that are correlated with f_1, and then replace f_1 with a weighted sum of them according to their correlation. This would have worked had that collection of functions been pairwise orthogonal. This is not true in our case, though. As a second attempt, one may find a single function $\chi \circ \sigma \cdot L$ that is correlated with f_1, and then iterate the argument with $f_1' = f_1 - \alpha \cdot \chi \circ \sigma \cdot L$ where $\alpha = \langle f_1, \chi \circ \sigma \cdot L\rangle$. This attempt also fails, this time due to the fact that f_1' may not be $O(1)$-bounded, and hence the result of [11] no longer applies.

To resolve this issue, we take inspiration from the invariance principle of [42, 41] and from the inverse Gowers' norms literature [29]. In the former, instead of harsh truncations, one performs "soft truncations" by applying a noise operator that gives a bounded function that is close in L_2 distance to the low-degree part of the function. In the latter, one defines an averaging operator with respect to a sigma-algebra induced by the collection of functions that were already found to be correlated with f_1. We combine these two solutions, and the bulk of our effort is devoted to showing it works.

2.2 The Noise Operator, the Regularity Lemma, and the Approximating Formula

Let ν be the marginal distribution of μ on the first coordinate, and consider the function $f_1\colon (\Sigma^n, \nu^{\otimes n}) \to \mathbb{C}$ above. Let $\mathcal{P} = \{P_1, \ldots, P_r\colon \Sigma^n \to H\}$ be a collection of functions into some discrete domain H, and let $\varepsilon > 0$. We define the function $\mathrm{T}_{\mathcal{P},1-\varepsilon} f_1\colon \Sigma^n \to \mathbb{C}$ in the following way. First, for each $x \in \Sigma^n$, we define the distribution $\mathrm{T}_{\mathcal{P},1-\varepsilon} x$, wherein a sample $y \sim \mathrm{T}_{\mathcal{P},1-\varepsilon} x$ is generated as follows: sample $I \subseteq [n]$ by including each element in it independently with probability $1-\varepsilon$, then sample $y \sim \nu$ conditioned on $y_I = x_I$ and $P_i(y) = P_i(x)$ for all $i = 1, \ldots, r$ (see Definition 5.1). The function $\mathrm{T}_{\mathcal{P},1-\varepsilon} f_1$ is then defined by

$$\mathrm{T}_{\mathcal{P},1-\varepsilon} f_1(x) = \mathop{\mathbb{E}}_{y\sim \mathrm{T}_{\mathcal{P},1-\varepsilon} x}\left[f_1(y)\right].$$

To get some intuition to the operator $\mathrm{T}_{\mathcal{P},1-\varepsilon}$ we recommend thinking of ε as very small. We begin by noting that the distribution ν is a stationary distribution of $\mathrm{T}_{\mathcal{P},1-\varepsilon}$, and that functions f for which $f(x) = g(P_1(x), \ldots, P_r(x))$ for some $g\colon H^r \to \mathbb{R}$ are eigenfunctions with eigenvalue 1. We also note (though this is a bit more tricky to prove) that low-degree functions are nearly eigenfunctions of $\mathrm{T}_{\mathcal{P},1-\varepsilon}$ of eigenvalue 1, and more precisely that $\|(I - \mathrm{T}_{\mathcal{P},1-\varepsilon})L\|_2 = o(\|L\|_2)$ provided that ε is small compared to the degree of L. In other words, when constructed for an appropriate collection $\mathcal{P}$, the operator $\mathrm{T}_{\mathcal{P},1-\varepsilon}$ almost doesn't affect functions of the form of the right-hand side of (10), and therefore it may be useful to detect such structures. We show that this is indeed the case.

2.2.1 The Regularity Lemma

With the new noise operator $\mathrm{T}_{\mathcal{P},1-\varepsilon}$ in hand, we may attempt to execute the idea of the list decoding argument in a different way. More specifically, starting with $f = f_1$, $\mathcal{P} = \emptyset$ and $\tilde{f}_1 = 0$, once we find a function of the form $\chi \circ \sigma \cdot L$ that f is correlated with, we insert $P = \chi \circ \sigma$ into $\mathcal{P}$, take $\tilde{f}_1 = \mathrm{T}_{\mathcal{P},1-\varepsilon} f$ and proceed the argument on $f = f_1 - \tilde{f}_1$. In the next step, we will find a new $\chi' \circ \sigma \cdot L'$ correlated with the updated f, insert $P' = \chi' \circ \sigma$ to $\mathcal{P}$, update $\tilde{f}_1 = \mathrm{T}_{\mathcal{P},1-\varepsilon} f$ and proceed the argument on $f = f_1 - \tilde{f}_1$. When the argument terminates, we will have that

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} \left[(f_1 - \tilde{f}_1)(x) f_2(y) f_3(z) \right] \right| = o(1),$$

meaning we can replace f_1 with $\tilde{f}_1$. Our proof goes along these lines, but to make it go through a great deal of care is required. The key issue is that if we inspect the previous inductive process closely, instead of updating $\tilde{f}_1$ at each step, we should have subtracted a new noised function. For instance, in the second step we would have subtracted $\mathrm{T}_{\{P'\},1-\varepsilon}(f - \mathrm{T}_{\{P\},1-\varepsilon} f)$ from $f - \mathrm{T}_{\{P\},1-\varepsilon} f$ to get the function

$$\tilde{f}_1 = f - \mathrm{T}_{\{P\},1-\varepsilon} f - \mathrm{T}_{\{P'\},1-\varepsilon}(f - \mathrm{T}_{\{P\},1-\varepsilon} f),$$

and at each iteration, the function we inspect gets even more complicated. After a large number of steps, the function we end up with is no longer $O(1)$-bounded, in which case the the iterative process gets stuck. Fortunately, we show that instead of doing this, one may update the approximating function $\tilde{f}_1$ as explained above. We remark that in the formal argument we are required to modify the noise rate ε at each step, which contributes to several technical challenges. Once this process is done appropriately, we find a function $\tilde{f}_1 = \mathrm{T}_{\mathcal{P},1-\varepsilon} f$ where $|\mathcal{P}|$ is constant and ε is bounded away from 0, such that

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} [f_1(x) f_2(y) f_3(z)] - \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} \left[\tilde{f}_1(x) f_2(y) f_3(z) \right] \right| = o(1).$$

Similarly, we show that one may replace f_2 and f_3 with noisy versions of them.

2.2.2 The Approximating Formula

The next step in the proof is to show that a function of the form $\mathrm{T}_{\mathcal{P},1-\varepsilon} f_1$ may be approximated by a function of the form of the right-hand side of (10) in L_2 distance.[10] We establish such a formula by fairly direct calculations, starting with

$$\mathrm{T}_{\mathcal{P},1-\varepsilon} f(x) = \frac{1}{A(x)} \mathbb{E}_{I \subseteq_{1-\varepsilon} [n], y\sim\nu} \left[f(y) 1_{P_i(y) = P_i(x)\ \forall i} \,\middle|\, y_I = x_I \right],$$

where $A(x) = \Pr_{I \subseteq_{1-\varepsilon}[n], y\sim\nu} [P_i(y) = P_i(x)\ \forall i \mid y_I = x_I]$. We show that on most inputs, both $1/A(x)$ and the expectation above are close to functions of the form of the right-hand side of (10). The intuition here is that first, the expectation above looks like the application of the standard noise operator on some function. Second, on inputs where $A(x)$ is bounded away from 0 (which we prove are almost all inputs), we have that

$$\frac{1}{A(x)} = \sum_{j=0}^{d} (1 - A(x))^d + 2^{-\Theta(d)},$$

and one may easily expand the definition of $A(x)$ and write down an explicit formula for it of the form (10).

[10] We remark that in fact, such a formula is necessary for us even to make the previous regularity lemma go through.

2.3 Decoupling the Abelian Part and the Low Degree Part

With the previous steps done, we managed to replace each one of the functions f_1, f_2, f_3 in the expectation (8) with the functions $\mathrm{T}_{\mathcal{P},1-\varepsilon}f_1$, $\mathrm{T}_{\mathcal{Q},1-\varepsilon}f_2$ and $\mathrm{T}_{\mathcal{R},1-\varepsilon}f_3$ respectively. Additionally, we have approximate formula for each one of these functions that seem to have the desired form. Namely, ignoring the aspect of this approximation for a moment, we have that

$$\mathrm{T}_{\mathcal{P},1-\varepsilon}f_1(x) = \sum_{P\in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P(x)L_P(x)$$

where $\mathsf{spn}_{\mathbb{N}}(\mathcal{P})$ consists of all possible products of functions from $\mathcal{P}$, and each one of the functions L_P is a low-degree function with bounded 2-norm. We would now like to decouple the input x into two copies, one of which will be fed to the functions P and the other one will be fed to the low-degree functions L_P (indeed, this is the only difference between the formulas in (10) and in (11)). It turns out that provided that the collection $\mathcal{P}$ is *well separated*, there is a coupling $X, (X', X'')$ between $\nu^{\otimes n}$ and $\nu^{\otimes n} \times \nu^{\otimes n}$ such that

$$\sum_{P\in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P(X)L_P(X) = \sum_{P\in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P(X')L_P(X'') + o(1)$$

in L_2 distance. By well-separated, we mean that besides the trivial all 1 function coming from the empty product, every function in $\mathsf{spn}_{\mathbb{N}}(\mathcal{P})$ has a high Fourier analytic degree; in fact, this degree should be much higher than the degree of any of the functions L_P. We do not elaborate on this point further but remark that we show how to achieve this property via an appropriate clean-up process (which in return necessitates the set-up made in Section 4).

The intuition for the existence of the coupling comes from the fact that if we look at the left-hand side above, and expose $1 - 1/D$ of the coordinates of X where D is much higher than the degrees of the functions L_P but much smaller than the separatedness of $\mathcal{P}$, then the values of the functions L_P are almost fully determined, whereas the values of the functions $P(X)$ still have the same distribution as the original one. This suggests that the behavior of the values of $P(X)$'s and $L_P(X)$'s is almost independent of each other, and hence, such a coupling should exist.

2.4 Deducing the Invariance Principle

Collecting the facts gathered so far, we have shown that the values of the noised function $\mathrm{T}_{\mathcal{P},1-\varepsilon}f_1$ can be coupled with the values of decoupled function from (11) in a way that is close in L_2 distance, implying that

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} [f_1(x)f_2(y)f_3(z)] - \mathop{\mathbb{E}}_{\substack{(x,y,z)\sim\mu^{\otimes n}\\ (x',y',z')\sim\mu^{\otimes n}}} \left[\tilde{f}_{1,\mathsf{decoupled}}(x,x')\tilde{f}_{2,\mathsf{decoupled}}(y,y')\tilde{f}_{3,\mathsf{decoupled}}(z,z')\right] \right| = o(1)$$

(strictly speaking, we need to apply a rounding function on the output of the decoupled functions; the goal is to make sure the output is a number bounded by 1 in absolute value). Multiplying the second expectation out, one may use the standard invariance principle to replace x', y', z' with Gaussian random variables. As for the x, y, z, one notes that the values of the functions only depend on $\sigma(x), \gamma(y)$ and $\phi(z)$, and so they can be replaced with elements from G_{master} that add up to 0. This finishes a proof sketch of Theorem 1.2.

3 Preliminaries

In this section, we collect a few basic notions and results that we need.

Notations. For a vector $x \in \Sigma^n$ and a subset $I \subseteq [n]$ of coordinates, we denote by x_I the vector in Σ^I which results by dropping from x all coordinates outside I. We denote by x_{-I} the vector in $\Sigma^{n-|I|}$ resulting from dropping from x all coordinates from I; if $I = \{i\}$ we often simplify the notation and write it as x_{-i}. For $I \subseteq [n]$, $a \in \Sigma^I$ and $b \in \Sigma^{n-|I|}$ we denote by $(x_I = a, x_{-I} = b)$ the point in Σ^n whose I-coordinates are filled according to a, and whose $\overline{I}$-coordinates are filled according to b. For two strings $x, y \in \Sigma^n$ we denote by $\Delta(x, y)$ the Hamming distance between x and y, that is, the number of coordinates $i \in [n]$ such that $x_i \neq y_i$. We denote by $\boldsymbol{i}$ the complex root of -1, and by $\overline{a}$ the complex conjugate of the number $a \in \mathbb{C}$.

We denote $A \lesssim B$ to refer to the fact that $A \leqslant C \cdot B$ for some absolute constant $C > 0$, and $A \gtrsim B$ to refer to the fact that $A \geqslant c \cdot B$ for some absolute constant $c > 0$. If this constant depends on some parameter, say m, the corresponding notation is $A \lesssim_m B$. We will also use standard big-O notations: we denote $A = O(B)$ if $A \lesssim B$, $A = \Omega(B)$ if $A \gtrsim B$; if there is a dependency of the hidden constant on some auxiliary parameter, say m, we denote $A = O_m(B)$ and $A = \Omega_m(B)$.

We denote $0 < A \ll B$ to refer the choice of A and B in the way that B is fixed, and then A is taken to be sufficiently small compared to B.

3.1 Discrete Fourier Analysis over Product Spaces

Let (Σ, ν) be a finite probability space, and consider the product space (Σ^n, ν^n). We will often consider the space $L_2(\Sigma^n, \nu^n)$ of complex-valued functions $f\colon \Sigma^n \to \mathbb{C}$ equipped with the standard inner product

$$\langle f, g\rangle_\nu = \mathop{\mathbb{E}}_{x\sim\nu^n}\left[f(x)\overline{g(x)}\right].$$

Often times, the underlying measure ν will omitted from the notation as it will be clear from context. The space $L_2(\Sigma^n, \nu^n)$ admits several types of decompositions that we will use throughout this paper. The coarsest decomposition is the degree decomposition; the Efron-Stein decomposition is a refinement of the degree decomposition; finally, the Fourier decomposition is a refinement of the Efron-Stein decomposition. We next present each one of these decompositions.

3.1.1 The Degree Decomposition

We first define the notion of d-juntas and the space $V_{\leqslant d}(\Sigma^n)$.

Definition 3.1. *For a subset $I \subseteq [n]$, function $f\colon \Sigma^n \to \mathbb{C}$ is called a I-junta if a function $f'\colon \Sigma^I \to \mathbb{C}$ such that $f(x) = f'(x_I)$ for all $x \in \Sigma^n$. A function f is called a d-junta for $d \in \mathbb{N}$ if it is an I-junta for I of size at most d. We define the space $V_{\leqslant d}(\Sigma^n)$ to be the linear span of all d-juntas.*

The space $V_{\leqslant d}(\Sigma^n)$ is often referred to as the space of degree d functions. Using our inner product, we may define the space of pure degree d functions as follows.

Definition 3.2. *Given a product space (Σ^n, ν^n), the space of pure degree d functions is defined as*

$$V_{=d}(\Sigma^n, \nu^n) = V_{\leqslant d}(\Sigma^n) \cap V_{\leqslant d}(\Sigma^n)^{\perp}.$$

It is easily seen that

$$L_2(\Sigma^n, \nu^n) = V_{\leqslant n}(\Sigma^n) = \bigoplus_{d=0}^{n} V_{=d}(\Sigma^n, \nu^n),$$

and thus any function $f\colon \Sigma^n \to \mathbb{C}$ may be uniquely written as

$$f(x) = \sum_{d=0}^{n} f^{=d}(x), \qquad\qquad \text{where } f^{=d} \in V_{=d}(\Sigma^n, \nu^n)\ \forall d.$$

This decomposition is called the degree decomposition of f, and the function $f^{=d}$ is referred to as the degree d part of f.

3.1.2 The Efron-Stein Decomposition

The Efron-Stein decomposition is a refinement of the degree decomposition. For each $d \in \mathbb{N}$ and $S \subseteq [n]$ of size d, one defines the space $V_{=S}(\Sigma^n, \nu^n) = V_{=d}(\Sigma^n, \nu^n) \cap \{S\text{-juntas}\}$. It is a standard fact that $\bigoplus_{|S|=d} V_{=S} = V_{=d}$, and thus one may write $f^{=d} \in V_{=d}$ as

$$f^{=d}(x) = \sum_{|S|=d} f^{=S}(x) \qquad\qquad \text{where } f^{=d} \in V_{=S}(\Sigma^n, \nu^n)\ \forall S.$$

Thus, one gets the decomposition of f as

$$f^{=d}(x) = \sum_{S\subseteq[n]} f^{=S}(x)$$

where $f^{=S} \in V_{=S}(\Sigma^n, \nu^n)$ for all S. This decomposition is known as the Efron-Stein Decomposition.

3.1.3 The Fourier Decomposition

The Fourier decomposition is the most refined and explicit decomposition among the decompositions discussed herein. One first looks at the base space, $L_2(\Sigma, \nu)$, and picks an orthonormal basis for it consisting of $m = |\Sigma|$ real-valued functions, $v_0, \ldots, v_{m-1}\colon \Sigma \to \mathbb{R}$. It is standard to take v_0 to be the all 1 function, and we will do so here. With these notations, we may construct an orthonormal basis for $L_2(\Sigma^n, \nu^n)$ consisting of the functions $\{v_{\vec{\alpha}}\}_{\vec{\alpha}\in\{0,\ldots,m-1\}^n}$ where

$$v_{\vec{\alpha}}(x) = \prod_{i=1}^{n} v_{\alpha_i}(x_i).$$

Thus, any function $f\colon \Sigma^n \to \mathbb{C}$ admits a unique decomposition as $f(x) = \sum\limits_{\vec{\alpha}\in[m]^n} \widehat{f}(\vec{\alpha}) v_{\vec{\alpha}}(x)$ where $\widehat{f}(\vec{\alpha}) = \langle f, v_{\vec{\alpha}}\rangle$. We remark that it is easily seen that the space $V_{=d}$ is the span of all $v_{\vec{\alpha}}$ with $|\mathsf{supp}(\alpha)| = d$, and furthermore that $V_{=S}$ is the span of all $v_{\vec{\alpha}}$ with $\mathsf{supp}(\alpha) = S$.

The Fourier definition presented herein can be somewhat arbitrary as there are many ways of defining the functions $v_1, \ldots, v_{m-1}$. Nevertheless, it will be useful for us in order to present the invariance principle [42]. Towards this end

3.1.4 Hypercontractivity

Definition 3.3. *For $q \geqslant 2$ and $f\colon (\Sigma^n, \nu^n) \to \mathbb{C}$, we define the q-norm of f as*

$$\|f\|_q = \left(\mathop{\mathbb{E}}_{x\sim\nu^n}\left[|f(x)|^q\right]\right)^{1/q}.$$

Theorem 3.4. *Suppose that (Σ, ν) is a probability space with $|\Sigma| = m$ such that $\nu(x) \geqslant \alpha$ for all $x \in \Sigma$, where $m \in \mathbb{N}$ and $\alpha > 0$. For all $q \geqslant 2$ there exists $C = C(q, \alpha, m) > 0$ such that if $f\colon (\Sigma^n, \nu^n) \to \mathbb{C}$ is a degree d function, then*

$$\|f\|_q \leqslant C^d \|f\|_2.$$

3.1.5 The Noise Operator

We will need the standard noise operator on product spaces.

Definition 3.5. *For a parameter $\rho \in [0,1]$ and an input $x \in \Sigma^n$, we define the distribution $\mathrm{T}_\rho x$ over ρ-correlated inputs with x in the following way. For each $i \in [n]$ independently, with probability ρ pick $y_i = x_i$, and otherwise sample y_i according to ν.*

The process T_ρ is described as a Markov chain, and as is standard we may consider it as an averaging operator over functions. That is, we may consider it as a linear operator $\mathrm{T}_\rho\colon L_2(\Sigma^n, \nu^n) \to L_2(\Sigma^n, \nu^n)$ defined as

$$\mathrm{T}_\rho f(x) = \mathop{\mathbb{E}}_{y \sim \mathrm{T}x}[f(y)].$$

3.2 Influences and Low-degree Influences

Next, we define a few basic notions from the analysis of Boolean functions.

Definition 3.6. *For a function $f\colon (\Sigma^n, \nu^n) \to \mathbb{C}$ and a coordinate $i \in [n]$, the influence of i is defined as*

$$I_i[f] = \mathop{\mathbb{E}}_{\substack{y \sim \nu^{n-1} \\ a,b \sim \nu}}\left[|f(x_{-i} = y, x_i = a) - f(x_{-i} = y, x_i = b)|^2\right].$$

Next, we define the notion of low-degree influences.

Definition 3.7. *For a function $f\colon (\Sigma^n, \nu^n) \to \mathbb{C}$, a parameter $d \in \mathbb{N}$ and a coordinate $i \in [n]$, the degree d influence of i is defined as $I_i^{\leqslant d}[f] = I_i[f^{\leqslant d}]$.*

3.3 The Invariance Principle

In this section, we present the invariance principle of [42], and we begin with some set-up. Suppose that Σ, Γ, Φ are finite alphabets of sizes m_1, m_2, m_3 respectively, and μ is a probability measure over $\Sigma \times \Gamma \times \Phi$ in which the probability of each atom is at least $\alpha > 0$. We set up Fourier bases for (Σ, μ_x), (Γ, μ_y) and (Φ, μ_z) given by $v_0, \ldots, v_{m_1-1}$, $u_0, \ldots, u_{m_2-1}$ and $w_0, \ldots, w_{m_3-1}$. Consider the ensemble of random variables

$$\mathcal{X} = \{v_1(x), \ldots, v_{m_1-1}(x), u_1(y), \ldots, u_{m_2-1}(y), w_1(z), \ldots, w_{m_3-1}(z)\}$$

where $(x,y,z) \sim \mu$. We define the covariance matrix $P \in \mathbb{R}^{(m_1+m_2+m_3)\times(m_1+m_2+m_3)}$ whose rows and columns correspond to the functions in $\mathcal{X}$, and the entry corresponding to tow random variables in $\mathcal{X}$. For example, for v_i and u_j, the corresponding entry is

$$P(v_i, u_j) = \mathop{\mathbb{E}}_{(x,y,z)\sim\mu}\left[v_i(x)\overline{u_j(y)}\right].$$

Let

$$\mathcal{G} = \{G_{1,x}, \ldots, G_{m_1-1,x}, \ldots, G_{1,z}, \ldots, G_{m_3-1}\}$$

be an ensemble of centered Gaussian random variables with covariance matrix P. The invariance principle relates the behavior of low-influence, multi-linear polynomials over $\mathcal{X}$ and over $\mathcal{G}$. Below, we state the version that we need from [41] specialized to our case of interest, but before that, we need a few definitions.

Denote $q = m_1 + m_2 + m_3 - 3$, and let $M\colon \mathbb{C}^{qn} \to \mathbb{C}$ be a multi-linear polynomial given as

$$M(a_{1,1}, \ldots, a_{1,q}, \ldots, a_{n,1}, \ldots, a_{n,q}) = \sum_{T \subseteq [n] \times [q]} m_T \prod_{(i,j) \in T} a_{i,j}.$$

Definition 3.8. *The influence of variable (i,j) on M as above is defined as*

$$I_{i,j}[M] = \sum_{T \ni (i,j)} |a_{i,j}|^2 .$$

We will also consider vector-valued multi-linear functions, which are functions $M\colon \mathbb{C}^{qn} \to \mathbb{C}^k$ wherein each M_s is a multi-linear function. The influence of (i,j) on M is defined as $I_{i,j}[M] = \max_s I_{i,j}[M_s]$.

Finally, define $\mathsf{trunc}\colon \mathbb{C} \to \mathbb{C}$ by $\mathsf{trunc}(a) = a$ if $|a| \leqslant 1$ and $\mathsf{trunc}(a) = a/|a|$ otherwise.

Theorem 3.9. *For all $\alpha > 0$, $k, m \in \mathbb{N}$, $d \in \mathbb{N}$, $C > 0$ and $\varepsilon > 0$, there exists $\tau > 0$ such that the following holds. Suppose that $|\Sigma|, |\Gamma|, |\Phi| \leqslant m$, that μ is a distribution over $\Sigma \times \Gamma \times \Phi$ in which the probability of each atom is at least α, and let $\mathcal{X}$ and $\mathcal{G}$ be the ensembles of random variables as above with the same covariance matrix. Let $M\colon \mathbb{C}^{qn} \to \mathbb{C}^k$ is a multi-linear polynomial with $\max_{i,j} I_{i,j}[M] \leqslant \tau$.*

1. *If $\Psi\colon \mathbb{C}^k \to \mathbb{C}$ is differentiable three times and its third order derivatives are at most C in absolute value, then*
$$\left| \mathbb{E}\left[\Psi(M(\mathcal{X}^n))\right] - \mathbb{E}\left[\Psi(M(\mathcal{G}^n))\right] \right| \leqslant \varepsilon.$$
2. *Define $\xi\colon \mathbb{C}^k \to \mathbb{R}$ by $\xi(a_1, \ldots, a_k) = \sqrt{\sum_{i=1}^{k} |\mathsf{trunc}(a_i) - a_i|^2}$. Then*
$$\left| \mathbb{E}\left[\xi(M(\mathcal{X}^n))\right] - \mathbb{E}\left[\xi(M(\mathcal{G}^n))\right] \right| \leqslant \varepsilon.$$

Last, we need the following elementary fact.

Fact 3.10. *Consider the function $\xi\colon \mathbb{C} \to [0,\infty)$ defined as $\xi(a) = |\mathsf{trunc}(a) - a|$. Then ξ is 2-Lipshitz function.*

Proof. Let $a, b \in \mathbb{C}$; we show that $|\xi(a) - \xi(b)| \leqslant 2|a - b|$. If a, b are both at most 1 in absolute value, then the left-hand size is 0 and the claim is trivial. If exactly one of a and b is at most 1 in absolute value, say a then

$$|\xi(a) - \xi(b)| = |\xi(b)| = |b| - 1 \leqslant |b| - |a| \leqslant |b - a|,$$

and we are done. It remains to consider the case that both a and b are at least 1 in absolute value. In that case, by the triangle inequality

$$\begin{aligned} |\xi(a) - \xi(b)| = \frac{|a\,|b| - b\,|a||}{|a|\,|b|} &\leqslant \frac{|a\,|b| - a\,|a|| + |a\,|a| - b\,|a||}{|a|\,|b|} \\ &= \frac{||b| - |a||}{|b|} + \frac{|a - b|}{|b|} \\ &\leqslant 2\,|a - b|, \end{aligned}$$

and the proof is concluded. □

3.4 The μ-norm and the CSP Stability Result

Definition 3.11. *For a distribution μ over $\Sigma \times \Gamma \times \Phi$ and a function $f\colon (\Sigma^n, \mu_x^{\otimes n}) \to \mathbb{C}$, we define the μ semi-norm of f as*

$$\|f\|_\mu = \sup_{\substack{g\colon \Gamma^n \to \mathbb{C} \\ h\colon \Phi^n \to \mathbb{C} \\ 1\text{-bounded}}} \left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} [f(x)g(y)h(z)] \right|.$$

For general distributions μ, $\|\|_\mu$ is actually a semi-norm; for instance, if the distribution μ is uniform over $\Sigma \times \Gamma \times \Phi$, then $\|f\|_\mu = |\mathbb{E}[f(x)]|$. For most distributions we will be concerned with, though, this semi-norm will actually be a norm.

In our applications, we will need to work with several distribution μ over triplets that have the same marginal distribution over x. A special collection of distributions μ that we care about is as follows:

Definition 3.12. *For alphabets Σ, Γ and Φ, a distribution ν over Σ and a parameter $\alpha > 0$, define the collections*

$$M_\nu = \{\, \mu \mid \textit{pairwise connected distribution over } \Sigma \times \Gamma \times \Phi \textit{ with no } (\mathbb{Z}, +)\textit{-embedding}, \mu_x = \nu \} ,$$

$$M_\alpha = \{\, \mu \mid \mu(x,y,z) \geqslant \alpha \; \forall (x,y,z) \in \mathsf{supp}(\mu) \} ,$$

and $M_{\nu,\alpha} = M_\nu \cap M_\alpha$.

Definition 3.13. *Let ν be a distribution over Σ, let Γ, Φ be alphabets and let M be a collection of distributions over $\Sigma \times \Gamma \times \Phi$ such that $\mu_x = \nu$ for all $\mu \in M$. For a function $f\colon (\Sigma^n, \mu_x^{\otimes n}) \to \mathbb{C}$, we define*

$$\|f\|_{M,\nu} = \sum_{\mu \in M} \|f\|_\mu.$$

In the special case that $M = M_{\nu,\alpha}$, we refer to the associated norm $\|f\|_{M,\nu}$ as the ν semi-norm of f, and denote it by

$$\|f\|_{\nu,\alpha} = \sup_{\mu \in M_{\nu,\alpha}} \|f\|_\mu.$$

Remark 3.14. *The results in Sections 4, 5 and 6.3 apply to the more general notion of semi-norm $\|f\|_{M,\nu}$ with suitable adaptations, and we will need such a result in Section 7. However, stating it in this generality would complicate the statements (that are already quantifier heavy) further, and hence we state all of the results for the most general semi-norm $\|f\|_{\nu,\alpha}$.*

It will be important for us to understand the type of functions f that have a significant μ-norm, and towards that end, we use the following result from [11]:

Theorem 3.1. *For all $m \in \mathbb{N}$, and alphabet Σ of size at most m, there is an Abelian group G and $\sigma\colon \Sigma \to G$ such that for all $\varepsilon, \alpha > 0$ there exists $\delta > 0$ and $d \in \mathbb{N}$ such that the following holds. Suppose that Σ, Γ, Φ are alphabets of size at most m, and μ is a pairwise connected distribution over $\Sigma \times \Gamma \times \Phi$ with no $\mathbb{Z}$ embeddings in which the probability of each atom is at least α. If $f\colon \Sigma^n \to \mathbb{C}$ is a 1-bounded function such that $\|f\|_\mu \geqslant \varepsilon$, then there is $\chi \in \hat{G}^{\otimes n}$ and $L\colon \Sigma^n \to \mathbb{C}$ of degree at most d and $\|L\|_2 \leqslant 1$ such that*

$$|\langle f, L \cdot \chi \circ \sigma \rangle| \geqslant \delta.$$

Throughout, a function of the type $\chi \circ \sigma$ will be referred to as a character function or a character embedding function.

4 Embeddings, Embedding Functions and Cyclic Embedding Functions

Character embedding functions arising in Theorem 3.1 will be important for us to study. The structure of a general character embedding function, however, is too abstract, and our arguments require more strict structure from them to go through. The main goal of this section is to process embeddings σ as arising in Theorem 3.1, as well as collections of character embedding functions. This process will produce sort-of refinements that have a more strict structure, which will be useful for us later.

4.1 Product Functions

Let Σ be an alphabet of size at most m, let G be an Abelian group of size $O_m(1)$ and let $\sigma\colon \Sigma \to G$ be a map.

Definition 4.1. *We define the class of product functions $\mathcal{P}(\Sigma, G, \sigma)$ to be the collection of functions of the form $P\colon \Sigma^n \to \mathbb{C}$ for which there are 1-bounded univariate functions $u_1, \ldots, u_n\colon G \to \mathbb{C}$ for which*

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} u_i(\sigma(x_i)).$$

Definition 4.2. *We say that a class of product functions $\mathcal{P}(\Sigma, G, \sigma)$ is τ-separated if for any univariate functions $u, v\colon \Sigma \to \mathbb{C}$ in it, it either holds that $|\langle u, v\rangle| = 1$ or else $|\langle u, v\rangle| \leqslant 1 - \tau$.*

We need to define a metric on product functions, which we refer to as the symbolic metric.

Definition 4.3. *Let $P, P'\colon \Sigma^n \to \mathbb{C}$ be product functions in $\mathcal{P}(\Sigma, G, \sigma)$. The symbolic metric $\Delta_{\mathsf{symbolic}}(P, P')$ is defined as the minimum number k such that there are $u_1, \ldots, u_n\colon G^n \to \mathbb{C}$, $v_1, \ldots, v_n\colon G^n \to \mathbb{C}$ for which $u_i = v_i$ for all but k of the indices $i \in [n]$, and*

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} u_i(\sigma(x_i)), \qquad P'(x_1, \ldots, x_n) = \prod_{i=1}^{n} v_i(\sigma(x_i)).$$

Definition 4.4. *For a set $\mathcal{P} = \{P_1, \ldots, P_r\colon \Sigma^n \to \mathbb{C}\}$, we define*

$$\mathsf{spn}_{\mathbb{N}}(\mathcal{P}) = \left\{ \prod_{i=1}^{r} P_i^{\alpha_i} \,\middle|\, \alpha_i \in \mathbb{N}, P^{\alpha_i} \neq 1 \text{ for at least one } i \right\}.$$

We note that in all applications, the functions P_i's will be a composition of a character over some finite Abelian group $\chi \in \widehat{A}^{\otimes n}$ with a map $\sigma^{\otimes n}$ where $\sigma\colon \Sigma \to A$ is some map, and as such there will be a power $m = O_{|A|}(1)$ such that $P_1^m = 1$. Thus, the set $\mathsf{sp}_{\mathbb{N}}(\mathcal{P})$ will always have a finite size $O_{|A|,r}(1)$.

Definition 4.5. *The rank of $\mathcal{P} = \{P_1, \ldots, P_r\colon \Sigma^n \to \mathbb{C}\}$ is defined as*

$$\mathsf{rk}(\mathcal{P}) = \min_{P \in \mathsf{sp}_{\mathbb{N}}, P \neq 1} \Delta_{\mathsf{symbolic}}(P, 1)$$

We have the following simple fact, asserting that product functions with large symbolic distance act like high-degree monomials.

Lemma 4.6. *Suppose that $P\colon \Sigma^n \to \mathbb{C}$ is a product function from $\mathcal{P}(\Sigma, G, \sigma)$ which is τ-separated. Then for all $\xi \in [0, 1)$ we have that $\|T_{1-\xi} P\|_2 \leqslant 2^{-\Omega_{\xi,m}(\Delta_{\mathsf{symbolic}}(P,1))}$.*

4.2 Embeddings, Master Embeddings and Reduced Embeddings

In this section we fix alphabets Σ, Γ, Φ and a distribution over ν. The primary goal of this section is to define the master embedding of ν and pre-process it to make it easier to work with. At a high level, the master embedding of ν is a group embedding that encapsulates all possible embeddings that may arise in Theorem 3.1 for Σ. This embedding is too abstract and hard to work with, and the goal of our preprocessing is to modify it so that it satisfies a few properties. Throughout, we let $m = \max(|\Sigma|, |\Gamma|, |\Phi|)$.

4.2.1 The Master Embedding

Let ν be a distribution over Σ, and note that the set M_ν is finite. For each $\mu \in M_\nu$ there are finitely many Abelian embeddings of μ, say $(\sigma_{i,\mu}, \gamma_{i,\mu}, \phi_{i,\mu})$ into $(G_i, +)$ for $i = 1, \ldots, k$ for some $k = O_m(1)$. Thus, we may define $\sigma_\mu(x) = (\sigma_{1,\mu}(x), \ldots, \sigma_{k,\mu}(x))$ into the group $G_\mu = G_1 \times \ldots \times G_k$. With this notation, the master embedding of ν is defined in the following way:

Definition 4.7. *For a distribution ν over Σ, define $\sigma_{\mathsf{master}}(x) = (\sigma_\mu(x))_{\mu \in M_\nu}$ and $G_{\mathsf{master}} = \prod_{\mu \in M_\nu} G_\mu$.*

We remark that strictly speaking, the set M_ν is infinite. However, as σ_μ depends only on $\mathsf{supp}(\mu)$ and the number of distinct sets that $\mathsf{supp}(\mu)$ may be for $\mu \in M_\nu$ is finite. Thus, we may think of σ_{master} as being of finite length of at most $O_m(1)$.

Remark 4.8. *We remark that if we wanted to work with the semi-norms $\|f\|_{M,\nu}$ (instead of $\|f\|_{\nu,\alpha}$) as suggested in Remark 3.14, then we have to modify the definition of the master embedding σ_{master} and the master group G_{master} in the natural way. Namely, we would only take product over $\mu \in M$, and then the content of this section remains as is.*

A key feature of the master embedding σ_{master} is that it captures all characters over any Abelian group that may arise via embeddings of any $\mu \in M_\nu$. Indeed, for any $\mu \in M_\mu$, for any Abelian group G that μ may be embedded into via, say, $\sigma\colon \Sigma \to G$, and for any $\chi \in \widehat{G}$, we may find $\chi' \in \widehat{G_{\mathsf{master}}}$ such that $\chi \circ \sigma(x) = \chi' \circ \sigma_{\mathsf{master}}(x)$ for all $x \in \Sigma$. This is true since σ is one of the coordinates in σ_{master}. In a sense, the master embedding σ_{master} encapsulates within it any character function.

4.2.2 Embedding Reduction

We would like to modify the embedding σ_{master} and the master group G_{master} so that they have more structure. The key feature of these modifications that will make them applicable for us is that they preserve the collection of character functions.

Definition 4.9. *Let ν be a probability distribution over Σ and let $\sigma'\colon \Sigma \to G'$ be a map.*

1. *We say that a map $\sigma\colon \Sigma \to G$ reduces to $\sigma'\colon \Sigma \to G'$ if for any $\chi \in \widehat{G}$ there is $\chi' \in \widehat{G'}$ and $\theta \in \mathbb{C}$ of absolute value 1 such that $\chi \circ \sigma \equiv \theta \chi' \circ \sigma'$.*
2. *We say that σ' is character-encapsulating if the master embedding of ν, $(\sigma_{\mathsf{master}}, G_{\mathsf{master}})$ reduces to χ'.*

Thus, the goal in our preprocessing step is to find an embedding that has more structural properties than the master embedding, yet it is still character-encapsulating.

4.2.3 Reducing the Master Embedding

We now begin with a series of modifications to $(\sigma_{\mathsf{master}}, G_{\mathsf{master}})$ so as to gain additional structural properties. First, we may assume without loss of generality that the group generated by $\mathsf{Image}(\sigma_{\mathsf{master}})$ is G_{master}, since otherwise, we may replace G_{master} with that sub-group. Second, as G_{master} is an Abelian group, by the fundamental theorem of finite Abelian groups, it may be written as $G_1 \times \ldots \times G_k$ where each one of the group G_i is a cyclic group of prime power order, say $(\mathbb{Z}_{{p_i}^{d_i}}, +)$. The fact that some primes p_i may repeat, and furthermore that we have to consider prime powers (i.e., the case $d_i > 1$) create complications, and the goal of this section is to obtain enough structure on our embedding so that we can tolerate them.

Of central interest to us will be the following consideration. Suppose that $\chi_1, \ldots, \chi_s$ are some character functions, and consider the partition over Σ^n defined as

$$P_{\theta_1,\ldots,\theta_s} = \{x \in \Sigma^n \mid (\chi_i \circ \sigma)(x) = \theta_i \ \forall i\}\,.$$

Suppose that $\chi_1, \ldots, \chi_s$ satisfy some relation such that $\chi_1^{b_1} \cdots \chi_s^{b_s} \equiv 1$. Can we always "simplify" the collection $\chi_1, \ldots, \chi_s$ without changing the partition it induces? The goal of our preprocessing is that the answer to the above would be positive, and to get some intuition as to how such simplifications look we first consider the basic case where the map σ is ignored.

4.2.4 Complications that May Arise in Prime Power Cyclic Groups

To get a sense of these potential complications, note that any character function of $(\mathbb{Z}_p^n, +)$ is a character function of $(\mathbb{Z}_{p^2}^n, +)$, and this create issues. For instance, suppose that for a prime p we have $\chi_1, \ldots, \chi_s \in \widehat{\mathbb{Z}_p^n}$ such that $\chi_1^b \cdots \chi_s^b \equiv 1$ for some $1 \leqslant b \leqslant p-1$. In our applications we will care about the partition of the domain, in this case, $\mathbb{Z}_p^n$, induced by the functions $\chi_1, \ldots, \chi_s$:

$$P_{\theta_1,\ldots,\theta_s} = \{x \mid \chi_i(x) = \theta_i \ \forall i\}\,.$$

In this language, the fact $\chi_1^b \cdots \chi_s^b \equiv 1$ means that there are redundancies in this partition. Indeed, in this case, we could drop χ_s and have the same partition: by raising both sides to the power a where $ab = 1 \pmod{p}$ it follows that $\chi_1 \cdots \chi_s \equiv 1$, so $\chi_s = \overline{\chi_1} \cdots \overline{\chi_{s-1}}$. Hence, the partition formed by $\{\chi_1, \ldots, \chi_{s-1}, \chi_s\}$ is the same as the partition formed by $\{\chi_1, \ldots, \chi_{s-1}\}$, and we have simplified our collection without affecting the partition. Things become more complicated in prime power cyclic groups.

Suppose for illustration that $\chi_1, \ldots, \chi_s \in \widehat{\mathbb{Z}_{p^2}^{\otimes n}}$ satisfy that $\chi_1^b \cdots \chi_s^b \equiv 1$ for some $1 \leqslant b \leqslant p^2 - 1$, then we can no longer always conclude that their product is also 1. If b is co-prime to p this would be true (by raising both sides again to a power a where $ab = 1 \pmod{p^2}$). If b is not co-prime to p, say $b = rp$, then we would be able to replace the characters $\{\chi_1, \ldots, \chi_s\}$ with $\{\chi_1, \ldots, \chi_{s-1}, \chi_s'\}$ where $\chi_s' \in \widehat{\mathbb{Z}_p^n}$ and have the same partition. Indeed, we may write $\chi_i(x) = e^{\frac{\langle c_i, x\rangle}{p^2} 2\pi \mathbf{i}}$ for some $c_i \in \mathbb{Z}_{p^2}^n$, and get that $p(c_1 + \ldots + c_s) = 0 \pmod{p^2}$. Thus, $c_s = -(c_1 + \ldots + c_{s-1}) \pmod{p}$, and so $c_s = -(c_1 + \ldots + c_{s-1}) + pt$ for $t \in \mathbb{N}^n$. We may take $\chi_{s'}(x) = e^{\frac{\langle t, x\rangle}{p} 2\pi\mathbf{i}}$ and get that

$$\chi_s(x) = e^{\frac{\langle c_s, x\rangle}{p^2} 2\pi\mathbf{i}} = e^{\frac{\langle -c_1 - \ldots - c_{s-1}, x\rangle}{p^2} 2\pi\mathbf{i}} e^{\frac{\langle t, x\rangle}{p} 2\pi\mathbf{i}} = \overline{\chi_1} \cdots \overline{\chi_{s-1}} \chi_s',$$

from which it follows that the partition induced by $\{\chi_1, \ldots, \chi_s\}$ is the same as the partition induced by $\{\chi_1, \ldots, \chi_{s-1}, \chi_s'\}$. Thus, we again got to simplify our system while keeping the partition the same.

Complications that may arise due to the map σ: re-introducing the map σ into the picture creates some further complications. Suppose for example that $d_1 = 2$ and the corresponding coordinate of σ, namely $\sigma_1\colon \Sigma \to \mathbb{Z}_{p_1^2}$ is a map of the form $\sigma_1(x) = pm(x)$ for some $m\colon \Sigma \to \mathbb{Z}_{p_1^2}$. Then character functions $\chi \circ \sigma_1$ for $\chi \in \widehat{\mathbb{Z}_{p_1^2}}$ yield functions $\chi \circ \sigma_1$ that are not really characters over $\mathbb{Z}_{p_1^2}$. Rather, they are trivial lifts of a character form $\mathbb{Z}_{p_1}$. We would like to avoid these lifted functions, and so we must clean up the map σ.

Definition 4.10. *Let p be a prime and let $d > 1$. We say that a map $m\colon \Sigma \to \mathbb{Z}_{p^d}$ is reducible if there is a map $m'\colon \Sigma \to \mathbb{Z}_{p^d}$ and $a, b \in \mathbb{Z}_{p^d}$ such that $m'(x) = p \cdot a \cdot m(x) + b$ for all $x \in \Sigma$. Otherwise, we say that m is irreducible.*

We note that if m is reducible, then any character function $\chi \circ m$ for $\chi \in \widehat{\mathbb{Z}_{p^d}}$ is equal to a constant times a character function $\chi' \circ m'$ for $\chi' \in \widehat{\mathbb{Z}_{p^{d-1}}}$. We also need to clean up redundancies between different coordinates of a master embedding. We begin with handling maps using a single prime p.

Definition 4.11. *Let $\sigma\colon \Sigma \to \prod_{k=1}^{s} \mathbb{Z}_{p^{d_k}}$ be a map, and let $d = \max d_k$. We say σ contains redundancies if there are $a_1, \ldots, a_s, b \in \mathbb{N}$ such that $0 \leqslant a_k < p^d$ are not all 0 and*

$$\sum_{k=1}^{s} a_k \sigma_k(x) = b \pmod{p^d}.$$

Otherwise, we say σ does not contain redundancies.

Suppose that σ is a map containing redundancies, and take $t\colon \Sigma \to \mathbb{Z}_{p^d}$ satisfying that

$$\sum_{k=1}^{s} a_k \sigma_k(x) = b + p^d t(x).$$

Write $a_k = p^{s_k} a'_k$ where a'_k is co-prime to p and take i minimizing $d_i + s_i$. Take the map σ' which results from dropping the coordinate i from σ and replacing it by t. We claim that if $\chi \circ \sigma$ is a character function, then we could write it as $\theta\chi' \circ \sigma'$ for some $\chi' \in \widehat{\prod_{j \neq i} \mathbb{Z}_{p^{d_j}}}$ and θ a complex number of absolute value 1. Indeed, we may write

$$\begin{aligned} e^{\frac{c\sigma_i(x)}{p^{d_i}} 2\pi\mathbf{i}} = e^{\frac{cp^{s_i}\sigma_i(x)}{p^{d_i+s_i}} 2\pi\mathbf{i}} &= e^{\frac{1}{p^{d_i+s_i}}\left(bc{a'_i}^{-1} + c{a'_i}^{-1} p^d t(x) - c{a'_i}^{-1}\sum\limits_{k\neq i} a_k p^{s_k}\sigma_k(x)\right) 2\pi\mathbf{i}} \\ &= e^{\frac{bc{a'_i}^{-1}}{p^{d_i+s_i}} 2\pi\mathbf{i}} e^{\frac{c{a'_i}^{-1} t(x)}{p^{d_i+s_i-d}} 2\pi\mathbf{i}} \cdot \prod_{k\neq i} e^{-\frac{c{a'_i}^{-1} a'_k p^{d_k+s_k-(d_i+s_i)}\sigma_k(x)}{p^{d_k}}}. \end{aligned}$$

Thus, we can take $t(x)$ modulo $\max(p^{d_i+s_i-d}, 1)$ and get that any character of the i coordinate of σ can be written as a product of a constant and characters over σ'. The benefit here is that we always have that $d_i + s_i - d < d_i$, hence we reduce the size of the group on the ith coordinate. For general embeddings, the definition proceeds as follows:

Definition 4.12. *Let $\sigma\colon \Sigma \to \prod_{i=1}^{k} \mathbb{Z}_{{p_i}^{d_i}}$. We say that σ contains redundancies if there is some prime p, such that looking at the coordinates of σ with $p_i = p$, the resulting map contains redundancies. Otherwise, we say that σ is redundancy-free.*

We summarize this section and the above discussion with the following immediate fact:

Fact 4.13. *Let $G = \prod_{i=1}^{k} \mathbb{Z}_{{p_i}^{d_i}}$ be an Abelian group, and let $\sigma\colon \Sigma \to G$ be a map. Then there exists an Abelian group $G' = \prod_{i=1}^{k} \mathbb{Z}_{{p_i}^{d'_i}}$ with $d'_i \leqslant d_i$ and a map $\sigma'\colon \Sigma \to G'$ such that:*

1. *The map σ is reducible to σ'.*
2. *Each coordinate of σ' is irreducible.*
3. *The map σ' contains no redundancies.*

Proof. As long as there is a coordinate of σ that is reducible, say $\sigma_i(x) = p \cdot a \cdot \sigma'_i(x) + b$, we replace σ_i with σ'_i and ${p_i}^{d_i}$ with ${p_i}^{d_i-1}$ to get the map σ'. We get from the above discussion that σ is reducible to σ'. As long as there are redundancies in σ, we find a coordinate i as above and replace the embedding σ there as explained above.

Clearly, this process must terminate, at which point we get to a σ' that is irreducible and does not contain redundancies. □

4.2.5 Reducing Our Master Embedding

Applying Fact 4.13 on the maser embedding σ_{master} we get an embedding which is irreducible and with no redundancies, and to simplify notations we assume henceforth that σ_{master} is irreducible to begin with. Furthermore, by applying an affine shift (which clearly preserves all the features of σ_{master}), we may assume that its image contains $0 \in G_{\mathsf{master}}$. We recall that we may assume that the image of σ_{master} generates the whole group G_{master}, otherwise we may shrink it (and, if necessary, we repeat the above process to get the embedding to once again be irreducible with no redundancies).

4.3 Cyclic Embedding Functions

As we explained earlier, embedding functions – namely functions of the form $P = \chi \circ \sigma_{\mathsf{master}}$ for $\chi \in \widehat{G^n_{\mathsf{master}}}$ – will be of interest to us. For convenience, it will be easier for us to handle cyclic embedding functions.

Definition 4.14. *Let $P = \chi \circ \sigma_{\mathsf{master}}$ be a function for $\chi \in \widehat{G^n_{\mathsf{master}}}$, and write $G_{\mathsf{master}} = \prod_{i=1}^{k} \mathbb{Z}_{{p_i}^{d_i}}$ and so $\chi = (\chi_1, \ldots, \chi_k)$ for $\chi_i \in \widehat{\mathbb{Z}_{{p_i}^{d_i}}}$ as above. We say P is a cyclic embedding function if there is $j \in \{1, \ldots, k\}$ such that for all $i \neq j$ we have that $\chi_i \equiv 1$.*

In words, a cyclic embedding function is a function $P = \chi \circ \sigma_{\mathsf{master}}$ that takes into account only one of the coordinates of σ_{master}. We have the following trivial fact:

Fact 4.15. *Suppose that $P = \chi \circ \sigma_{\mathsf{master}}$ is an embedding function. Then there are $O_m(1)$ functions cyclic embedding functions $P_1, \ldots, P_k$ such that $P(x) = P_1(x) \cdots P_k(x)$ for all x.*

Proof. Write $\chi = (\chi_1, \ldots, \chi_k)$ for $\chi_i \in \widehat{\mathbb{Z}_{{p_i}^{d_i}}}$ and take $P_i(x) = \chi_i(\sigma_{\mathsf{master}}(x))$ for each $i = 1, \ldots, k$. □

Fact 4.15 will allow us to reduce ourselves to always working with cyclic embedding functions. The main benefit of working with cyclic embedding functions is that it allows one to isolate calculation to be only with respect to some prime number p and its powers (as opposed to multiple different primes). Note that if P is a cyclic embedding function using embedding into $\mathbb{Z}_{p^d}$, then $P^{p^d} \equiv 1$. However, as the image of σ_{master} may not include the whole group $\mathbb{Z}_{p^d}$, there may be a smaller power k such that $P^k \equiv 1$. The smallest such power is defined to be the order of P:

Definition 4.16. *Let P be a cyclic embedding function on $\mathbb{Z}_{p^d}$. The order of P, denoted by $\mathsf{ord}(P)$ is the smallest number k such that $P^k \equiv 1$.*

The following facts will be useful for us:

Fact 4.17. *Suppose that P is a cyclic embedding function into $\mathbb{Z}_{p^d}$, suppose that $\mathsf{ord}(P) = t$ and write $P(x) = e^{\frac{\langle c, \sigma_{\mathsf{master},j}(x)\rangle}{p^d} 2\pi \mathbf{i}}$. Then:*

1. *We have that $t = p^s$ for some integer $1 \leqslant s \leqslant d$.*
2. *We may write $c = p^{d-s}c'$ for some $c' \in \mathbb{Z}_{p^d}^n$.*

Proof. Take s to be the smallest integer so that p^{d-s} divides each one of the entries of c. Then the second item holds and we write $c = p^{d-s}c'$. Note that

$$P^{p^s} = e^{p^s \frac{\langle p^{d-s}c', \sigma_{\mathsf{master},j}(x)\rangle}{p^d} 2\pi \mathbf{i}} e^{\langle c', \sigma_{\mathsf{master},j}(x_i)\rangle 2\pi \mathbf{i}} = 1,$$

so $\mathsf{ord}(P) \leqslant p^s$. Suppose for contradiction that $t = \mathsf{ord}(P) < p^s$. Then by the above computation, we get that $\langle p^{d-s}tc', \sigma_{\mathsf{master},j}(x)\rangle = 0 \pmod{p^d}$ for all x, and as $\mathsf{gcd}(p^{d-s}t, p^d) < p^d$ it follows that $\langle c', \sigma_{\mathsf{master},j}(x)\rangle = 0 \pmod{p}$ for all x. By choice of s there is i such that c'_i is co-prime to p, and we fix this i. As 0 is in the image of σ_{master} (see Section 4.2.5), we may consider fixings of the input x outside the coordinate i such that $\sigma_{\mathsf{master},j}(x_k) = 0$ for all $k \neq i$ and get that $c'_i \sigma_{\mathsf{master},j}(x_i) = 0 \pmod{p}$ for all i. It follows that $\sigma_{\mathsf{master},j}(x_i) = 0 \pmod{p}$ for all $x_i \in \Sigma$, and contradiction to the fact that σ_{master} is irreducible. □

4.4 Utilizing Redundancies in Collections of Simple Cyclic Embedding Functions

We are now ready to formalize the simplification process as outlined in Section 4.2.4. More specifically, the general flavor of statements in this section will be of the following type: we have a collection of cyclic functions that satisfy that equality (or near equality), and we would like to produce a new collection of embedding functions that simplifies it while producing the same partition over inputs (or nearly the same in the case of near equality).

We begin with the case of exact equality:

Lemma 4.18. *Fix a prime p and suppose that $P_1, \ldots, P_r$ are cyclic embedding functions of $\mathbb{Z}_{p^{d_1}}, \ldots, \mathbb{Z}_{p^{d_r}}$ respectively. Further suppose that there are powers $1 \leqslant \alpha_i < \mathsf{ord}(P_i)$ for $i = 1, \ldots, r$ such that*

$$P_1^{\alpha_1}(x) \cdots P_r^{\alpha_r}(x) = 1$$

for all $x \in \Sigma^n$. Then there exists i and powers $\{\beta_j\}_{j \neq i}$ such that $P_i = \prod_{j \neq i} P_j^{\beta_j}$.

Proof. Consider the powers $\alpha_1, \ldots, \alpha_r$. If there is an i such that α_i is co-prime to p, then we may find $\beta_i \in \mathbb{N}$ such that $\alpha_i \beta_i = 1 \pmod{p^{d_i}}$. Taking the equality in the statement of the lemma to the power β_i and re-arranging gives that

$$P_i = \prod_{j \neq i} P_j^{-\alpha_j \beta_j}.$$

We assume henceforth that each one of the α_i's is divisible by p, and write $\alpha_i = p^{a_i} \alpha'_i$ where α'_i is co-prime to p. Write $\mathsf{ord}(P_i) = p^{d_i - s_i}$ and

$$P_i = e^{\frac{\langle p^{s_i} c(i), \sigma_{\mathsf{master}, j_i}(x) \rangle}{p^{d_i}} 2\pi \mathbf{i}}$$

where j_i is some index, d_i is an integer and $c(i)$ is not divisible by p. Let $d = \max d_i$; we get that

$$\sum_{i=1}^{m} \alpha'_i p^{a_i + s_i + d - d_i} \langle c(i), \sigma_{\mathsf{master}, j_i}(x) \rangle = 0 \pmod{p^d}. \tag{13}$$

We note that $0 < a_i + s_i + d - d_i < d$ for all i as $\alpha_i < \mathsf{ord}(P_i)$ and α_i is divisible by p. Next, we argue that all of the functions P_i use the same coordinate of the master embedding.

Claim 4.19. *There is j such that $j_i = j$ for all i.*

Proof. Suppose this is not the case, and without loss of generality $j_1 \neq j_2$. Pick a coordinate k such that $c(1)_k$ is co-prime to k, and fix the input x in (13) on all other coordinates so that we get from (13) that

$$Ac(1)_k \sigma_{\mathsf{master}, j_1}(x_k) + Bc(2)_k \sigma_{\mathsf{master}, j_2}(x_k) = C \pmod{p^d}$$

for all $x_k \in \Sigma$, where A and B are constants smaller than p^d and C is some constant. We get that σ_{master} contains redundancies, and contradiction. ☐

Thus, (13) simplifies to

$$\sum_{i=1}^{r} \alpha'_i p^{a_i + s_i + d - d_i} \langle c(i), \sigma_{\mathsf{master}, j}(x) \rangle = 0 \pmod{p^d}. \tag{14}$$

Using the fact that the image of σ_{master} contains the 0-element and that the image of σ_{master} generates the whole group (see Section 4.2.5), we get that

$$\sum_{i=1}^{r} \alpha'_i p^{a_i + s_i + d - d_i} c(i) = 0 \pmod{p^d}.$$

Take i minimizing a_i, and denote by $(\alpha'_i)^{-1}$ an integer such that $(\alpha'_i)(\alpha'_i)^{-1} = 1 \pmod{p^d}$. Then

$$\begin{aligned}
P_i = e^{\frac{p^{a_i + s_i + d - d_i} \langle c(i), \sigma_{\mathsf{master}, j}(x) \rangle}{p^{a_i + d}} 2\pi \mathbf{i}} &= \prod_{k \neq i} e^{-\frac{p^{a_k + s_k + d - d_k} \alpha'_k (\alpha'_i)^{-1} \langle c(k), \sigma_{\mathsf{master}, j}(x) \rangle}{p^{a_i + d}} 2\pi \mathbf{i}} \\
&= \prod_{k \neq i} e^{-\frac{p^{a_k + s_k - a_i} \alpha'_k (\alpha'_i)^{-1} \langle c(k), \sigma_{\mathsf{master}, j}(x) \rangle}{p^{d_k}} 2\pi \mathbf{i}} \\
&= \prod_{k \neq i} P_k^{-p^{a_k - a_i} \alpha_k (\alpha'_i)^{-1}},
\end{aligned}$$

concluding the proof. ☐

Next, the following lemma handles the case of approximate equality. In this case, things are more complicated, and there are two options. A first conclusion may be, as in the exact equality case, that there is a P_i in our collection that is very close to some product of the other functions in our collection. This may fail however, and another possibility is that we may replace one of the P_i's in our system with a different cyclic embedding function P_i' such that the partitions induced by the two collections are very similar, and additionally $\mathsf{ord}(P_i') < \mathsf{ord}(P_i)$.

Lemma 4.20. *Suppose that $\mathcal{P} = \{P_1, \ldots, P_r\}$ is a collection of cyclic embedding functions of the additive groups $\mathbb{Z}_{p^{d_1}}, \ldots, \mathbb{Z}_{p^{d_r}}$ respectively, and suppose that $\mathsf{rk}(\mathcal{P}) \leqslant T$. Then at least one of the following holds:*

1. *There exists i and powers $\{\beta_j\}_{j \neq i}$ such that*

$$\Delta_{\mathsf{symbolic}}(P_i, \prod_{j \neq i} P_j^{\beta_j}) \leqslant T.$$

2. *There exists i and a cyclic embedding function P_i' such that $\mathsf{ord}(P_i') < \mathsf{ord}(P_i)$ and*

$$\Delta_{\mathsf{symbolic}}(P_i, P_i') \leqslant T.$$

Proof. By definitions, there are powers $\alpha_1 < \mathsf{ord}(P_1), \ldots, \alpha_r < \mathsf{ord}(P_r)$ not all 0 and a subset $I \subseteq [n]$ of size at least $n - T$ such that

$$P_1^{\alpha_1}|_I \cdots P_1^{\alpha_r}|_I = 1.$$

Define $Q_i = P_i|_I$ and note that $\mathsf{ord}(Q_i) \leqslant \mathsf{ord}(P_i)$. Define $\alpha_i' = \alpha_i \pmod{\mathsf{ord}(Q_i)}$, and note that we get that $Q_1^{\alpha_1'} \cdots Q_r^{\alpha_r'}$. There are two cases:

1. If there is some α_i' that is 0, then we conclude that $\mathsf{ord}(Q_i) < \mathsf{ord}(P_i)$. Thus, we may replace P_i by Q_i and get that the second conclusion holds.

2. Else, all of the α_i''s are non-zero. Applying Lemma 4.18 we conclude that there is i and powers $\{\beta_j\}_{j \neq i}$ such that

$$Q_i = \prod_{j \neq i} Q_j^{\beta_j},$$

   and so by definition $\Delta_{\mathsf{symbolic}}(P_i, \prod_{j \neq i} P_j^{\beta_j}) \leqslant T$. □

With Lemma 4.20 in hand, we can now state and prove a version of it that applies to general collections of cyclic functions (and not only ones that are associated with the same prime p).

Lemma 4.21. *Suppose that $\mathcal{P} = \{P_1, \ldots, P_r\}$ is a collection of cyclic embedding functions of $\mathbb{Z}_{{p_1}^{d_1}}, \ldots, \mathbb{Z}_{{p_r}^{d_r}}$ respectively, and suppose that $\mathsf{rk}(\mathcal{P}) \leqslant T$. Then at least one of the following holds:*

1. *There exists i and powers $\{\beta_j\}_{j \neq i}$ such that*

$$\Delta_{\mathsf{symbolic}}(P_i, \prod_{j \neq i} P_j^{\beta_j}) \leqslant T.$$

2. *There exists i and a cyclic embedding function P_i' such that $\mathsf{ord}(P_i') < \mathsf{ord}(P_i)$ and*

$$\Delta_{\mathsf{symbolic}}(P_i, P_i') \leqslant T.$$

Proof. Pick a prime p such that $p_i = p$ for at least a single i, let $I = \{\, i \mid p_i = p\}$ and define $a = \prod_{i : p_i \neq p} {p_i}^{d_i}$. Then

$$\Delta_{\mathsf{symbolic}}(P_1^{\alpha_1 a} \cdots P_r^{\alpha_r a}, 1) \leqslant \Delta_{\mathsf{symbolic}}(P_1^{\alpha_1} \cdots P_r^{\alpha_r}, 1) \leqslant T.$$

On the other hand, note that if $i \notin I$, then

$$P_i^{\alpha_i a} = (P_i^{p_i^{d_i}})^{\alpha_i a / p_i^{d_i}} = 1,$$

and it follows that

$$\Delta_{\mathsf{symbolic}}(\prod_{i \in I} P_i^{\alpha_i a}, 1).$$

Let $d = \max_{i \in I} d_i$. Note that as a is co-prime to p we may find an integer b such that $ab = 1 \pmod{p^d}$. It follows that

$$\Delta_{\mathsf{symbolic}}(\prod_{i \in I} P_i^{\alpha_i}, 1) = \Delta_{\mathsf{symbolic}}(\prod_{i \in I} P_i^{\alpha_i ab}, 1) \leqslant \Delta_{\mathsf{symbolic}}(\prod_{i \in I} P_i^{\alpha_i a}, 1) \leqslant T.$$

The proof is concluded by applying Lemma 4.20. □

4.5 Low Degree Functions

In this section, we collect a few facts about low-degree functions and related notions that we shall need. The first fact asserts that if a function f is close to a low-degree function, then the values of $f(x)$ and $f(y)$ are close if we sample x and $y \sim \mathrm{T}_{1-\delta}$; also, the statement holds in the other direction.

Fact 4.22. *Let $f\colon \Sigma^n \to \mathbb{C}$ be a function.*

1. *For all $\delta > 0$ and $d \in \mathbb{N}$,*

$$\mathop{\mathbb{E}}_{\substack{x \\ y \sim \mathrm{T}_{1-\delta} x}} \left[|f(x) - f(y)|^2 \right] \leqslant 2d\delta + 2W_{\geqslant d}[f].$$

2. *Suppose that $\mathop{\mathbb{E}}_{\substack{x \\ y \sim \mathrm{T}_{1-\delta} x}} \left[|f(x) - f(y)|^2 \right] \leqslant \xi$ for some δ. Then $W_{\geqslant 1/\delta}[f] \leqslant \xi$.*

Proof. We begin with the first item. Writing the degree decomposition $f = \sum_{i=0}^{n} f^{=i}$ and using Parseval, we ge that

$$\mathop{\mathbb{E}}_{\substack{x \\ y \sim \mathrm{T}_{1-\delta} x}} \left[|f(x) - f(y)|^2 \right] = 2\langle (I - \mathrm{T}_{1-\delta}) f, f \rangle = \sum_{i=0}^{n} 2(1 - (1-\delta)^i) \|f^{=i}\|_2^2,$$

and separating the sum into $i \leqslant d$ and $i > d$ finishes the proof.

For the second item, doing the same computation we get that for $d \geqslant 1/\delta$ it holds that

$$W_{\geqslant d}[g] \leqslant \frac{1}{2(1 - e^{-1})} \sum_{i=0}^{n} 2(1 - (1-\delta)^i) \|f^{=i}\|_2^2 \leqslant \xi,$$

and the proof is concluded. □

If $f_1, \dots, f_s \colon \Sigma^n \to \mathbb{C}$ are functions of degree at most d, then it is clear that $\prod_{i=1}^{s} f_i$ is a function of degree at most sd. What if the functions $f_1, \dots, f_s$ are not degree d functions, but instead are very close to degree d functions in L_2-norm? In the following lemma, we show that if $f_1, \dots, f_s$ are bounded, then $f_1 \cdots f_s$ is still close to a low-degree function.

Lemma 4.23. *Suppose that $f_1, \dots, f_s \colon \Sigma^n \to \mathbb{C}$ are 1-bounded functions such that $W_{\geqslant d}[f_i] \leqslant \xi$ for all i. Then $W_{\geqslant d/\xi}\left[\prod_{i=1}^{s} f_i\right] \leqslant 4s^2\xi$.*

Proof. Define $g = \prod_{i=1}^{s} f_i$, and let $\delta > 0$ be a parameter to be determined. We show that

$$\|g - \mathrm{T}_{1-\delta} g\|_2^2 \leqslant \mathop{\mathbb{E}}_{\substack{x \\ y \sim \mathrm{T}_{1-\delta} x}} \left[|g(x) - g(y)|^2\right].$$

Define $h_i(x, y) = \prod_{j=1}^{i} f_j(x) \prod_{j=i+1}^{s} f_j(y)$. Then we get that

$$\|g - \mathrm{T}_{1-\delta} g\|_2^2 \leqslant \mathop{\mathbb{E}}_{\substack{x \\ y \sim \mathrm{T}_{1-\delta} x}} \left[\left|\sum_{i=0}^{s-1} h_{i+1}(x,y) - h_i(x,y)\right|^2\right] \leqslant s \sum_{i=0}^{s-1} \mathop{\mathbb{E}}_{\substack{x \\ y \sim \mathrm{T}_{1-\delta} x}} \left[|h_{i+1}(x,y) - h_i(x,y)|^2\right],$$

where we used Cauchy-Schwarz. Using the fact that each f_i is 1-bounded, we get that $|h_{i+1}(x,y) - h_i(x,y)| \leqslant |f_{i+1}(x) - f_{i+1}(y)|$, and so we get that

$$\|g - \mathrm{T}_{1-\delta} g\|_2^2 \leqslant s^2 \max_i \mathop{\mathbb{E}}_{\substack{x \\ y \sim \mathrm{T}_{1-\delta} x}} \left[|f_i(x) - f_i(y)|^2\right] \leqslant 2s^2(d\delta + \xi),$$

where in the last inequality we used Fact 4.22. By Fact 4.22 it follows that $W_{\geqslant 1/\delta}[g] \leqslant 2s^2(d\delta + \xi)$, and the proof is concluded by picking $\delta = \xi/d$. □

5 A Regularity Lemma

In this section, we state and prove a regularity lemma that will be crucial for our mixed invariance principle.

5.1 The Noise Operator with Respect to Sigma Algebras

In the following definition, we give a variant of the standard noise operator over product spaces that will be crucial for our regularity lemma.

Definition 5.1. *Let Σ be a finite alphabet, let ν be a distribution over Σ, let $\mathcal{P} = \{P_1, \dots, P_r \colon \Sigma^n \to \mathbb{C}\}$ be a collection of functions, and let $\varepsilon > 0$. For each $x \in \Sigma^n$ we define the distribution $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon} x$ to be the distribution sampled as:*

1. *Sample $I \subseteq_{\varepsilon} [n]$.*
2. *Sample $y \sim \nu^{\otimes n}$ conditioned on $y_{\overline{I}} = x_{\overline{I}}$ and $P_i(y) = P_i(x)$ for all i.*

3. *Output* y.

Often times the distribution ν *will be clear from context and we will omit it from the notation.*

As is usually the case, we will associated with $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$ an averaging operator over functions, which by abuse of notation we denote as $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}\colon L_2(\Sigma^n,\nu^{\otimes n})\to L_2(\Sigma^n,\nu^{\otimes n})$ and define as

$$\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}f(x)=\mathop{\mathbb{E}}_{y\sim \mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}x}[f(y)].$$

We will use noise operators as a replacement for harsh truncations, as it is important for us to come up with an operation that is truncation-like yet preserves the boundedness of functions. In this section, we collect a few basic properties of the noise operator $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$.

5.1.1 A Stationary Distribution of $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$

We begin by stating a few basic properties of the operator $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$. The following claim asserts that $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$ is a Markov chain over Σ^n and $\nu^{\otimes n}$ is a stationary distribution of it. We remark that whenever the set $\mathcal{P}$ is non-empty, the Markov chain $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$ is disconnected.

Fact 5.2. *Sampling* $x\sim\nu^{\otimes n}$ *and then* $y\sim \mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}x$*, the distribution of* y *is* $\nu^{\otimes n}$*.*

Proof. Write $\mathcal{P}=\{P_1,\ldots,P_m\}$, fix a point $w\in\Sigma^n$ and calculate the probability that $y=w$. Denote $a_i=P_i(w_i)$, and fix I a choice of a subset of coordinates in the process that samples y conditioned on x. Then

$$\Pr_{x,y}[y=w\mid I]=\Pr_{x,y}[P_i(x)=a_i\forall i,x_I=w_I\mid I]\Pr_{x,y}[y=w\mid I,P_i(x)=a_i\forall i,x_I=w_I].$$

Let A be the set of $x\in\Sigma^n$ such that $P_i(x)=a_i$ for all i and $x_I=w_I$. Then by sampling of y we have that $\Pr_{x,y}[y=w\mid P_i(x)=a_i\forall i,x_I=w_I]=\nu(w)/\nu(A)$, as y is sampled according to ν conditioned on $y\in A$. Also, $\Pr_{x,y}[P_i(x)=a_i\forall i,x_I=w_I]=\nu(A)$, and multiplying gives that $\Pr_{x,y}[y=w]=\nu(w)$. Since this is true for every I, it follows that it is true for any distribution over I, and thus we get that $\Pr_{x,y}[y=w]=\nu(w)$. ☐

5.1.2 Relating Different Noise Operators

In the next fact, we show that if P' is close to $\mathsf{spn}_{\mathbb{N}}(\mathcal{P})$, then the operators $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$ and $\mathrm{T}_{\nu,\mathcal{P}\cup\{P'\},1-\varepsilon}$ are close.

Fact 5.3. *Let* $\mathcal{P}$ *be a collection, let* $\varepsilon>0$*, let* $P'\colon\Sigma^n\to\mathbb{C}$ *and suppose that* $k=\min_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})}\Delta(P,P')$*. Then*

1. *There is a coupling of* (x,y,y') *such that* (x,y) *is distributed according to* $(x,\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}x)$*,* (x,y') *is distributed according to* $(x,\mathrm{T}_{\nu,\mathcal{P}\cup\{P'\},1-\varepsilon}x)$ *and* $\Pr[y\neq y']\leqslant k\varepsilon$*.*

2. *For any* 1*-bounded function* $f\colon\Sigma^n\to\mathbb{N}$*,*

$$\|\mathrm{T}_{\nu,\mathcal{P}\cup\{P'\},1-\varepsilon}f-\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}f\|_2\leqslant 2\sqrt{k\varepsilon}.$$

Proof. We begin with the first item. Let $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$ be the P achieving the minimum, and let $K \subseteq [n]$ be the set of coordinates where P and P' differ. Sample $x \sim \nu$, then $I \subseteq_{1-\varepsilon} [n]$ and then $y \sim \nu$ conditioned on $y_I = x_I$ and $P_i(y) = P_i(x)$ for all i; note that y is distributed according to $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}x$. Let $y' \sim \mathrm{T}_{\nu,\mathcal{P}\cup\{P'\},1-\varepsilon}x$. Note that if $I \cap K = \emptyset$, then $P'(y) = P'(x)$. Thus, letting E be the event that $I \cap K = \emptyset$, we have that $\Pr[E] \geqslant 1 - k\varepsilon$ and that the distributions of $y \mid E$ and $y' \mid E$ are identical. Thus, the statistical distance between y and y' is at most $\Pr[E] \leqslant k\varepsilon$. It follows that there is a coupling between y and y' such that $\Pr[y \neq y'] \leqslant k\varepsilon$.

For the second item, fix the coupling (x, y, y') so that we may write

$$\|\mathrm{T}_{\nu,\mathcal{P}\cup\{P'\},1-\varepsilon}f - \mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}f\|_2^2 = \mathop{\mathbb{E}}_{x}\left[\left|\mathop{\mathbb{E}}_{y,y'}\left[f(y) - f(y')\right]\right|^2\right] \leqslant 4\mathop{\mathbb{E}}_{x}\left[1_{y\neq y'}\right] \leqslant 4k\varepsilon. \qquad \square$$

5.1.3 Nearly Preserving Low Degree Functions

The first of which says that $\mathrm{T}_{\mathcal{P},1-\varepsilon}$ roughly preserves low-degree functions. Toward this end, we introduce the function

$$A_{\mathcal{P}}(x) = \mathop{\Pr}_{\substack{I\subseteq_{\varepsilon}[n]\\ x'\sim\nu^{\otimes n}}}\left[P_i(x') = P_i(x)\ \forall i \,\middle|\, x'_{\overline{I}} = x_{\overline{I}}\right] = \mathop{\Pr}_{x'\sim \mathrm{T}_{1-\varepsilon}x}\left[P_i(x') = P_i(x)\ \forall i\right]$$

that measures how likely a random noisy step from x is to leave the values of all $P_i \in \mathcal{P}$ unchanged. We have the following fact:

Fact 5.4. *Suppose that Σ is an alphabet of size at most m, $\mathcal{P} = \{P_1, \ldots, P_r\}$ is a collection of r product functions and ν is a distribution over Σ. Then for all $\tau > 0$ we have that*

$$\mathop{\Pr}_{x\sim\nu^{\otimes n}}\left[A_{\mathcal{P}}(x) \leqslant \tau\right] \leqslant O_{r,m}(\tau).$$

Proof. For each $a \in \prod_{i=1}^{r} \mathsf{Image}(P_i)$ define

$$B_{\vec{a}} = \{x \in \Sigma^n \mid P_i(x) = a_i\ \forall i = 1, \ldots, r\}, \quad B'_{\vec{a}} = \left\{x \in B_{\vec{a}} \mid \mathop{\Pr}_{y\sim\mathrm{T}_{1-\varepsilon}}\left[P_i(y) = a_i\ \forall i = 1, \ldots, r\right] \leqslant \tau\right\}.$$

In words, $B'_{\vec{a}}$ is a subset of $B_{\vec{a}}$ of elements from which a step according to the random walk of $\mathrm{T}_{1-\varepsilon}$ escapes outside the set with noticeable probability. Note that

$$\begin{aligned}\mathop{\Pr}_{x,y\sim\mathrm{T}_{1-\varepsilon}}\left[x, y \in B_{\vec{a}'}\right] = \langle 1_{B_{\vec{a}'}}, \mathrm{T}_{1-\varepsilon}1_{B_{\vec{a}'}}\rangle = \langle \mathrm{T}_{\sqrt{1-\varepsilon}}1_{B_{\vec{a}'}}, \mathrm{T}_{\sqrt{1-\varepsilon}}1_{B_{\vec{a}'}}\rangle &= \|\mathrm{T}_{\sqrt{1-\varepsilon}}1_{B_{\vec{a}'}}\|_2^2 \\ &\geqslant \|\mathrm{T}_{\sqrt{1-\varepsilon}}1_{B_{\vec{a}'}}\|_1^2,\end{aligned}$$

which is equal to $\nu(B_{\vec{a}'})^2$. On the other hand,

$$\mathop{\Pr}_{x,y\sim\mathrm{T}_{1-\varepsilon}}\left[x, y \in B_{\vec{a}'}\right] = \nu(B_{\vec{a}'})\mathop{\Pr}_{x\in B_{\vec{a}'},y\sim\mathrm{T}_{1-\varepsilon}}\left[y \in B_{\vec{a}'}\right] \leqslant \nu(B_{\vec{a}'})\mathop{\Pr}_{x\in B_{\vec{a}'},y\sim\mathrm{T}_{1-\varepsilon}}\left[y \in B_{\vec{a}}\right] \leqslant \nu(B_{\vec{a}'})\tau,$$

and combining the two inequalities we get that $\nu(B_{\vec{a}'}) \leqslant \tau$. We define $X_{\mathsf{bad}} = \bigcup_{\vec{a}\in\prod_{i=1}^{r}\mathsf{Image}(P_i)} B_{\vec{a}'}$, and get that as the size of the image of each one of the P_i's is $O_m(1)$ it holds that

$$\mathop{\Pr}_{x\sim\nu^{\otimes n}}\left[A_{\mathcal{P}}(x) \leqslant \tau\right] \leqslant \nu(X_{\mathsf{bad}}) \lesssim_m \tau. \qquad \square$$

Claim 5.5. *Let $d, r, m \in \mathbb{N}$ and $\alpha, \varepsilon > 0$. Suppose that Σ has size at most m, ν is a distribution over Σ in which the probability of each atom is at least α, and $\mathcal{P} = \{P_1, \ldots, P_r\}$ is a collection of product functions from $\mathcal{P}(\Sigma, G, \sigma)$ where G is a group of size at most m. Then for any $L\colon \Sigma^n \to \mathbb{C}$ of degree at most d we have that*

$$\|(I - \mathrm{T}_{\mathcal{P},1-\varepsilon})L\|_2^2 \lesssim_{d,m,r,\alpha} \varepsilon^{1/3}\|L\|_2^2.$$

Proof. Normalizing, we assume that $\|L\|_2 = 1$. The left-hand side is equal to

$$\mathop{\mathbb{E}}_{x}\left[\left|\mathop{\mathbb{E}}_{y\sim \mathrm{T}_{\mathcal{P},1-\varepsilon}x}[L(x) - L(y)]\right|^2\right] \leqslant \mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{y\sim \mathrm{T}_{\mathcal{P},1-\varepsilon}x}[|L(x) - L(y)|]^2\right], \tag{15}$$

and in the rest of the proof we bound the right-hand side. Let $\tau > 0$ be a parameter to be chosen. We break right hand size of (15) and break it to x such that $A_{\mathcal{P}}(x) \geqslant \tau$ and x such that $A_{\mathcal{P}}(x) < \tau$:

$$\underbrace{\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{y\sim \mathrm{T}_{\mathcal{P},1-\varepsilon}x}[|L(x) - L(y)|]^2 1_{A_{\mathcal{P}}(x)<\tau}\right]}_{(I)} + \underbrace{\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{y\sim \mathrm{T}_{\mathcal{P},1-\varepsilon}x}[|L(x) - L(y)|]^2 1_{A_{\mathcal{P}}(x)\geqslant\tau}\right]}_{(II)}.$$

For (I), we have by Cauchy-Shcwartz that

$$(I) \leqslant \sqrt{\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{y\sim \mathrm{T}_{\mathcal{P},1-\varepsilon}x}[|L(x) - L(y)|]^4\right]}\sqrt{\Pr_{x\sim\nu^{\otimes n}}[A_{\mathcal{P}}(x) \leqslant \tau]} \lesssim \|L\|_4^2\sqrt{\Pr_{x\sim\nu^{\otimes n}}[A_{\mathcal{P}}(x) \leqslant \tau]}.$$

By hypercontractivity $\|L\|_4^2 \lesssim_{\alpha,d} \|L\|_2^2 = 1$, and combining with Fact 5.4 we conclude that $(I) \lesssim_{\alpha,m,d} \sqrt{\tau}$. For (II), we have that

$$\begin{aligned}(II) &= \mathop{\mathbb{E}}_{x}\left[\frac{\mathbb{E}_{I\subseteq_\varepsilon[n], y\sim\nu^{\otimes n}}\left[|L(y) - L(x)|^2 1_{P_i(y)=P_i(x)\ \forall i} \,\middle|\, y_{\overline{I}} = x_{\overline{I}}\right]}{\Pr_{I\subseteq_\varepsilon[n], y\sim\nu^{\otimes n}}\left[P_i(y) = P_i(x)\ \forall i \mid y_{\overline{I}} = x_{\overline{I}}\right]} 1_{A_{\mathcal{P}}(x)\geqslant\tau}\right] \\ &= \mathop{\mathbb{E}}_{x}\left[\frac{\mathbb{E}_{I\subseteq_\varepsilon[n], y\sim\nu^{\otimes n}}\left[|L(y) - L(x)|^2 1_{P_i(y)=P_i(x)\ \forall i} \,\middle|\, y_{\overline{I}} = x_{\overline{I}}\right]}{A_{\mathcal{P}}(x)} 1_{A_{\mathcal{P}}(x)\geqslant\tau}\right]\end{aligned}$$

We conclude that (II) is upper bounded by

$$\frac{1}{\tau}\mathop{\mathbb{E}}_{x}\left[\mathop{\mathbb{E}}_{y\sim\mathrm{T}_{1-\varepsilon}x}\left[|L(y) - L(x)|^2\right]\right] = \frac{1}{\tau}\left(2 - 2\langle L, \mathrm{T}_{1-\varepsilon}L\rangle\right) = \frac{2}{\tau}\left(1 - \|\mathrm{T}_{\sqrt{1-\varepsilon}}L\|_2^2\right) \leqslant \frac{2}{\tau}\left(1 - (1-\varepsilon)^d\right),$$

which is at most $\frac{2d}{\tau}\varepsilon$. Combining, we get that

$$(15) \lesssim_{\alpha,m,d} \sqrt{\tau} + \frac{\varepsilon}{\tau},$$

and choosing $\tau = \varepsilon^{2/3}$ finishes the proof. ☐

5.2 An Approximating Formula for $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon} f$

Note that any function $P \in \mathsf{sp}_{\mathbb{N}}(\mathcal{P})$ is an eigenvector of $T_{\mathcal{P},1-\varepsilon}$ of eigenvalue 1. Thus, by Claim 5.5 functions of the form $P \cdot L$ are nearly preserved under the operator $T_{\mathcal{P},1-\varepsilon}$ (provided that the degree of L is sufficiently small compared to $1/\varepsilon$). In this section, we show that for any function f, $\mathrm{T}_{\mathcal{P},1-\varepsilon} f$ may be approximated as a linear combination of such functions. This statement, which is Lemma 5.8 below, will be very important for us in at least two contexts:

1. First, in Section 5 we will use this approximate formula in the proof of our μ-regularity lemma.
2. Second, in Section 6.3 we will use it to state and prove our mixed invariance principle.

Towards getting such an approximation we need more basic properties of the function $A_{\mathcal{P}}(x)$. In Claim 5.4 we proved that its values are almost always positive numbers bounded away from 0. In the next fact, we show that $A_{\mathcal{P}}(x)$ is close to a linear combination of functions of the form $P \cdot L$ for low-degree functions L. For technical purposes, we need a slightly different claim, giving low-degree-like functions L that are still bounded:

Fact 5.6. *For all $r, m \in \mathbb{N}$, $\varepsilon > 0$ and $\xi > 0$ and any collection of product functions $\mathcal{P}$, the function $A_{\mathcal{P}}$ may be written as*

$$A_{\mathcal{P}}(x) = \sum_{P \in \mathsf{sp}_{\mathbb{N}}(\mathcal{P})} P \cdot \tilde{L}_P,$$

where for every P the function $\tilde{L}_P$ is $O_{m,r}(1)$ and for all d, $W_{\geqslant d}(\tilde{L}_P) \leqslant O_{m,r}((1-\varepsilon)^d)$.

Proof. Note that as all of the functions P_i's get the values of characters over an Abelian group of size at most m, we have that in the set $I = \bigcup_i \mathsf{Image}(P_i)$ the distance between any two distinct points is $\Omega_m(1)$. Thus, we may find a bi-variate polynomial $Q(z_1, z_2)$ of degree at most $O_m(1)$ and coefficients that are at most $O_m(1)$ in absolute value such that $Q(z_1, z_2) = 1$ if $z_1 = z_2 \in I$, $Q(z_1, z_2) = 0$ if $z_1, z_2 \in I$ are distinct.[11] Using Q, we may write

$$A_{\mathcal{P}}(x) = \mathop{\mathbb{E}}_{y \sim \mathrm{T}_{\mathcal{P},1-\varepsilon}} \left[\prod_{i=1}^{r} Q(P_i(y), P_i(x)) \right].$$

Write $Q(z_1, z_2) = \sum_{j,k=0}^{d} a_{j,k} z_1^j z_2^k$ for $d = O_m(1)$ and plug it in above to get that

$$\begin{aligned} A_{\mathcal{P}}(x) &= \sum_{j_1,\ldots,j_r,k_1,\ldots,k_r=0}^{d} a_{j_1,k_1} \cdots a_{j_r,k_r} P_1(x)^{k_1} \cdots P_r(x)^{k_r} \mathop{\mathbb{E}}_{y \sim \mathrm{T}_{1-\varepsilon}} \left[P_1(y)^{j_1} \cdots P_r(y)^{j_r} \right] \\ &= \sum_{j_1,\ldots,j_r,k_1,\ldots,k_r=0}^{d} a_{j_1,k_1} \cdots a_{j_r,k_r} P_1(x)^{k_1} \cdots P_r(x)^{k_r} \mathrm{T}_{1-\varepsilon}(P_1^{j_1} \cdots P_r^{j_r})(x). \end{aligned}$$

□

[11] A polynomial Q satisfying these properties may be constructed via Lagrange interpolation for example.

In our argument, we will actually want to show that the function $1/A_{\mathcal{P}}(x)$ is close to being a linear combination of functions of the type $P \cdot L$. A slight concern is about values of x where $A_{\mathcal{P}}(x)$ since such behavior is not typical for low-degree functions. The following fact asserts that when we discount the contribution of such inputs x, $1/A_{\mathcal{P}}(x)$ may be approximated by a linear combination of functions of the type $P \cdot L$; this fact is a relatively easy consequence of Fact 5.6.

Fact 5.7. *For all $m \in \mathbb{N}$, $\tau, \varepsilon > 0$ and $\xi > 0$, the following holds for sufficiently large $D \in \mathbb{N}$. There is $A'' \colon \Sigma^n \to \mathbb{C}$ with the following properties:*

1. *A'' may be written as $A''(x) = \sum_{P \in \mathsf{sp}_{\mathbb{N}}(\mathcal{P})} P \cdot \tilde{L}_P$ where for all P, $\tilde{L}_P$ is $O_{m,\xi,\tau,r}(1)$ bounded and $W_{\geqslant 2^{2D}}[\tilde{L}_P] \leqslant 2^{-\Omega(D)}$.*
2. *$\mathbb{E}_x\left[\left|\frac{1}{A_{\mathcal{P}}(x)} - A''(x)\right|^2 1_{A_{\mathcal{P}}(x) \geqslant \tau}\right] \leqslant \xi$.*

Proof. Fix x such that $A_{\mathcal{P}}(x) \geqslant \tau$, and note that clearly $A_{\mathcal{P}}(x) \leqslant 1$. Thus, we may write

$$\frac{1}{A_{\mathcal{P}}(x)} = \frac{1}{1-(1-A_{\mathcal{P}}(x))} = \sum_{k=0}^{\infty}(1-A_{\mathcal{P}}(x))^k = \sum_{k=0}^{T}(1-A_{\mathcal{P}}(x))^k + e^{-\tau T},$$

where we take $T = \log(2/\xi)/\tau$. Define $A''(x) = \sum_{k=0}^{T}(1-A_{\mathcal{P}}(x))^k$; then the second item is clear.

Using the formula for $A_{\mathcal{P}}$ from Fact 5.6, expanding A'' and re-grouping terms we get that it may be written as a linear combination of functions of the form $P \cdot \tilde{L}_1 \cdots \tilde{L}_{T+1}$ where $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$ and each $\tilde{L}_j$ is $O_{m,r}(1)$ bounded and $W_{\geqslant d}[\tilde{L}_j] \ll_{m,r} (1-\varepsilon)^d$ for all d. Take $d = D/\varepsilon$ for D to be determined; it follows from Lemma 4.23 that $W_{\geqslant d2^D}[\tilde{L}_1 \cdots \tilde{L}_{T+1}] \lesssim_T 2^{-\Omega(D)}$. Re-grouping, we may write

$$A''(x) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P\tilde{L}_P$$

where $\tilde{L}_P$ is $O_{m,T,r}(1) = O_{m,\xi,\tau,r}(1)$-bounded and $W_{\geqslant d2^D}[\tilde{L}_P] \lesssim_{m,\xi,\tau,r} 2^{-\Omega(D)}$. ☐

Lemma 5.8. *For all $m, r \in \mathbb{N}$, $\alpha, \varepsilon > 0$ and $\xi > 0$ there exist $C, D \in \mathbb{N}$ such that the following holds. Suppose that Σ has size at most m, ν is a distribution over Σ in which the probability of each atom is at least α, and $\mathcal{P} = \{P_1, \ldots, P_r\}$ is a collection of product functions from $\mathcal{P}(\Sigma, G, \sigma)$ where G is a group of size at most m. Then for all 1-bounded $f \colon \Sigma^n \to \mathbb{C}$, there exists a function $f' \colon \Sigma^n \to \mathbb{C}$ such that:*

1. *The function f' approximates $\mathrm{T}_{\mathcal{P},1-\varepsilon} f$: $\|\mathrm{T}_{\mathcal{P},1-\varepsilon} f - f'\|_2 \leqslant \xi$.*
2. *The function f' can be written as*

$$f'(x) = \sum_{P \in \mathsf{sp}_{\mathbb{N}}(\mathcal{P})} P(x) \cdot L_P(x)$$

where for all P, $\mathsf{deg}(L_P) \leqslant D$ and $\|L_P\|_2 \leqslant C$.

Proof. Take the parameters

$$0 < D^{-1} \ll \zeta \ll \tau \ll r^{-1}, m^{-1}, \alpha, \varepsilon, \xi \leqslant 1.$$

Let $\tau' \in [\tau, \tau + \sqrt{\tau}]$ be a parameter to be determined, define $F_1(x) = \mathrm{T}_{\mathcal{P},1-\varepsilon} f(x) 1_{A_{\mathcal{P}}(x) \geqslant \tau'}$ and note that

$$\|\mathrm{T}_{\mathcal{P},1-\varepsilon} f - F_1\|_2 = \|\mathrm{T}_{\mathcal{P},1-\varepsilon} f(x) 1_{A_{\mathcal{P}}(x) \leqslant \tau}\|_2 \leqslant \|1_{A_{\mathcal{P}}(x) \leqslant \tau'}\|_2 \lesssim_{m,r} \sqrt{\tau'} \leqslant \tau^{1/8} \tag{16}$$

where we used Fact 5.4. Next, expanding the definition we get that

$$\begin{aligned}\mathrm{T}_{\mathcal{P},1-\varepsilon} f(x) &= \mathbb{E}_{\substack{I \subseteq_\varepsilon [n] \\ x' \sim \nu^{\otimes n}}} \left[f(x') \,\middle|\, x'_{\overline{I}} = x_{\overline{I}}, P_i(x') = P_i(x)\ \forall i \right] \\ &= \frac{1}{A_{\mathcal{P}}(x)} \underbrace{\mathbb{E}_{\substack{I \subseteq_\varepsilon [n] \\ x' \sim \nu^{\otimes n}}} \left[f(x') 1_{P_i(x') = P_i(x)\ \forall i} \,\middle|\, x'_{\overline{I}} = x_{\overline{I}} \right]}_{(I)}.\end{aligned}$$

Take $A''(x)$ from Fact 5.7 with the parameter ξ therein being τ in the current setting. Then $\|A'' 1_{A_{\mathcal{P}} \geqslant \tau'} - \frac{1}{A_{\mathcal{P}}(x)} 1_{A_{\mathcal{P}} \geqslant \tau'}\|_2 \leqslant \tau$. Thus, as (I) is 1-bounded we conclude that for $F_2 = (I)(x) A''(x) 1_{A_{\mathcal{P}}(x) \geqslant \tau'}$ it holds that

$$\|F_1 - F_2\|_2 \leqslant \tau. \tag{17}$$

Next, note that $\|F_2\|_2^2 \leqslant 2$, therefore there exists $1 \leqslant j \leqslant 1/\sqrt{\tau}$ such that

$$\mathop{\mathbb{E}}_{x} \left[(I)(x)^2 A''(x)^2 1_{A_{\mathcal{P}}(x) \in [\tau + j\tau, \tau + (j+1)\tau]} \right] \leqslant \tau^{1/4},$$

and we fix such j. Define the continuous function $h\colon [0,1] \to [0,1]$ so that $h(t) = 0$ for $t \leqslant \tau + j\tau$, $h(t) = 1$ for $t \geqslant \tau + (j+1)\tau$ and we linearly interpolate between the two ranges. Also, take $\tau' = \tau + j\tau$. Defining $F_3(x) = (I)(x) A''(x) h(A_{\mathcal{P}}(x))$, we get that

$$\|F_2 - F_3\|_2 \leqslant \sqrt{\mathop{\mathbb{E}}_{x} \left[(I)(x)^2 A''(x)^2 1_{A_{\mathcal{P}}(x) \in [\tau + j\zeta, \tau + (j+1)\zeta]} \right]} \leqslant \tau^{1/8}. \tag{18}$$

At this point, our approximating function F_3 almost fits the form as needed for f', but we still need some modifications. First, we need to explore the structure of $(I)(x)$ and of $h(A_{\mathcal{P}}(x))$ further, and secondly we need to perform degree truncation.

For $h(A_{\mathcal{P}}(x))$, by Weirstrass approximation theorem we may find a polynomial $W\colon [0,1] \to \mathbb{R}$ such that $|W(t) - h(t)| \leqslant \zeta$ for all $t \in [0,1]$. We define $F_4(x) = (I)(x) A''(x) W(A_{\mathcal{P}}(x))$ and get that

$$\|F_3 - F_4\|_2 \leqslant \zeta \sqrt{\mathop{\mathbb{E}}_{x} [(I)(x)^2 A''(x)^2]} \leqslant \zeta \sqrt{\mathop{\mathbb{E}}_{x} [A''(x)^2]} \lesssim_{m,r,\tau,\varepsilon} \zeta \leqslant \sqrt{\zeta}, \tag{19}$$

where we used the fact that A'' is $O_{m,r,\tau,\varepsilon}(1)$ bounded.

The Structure of (I). As in the proof of Fact 5.6, take a bi-variate polynomial $Q(z_1, z_2)$ of degree $O_{m,r}(1)$ whose coefficients are all bounded by $O_{m,r}(1)$ in absolute value such that $Q(z_1, z_2) = 1_{z_1 = z_2}$ for every $z_1, z_2 \in \bigcup_{i=1}^{r} \mathsf{Image}(P_i)$. Then

$$(I) = \mathbb{E}_{\substack{I \subseteq_\varepsilon [n] \\ x' \sim \nu^{\otimes n}}} \left[f(x') \prod_{i=1}^{r} Q(P_i(x'), P_i(x)) \,\middle|\, x'_{\overline{I}} = x_{\overline{I}} \right].$$

Writing $Q(z_1,z_2)=\sum_{i,j=0}^{d} a_{i,j}z_1^i z_2^j$ where $|a_{i,j}|=O_{m,r}(1)$, we get that

$$(I)=\sum_{j_1,\ldots,j_r=0}^{d} P_1(x)^{i_1}\cdots P_r(x)^{i_r}\mathbb{E}_{\substack{I\subseteq_{\varepsilon}[n]\\ x'\sim\nu^{\otimes n}}}\left[\sum_{i_1,\ldots,i_r=0}^{d} a_{i_1,j_1}\cdots a_{i_r,j_r}f(x')P_1(x')^{j_1}\cdots P_r(x')^{j_r}\,\middle|\, x'_{\overline{I}}=x_{\overline{I}}\right]$$
$$=\sum_{j_1,\ldots,j_r} P_1(x)^{i_1}\cdots P_r(x)^{i_r}F_{j_1,\ldots,j_r},$$

where $F_{j_1,\ldots,j_r}(x')=T_{1-\varepsilon}\sum_{i_1,\ldots,i_r=0}^{d} a_{i_1,j_1}\cdots a_{i_r,j_r}f(x')P_1^{j_1}(x')\cdots P_r^{j_r}(x')$. Note that F_j is $O_{m,r}(1)$-bounded by the triangle inequality, and $W_{\geqslant d}[F_{j_1,\ldots,j_r}]\ll_{m,r}(1-\varepsilon)^d$ for all d. Thus, after re-arranging we may write

$$(I)=\sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P\cdot\tilde{L}_P,$$

where $W_{\leqslant 2^{2D}}[\tilde{L}_P]\leqslant 2^{-\Omega(D)}$.

Combining the Structures. Write $W(t)=\sum_{i=0}^{q} a_i t^i$, where $q,|a_i|\leqslant O_{\zeta}(1)$ for all i, so that

$$W(A_{\mathcal{P}}(x))=\sum_{i=0}^{q} a_i A_{\mathcal{P}}(x)^i.$$

Plugging in the formula for $A_{\mathcal{P}}(x)$ from Fact 5.6, we get that $W(A_{\mathcal{P}}(x))$ may be written as a linear combination of at most $O_q(1)$ may terms each of the form $aP\cdot\tilde{L}_1(x)\cdots\tilde{L}_q(x)$, where $|a|=O_{\zeta}(1)$ and for all i, $\tilde{L}_i$ is $O_{m,r}(1)$ bounded with $W_{\geqslant d}[\tilde{L}_i]\ll_{m,r}(1-\varepsilon)^d$ for each i. By Lemma 4.23, it follows that $W_{\geqslant\frac{D}{\varepsilon}2^D}[\tilde{L}_1(x)\cdots\tilde{L}_q]\lesssim_{\zeta,m,r}2^{-\Omega(D)}$, so for large enough D we have $W_{\geqslant 2^{2D}}[\tilde{L}_1(x)\cdots\tilde{L}_q]\leqslant 2^{-\Omega(D)}$. Re-arranging the resulting expression for $W(A_{\mathcal{P}}(x))$ gives that it may be written as

$$W(A_{\mathcal{P}}(x))=\sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P\cdot\tilde{L}'_P,$$

where $\tilde{L}'_P$ is $O_{\zeta}(1)$ bounded and $W_{\geqslant 2^{2D}}[\tilde{L}'_P]\lesssim_{m,r,\zeta,q}2^{-\Omega(D)}$.

By choice of A'' from Fact 5.7, we may write

$$A''(x)=\sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P\cdot\tilde{L}''_P$$

where $\tilde{L}''_P$ is $O_{m,r,\tau}(1)$ bounded and $W_{\geqslant 2^{2D}}[\tilde{L}''_P]\leqslant 2^{-\Omega(D)}$.

Combining the formulas for (I), $W(A_{\mathcal{P}}(x))$ and $A''(x)$ and multiplying out, we get that

$$F_4(x)=\sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P\cdot\tilde{L}'''_P,$$

where each $\tilde{L}'''_P$ is a linear combination of at most $O_{m,r}(1)$ products of at most 3 of the functions $\tilde{L}_Q$, $\tilde{L}'_{Q'}$ and $\tilde{L}''_{Q''}$. Thus, $\tilde{L}'''_P$ is $O_\zeta(1)$-bounded. Also, For each Q, Q' and Q'' we get by Lemma 4.23 that $W_{\geqslant 2^{3D}}[\tilde{L}_Q\tilde{L}'_{Q'}\tilde{L}''_{Q''}] \leqslant 2^{-\Omega(D)}$, and so $W_{\geqslant 2^{3D}}[\tilde{L}'''_P] \leqslant 2^{-\Omega(D)}$. Thus, defining

$$f' = \sum_{P\in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P \cdot (\tilde{L}'''_P)^{\leqslant 2^{3D}},$$

we get that

$$\|f' - F_4\|_2 \lesssim_{m,r} 2^{-\Omega(D)}. \tag{20}$$

Combining (16), (17), (18), (19) and (20) gives the second item of the lemma. □

5.3 The μ Regularity Lemma

With the tools from the previous sections in hand, we may now state a regularity lemma for the norm $\|\|\|_\mu$. In words, the regularity lemma says that we can approximate any 1-bounded function f with a function of the form $f' = \mathrm{T}_{\mathcal{P},1-\varepsilon}$, where $\mathcal{P}$ has a bounded size and ε is bounded away from 0.

5.3.1 A Basic Version of the Regularity Lemma

We begin with the following basic version of our regularity lemma, which already contains all of the essential ideas.

Lemma 5.9. *For all $\alpha > 0$, $m \in \mathbb{N}$ and $\xi > 0$ there exist $\varepsilon_0 > 0$ and $r \in \mathbb{N}$ such that the following holds. Let Σ be an alphabet of size at most m, let ν be a distribution over Σ in which the probability of each atom is at least α and let $f\colon \Sigma^n \to \mathbb{C}$ be a 1-bounded function. Then there exists a collection $\mathcal{P}$ of cyclic embedding functions of size at most r such that*

$$\|f - \mathrm{T}_{\mathcal{P},1-\varepsilon} f\|_{\nu,\alpha} \leqslant \xi.$$

The rest of this section is devoted to the proof of Lemma 5.9.

We look at all $\mu \in M_{\nu,\alpha}$ and apply Theorem 3.1 on them with the parameters α, m as here and ε there being $\xi/100$. As $M_{\nu,\alpha}$ has size $O_m(1)$, we may take δ to be the minimum over all the δ's in Theorem 3.1 and d to be the maximum of all the d's, take the Abelian group G to be the product of all the Abelian group in the applications in Theorem 3.1 and σ to be the concatenation of all of σ's from the applications of Theorem 3.1.

We fix these parameters, set $R = \lceil \frac{100}{\delta^2} \rceil$ and further use the following parameters:

$$0 < d_R^{-1} \ll \zeta_R \ll \varepsilon_R \ll \ldots \ll d_1^{-1} \ll \zeta_1 \ll \varepsilon_1 \ll d^{-1}, \delta \ll \alpha, \xi, m^{-1} \leqslant 1. \tag{21}$$

We now proceed with the following iterative process. Starting with $g_0 = \mathbb{E}[f]$, $f_0 = f - g_0$, $\mathcal{P}_0 = \emptyset$ and $i = 0$, if f_i has correlation at least δ with a function of the form $P_i \cdot L_i$ for $P_i = \chi \circ \sigma$ and $L_i\colon \Sigma^n \to \mathbb{C}$ of 2-norm 1, we do the following:

1. Using Fact 4.15, write $P_i = P_{i,1} \cdots P_{i,k}$ where each $P_{i,j}$ is a cyclic embedding function and $k = O_m(1)$. Define $\mathcal{P}_{i+1} = \mathcal{P}_i \cup \{P_{i,1}, \ldots, P_{i,k}\}$, $g_{i+1} = \mathrm{T}_{\mathcal{P}_{i+1},1-\varepsilon_{i+1}} f$ and $f_{i+1} = f - g_{i+1}$.
2. Increase i by 1 and repeat.

The following claim, stating that $\|g_{i+1}\|_2^2 \geqslant i\delta^2/2$, is the key in the proof of Lemma 5.9. Once it is established, the proof is concluded quickly.

Claim 5.10. *For all $i \leqslant R$ we have that $\|g_{i+1}\|_2^2 \geqslant i\delta^2/10$.*

Proof. Fix i, and for each of notation denote $\mathrm{T}_i = \mathrm{T}_{\mathcal{P}_i,1-\varepsilon_i}$. Thus, for $j = 1,\dots,i+1$ we have that $|\langle f_j, P_jL_j\rangle| \geqslant \delta$. Thus, we may pick a complex number θ_j of absolute value 1 such that $\langle f_j, \theta_jP_jL_j\rangle \geqslant \delta$. Noting that $f_i = f - g_i = (I-\mathrm{T}_j)f$ and using the fact that $I - \mathrm{T}_j$ is self adjoint, we conclude that

$$\langle f, \theta_j(I-\mathrm{T}_j)P_jL_j\rangle \geqslant \delta \qquad\qquad \forall j = 1,\dots,i+1. \tag{22}$$

We next inspect that quantity $(I) = \langle \mathrm{T}_{i+2}f, \sum_{j=1}^{i+1}\theta_j(I-\mathrm{T}_j)P_jL_j\rangle$, and prove an upper bound as well as a lower bound on it.

The Lower Bound: as T_{i+2} is self adjoint, we conclude that $(I) = \langle f, \sum_{j=1}^{i+1}\theta_j\mathrm{T}_{i+2}(I-\mathrm{T}_j)P_jL_j\rangle$, and we next argue that $\mathrm{T}_{i+2}(I-\mathrm{T}_j)P_jL_j$ is very close to $(I-\mathrm{T}_j)P_jL_j$ for each j. Clearly,

$$\mathrm{T}_{i+2}(I-\mathrm{T}_j)P_jL_j = \mathrm{T}_{i+2}P_jL_j - \mathrm{T}_{i+2}\mathrm{T}_jP_jL_j.$$

For the first term on the right-hand side, note that $P_j \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{i+2})$ and so $\mathrm{T}_{i+2}P_jL_j = P_j\mathrm{T}_{i+2}L_j$. It follows from Claim 5.5 that

$$\|\mathrm{T}_{i+2}P_jL_j - P_jL_j\|_2 \leqslant \|\mathrm{T}_{i+2}L_j - L_j\|_2 = \|(I-\mathrm{T}_{i+2})L_j\| \lesssim_{d,m,i,\alpha} \varepsilon_{i+2}^{1/3},$$

and so $\|\mathrm{T}_{i+2}P_jL_j - P_jL_j\|_2 \leqslant \|\mathrm{T}_{i+2}L_j - L_j\|_2 \leqslant \varepsilon_{i+2}^{1/6}$. For the second term, namely for $\mathrm{T}_{i+2}\mathrm{T}_jP_jL_j$, first apply Lemma 5.8 to get that $\|\mathrm{T}_jP_jL_j - h_j\|_2 \leqslant \zeta_j$ where h_j is a function of the form $\sum_{P'\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}_j)} P'L_{P'}$ where $L_{P'}$ has degree at most d_j and $\|L_{P'}\|_2 = O_{m,\zeta_j,\alpha,j,\varepsilon_j}(1)$. Combining with Claim 5.5 again it follows that

$$\|\mathrm{T}_{i+2}\mathrm{T}_jP_jL_j - \mathrm{T}_jP_jL_j\|_2 \leqslant 2\zeta_j + \|\mathrm{T}_{i+2}h_j - h_j\|_2 = 2\zeta_j + \|(I-\mathrm{T}_{i+2})h_j\|_2 \leqslant 2\zeta_j + O_{m,\zeta_j,\alpha,j,\varepsilon_j,d_j}(\varepsilon_{i+2}^{1/3}),$$

and so $\|\mathrm{T}_{i+2}\mathrm{T}_jP_jL_j - \mathrm{T}_jP_jL_j\|_2 \leqslant \varepsilon_{i+2}^{1/6}$.

Concluding, we get that $\|\mathrm{T}_{i+2}(I-\mathrm{T}_j)P_jL_j - (I-\mathrm{T}_j)P_jL_j\| \leqslant 2\varepsilon_{i+2}^{1/6}$, and plugging this into (I) gives that

$$(I) \geqslant \langle f, \sum_{j=1}^{i+1}\theta_j(I-\mathrm{T}_j)P_jL_j\rangle - 2(i+1)\varepsilon_{i+2}^{1/6} \geqslant (i+1)\delta - 2(i+1)\varepsilon_{i+2}^{1/6},$$

where in the last inequality we used (22). Thus, $(I) \geqslant 0.99(i+1)\delta$.

The Upper Bound: by Cauchy-Schwarz we have that

$$(I) \leqslant \|\mathrm{T}_{i+2}f\|_2\|\sum_{j=1}^{i+1}\theta_j(I-\mathrm{T}_j)P_jL_j\|_2,$$

and we upper bound the second norm. Taking a square and expanding, we have that

$$\|\sum_{j=1}^{i+1}\theta_j(I-\mathrm{T}_j)P_jL_j\|_2^2 = \sum_j \|(I-\mathrm{T}_j)P_jL_j\|_2^2 + 2\sum_{j'<j}\langle (I-\mathrm{T}_j)P_jL_j, (I-\mathrm{T}_{j'})P_{j'}L_{j'}\rangle. \tag{23}$$

We bound the first sum on the right-hand side by the trivial bound of $2(i+1)$ (as each one of the norms individually is at most 2), and next we show that the off-diagonal terms are negligible. Fix $j' < j$ and inspect the corresponding summand. Then by self-adjointness and Cauchy-Schwarz

$$\left|\langle (I-\mathrm{T}_j)P_jL_j, (I-\mathrm{T}_{j'})P_{j'}L_{j'}\rangle\right| = \left|\langle P_jL_j, (I-\mathrm{T}_j)(I-\mathrm{T}_{j'})P_{j'}L_{j'}\rangle\right| \leqslant \|(I-\mathrm{T}_j)(I-\mathrm{T}_{j'})P_{j'}L_{j'}\|_2.$$

Using the same argument as in the upper bound section, we have that $\|\mathrm{T}_j(I-\mathrm{T}_{j'})P_{j'}L_{j'} - (I-\mathrm{T}_{j'})P_{j'}L_{j'}\| \leqslant \varepsilon_j^{1/6}$, implying that $\|(I-\mathrm{T}_j)(I-\mathrm{T}_{j'})P_{j'}L_{j'}\|_2 \leqslant \varepsilon_j^{1/6}$. Plugging this into (23) gives that

$$(23) \leqslant 2(i+1) + (i+1)^2\varepsilon_1^{1/6} \leqslant 2(i+2),$$

and so $(I) \leqslant \|\mathrm{T}_{i+2}f\|_2\sqrt{2(i+2)}$.

Combining the Upper and Lower Bounds: combining the upper and lower bounds for (I), we conclude that $\|\mathrm{T}_{i+2}f\|_2^2 \cdot 2(i+2) \geqslant 0.99^2(i+1)^2\delta^2$, and simplifying we get that $\|\mathrm{T}_{i+2}f\|_2^2 \geqslant i\delta^2/10$, concluding the proof. □

Using Claim 5.10, the process we designed terminates within at most R steps at step $i \leqslant R$, at which point we have that f_i has correlation at most δ with any function of the form $P \cdot L$. Applying Theorem 3.1 and the fact that f_i is 2-bounded we conclude that $\|f_i\|_{\nu,\alpha} \leqslant 2 \cdot \frac{\xi}{100} \leqslant \xi$, concluding the proof. □

5.3.2 A Version of the Regularity Lemma Allowing Noise Modification

In this section, we state and prove a variant of Lemma 5.9 in which we have a lot of freedom in picking the noise parameter ε. Naively, and borrowing intuition from the case of the standard noise operator, one expects that the first item in Lemma 5.9 to hold not only for ε but rather for any $0 < \varepsilon' < \varepsilon$. Indeed, for the standard noise operator the left-hand side may be interpreted as the weight of f on the high degrees, and taking smaller epsilon amounts to looking at even higher degrees (thus trivially looking at less weight). While we do not know how to make such an argument go through, below we show how to circumvent this issue in a relatively easy way.

In the formulation of the lemma below, we will use a decay function $w\colon [0,1] \to [0,1]$, meaning a function w satisfying that $w(\varepsilon) \leqslant \varepsilon$. The reader should have in mind w of the forms $w(\varepsilon) = \varepsilon^{10}$ or $w(\varepsilon) = 2^{-1/\varepsilon}$.

Lemma 5.11. *For all $\alpha > 0$, $m \in \mathbb{N}$ and $\xi > 0$ there exists $r \in \mathbb{N}$ such that for any decay function $w\colon [0,1] \to [0,1]$ there is $\varepsilon_0 > 0$ for which the following holds. Let Σ be an alphabet of size at most m, let ν be a distribution over Σ in which the probability of each atom is at least α. Then there exists an Abelian group G whose size depends only on m, $\sigma\colon \Sigma \to G$, such that for any 1-bounded function $f\colon \Sigma^n \to \mathbb{C}$ there is $r' \leqslant r$, a collection $\mathcal{P}$ of at most r' cyclic embedding functions and $\varepsilon \geqslant \varepsilon_0$ such that*

$$\|f - \mathrm{T}_{\mathcal{P},1-\varepsilon'}f\|_{\nu,\alpha} \leqslant \xi$$

for all $w(\varepsilon) \leqslant \varepsilon' \leqslant \varepsilon$.

Proof. We run the same process as in the proof of Lemma 5.9. Once the process terminates, say at step i, we know that $(I - \mathrm{T}_{\mathcal{P}_i,1-\varepsilon})f$ has correlations at most δ with any function of the form $P \cdot L$ for $\varepsilon = \varepsilon_i$. If this holds for all $\varepsilon \in (w(\varepsilon_i), \varepsilon_i)$, we are done as then by Theorem 3.1 it follows that $\|(I-\mathrm{T}_{\mathcal{P}_i,1-\varepsilon})f\|_{\nu,\alpha} \leqslant \xi$ for all $\varepsilon \in (w(\varepsilon_i), \varepsilon_i)$. Otherwise, we may find $\varepsilon_i' \in (w(\varepsilon_i), \varepsilon_i)$ such that $(I-\mathrm{T}_{\mathcal{P}_i,1-\varepsilon_i'})f$ has correlation at least δ with some function $P_{i+1}L_{i+1}$, and we may continue the argument therein by modifying ε_i to be ε_i' (ε_i' still satisfies the same requirements from the parameters as ε_i). Carrying out the same analysis as in Lemma 5.9, we conclude that this modified process also terminates within R steps, and the proof is concluded. □

5.3.3 A Regularity Lemma with High Rank

We finish this section off by presenting our final regularity lemma which will be used in subsequent sections. This lemma is a strict strengthening of Lemma 5.11, and it asserts that the collection of cyclic embedding functions can additionally be guaranteed to have a high rank.

Lemma 5.12. *For all $\alpha > 0$, $m \in \mathbb{N}$ and $\xi > 0$ there exist $r \in \mathbb{N}$ and $\varepsilon_0 > 0$ such that the following holds for any decay function $w\colon [0,1] \to [0,1]$. Let Σ be an alphabet of size at most m, let ν be a distribution over Σ in which the probability of each atom is at least α. Then there exists an Abelian group G whose size depends only on m, $\sigma\colon \Sigma \to G$, such that for any 1-bounded function $f\colon \Sigma^n \to \mathbb{C}$ there are $r' \leqslant r$, $\varepsilon \geqslant \varepsilon_0$ and a collection $\mathcal{P}$ of cyclic embedding functions of size at most r' such that:*

1. *We have $\mathsf{rk}(\mathcal{P}) \geqslant \frac{1}{w(\varepsilon)}$.*
2. *For all $\varepsilon' \in (w(\varepsilon), \varepsilon)$ we have that $\|f - \mathrm{T}_{\mathcal{P},1-\varepsilon'} f\|_{\nu,\alpha} \leqslant \xi$.*

Proof. Fix α, m, ξ, pick r from Lemma 5.11 and take $R = R(m,r)$ to be sufficiently large. Take parameters

$$0 < \varepsilon_R \ll M_R^{-1} \ll \varepsilon_{R-1} \ll M_{R-1}^{-1} \ll \ldots \ll M_2^{-1} \ll \varepsilon_1 \ll M_1^{-1} \ll \varepsilon_0 \leqslant \varepsilon, \tag{24}$$

and pick the decay function $w(\varepsilon) = \varepsilon_r$, and then take ε_0 from Lemma 5.11.

Applying Lemma 5.11, we get that there is r', a collection $\mathcal{P}$ of cyclic embedding functions of size at most r' and $\varepsilon > \varepsilon_0$ such that for all $\varepsilon' \in (w(\varepsilon), \varepsilon)$ it holds that

$$\|f - \mathrm{T}_{\mathcal{P},1-\varepsilon'} f\|_{\nu,\alpha} \leqslant \xi.$$

If $\mathsf{rk}(\mathcal{P}) \geqslant M_1$, we are done by picking $\varepsilon' = \varepsilon$. Otherwise, we apply Lemma 4.21 and there are two cases:

1. In the first case, there is $P \in \mathcal{P}$ such that writing $\mathcal{P}_1 = \mathcal{P} \setminus \{P\}$ we have $\Delta_{\mathsf{symboblic}}(P, \mathsf{spn}_{\mathbb{N}}(\mathcal{P}')) \leqslant M_1$. We get by Fact 5.3 that for all $\varepsilon' \leqslant \varepsilon_1$

$$\|\mathrm{T}_{\mathcal{P},1-\varepsilon'} f - \mathrm{T}_{\mathcal{P}_1,1-\varepsilon'} f\|_2 \leqslant 2\sqrt{M_1 \varepsilon_1} \leqslant \frac{\xi}{R}.$$

2. Else, there is $P \in \mathcal{P}$ and a cyclic embedding function P' such that $\Delta_{\mathsf{symbolic}}(P, P') \leqslant M_1$ and $\mathsf{ord}(P') < \mathsf{ord}(P)$. In that case, we take $\mathcal{P}_1 = (\mathcal{P} \setminus \{P\}) \cup \{P'\}$ and get by applying Fact 5.3 twice that

$$\begin{aligned}\|\mathrm{T}_{\mathcal{P},1-\varepsilon'} f - \mathrm{T}_{\mathcal{P}_1,1-\varepsilon'} f\|_2 &\leqslant \|\mathrm{T}_{\mathcal{P}\cup\{P'\},1-\varepsilon'} f - \mathrm{T}_{\mathcal{P},1-\varepsilon'} f\|_2 + \|\mathrm{T}_{\mathcal{P}\cup\{P'\},1-\varepsilon'} f - \mathrm{T}_{\mathcal{P}_1,1-\varepsilon'} f\|_2 \\ &\lesssim \sqrt{M_1 \varepsilon_1},\end{aligned}$$

which is again at most $\frac{\xi}{R}$.

Continuing, if $\mathsf{rk}(\mathcal{P}_1) \geqslant M_2$ we are done, and else we repeat the above argument to produce a new collection of cyclic embedding functions $\mathcal{P}_2$ satisfying that for all $\varepsilon' \leqslant \varepsilon_2$ it holds that $\|\mathrm{T}_{\mathcal{P}_2,1-\varepsilon'} f - \mathrm{T}_{\mathcal{P}_1,1-\varepsilon'} f\|_2 \leqslant \frac{\xi}{R}$. Denote by $\mathcal{P}_i$ the collection formed after i steps of this process.

Note that $\sum\limits_{P \in \mathcal{P}_{i+1}} \mathsf{ord}(P) \leqslant \sum\limits_{P \in \mathcal{P}_i} \mathsf{ord}(P) - 1$ for all i. Indeed, in the first case above we drop an element from the collection and that element has order at least 1. In the second case above we replace an element in the collection with an element whose order is strictly smaller. Thus, as originally we have that $\sum\limits_{P \in \mathcal{P}} \mathsf{ord}(P)$

is at most $r \cdot O_m(1)$ (as each $P \in \mathcal{P}$ has order which is at most $O_m(1)$), it follows that the process terminates within $r \cdot O_m(1) < R$ steps.

Let i be the step in which the process is finished. Thus, $\mathsf{rk}(\mathcal{P}_i) \geqslant M_{i+1}$, and by the triangle inequality get that

$$\|\mathrm{T}_{\mathcal{P}_i,1-\varepsilon'} f - \mathrm{T}_{\mathcal{P},1-\varepsilon'} f\|_2 \leqslant \sum_{j=0}^{i-1} \|\mathrm{T}_{\mathcal{P}_{j+1},1-\varepsilon'} f - \mathrm{T}_{\mathcal{P}_j,1-\varepsilon'} f\| \leqslant i \cdot \frac{\xi}{R} \leqslant \xi$$

for all $\varepsilon' \leqslant \varepsilon_i$. It follows that for any such ε' it holds that

$$\begin{aligned}
\|f - \mathrm{T}_{\mathcal{P}_i,1-\varepsilon'} f\|_{\nu,\alpha} &\leqslant \|f - \mathrm{T}_{\mathcal{P},1-\varepsilon'} f\|_{\nu,\alpha} + \|\mathrm{T}_{\mathcal{P},1-\varepsilon'} f - \mathrm{T}_{\mathcal{P}_i,1-\varepsilon'} f\|_{\nu,\alpha} \\
&\leqslant \|f - \mathrm{T}_{\mathcal{P},1-\varepsilon'} f\|_{\nu,\alpha} + \|\mathrm{T}_{\mathcal{P},1-\varepsilon'} f - \mathrm{T}_{\mathcal{P}_i,1-\varepsilon'} f\|_2 \\
&\leqslant 2\xi,
\end{aligned}$$

and the proof is complete. □

6 Moving to the Mixed Space

In this section, we make another step towards proving our mixed invariance principle: we show how to associate with a given product space (Σ^n, ν^ν) a mixed space $(\mathbb{R}^{n'} \times G^{n'}, \mathcal{D}^{n'})$ where $n', n'' = \Theta(n)$, which is a product of a Gaussian space with an Abelian group (equipped with some product measure). In particular, we will need this association to allow us to transfer bounded functions from the product space to the mixed space, and in this section, we explain this transference.

6.1 Motivation

To begin the discussion, we recall that in Section 5.3, we have shown that if $f\colon (\Sigma^n, \nu^n) \to \mathbb{C}$ is a 1-bounded function, then we may find a collection $\mathcal{P}$ of cyclic embedding functions such that (we omit the various quantifiers and parameters for the sake of clarity):

1. The size of $\mathcal{P}$ is constant.
2. The functions f and $\mathrm{T}_{\mathcal{P},1-\varepsilon}$ are close in $\|\cdot\|_{\nu,\alpha}$-norm.
3. $\mathcal{P}$ has arbitrarily large rank.

Similarly, we may apply the statement for functions $g\colon \Gamma^n \to \mathbb{C}$ and $h\colon \Phi^n \to \mathbb{C}$ to get collections $\mathcal{Q}$ and $\mathcal{R}$ of cyclic embedding functions such that

$$\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n} [f(x)g(y)h(z)] = \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n} [\mathrm{T}_{\mathcal{P},1-\varepsilon} f(x) \mathrm{T}_{\mathcal{Q},1-\varepsilon} g(y) \mathrm{T}_{\mathcal{R},1-\varepsilon} h(z)] + o(1).$$

Thus, to achieve our goal it suffices to come up with a mixed-space on which there are functions f', g', h' that can be associated with $\mathrm{T}_{\mathcal{P},1-\varepsilon} f$, $\mathrm{T}_{\mathcal{P},1-\varepsilon} g$ and $\mathrm{T}_{\mathcal{P},1-\varepsilon} h$ such that

$$\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n} [\mathrm{T}_{\mathcal{P},1-\varepsilon} f(x) \mathrm{T}_{\mathcal{Q},1-\varepsilon} g(y) \mathrm{T}_{\mathcal{R},1-\varepsilon} h(z)] = \mathop{\mathbb{E}}_{(x,y,z)\sim\mathcal{D}^{n'}} \left[f'(x) g'(y) h'(z)\right] + o(1).$$

To motivate the construction of f', g' and h' we consider the following two simplified scenarios:

1. **The case that $\mathcal{P} = \mathcal{Q} = \mathcal{R} = \emptyset$:** In this case, we have the functions $\mathrm{T}_{1-\varepsilon}f, \mathrm{T}_{1-\varepsilon}g, \mathrm{T}_{1-\varepsilon}h$ which are morally low-degree functions. For these types of functions, the invariance principle of [42] tells us that we can transfer them to Gaussian space. This is done by looking at $\mathrm{T}_{1-\varepsilon}f, \mathrm{T}_{1-\varepsilon}g, \mathrm{T}_{1-\varepsilon}h$ as real-valued polynomials of an orthonormal basis of the basic space, and then plugging in correlated Gaussian random variables (g_x, g_y, g_z) whose pairwise correlations match those in the base space.

2. **The case that $\varepsilon = 1$:** In that case, the value of $\mathrm{T}_{\mathcal{P},1-\varepsilon}f(x)$ depends only on the value of $(P(x))_{P\in\mathcal{P}}$. In other words, writing $\mathcal{P} = \{P_1, \ldots, P_r\}$, there is a function $f' \colon \mathbb{C}^r \to \mathbb{C}$ such that $\mathrm{T}_{\mathcal{P},1-\varepsilon}f(x) = f'(P_1(x), \ldots, P_r(x))$. Since each P_i is a cyclic embedding function, for each i there is $\chi_i \in \widehat{G_{\mathsf{master}}}$ such that $P_i(x) = \chi_i \circ \sigma(x)$, and so there is a function $f'' \colon G_{\mathsf{master}}^n \to \mathbb{C}$ such that $\mathrm{T}_{\mathcal{P},1-\varepsilon}f(x) = f''(\sigma_{\mathsf{master}}(x))$.

Summarizing, in the first case we can move to Gaussian space, whereas in the second case, we can move to the master group. In both cases, one can show that the associated function is (essentially) 1-bounded and has roughly the same distribution of values. The goal of this section is to show that in the general case that $\mathcal{P}$ may be non-empty and ε is smaller than 1 (but bounded away from 0), we may find an associated function with $\mathrm{T}_{\mathcal{P},1-\varepsilon}f$ that "mixes" between these two possibilities.

6.2 The Decoupled Function

We now move to the formal description of the mixed space and the associated function with f on it.

Our transference works with a function $\mathrm{T}_{\mathcal{P},1-\varepsilon}f$ satisfying the following properties:

1. $|\mathcal{P}| = O_{m,\alpha,\eta}(1)$.
2. $\|f - \mathrm{T}_{\mathcal{P},1-\varepsilon}f\|_{\nu,\alpha} \leqslant \eta$ where ε is arbitrarily small compared to η.
3. $\mathsf{rk}(\mathcal{P}) \geqslant T$ where $T \gg 1/\eta, 1/\varepsilon, 1/\alpha, m$.

A function $\mathrm{T}_{\mathcal{P},1-\varepsilon}f$ of this form is guaranteed to exist by Lemma 5.12, and for simplicity of notation we denote $f' = \mathrm{T}_{\mathcal{P},1-\varepsilon}f$. By Lemma 5.8 we may find a function $\tilde{f}$ of the form

$$\tilde{f}(x) = \sum_{P\in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P(x)\cdot L_P(x)$$

such that:

1. $\|f' - \tilde{f}\|_2 \leqslant \zeta$.
2. $\mathsf{deg}(L_P)$ and $\|L_P\|_2$ are both $O_{m,\alpha,\eta,\zeta,\varepsilon}(1)$.

Define $\tilde{f}_{\mathsf{decoupled}} \colon (\Sigma^n \times \Sigma^n, \nu^n \times \nu^n) \to \mathbb{C}$ by

$$\tilde{f}_{\mathsf{decoupled}}(x, x') = \sum_{P\in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P(x)\cdot L_P(x').$$

In words, we take two independent copies of the input of $\tilde{f}$, plug one copy of it to the embedding functions and another copy to the low-degree functions.

We wish to prove a relation between the functions $\tilde{f}_{\mathsf{decoupled}}$ and $\tilde{f}$. Intuitively, the distribution of values of $\tilde{f}_{\mathsf{decoupled}}(x, x')$ when x, x' are sampled from ν^n independently is close to the distribution of values of

$\tilde{f}(X)$ when $X \sim \nu^n$. To see that, looking at the formula of $\tilde{f}(X)$, the values of the low-degree functions L_P are mostly determined after we expose $1 - \delta/d$ of the coordinates of X, where d is an upper bound on the degrees of L_P. On the other hand, since $\mathcal{P}$ has a high rank, conditioning on exposing these coordinates of X the values of $(P(X))_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})}$ still has the same distribution. Thus, the values of $(L_P(X))_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})}$ and $(P(X))_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})}$ are almost independent of each other, which matches the values of the function $\tilde{f}_{\mathsf{decoupled}}$.

We now move on to stating and proving a formal relation between $\tilde{f}_{\mathsf{decoupled}}$ and $\tilde{f}$, and begin with a few auxiliary facts. The first of which asserts that if $\mathcal{P} = \{P_1, \ldots, P_r\}$ is a collection of cyclic embedding functions of high rank, then for all $a_i \in \mathsf{Image}(P_i)$ the event that $P_i(x) = a_i$ for all i has probability at least $\Omega_{m,r}(1)$.

Fact 6.1. *For all $m, r \in \mathbb{N}$ there exists $M \in \mathbb{N}$ such that if $\mathcal{P} = \{P_1, \ldots, P_r \colon \Sigma^n \to \mathbb{C}\}$ is a collection of cyclic embedding functions with $\mathsf{rk}(\mathcal{P}) \geqslant M$, then for all $a_1 \in \mathsf{Image}(P_1), \ldots, a_r \in \mathsf{Image}(P_r)$ it holds that either* $\Pr_{x \sim \nu^n}[P_i(x) = a_i] = 0$ *or*

$$\Pr_{x \sim \nu^n}[P_i(x) = a_i] \geqslant \Omega_{m,r}(1).$$

Proof. We write $1_{P_i(x)=a_i} = \prod_{b_i \in \mathsf{supp}(P_i), b_i \neq a_i} \frac{b_i - P_i(x)}{b_i - a_i}$, so that the left-hand side is equal to

$$C \mathop{\mathbb{E}}_{x}\left[\prod_{i} \prod_{b_i \in \mathsf{supp}(P_i) \setminus \{a_i\}} (b_i - P_i(x))\right],$$

where $C = \prod_{i} \prod_{b_i \in \mathsf{supp}(P_i) \setminus \{a_i\}} \frac{1}{b_i - a_i}$, thus $|C| \geqslant \Omega_r(1)$. To compute the expectation, we expand it out and get that

$$\mathop{\mathbb{E}}_{x}\left[\prod_{i} \prod_{b_i \in \mathsf{supp}(P_i) \setminus \{a_i\}} (b_i - P_i(x))\right] = C' + \sum_{(\alpha_1, \ldots, \alpha_r) \in \mathcal{A}} C(\alpha_1, \ldots, \alpha_r) \mathop{\mathbb{E}}_{x}\left[\prod_{i=1}^{r} P_i(x)^{\alpha_i}\right],$$

where

$$\mathcal{A} = \{(\alpha_1, \ldots, \alpha_r) \mid 0 \leqslant \alpha_i \leqslant |\mathsf{Image}(P_i)| \ \forall i \text{ and } \alpha_i < |\mathsf{Image}(P_i)| \text{ for some } i\}.$$

Note that $|\mathsf{Image}(P_i)| \leqslant \mathsf{ord}(P_i)$ hence for each $(\alpha_1, \ldots, \alpha_r) \in \mathcal{A}$ it holds that $\left|\mathbb{E}_x\left[\prod_{i=1}^{r} P_i(x)^{\alpha_i}\right]\right| = O(2^{-\Omega_{r,m}(M)})$. We conclude that

$$\Pr_{x \sim \nu^n}[P_i(x) = a_i] \geqslant C - O(2^{-\Omega_{r,m}(M)}) \geqslant O_{m,r}(1)$$

for sufficiently large M. □

Next, suppose that $\mathcal{P} = \{P_1, \ldots, P_r\}$ is a collection of cyclic embedding functions, for each i take $a_i \in \mathsf{Image}(P_i)$ and consider the set of inputs $S = \{x \in \Sigma^n \mid P_i(x) = a_i\}$. By Fact 6.1 we know that S has a fairly large density. The next statement asserts that if the rank of $\mathcal{P}$ is sufficiently large, then a noisy process starting from a random point in S generates a probability distribution close to ν^n.

Fact 6.2. *For all $r, m \in \mathbb{N}$, $\xi > 0$ and $M \in \mathbb{N}$, if $\mathcal{P} = \{P_1, \ldots, P_r : \Sigma^n \to \mathbb{C}\}$ is a collection of cyclic embedding functions with $\mathsf{rk}(\mathcal{P}) \geqslant M$, take $a_1 \in \mathsf{Image}(P_1), \ldots, a_r \in \mathsf{Image}(P_r)$ and define*

$$S = \{ x \in \Sigma^n \mid P_i(x) = a_i \}.$$

Then

$$\left\| \frac{T_{1-\varepsilon} 1_S}{\nu(S)} - 1 \right\|_2 \leqslant 2^{-\Omega_{m,\varepsilon,r}(M)}.$$

Proof. Expanding 1_S as in the proof of Fact 6.1, we see that $1_S = A + \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}), P \neq 1} C(P)P$ where $|C(P)| = O_{m,r}(1)$ for all P. Taking expectation of both sides shows that $\nu(S) = A + 2^{-\Omega_{m,r}(M)}$ and so $A = \nu(S)(1 + 2^{-\Omega_{m,r}(M)})$. We get that

$$\frac{1_S}{\nu(S)} - 1 = 2^{-\Omega_{m,r}(M)} + \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}), P \neq 1} C'(P)P.$$

Applying the operator $T_{1-\xi}$ on both sides we conclude that

$$\begin{aligned}\left\| \frac{T_{1-\xi} 1_S}{\nu(S)} - 1 \right\|_2 &\leqslant 2^{-\Omega_{m,r}(M)} + O_{m,r}\left(\max_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}), P \neq 1} \|T_{1-\varepsilon} P\|_2 \right) \\ &\leqslant 2^{-\Omega_{m,r}(M)} + O_{m,r}\left(\max_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}), P \neq 1} 2^{-\Omega_{m,r,\varepsilon}(M)} \right) \\ &\leqslant 2^{-\Omega_{m,\varepsilon,r}(M)},\end{aligned}$$

where we used Lemma 4.6. □

We are now ready to prove the main technical ingredient in the relation between $\tilde{f}$ and $\tilde{f}_{\mathsf{decoupled}}$.

Lemma 6.3. *Let $m, r \in \mathbb{N}$, $\xi > 0$ and $M \in \mathbb{N}$. Let $\mathcal{P} = \{P_1, \ldots, P_r : \Sigma^n \to \mathbb{C}\}$ be a collection of cyclic embedding functions such that $\mathsf{rk}(\mathcal{P}) \geqslant M$. Consider the joint distribution of (x, x') and X sampled as: sample $X \sim \nu^n$ and independently sample $x \sim \mathrm{T}_{\mathcal{P}} X$ and $x' \sim \mathrm{T}_{1-\varepsilon} X$. Then then distribution of (x, x') is $2^{-\Omega_{r,m,\varepsilon}(M)}$ close to $\nu^n \times \nu^n$ in statistical distance.*

Proof. Fix $w, w' \in \Sigma^n$ and define $q_{w,w'} = \Pr_{X,x,x'}[x = w, x' = w']$ and $q_{w'|w} = \Pr_{X,x,x'}[x' = w' \mid x = w]$. By Fact 5.2 the marginal distribution of x is ν^n and so $q_{w,w'} = \nu(w) q_{w'|w}$. Denote

$$S = \{ u \in \Sigma^n \mid P_i(u) = P_i(w)\ \forall i \},$$

and note that conditioned on $x = w$, the distribution of X is ν conditioned on S, i.e.

$$\Pr_{X,x,x'}[X = u \mid x = w] = \nu(u) \frac{1_S(u)}{\nu(S)}.$$

Thus,

$$q_{w'|w} = \sum_u \nu(u) \frac{1_S(u)}{\nu(S)} \Pr_{u' \sim T_{1-\varepsilon} u}[u' = w'] = \sum_u \nu(u) \frac{1_S(u)}{\nu(S)} \mathrm{T}_{1-\varepsilon} 1_{w'}(u),$$

where $1_{w'}(v) = 1_{w'=v}$. Writing this last expression as an inner product, we get that

$$q_{w'|w} = \langle \frac{1_S}{\nu(S)}, \mathrm{T}_{1-\varepsilon} 1_{w'} \rangle = \langle \mathrm{T}_{1-\varepsilon} \frac{1_S}{\nu(S)}, 1_{w'} \rangle = \nu(w') \mathrm{T}_{1-\varepsilon} \frac{1_S}{\nu(S)}(w'),$$

and so $q_{w,w'} = \nu(w)\nu(w')T_{1-\varepsilon}\frac{1_S}{\nu(S)}(w')$. Computing, we get that

$$\sum_{w,w'} \left|q_{w,w'} - \nu(w)\nu(w')\right| = \sum_{w,w'} \nu(w)\nu(w') \left|\mathrm{T}_{1-\varepsilon}\frac{1_S}{\nu(S)}(w') - 1\right| = \left\|\mathrm{T}_{1-\varepsilon}\frac{1_S}{\nu(S)} - 1\right\|_1 \leqslant \left\|\mathrm{T}_{1-\varepsilon}\frac{1_S}{\nu(S)} - 1\right\|_2,$$

and the proof is concluded by applying Fact 6.2. □

With Lemma 6.3 in hand, we can now state and prove the relation between the functions $\tilde{f}$ and $\tilde{f}_{\mathsf{decoupled}}$. We begin with a basic version of it:

Lemma 6.4. *Let $m, r, d \in \mathbb{N}$, and $M \in \mathbb{N}$. Let $\mathcal{P} = \{P_1, \ldots, P_r : \Sigma^n \to \mathbb{C}\}$ be a collection of cyclic embedding functions such that $\mathsf{rk}(\mathcal{P}) \geqslant M$, and consider the joint distribution of (x, x') and X from Lemma 6.3. Then*

$$\mathop{\mathbb{E}}_{x,x',X}\left[\left|\tilde{f}_{\mathsf{decoupled}}(x, x') - \tilde{f}(X)\right|^2\right] \lesssim_{d,m,r} \varepsilon.$$

Proof. Writing $\tilde{f} = \sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P(x)L_P(x)$ where $\mathsf{deg}(L_P) \leqslant d$, we get that the left-hand side in the lemma is equal to

$$\mathop{\mathbb{E}}_{x,x',X}\left[\left|\sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P(x)L_P(x') - \sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P(X)L_P(X)\right|^2\right]$$
$$= \mathop{\mathbb{E}}_{x,x',X}\left[\left|\sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P(X)(L_P(x') - L_P(X))\right|^2\right].$$

By Cauchy-Schwarz, we may upper bound the last quantity by

$$\lesssim_{m,r} \sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} \mathop{\mathbb{E}}_{x,x',X}\left[\left|L_P(x') - L_P(X)\right|^2\right] = \sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} \|L_P - \mathrm{T}_{1-\varepsilon}L_P\|_2^2 \lesssim_{m,r} d\varepsilon.$$ □

We next state the version of our relation that will be used later on.

Lemma 6.5. *Let $m, r, d \in \mathbb{N}$, $\varepsilon > 0$ and $M \in \mathbb{N}$. Let $\mathcal{P} = \{P_1, \ldots, P_r : \Sigma^n \to \mathbb{C}\}$ be a collection of cyclic embedding functions such that $\mathsf{rk}(\mathcal{P}) \geqslant M$. Then there is a coupling (X', X'') and X between $\nu^n \times \nu^n$ and ν^n such that*

$$\mathop{\mathbb{E}}_{X,X',X''}\left[\left|\tilde{f}_{\mathsf{decoupled}}(X', X'') - \tilde{f}(X)\right|^2\right] \leqslant O_{m,r,d}(\varepsilon) + O_{m,r,d,\alpha,\eta,\varepsilon}(2^{-\Omega_{r,m,\varepsilon}(M)}).$$

Proof. Let (x, x') and X be a coupling as in Lemma 6.3. As the statistical distance between (x, x') and $\nu^n \times \nu^n$ is at most $2^{-\Omega_{r,m,\varepsilon}(M)}$, we may couple (x, x') with (X', X'') such that $\Pr[(x, x') \neq (X', X'')] \leqslant 2^{-\Omega_{r,m,\varepsilon}(M)}$ and (X', X'') is distributed according to $\nu^n \times \nu^n$. We take the coupling of (X', X'') and X and prove that it satisfies the conclusion of the lemma. Note that the left-hand side of the lemma is at most

$$\lesssim \mathop{\mathbb{E}}_{x,x',X,X',X''}\left[\left|\tilde{f}_{\mathsf{decoupled}}(x, x') - \tilde{f}(X)\right|^2\right] + \mathop{\mathbb{E}}_{x,x',X,X',X''}\left[\left|\tilde{f}_{\mathsf{decoupled}}(x, x') - \tilde{f}_{\mathsf{decoupled}}(X', X'')\right|^2\right].$$

The first expectation above is at most $O_{r,m,d}(\varepsilon)$. For the second expectation, let E be the event that $(x, x') \neq (X', X'')$. Then by Cauchy-Schwarz

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,x',X,X',X''}\left[\left|\tilde{f}_{\mathsf{decoupled}}(x, x') - \tilde{f}_{\mathsf{decoupled}}(X', X'')\right|^2 1_E\right]^2\\
&\leqslant \mathop{\mathbb{E}}_{x,x',X,X',X''}\left[\left|\tilde{f}_{\mathsf{decoupled}}(x, x') - \tilde{f}_{\mathsf{decoupled}}(X', X'')\right|^4\right]\Pr\left[E\right]\\
&\leqslant \mathop{\mathbb{E}}_{x,x',X,X',X''}\left[\left|\tilde{f}_{\mathsf{decoupled}}(x, x')\right|^4 + \left|\tilde{f}_{\mathsf{decoupled}}(X', X'')\right|^4\right]\Pr\left[E\right].
\end{aligned}$$

We upper bound the first expectation above. By the triangle inequality, $\left|\tilde{f}_{\mathsf{decoupled}}(x, x')\right| \leqslant \sum_P |L_P(x')|$, and hence by Holder's inequality and hypercontractivity we get that

$$\mathop{\mathbb{E}}_{x,x'}\left[\left|\tilde{f}_{\mathsf{decoupled}}(x, x')\right|^4\right] \lesssim_r \sum_P \|L_P\|_4^4 \lesssim_{r,d} \sum_P \|L_P\|_2^4 \lesssim_{r,m,d,\alpha,\eta,\varepsilon} 1.$$

The expectation of $\left|\tilde{f}_{\mathsf{decoupled}}(X', X'')\right|^4$ can be bounded in the same way by $O_{r,m,d,\alpha,\eta,\varepsilon}(1)$. As $\Pr\left[E\right] \leqslant 2^{-\Omega_{r,m,\varepsilon}(M)}$ we conclude that

$$\mathop{\mathbb{E}}_{x,x',X,X',X''}\left[\left|\tilde{f}_{\mathsf{decoupled}}(x, x') - \tilde{f}_{\mathsf{decoupled}}(X', X'')\right|^2\right] \lesssim_{r,m,d,\alpha,\eta,\varepsilon} 2^{-\Omega_{r,m,\varepsilon}(M)}. \qquad \square$$

6.3 The Mixed Invariance Principle

In this section, we prove the mixed invariance principle. We will use the notations of the beginning of Section 6.2; the notation therein is presented in terms of the function f, and we use analogous notations for the functions g and h below. The notion of shifted low-degree influences will be important for us, and below we define it formally.

Definition 6.6. *We say that a function $\tilde{f}$ of the form above has τ-small shifted low-degree influences if for every $i \in [n]$ and every $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$ it holds that $I_i[L_P] \leqslant \tau$.*

Let $\Psi\colon \mathbb{C}^3 \to \mathbb{C}$ be any smooth function such that $\Psi(a, b, c) = abc$ for a, b, c that have absolute value at most 1, which additionally satisfies that

$$\left|\Psi(a, b, c) - \Psi(a', b', c')\right| \lesssim \sqrt{|a - a'|^2 + |b - b'|^2 + |c - c'|^2}$$

for all complex numbers $a, b, c \in \mathbb{C}$. Denote $m_1 = |\Sigma|$, $m_2 = |\Gamma|$ and $m_3 = |\Phi|$. Given 1-bounded functions $f\colon (\Sigma^n, \nu_x^n) \to \mathbb{C}$, $g\colon (\Gamma^n, \nu_y^n) \to \mathbb{C}$ and $h\colon (\Phi^n, \nu_z^n) \to \mathbb{C}$, define the functions $F\colon \Sigma^n \times \mathbb{R}^{(m_1-1)n} \to \mathbb{C}$, $G\colon \Gamma^n \times \mathbb{R}^{(m_2-1)n} \to \mathbb{C}$, $H\colon \Phi^n \times \mathbb{R}^{(m_3-1)n} \to \mathbb{C}$ by

$$F(x, G_x) = \mathsf{trunc}(\tilde{f}_{\mathsf{decoupled}}(x, G_x)), \qquad G(y, G_y) = \mathsf{trunc}(\tilde{g}_{\mathsf{decoupled}}(y, G_y)),$$
$$H(z, G_z) = \mathsf{trunc}(\tilde{h}_{\mathsf{decoupled}}(z, G_z)),$$

where $\mathsf{trunc}\colon \mathbb{C} \to \mathbb{C}$ was defined in Section 3.3. The main result of this section takes the following form:

Theorem 6.7. *Let $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$ and $h\colon \Phi^n \to \mathbb{C}$ be 1-bounded functions, and consider the functions $\tilde{f}, F, \tilde{g}, G, \tilde{h}, H$ defined as above. Then for every $\xi > 0$ there exists $\tau > 0$ such that if $\tilde{f}, \tilde{g}, \tilde{h}$ have τ-small shifted low-degree influences*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}[\Psi(f(x), g(y), h(z))] - \mathop{\mathbb{E}}_{\substack{(x,y,z)\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}}[\Psi(F(x,G_x), G(y,G_y), H(z,G_z))] \right| \leqslant \xi. \tag{25}$$

Proof. We keep the parameters as in the setup preceding the theorem. These parameters satisfy the hierarchy:

$$0 < \tau \ll T^{-1} \ll \varepsilon \ll d^{-1} \ll \zeta \ll \eta, r^{-1} \ll m^{-1}, \alpha, \xi < 1. \tag{26}$$

First, as $\|f - f'\|_{\nu,\alpha} \leqslant \eta$ and similarly for g and h, we have that

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}[\Psi(f(x), g(y), h(z))] - \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}\left[\Psi(f'(x), g'(y), h'(z))\right] \right| \lesssim \xi. \tag{27}$$

Second, by the smoothness of Ψ it follows that

$$\begin{aligned} &\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}\left[\Psi(f'(x), g'(y), h'(z))\right] - \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}\left[\Psi(\tilde{f}(x), \tilde{g}(y), \tilde{h}(z))\right] \right| \\ &\quad\lesssim \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}\left[\sqrt{\left|f'(x) - \tilde{f}(x)\right|^2 + |g'(y) - \tilde{g}(y)|^2 + \left|h'(z) - \tilde{h}(z)\right|^2}\right] \\ &\quad\lesssim \xi, \end{aligned} \tag{28}$$

where we used Cauchy-Schwarz and the fact that f' and $\tilde{f}$ are η-close in L_2 distance, and similarly $g', \tilde{g}$ and $h', \tilde{h}$. Third, using the coupling from Lemma 6.5 and the smoothness of Ψ it follows that

$$\begin{aligned} &\left| \mathop{\mathbb{E}}_{(X,Y,Z)}\left[\Psi(\tilde{f}(X), \tilde{g}(Y), \tilde{h}(Z))\right] - \mathop{\mathbb{E}}_{\substack{(X,Y,Z)\\(X',Y',Z')\\(X'',Y'',Z'')}}\left[\Psi(\tilde{f}_{\mathsf{decoupled}}(X',X''), \tilde{g}_{\mathsf{decoupled}}(Y',Y''), \tilde{h}_{\mathsf{decoupled}}(Z',Z''))\right] \right| \\ &\lesssim \mathop{\mathbb{E}}_{\substack{(X,Y,Z)\\(X',Y',Z')\\(X'',Y'',Z'')}}\left[\sqrt{|\Delta_1(X,X',X'')|^2 + |\Delta_2(Y,Y',Y'')|^2 + |\Delta_3(Z,Z',Z'')|^2}\right]. \end{aligned} \tag{29}$$

Here, $\Delta_1(X,X',X'') = \tilde{f}(X) - \tilde{f}_{\mathsf{decoupled}}(X',X'')$ and Δ_2, Δ_3 are defined analogously for g and h. By Cauchy-Schwarz and Lemma 6.5, we get that (29) is at most ξ. Fix X', Y', Z', and note that the 2-norm of $\tilde{f}_{\mathsf{decoupled}}(X',X'')$ over the choice of X'' is at most $O(1)$ and the influences are at most $O(\tau)$. We apply Theorem 3.3 and average over $(X'',Y'',Z'') \sim \mu^n$, to get that, provided that τ is small enough

$$\begin{aligned} &\Bigg| \mathop{\mathbb{E}}_{\substack{(X',Y',Z')\\(X'',Y'',Z'')}}\left[\Psi(\tilde{f}_{\mathsf{decoupled}}(X',X''), \tilde{g}_{\mathsf{decoupled}}(Y',Y''), \tilde{h}_{\mathsf{decoupled}}(Z',Z''))\right] \\ &\quad - \mathop{\mathbb{E}}_{\substack{(X',Y',Z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}}\left[\Psi(\tilde{f}_{\mathsf{decoupled}}(X',G_x), \tilde{g}_{\mathsf{decoupled}}(Y',G_y), \tilde{h}_{\mathsf{decoupled}}(Z',G_z))\right] \Bigg| \leqslant \xi. \end{aligned} \tag{30}$$

Next, by the smoothness of Ψ it follows that

$$\Bigg| \mathop{\mathbb{E}}_{\substack{(X',Y',Z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}} \left[\Psi(\tilde{f}_{\mathsf{decoupled}}(X',G_x), \tilde{g}_{\mathsf{decoupled}}(Y',G_y), \tilde{h}_{\mathsf{decoupled}}(Z',G_z))\right] - \mathop{\mathbb{E}}_{\substack{(X',Y',Z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}} \left[\Psi(F(X',G_x), G(Y',G_y), H(Z',G_z))\right]\Bigg| \leqslant E, \tag{31}$$

where

$$E = \mathop{\mathbb{E}}_{\substack{(X',Y',Z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}} \left[\sqrt{\xi(\tilde{f}_{\mathsf{decoupled}}(X',G_x)) + \xi(\tilde{g}_{\mathsf{decoupled}}(Y',G_y)) + \xi(\tilde{h}_{\mathsf{decoupled}}(Z',G_z))}\right],$$

and we recall that the function ξ is $\xi(a_1,\ldots,a_s) = \sqrt{\sum_i |\mathsf{trunc}(a_i) - a_i|^2}$.

Claim 6.8. $E \lesssim \sqrt{\zeta}$.

Proof. By Cauchy-Schwarz, $E \leqslant \sqrt{E_1+E_2+E_3}$ where

$$E_1 = \mathop{\mathbb{E}}_{\substack{(X',Y',Z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}} \left[\xi(\tilde{f}_{\mathsf{decoupled}}(X',G_x))\right], \qquad E_2 = \mathop{\mathbb{E}}_{\substack{(X',Y',Z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}} \left[\xi(\tilde{g}_{\mathsf{decoupled}}(Y',G_y))\right],$$

$$E_3 = \mathop{\mathbb{E}}_{\substack{(X',Y',Z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}} \left[\xi(\tilde{h}_{\mathsf{decoupled}}(Z',G_z))\right],$$

and we upper bound each one of E_1, E_2 and E_3 separately. As the arguments are identical, we show it only for E_1. By Theorem 3.9, provided that τ is small enough

$$E_1 \leqslant \mathop{\mathbb{E}}_{\substack{(X',Y',Z')\sim\mu^n\\(X'',Y'',Z'')\sim\mu^n}} \left[\xi(\tilde{f}_{\mathsf{decoupled}}(X',X''))\right] + \zeta.$$

By Fact 3.10 the function ξ is $O(1)$-Lipshitz, and so

$$\mathop{\mathbb{E}}_{X',X''} \left[\xi(\tilde{f}_{\mathsf{decoupled}}(X',X''))\right] \lesssim \mathop{\mathbb{E}}_{X,X',X''} \left[\xi(f'(X)) + \left|\tilde{f}_{\mathsf{decoupled}}(X',X'') - f'(X)\right|\right].$$

Note that as f' is 1-bounded, $\xi(f'(X)) = 0$. Also, note that f' is ζ-close in ℓ_2 distance to $\tilde{f}$, and so we get that

$$E_1 \leqslant \mathop{\mathbb{E}}_{X,X',X''} \left[\left|\tilde{f}_{\mathsf{decoupled}}(X',X'') - \tilde{f}(X)\right|\right] + O(\sqrt{\zeta}).$$

Applying Cauchy-Schwarz and Lemma 6.5 gives that $E_1 \lesssim \sqrt{\zeta}$, as desired. □

Combining all of the inequalities (27), (28), (29), (30) and (31) and Claim 6.8 proves (25). □

6.4 A Version of the Mixed Invariance Principle for CSPs

In this section, we prove a variant of Theorem 6.7 which is tailored to the application in approximating constraint satisfaction problems. The core issue we have to address is that we need to allow an algorithm designer the capabilities of making samples according to the mixed space we invariance into. Recall that an algorithm designer may generate Gaussian samples as G_x, G_y, G_z as in the theorem statement, and this is done via solving an SDP relaxation of the CSP. However, the same cannot be said about the inputs x, y, z which are also needed to evaluate the functions F, G and H. A closer inspection shows that we do not actually need the samples x, y, z to evaluate F, G and H; instead, we only need to know the values of $\sigma(x), \gamma(y)$ and $\phi(z)$ where σ, γ, ϕ are master embeddings of all distributions that we look at.

Generating samples of $\sigma(x), \gamma(y)$ and $\phi(z)$ to plug into F, G and H sounds like a feasible task. Consider the case that our master embedding has a particularly simple structure, and that they form embeddings into the Abelian group $(\mathbb{F}_p, +)$. In that case, the embeddings σ, γ, ϕ can be computed by solving the system of linear equations $\sigma(x) + \gamma(y) + \phi(z) = 0$ for all $(x, y, z) \in \mathsf{supp}(\mu)$. The solution to this system of equations is a subspace of $\mathbb{F}_p^{|\Sigma|+|\Gamma|+|\Phi|}$, and we may generate uniform samples from it so that we may plug them in place of the values of $\sigma(x), \gamma(y), \phi(z)$. Intuitively, this makes sense: the algorithm designer uses their capability of doing linear algebra (on top of their capabilities to solve SDP programs) in order to realize rounding schemes corresponding to the functions F, G and H in Theorem 6.7.

Indeed, ultimately we show that this is in fact the case, but some care is needed due to the mismatch in the distributions of group elements the algorithm designer generates, versus what the functions F, G and H expect. The issue boils down to the fact that the distribution of $\sigma(x), \gamma(y), \phi(z)$ where $(x, y, z) \sim \mu$ need not be the same as the one generated by the algorithm designer, and thus we are not able to say that, for any triplet of product functions

$$P(x) = \prod_{i=1}^{n} \chi_i(\sigma(x_i)), \qquad Q(y) = \prod_{i=1}^{n} \chi'_i(\gamma(y_i)), \qquad R(z) = \prod_{i=1}^{n} \chi''_i(\phi(z_i))$$

it holds that the expectation of $P(x)Q(y)R(z)$ is roughly the same under the two distributions. Having said that, there are two cases wherein we are able to make such an assertion:

1. If we have that $P(x)Q(y)R(z) \equiv 1$ pointwise on the support of μ^n, then the two expectations are of course the same.
2. If we have that $\Delta_{\mathsf{symbolic}}(PQR, 1) \geqslant T$, then each one of the expectations above is exponentially small in T, and in particular, they are roughly the same.

Thus, to make the shift from the distribution of $(\sigma(x), \gamma(y), \phi(z))$ in Theorem 6.7 to the distribution generated by the algorithm designer, we must ensure that any 3 triplet of functions associated with the decompositions used for f, g and h fall into one of the types above. We achieve this property in the context of dictatorship tests by a simple preprocessing step.

Towards this end, let H_{master} be the master group, and define the functions $F' \colon H_{\mathsf{master}}^n \times \mathbb{R}^{n'} \to \mathbb{C}$, $G' \colon H_{\mathsf{master}}^n \times \mathbb{R}^{n'} \to \mathbb{C}$ and $H' \colon H_{\mathsf{master}}^n \times \mathbb{R}^{n'} \to \mathbb{C}$ in the following way. We demonstrate it for F', and the same goes for G' and H'. Let $\mathcal{P}$ be the collection of product functions associated with f, and write each $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$ as $P = \prod_{i=1}^{n} \chi_{P,i}(\sigma(x_i))$. Thus,

$$\tilde{f}_{\mathsf{decoupled}}(x, G_x) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})} \prod_{i=1}^{n} \chi_{i,P}(\sigma(x_i)) L_P(G_x).$$

We define F' by replacing $\sigma(x_i)$ by the group element input. Namely,

$$F'(a, G_x) = \mathsf{trunc}\left(\sum_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})}\prod_{i=1}^{n}\chi_{i,P}(a_i)L_P(G_x)\right).$$

Theorem 6.9. *For all η, $r \in \mathbb{N}$ the following holds for sufficiently small τ. In the set up of Theorem 6.7, suppose that $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$ and $h\colon \Phi^n \to \mathbb{C}$ are 1-bounded functions, and that we associate with them collections of cyclic embedding embedding functions $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ respectively of size at most r such that for any $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$, $Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{Q})$ and $R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{R})$ it either $P(x)Q(y)R(z) \equiv 1$ in the support of μ^n, or $\Delta_{\mathsf{symbolic}}(PQR, 1) \geqslant T'$. Then*

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}\left[\Psi(f(x), g(y), h(z))\right] - \mathop{\mathbb{E}}_{\substack{a,b,c\in H^n_{\mathsf{master}}\\ a+b+c=0\\ (G_x,G_y,G_z)\sim\mathcal{G}^n}}\left[\Psi(F'(a, G_x), G'(b, G_y), H'(c, G_z))\right]\right| \lesssim \eta + 2^{-\Omega_{m,r,\alpha}(T')}. \tag{32}$$

6.4.1 Auxiliary Facts

The proof of Theorem 6.9 requires a few auxiliary facts that we record here.

Fact 6.10. *Let H be an Abelian group, Σ, Γ, Φ be finite alphabets and let μ be a pairwise connected distribution over $\Sigma \times \Gamma \times \Phi$. Suppose that $\sigma\colon \Sigma \to H$, $\gamma\colon \Gamma \to H$ and $\phi\colon \Phi \to H$ be maps such that (σ, γ, ϕ) is an Abelian embedding of μ into H and each one of $\mathsf{Image}(\sigma)$, $\mathsf{Image}(\gamma)$ and $\mathsf{Image}(\phi)$ generates H. Then, for any $\chi, \chi', \chi'' \in \hat{H}$, we have that the following conditions are equivalent:*

1. *For a, b, c such that $a + b + c = 0$ it holds that $\chi(a)\chi'(b)\chi''(c) = 1$.*
2. *For all $(x, y, z) \in \mathsf{supp}(\mu)$ we have that $\chi(\sigma(x))\chi'(\gamma(y))\chi''(\phi(z)) = 1$.*
3. *$\chi \equiv \chi' \equiv \chi''$.*

Proof. It is clear that the third item implies the first item, and that the first item implies the second item. Thus, it suffices to prove that the second item implies the third item.

Plugging in $\gamma(y) = -\sigma(x) - \phi(z)$, we get that for all $(x, z) \in \mathsf{supp}(\mu_{x,z})$ it holds that

$$(\chi\overline{\chi'})(\sigma(x)) = (\chi'\overline{\chi''})(\phi(z)).$$

In particular, we get that if x, x' have a common z such that (x, z) and (x', z) are both in the support of $\mu_{x,z}$, then $(\chi\overline{\chi'})(\sigma(x)) = (\chi\overline{\chi'})(\sigma(x'))$. Thus, the function $(\chi\overline{\chi'})(\sigma(x))$ is constant, and plugging in an x such that $\sigma(x) = 0$ gives it is the constant 1 function. As the image of σ is H and $\chi\overline{\chi'}$ is a character, it follows that $\chi\overline{\chi'} \equiv 1$ on H, and so $\chi = \chi'$. In a similar fashion, one can argue that $\chi' = \chi''$, and the proof is concluded. ☐

Fact 6.11. *In the setting of Fact 6.10, suppose that $\chi, \chi', \chi'' \in \widehat{H}^n$ are such that $(\chi\chi'\chi'')_i \neq 1$ for at least T' coordinates. Then letting α be the minimum probability of an atom in μ, we have that*

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}\left[\chi(\sigma(x))\chi'(\gamma(y))\chi''(\phi(z))\right]\right| \leqslant (1 - \Omega_{\alpha,|H|}(1))^{T'}.$$

Proof. Let $\mathcal{T}'$ be the set of i's such that $(\chi\chi'\chi'')_i \neq 1$. Then

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}\left[\chi(\sigma(x))\chi'(\gamma(y))\chi''(\phi(z))\right]\right| = \prod_{i\in\mathcal{T}'}\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}\left[\chi_i(\sigma(x_i))\chi_i'(\gamma(y_i))\chi_i''(\phi(z_i))\right]\right|,$$

and it suffices to upper bound each one of the terms by $1-\Omega_{\alpha,|H|}(1)$. Fix $i \in \mathcal{T}'$, and note that the values that $\chi_i\chi_i'\chi_i''$ may receive are discrete, all have absolute value 1 and they are $\Omega_{|H|}(1)$ far apart in absolute value. Thus, there is a fixed constant $c_{|H|,\alpha} > 0$ such that either $\left|\mathbb{E}_{(x,y,z)\sim\mu^n}\left[\chi_i(\sigma(x_i))\chi_i'(\gamma(y_i))\chi_i''(\phi(z_i))\right]\right| \leqslant 1 - c_{|H|,\alpha}$ or else $\left|\mathbb{E}_{(x,y,z)\sim\mu^n}\left[\chi_i(\sigma(x_i))\chi_i'(\gamma(y_i))\chi_i''(\phi(z_i))\right]\right| = 1$. Suppose for the sake of contradiction the latter holds. Then we get that there is a constant $\theta \in \mathbb{C}$ of absolute value 1 such that

$$\chi_i(\sigma(x_i))\chi_i'(\gamma(y_i))\chi_i''(\phi(z_i)) \equiv \theta.$$

Plugging in $(x_i, y_i, z_i) \in \mathsf{supp}(\mu)$ that maps to $(0,0,0) \in H^3$ we get that $\theta = 1$. It follows from Fact 6.10 that $\chi_i \equiv \chi_i' \equiv \chi_i''$, and contradiction. $\square$

Next, suppose we have collections $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{R}$ as in the setting of Theorem 6.9. For notational convenience, for each $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$ we pick $\chi_P \in \widehat{H}$ such that $P(x) = \chi_P(\sigma(x))$. We wish to consider the following two distributions:

1. The distribution $\mathcal{D}_H$: sample $(a,b,c) \in H^n$ uniformly such that $a+b+c=0$ and output

$$(\chi_P(a), \chi_Q(b), \chi_R(c))_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}), Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{Q}), R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{R})}.$$

2. The distribution $\mathcal{D}_\mu$: sample $(x,y,z) \sim \mu^n$ and output

$$(\chi_P(\sigma(x)), \chi_Q(\gamma(y)), \chi_R(\phi(z)))_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}), Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{Q}), R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{R})}.$$

The following lemma asserts that the distributions $\mathcal{D}_H$ and $\mathcal{D}_\mu$ are close in statistical distance.

Lemma 6.12. *Suppose the sizes of each one of $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ is at most r, $|H| \leqslant m$ and that for any $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$, $Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{Q})$ and $R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{R})$ it holds that either $PQR \equiv 1$ or else $P_iQ_iR_i \not\equiv 1$ for at least T' of the coordinates $i \in [n]$. Then*

$$\mathsf{SD}(\mathcal{D}_\mu, \mathcal{D}_H) \lesssim_{m,r} (1 - \Omega_{m,\alpha}(1))^{T'}.$$

Proof. The proof is similar to the proof of Fact 6.1. Note that the support of $\mathcal{D}_H, \mathcal{D}_\mu$ has size at most $O_{m,r}(1)$, and take $S = (a_P, b_Q, c_R)_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}), Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{Q}), R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{R})}$ in either one of the supports. Note that

$$\mathcal{D}_H(S) = \mathop{\mathbb{E}}_{\substack{a,b,c\in H^n\\ a+b+c=0}}\left[\prod_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} 1_{\chi_P(a)=a_P} \prod_{Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{Q})} 1_{\chi_Q(b)=b_Q} \prod_{R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{R})} 1_{\chi_R(c)=c_R}\right]$$

and

$$\mathcal{D}_\mu(S) = \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n}\left[\prod_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})} 1_{\chi_P(\sigma(x))=a_P} \prod_{Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{Q})} 1_{\chi_Q(\gamma(y))=b_Q} \prod_{R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{R})} 1_{\chi_R(\phi(z))=c_R}\right].$$

Arithmetizing the indicators as $1_{\chi_P(a)=a_P} = \prod_{\substack{a'\neq a_P \\ a'\in \mathsf{Image}(P)}} \frac{\chi_P(a)-a'}{a_P-a'}$ and similarly for the other ones, then opening things up, we get that there are coefficients $C(P,Q,R)$ that are at most $O_{m,r}(1)$ in absolute value such that

$$\mathcal{D}_H(S) = \sum_{\substack{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})\\ Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{Q})\\ R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{R})}} C(P,Q,R) \mathop{\mathbb{E}}_{\substack{a,b,c\in H^n\\ a+b+c=0}} [\chi_P(a)\chi_Q(b)\chi_R(c)]$$

and

$$\mathcal{D}_\mu(S) = \sum_{\substack{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})\\ Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{Q})\\ R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{R})}} C(P,Q,R) \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^n} [\chi_P(\sigma(x))\chi_Q(\gamma(y))\chi_R(\phi(z))].$$

Fix P,Q,R and consider their contribution to $\mathcal{D}_\mu(S)$ and $\mathcal{D}_H(S)$. For P,Q,R such that $PQR\equiv 1$, the two contributions are the same. Else, by assumption $P_iQ_iR_i \not\equiv 1$ for at least T' coordinates, and by fact each one of the expectations is at most $(1-\Omega_{m,\alpha}(1))^{T'}$ in absolute value. It follows that

$$|\mathcal{D}_H(S)-\mathcal{D}_\mu(S)| \leqslant \sum_{\substack{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P})\\ Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{Q})\\ R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{R})}} |C(P,Q,R)| \left(1-\Omega_{m,\alpha}(1)\right)^{T'} \lesssim_{m,r} \left(1-\Omega_{m,\alpha}(1)\right)^{T'}.$$

Therefore we get that $\mathsf{SD}(\mathcal{D}_\mu,\mathcal{D}_H) \lesssim_{m,r} \sum_S (1-\Omega_{m,\alpha}(1))^{T'} \lesssim_{m,r} (1-\Omega_{m,\alpha}(1))^{T'}$. □

6.4.2 Proof of Theorem 6.9

By Theorem 6.7 and the triangle inequality, it suffices to show that the difference between

$$(I) = \mathop{\mathbb{E}}_{\substack{(x,y,z)\sim\mu^n\\ (G_x,G_y,G_z)\sim\mathcal{G}^n}} [\Psi(F(x,G_x),G(y,G_y),H(z,G_z))]$$

and

$$(II) = \mathop{\mathbb{E}}_{\substack{a,b,c\in H^n_{\mathsf{master}}\\ a+b+c=0\\ (G_x,G_y,G_z)\sim\mathcal{G}^n}} \left[\Psi(F'(a,G_x),G'(b,G_y),H'(c,G_z))\right]$$

is at most $\lesssim_{r,m} 2^{-\Omega_{m,r,\alpha}(T')}$. Let $S\in\mathsf{supp}(\mathcal{D}_H)$, and define

$$M(S) = \mathbb{E}_{\substack{(x,y,z)\sim\mu^n\\ (G_x,G_y,G_z)\sim\mathcal{G}^n}} [\Psi(F(x,G_x),G(y,G_y),H(z,G_z)) \mid S];$$

the conditioning on S means $(\chi_P(\sigma(x)),\chi_Q(\gamma(y)),\chi_R(\phi(z)))_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}),Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{Q}),R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{R})} = S$. Note that $|M(S)|\leqslant 1$ always. Note that $(I) = \mathbb{E}_{S\sim\mathcal{D}_\mu}[M(S)]$ and $(II) = \mathbb{E}_{S\sim\mathcal{D}_H}[M(S)]$, and it follows that

$$|(I)-(II)| \leqslant \mathsf{SD}(\mathcal{D}_\mu,\mathcal{D}_H) \lesssim_{m,r} 2^{-\Omega m,r(T')}$$

where we used Lemma 6.12. □

7 Applying the Mixed Invariance in CSPs

The goal of this section is to prove Theorem 1.1.

One of the main difficulties in coming up with a dictatorship test with completeness 1 and analyzing its soundness is the following. Following Raghavendra [46], and as shown in [7], one way to construct a dictatorship test is to start with an instance of a Max-P-CSP such that its SDP value is 1. Such an SDP solution gives a set of distributions on the set of satisfying assignments of the predicate P which then can be used in constructing the queries of the dictatorship test. One issue with this approach is that even if we know that the SDP value is 1, there is no guarantee that a given local distribution is fully supported on $P^{-1}(1)$. Therefore, the support of a local distribution might not enjoy the pairwise connectedness and the non-$\mathbb{Z}$-embeddability even if the predicate P satisfies both properties. In this case, we cannot apply the analytical lemma from [11], in the soundness analysis of the dictatorship test.

In order to handle this issue, we demonstrate the use of our mixed invariance principle (which uses the analytical lemma from [11]) by focusing on a MILDLY-SYMMETRIC predicates. These predicates enjoy interesting *symmetry* properties. More specifically, any SDP solution with value 1 can be converted into another SDP solution with value 1 such that in the modified SDP solution, the support of every local distribution satisfies the pairwise connectedness and the non-$\mathbb{Z}$-embeddability conditions.

We start by setting up a few notations.

Instance of Max-$\mathcal{P}$-CSP Fix any collection $\mathcal{P}$ of MILDLY-SYMMETRIC predicates. An instance $\Upsilon = (\mathcal{V}, \mathcal{C})$ of Max-$\mathcal{P}$-CSP consists of a variable set $\mathcal{V}$ where the variables take values from Σ, and a distribution $\mathcal{C}$ on the constraint set. We will associate $\mathcal{V}$ with the set $[N] = \{1, \cdots, N\}$ for $N = |\mathcal{V}|$. Each constraint $C \in \mathsf{supp}(\mathcal{C})$ is over a tuple of 3 variables, denoted by $\mathcal{V}(C) = (s_1, s_2, s_3)$, and consists of a predicate $P_C : \Sigma^3 \to \{0, 1\}$. An assignment (x, y, z) to the tuple $\mathcal{V}(C)$ satisfies the constraint C iff $(x, y, z) \in P_C^{-1}(1)$ where $P_C \in \mathcal{P}$. For ease of notation, we will use C to refer to the constraint as well as the underlying predicate P_C, and simply write $C(x, y, z) = 1$ or $(x, y, z) \in C^{-1}(1)$ if $P_C(x, y, z) = 1$.

7.1 The SDP Program

Given an instance $\Upsilon = (\mathcal{V}, \mathcal{C})$ of Max-$\mathcal{P}$-CSP, the basic semidefinite programming relaxation of the instance is given in Figure 2. Given a set T, the set of all the distributions on T is denoted by $\blacktriangle(T)$. The SDP formulation consists of vectors $\{\boldsymbol{b}_{i,a}\}_{i\in\mathcal{V}, a\in\Sigma}$, distributions $\{\mu_C\}_{C\in\mathsf{supp}(\mathcal{C})}$ over local assignments (i.e., on $\Sigma^{\mathcal{V}(C)}$) and a unit vector $\boldsymbol{b}_0$. Let $\mathbf{val}(\boldsymbol{V}, \boldsymbol{\mu})$ be the objective value of the solution $(\boldsymbol{V}, \boldsymbol{\mu})$.

For every $\eta > 0$, the SDP can be solved up to an additive accuracy of η in time $\mathsf{poly}(n, \log(1/\eta))$. We will ignore this issue of approximation and assume that the SDP can be solved optimally in polynomial time.

During the execution of our algorithm, we will modify the SDP by imposing additional linear conditions on the local distributions. We define a valid *integral solution* to the SDP program as follows:

Definition 7.1. *Fix any assignment $\boldsymbol{\alpha}$ to Υ. The vector assignment*

$$b_{i,a} = \begin{cases} \boldsymbol{b}_0 & a = \alpha|_i \\ 0 & \text{otherwise} \end{cases}$$

along with $\boldsymbol{\mu}$ where for every $C \in \mathsf{supp}(\mathcal{C})$, $\mu_C(\alpha|_{\mathcal{V}(C)}) = 1$ and $\mu_C(\boldsymbol{d}) = 0$ for every $\boldsymbol{d} \neq \alpha|_{\mathcal{V}(C)}$ is called an integral assignment to the SDP. Such an assignment is called a valid integral solution if it is a feasible solution to the SDP.

$$
\begin{aligned}
&\text{maximize} && \mathop{\mathbb{E}}_{C\sim\mathcal{C}}\mathop{\mathbb{E}}_{x\in\mu_C}[C(x)] && \\
&\text{subject to} && \langle \boldsymbol{b}_{i,a}, \boldsymbol{b}_{j,b}\rangle = \Pr_{x\sim\mu_C}[x_i = a, x_j = b] && C \in \mathtt{supp}(\mathcal{C}), \quad i,j \in \mathcal{V}(C), \quad a,b\in\Sigma \quad (33)\\
& && \langle \boldsymbol{b}_{i,a}, \boldsymbol{b}_0\rangle = \|\boldsymbol{b}_{i,a}\|_2^2 && \forall i \in \mathcal{V}, a\in\Sigma \quad (34)\\
& && \|\boldsymbol{b}_0\|_2^2 = 1 && (35)\\
& && \mu_C \in \blacktriangle(\Sigma^{\mathcal{V}(C)}) && C\in\mathtt{supp}(\mathcal{C}) \quad (36)
\end{aligned}
$$

Figure 2: Basic SDP relaxation of a Max-CSP instance $\Upsilon = (\mathcal{V}, \mathcal{C})$.

7.2 Setting up a System of Linear Equations

Once the SDP relaxation is solved, we construct an initial system of linear equations over a certain abelian group. In this section, we describe an algorithm for formulating this system of linear equations. For convenience, let us call this system of linear equations GE System.

Fix an arbitrary SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ with value 1. The solution induces *local distributions* μ_C over $\Sigma^{\mathcal{V}(C)}$ where $C \in \mathsf{supp}(\mathcal{C})$. We assume that for every $C \in \mathsf{supp}(\mathcal{C})$, the support of μ_C is pairwise connected and has no $\mathbb{Z}$-embedding.

7.2.1 Setting up the Variables associated to $v \in \mathcal{V}$ in GE System

In our GE System, there will be many variables associated with a given variable $v \in \mathcal{V}$ from the CSP instance Υ. Here, we describe a polynomial-time procedure that first constructs a matrix M_v with $|\Sigma|$ rows associated with the variable v. The columns of the matrix are all the embedding functions associated with all the constraints v is involved in.

In order to be consistent across different embeddings, we will need to work with embeddings that assign the identity element to a special element from Σ. It will be convenient to treat Σ as $[q] = \{1, 2, \dots, q\}$ where $q = |\Sigma|$ and let $w^\star = 1$.

Definition 7.2. *An embedding $\sigma_1, \sigma_2, \sigma_3 : \Sigma \to G$ of a subset $S \subseteq \Sigma \times \Sigma \times \Sigma$ is called a standard embedding if $\sigma_1(w^\star) = \sigma_2(w^\star) = \sigma_3(w^\star) = 0_G$ and there exists $g \in G$ such that for every $(x, y, z) \in S$, $\sigma_1(x) + \sigma_2(y) + \sigma_3(z) = g$. We will denote such embeddings by $((\sigma_1, \sigma_2, \sigma_3), g)$.*

Note that any embedding $(\sigma_1, \sigma_2, \sigma_3)$ defined in Definition 1.2 can be converted into a standard embedding by adding $-\sigma_1(w^\star), -\sigma_2(w^\star), -\sigma_3(w^\star)$ to the maps $\sigma_1, \sigma_2, \sigma_3$ respectively.

An informal description of the algorithm (presented in Algorithm 1) for computing the matrix M_v is as follows. Fix a variable $v \in \mathcal{V}$. The matrix M_v is generated in the following three steps.

Step 1. In this step, for every constraint C such that $v \in \mathcal{V}(C)$, and every embedding $(\sigma_1, \sigma_2, \sigma_3, g)$ of $\mathsf{supp}(\mu_C)$ into an ableian group G, we add columns corresponding to the evaluations of characters of G on $(\sigma_j(x))_{x\in\Sigma}$ where j is the index of v in C.

Step 2. Next, for every subset S of columns in the matrix M_v generated in Step 1, we add a column, which is a pointwise multiplication of columns S. If we treat each complex number in the polar form as $e^{i\theta}$, then in this step, we add all the linear combinations of columns when viewed as vectors of exponents.

Step 3. Finally, we add rows that are pointwise multiplication of the subsets of rows of M_v generated in Step 2. Similar to Step 2., here we add all the linear combinations of the exponents of the rows.

This completes the description of the algorithm for computing M_v. In the actual algorithm, we also keep track all the group elements mapped to a given element from Σ across all the embeddings. This is stored in the variable r_ℓ^v. Finally, we denote by $H_v^\star$ a subgroup generated by $\{r_\ell^v\}_{\ell\in\Sigma}$.

Algorithm 1 Computing the matrix M_v

1: Start with $q \times 1$ matrix M_v with $\vec{1}$ as a column.
2: Instantiate r_ℓ^v to be an empty tuple for every $\ell \in [q]$. Set H to be the trivial group.
3: Suppose v is involved in the constraints $C_1, C_2, \ldots, C_t$.
4: **for** $i \leftarrow 1$ to t **do**
5: Let $j \in \{1,2,3\}$ be the index of v in $\mathcal{V}(C_i)$.
6: **for** each standard irreducible embedding $((\sigma_1,\sigma_2,\sigma_3),g)$[12]of $\mathtt{supp}(\mu_{C_i})$ into an abelian group G **do**
7: Add a column $(\chi(\sigma_j(x))_{x\in\Sigma})$ to M_v for every $\chi \in \hat{G}$ such that $\chi \not\equiv 1$.
8: $H \leftarrow H \times G$.
9: $r_\ell^v \leftarrow (r_\ell^v, \sigma_j(\ell))$ for every $\ell \in [k]$. ▷ $r_\ell^v \in H$.
10: **end for**
11: **end for**
12: Set $G^v_{\mathsf{master}} \leftarrow H$.
13: Let L' be the number of columns in M_v
14: **for** every subset $S \subseteq [L']$ **do** ▷ Adding more columns
15: **if** $\circ_{i\in S} M_v[.][i]$ is not present as a column in M_v **then** ▷ $M_v[.][i]$ is the i^{th} column of M_v and $\circ$ is pointwise multiplication
16: Add a column $\circ_{i\in S} M_v[.][i]$ to M_v.
17: **end if**
18: **end for**
19: Let L_v be the number of columns of M_v.
20: Let $H_v^\star$ be the group generated by $\{r_\ell^v\}_{\ell\in[q]}$. Let $i = q$
21: **for** each $h \in H_v^\star$ **do** ▷ Adding more rows
22: Suppose $h = \sum_{\ell\in S} r_\ell^v$ where S is a multi-subset of $[q]$ and addition is a group operation G^v_{master}.
23: Add a row $\circ_{\ell\in S} M_v[\ell][.]$ to M_v and set $i \leftarrow i+1$.
24: Set $r_i^v = h$.
25: **end for**

We start with a few simple observations.

Claim 7.3. *The group G^v_{master} generated in Step 12. of Algorithm 1 is the Master group associated with the set of distributions μ_{C_i} where $v \in \mathcal{V}(C_i)$.*

Proof. First, we add all the embedding of v in lines 4-11 of the algorithm. While doing this, we update the group H appropriately by 'adding' the current group G to it. The variable r_ℓ^v corresponds to the element of H associated with the row ℓ. After the algorithm exits from the for loop at line 12, the value of r_ℓ^v is the master embedding of the symbol $\ell \in \Sigma$ in G^v_{master}. □

Claim 7.4. *The columns of the matrix M_v contain evaluations of all the characters of G^v_{master} on a subgroup generated by $\{r_1^v, r_2^v, \ldots, r_q^v\}$.*

Proof. One should think of the column generated by $\chi \in \hat{G}$ in step 7 as the evaluation of the character $(1, \ldots, 1, \chi, 1, \ldots, 1) \in \widehat{G^v_{\mathsf{master}}}$ on r^v_ℓ. With this view in mind, in steps 14-18, we add columns that correspond to every other character from $\widehat{G^v_{\mathsf{master}}}$ evaluated at $\{r^v_\ell\}_{\ell \in [q]}$. This follows as the character $\boldsymbol{\chi} := (\chi_1, \chi_2, \ldots, \chi_t) \in \widehat{G^v_{\mathsf{master}}}$ is, by definition, $\boldsymbol{\chi}(\boldsymbol{g}) = \prod_{i=1}^t \chi_i(g_i)$, and Step 16. is precisely adding such columns.

Finally, in lines 20-14, we are adding rows that correspond to the evaluations of all the characters from $\widehat{G^v_{\mathsf{master}}}$ on every element of G^v_{master} generated by $\{r^v_1, r^v_2, \ldots, r^v_q\}$. This follows as $\boldsymbol{\chi} \in \widehat{G^v_{\mathsf{master}}}$ is a group homomorphism $\boldsymbol{\chi} : G^v_{\mathsf{master}} \to \mathbb{C}$, i.e., $\boldsymbol{\chi}(\boldsymbol{g}_1 + \boldsymbol{g}_2) = \boldsymbol{\chi}(\boldsymbol{g}_1) \cdot \boldsymbol{\chi}(\boldsymbol{g}_2)$ for every $\boldsymbol{g}_1, \boldsymbol{g}_2 \in G^v_{\mathsf{master}}$. □

The variables in the GE System associated to $v \in \mathcal{V}$ are $\{y_v^{\vec{\chi}}\}$ where $\vec{\chi} \in \widehat{G^v_{\mathsf{master}}}$. It would be convenient to think of the variable set as a row of the matrix M_v.

7.2.2 Adding Equations to the GE System

Our next step is to set up a system of linear equations. Ideally, we would like to assign a group element from $\{r^v_1, r^v_2, \ldots, r^v_q\}$ to a variable v (which corresponds to assigning an element from Σ). However, this may not be enforced by a set of linear equations over the group G^v_{master}. Instead, we relax this requirement and expect an assignment from $H^\star_v$ to the variable v.

In a GE System that we describe next, a solution to the GE System corresponds to assigning a group element from $H^\star_v$ to v. We do this by relating values to the variables associated with $v \in \mathcal{V}$ to a group element. In other words, every $h \in H^\star_v$ is identified by the respective row of the matrix M_v.

The following two types of sets of linear equations are added to the GE System.

1. **Valid character constraints.** These equations enforce that the variables $\{y_v^{\vec{\chi}}\}_{\vec{\chi} \in \widehat{G^v_{\mathsf{master}}}}$ correspond to a row of the matrix M_v, and hence, the vector assignment corresponds to a group element from $H^\star_v$. Towards this, we add the following set of equations for every $v \in \mathcal{V}$,
   - $y_v^{\mathsf{triv}} = 1$, where triv corresponds to the first column of the matrix M_v.
   - For every column $\vec{\chi}' := (1, \ldots, 1, \chi, 1, \ldots, 1)$ added in step 7, if $G \cong \mathbb{Z}_{p^d}$, add a constraint $(y_v^{\vec{\chi}'})^{p^d} = 1$.
   - For every $\vec{\chi}, \vec{\chi}', \vec{\chi}''$ such that $\vec{\chi}'' = \vec{\chi} \cdot \vec{\chi}'$, we add the equation $y_v^{\vec{\chi}''} = y_v^{\vec{\chi}} \cdot y_v^{\vec{\chi}'}$.
   - As $H^\star_v$ can be a subgroup of G^v_{master}, there will be certain columns in M_v that are constant. If the column corresponding to $\vec{\chi}$ is constant c, then add the equation $y_v^{\vec{\chi}} = c$.
2. **Valid satisfying assignments constraints.** These equations make sure that the solution to the GE System is consistent with the SDP local distributions (the latter are supported on the set of satisfying assignments). Towards this, we add the following set of equations
   - For every constraint $C \in \mathsf{supp}(\mathcal{C})$ such that $\mathcal{V}(C) = (s_1, s_2, s_3)$, and for every embedding $((\sigma_1, \sigma_2, \sigma_3), g)$ of $\mathsf{supp}(\mu_C)$ in G, we add the equation $y_{s_1}^{\vec{\chi}_1} \cdot y_{s_2}^{\vec{\chi}_2} \cdot y_{s_3}^{\vec{\chi}_3} = \chi(g)$ where $\chi \in \hat{G}$ and $\vec{\chi}_1, \vec{\chi}_2, \vec{\chi}_3$ are the respective columns added in step 7 in $M_{s_1}, M_{s_2}, M_{s_3}$, for this embedding and the character χ.

The equations are linear equations over the circle group $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$ and hence a solution to the GE System can be found in polynomial time.

The justification for these constraints is given by the following observation, which states that a valid integral solution to the SDP gives a solution to the GE System.

Observation 7.5. *Suppose $\boldsymbol{x} \in [k]^N$ is a satisfying assignment to the instance Υ and suppose that this assignment survives in the SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$, i.e., for every $C \in \mathsf{supp}(\mathcal{C})$, the local distribution μ_C assigns a non-zero probability mass to $\boldsymbol{x}|_{\mathcal{V}(C)}$. Then assigning the $\boldsymbol{x}_v^{th}$ row from M_v to the variables $\{y_v^{\vec{X}}\}$ satisfies all the equations from the* GE System.

There is a natural way in which one can view a solution to the system of equations above as an element of $\prod_{v \in \mathcal{V}} H_v^\star$. With this map, we denote the set of all the solutions to the GE System as a subset of $\prod_{v \in \mathcal{V}} H_v^\star$.

7.3 Set up towards the Hybrid Algorithm: Modifying the SDP and the GE System

Jumping ahead in the soundness analysis of our dictatorship test that matches the guarantees of our (to be stated) approximation algorithm, we are interested in computing the expectation

$$\mathop{\mathbb{E}}_{C(s_1,s_2,s_3)\sim\mathcal{C}}\left[\mathop{\mathbb{E}}_{(x,y,z)\sim\mu_C}[F_{s_1}(\sigma(x),x)\cdot F_{s_2}(\gamma(y),y)\cdot F_{s_3}(\phi(z),z)]\right]$$

where the first input goes to the product functions from the decomposition of $F_{v_i}s$ and the second input goes to the low-degee functions. During the application of our mixed invariance principle, we want to replace the local distribution $(\sigma(x), \gamma(y), \phi(z))$ by some global distribution. In our analysis, the global distribution will be a random solution to our GE System.

One important technical condition that is required to apply the mixed invariance principle is that when we make this switch, then the product functions coming from F_{v_i}s should satisfy the following property.

- For every tuple of product functions P, Q and R coming from the decompositions of F_{s_1}, F_{s_2} and F_{s_3}, respectively, $PQR \equiv 1$ under the local distribution iff $PQR \equiv 1$ under the global distribution.

In order to make sure this technical condition is satisfied, we modify our SDP solution as well as the GE System iteratively. This procedure is given in Algorithm 2. This algorithm has two main components.

1. ModifyGESystem: In this procedure, we modify the GE System so that if $PQR \equiv 1$ under the local distribution then $PQR \equiv 1$ under the global distribution.

2. ModifySDP: In this procedure, we modify the SDP formulation so that if $PQR \equiv 1$ under the global distribution then $PQR \equiv 1$ under the SDP local distribution.

The following claim shows the correctness of the procedure ModifyGESystem given in Algorithm 4.

Claim 7.6. *Suppose for cyclic embedding function P, Q, and R, $PQR \equiv 1$ under the SDP local distribution, then the* GE System *can be modified so that $PQR \equiv 1$ under the modified global distribution (i.e., under a random solution to the* GE System*).*

Proof. Fix a constraint $C \in \mathtt{supp}(\mathcal{C})$ and let $\mathcal{V}(C) = (s_1, s_2, s_3)$. In the algorithm for computing M_{s_i}s, we add columns corresponding to all the embeddings $((\sigma_j, \sigma'_j, \sigma''_j), g_j)$ of $\mathtt{supp}(\mu_C)$ over the group G_j for $j = 1, 2, \ldots, t$ for some $t \geqslant 1$. Let $H_C^\star := \prod_{j=1}^t G_j$ be the product group. Note that $H_C^\star$ is a subgroup of $G_{\mathsf{master}}^{v_i}$ for $i \in [3]$.

Now, if $PQR \equiv 1$ under the local distribution then the product functions must be from the $\mathsf{sp}_{\mathbb{N}}$ of the characters over $H_C^\star$, as these product functions correspond to an embedding of $\mathtt{supp}(\mathcal{C})$ and $H_C^\star$ is the master group of $\mathsf{supp}(\mu_C)$ by definition. Let $\mathsf{Hspan}(\mathsf{supp}(\mu_C))$ be the subgroup of $H_C^\star \times H_C^\star \times H_C^\star$ generated by $\{(\sigma_j(a), \sigma'_j(b), \sigma''_j(c)) \mid (a,b,c) \in \mathsf{supp}(\mu_C)\}$. Using the fact that each of P, Q, R belongs to the $\mathsf{sp}_{\mathbb{N}}$ of the characters over $H_C^\star$, it can be easily checked that $PQR \equiv 1$ on the subgroup $\mathsf{Hspan}(\mathsf{supp}(\mu_C))$. This means that there is a unique collection of characters $\{(\chi_1^i, \chi_2^i, \chi_3^i)\}_i$ of the group $H_C^\star \times H_C^\star \times H_C^\star$ that evaluates to the constant 1 on $\mathsf{Hspan}(\mathsf{supp}(\mu_C))$ and are not constant on any subgroup containing it. Using these characters, we can add an additional equation to the GE System that enforces $PQR \equiv 1$ on every solution to the modified GE System. More specifically, we add the equation $y_{s_1}^{\chi_1^i} \cdot y_{s_2}^{\chi_2^i} \cdot y_{s_3}^{\chi_3^i} = 1$ for all i to the GE System. $\square$

The above claim fixes the problem in one direction. We also require that $PQR \equiv 1$ under a random solution to the GE System implies that $PQR \equiv 1$ under the local distribution. This is easy to fix by imposing conditions on the local distributions in the SDP. This is done in Algorithm 3 by simply excluding the assignments from a local distribution not implied by the GE System.

Claim 7.7. *Suppose $PQR \equiv 1$ for every solution to the* GE System*, then the SDP can be modified so that $PQR \equiv 1$ under the SDP local distribution.*

Thus, we run the procedures modifyGESystem and modifySDP towards achieving the condition that $PQR \equiv 1$ under a random solution to the GE System iff $PQR \equiv 1$ under the local distribution. The procedure modifyGESystem depends on the SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$. We need to make sure that after running the procedure on a satisfiable instance, the final SDP and the GE System should have all the satisfying assignments preserved (i.e., they will be valid integral solutions to the final SDP and the master embedding of any valid satisfiable assignment is a solution to the GE System).

Preserving all the integral solutions. To make sure we do not exclude any satisfying assignment from the subsequent SDP formulation during the execution of Algorithm 2, we will need to work with a SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ in step 3 with the following property: For every satisfying assignment $\sigma \in \Sigma^n$ to the instance, and any constraint $C \in \mathtt{supp}(\mathcal{C})$, the support of the local distribution μ_C contains $\sigma|_{\mathcal{V}(C)}$. To see the necessity of this condition on $(\boldsymbol{V}, \boldsymbol{\mu})$, suppose there are multiple satisfying assignments to Υ, but the SDP solution $\mathcal{S}$ is supported on a specific satisfying assignment $\sigma^\star$. During the execution of ModifyGEStstem in step 15 of the algorithm, the GE System will exclude every other assignment from its solution space. Thus, if we end up running ModifySDP in the subsequent iterations, then the SDP formulation will no longer support satisfying assignments other than $\sigma^\star$.

Lemma 7.9 below states that we can get an SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ with value 1 such that for every satisfiable assignment $\boldsymbol{\alpha}$ to the instance Υ, and for every $C \in \mathtt{supp}(\mathcal{C})$, $\mu_C(\alpha|_{\mathcal{V}(C)}) > 0$. Before we state and prove the lemma, we first prove the following simple claim.

Claim 7.8. *If $(\boldsymbol{V}, \boldsymbol{\mu})$ and $(\boldsymbol{V}', \boldsymbol{\mu}')$ be any two SDP solutions with value 1. Then there is a SDP solution $(\boldsymbol{V}'', \boldsymbol{\mu}'')$ with value 1 such that for every constraint $C \in \mathtt{supp}(\mathcal{C})$,*

$$\mathtt{supp}(\mu''_C) = \mathtt{supp}(\mu_C) \cup \mathtt{supp}(\mu'_C).$$

Proof. Suppose the SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ consists of vectors $\{\boldsymbol{b}_{i,a}\} \cup \{\boldsymbol{b}_0\}$ and the SDP solution $(\boldsymbol{V}', \boldsymbol{\mu}')$ consists of vectors $\{\boldsymbol{b}'_{i,a}\} \cup \{\boldsymbol{b}'_0\}$ from $\mathbb{R}^{qn+1}$. Let $\boldsymbol{e}_1 = (1,0)$ and $\boldsymbol{e}_2 = (0,1)$ be the vectors in $\mathbb{R}^2$. Consider the following vectors

$$\boldsymbol{b}''_{i,a} = \frac{1}{\sqrt{2}}(\boldsymbol{e}_1 \otimes \boldsymbol{b}_{i,a}) + \frac{1}{\sqrt{2}}(\boldsymbol{e}_2 \otimes \boldsymbol{b}'_{i,a}), \qquad \boldsymbol{b}''_0 = \frac{1}{\sqrt{2}}(\boldsymbol{e}_1 \otimes \boldsymbol{b}_0) + \frac{1}{\sqrt{2}}(\boldsymbol{e}_2 \otimes \boldsymbol{b}'_0)$$

It can be easily verified that

$$\begin{aligned}\langle \boldsymbol{b}''_{i,a}, \boldsymbol{b}''_0\rangle &= \langle \frac{1}{\sqrt{2}}(\boldsymbol{e}_1 \otimes \boldsymbol{b}_{i,a}) + \frac{1}{\sqrt{2}}(\boldsymbol{e}_2 \otimes \boldsymbol{b}'_{i,a}), \boldsymbol{b}''_0\rangle \\ &= \frac{1}{\sqrt{2}}\langle \boldsymbol{e}_1 \otimes \boldsymbol{b}_{i,a}, \boldsymbol{b}''_0\rangle + \frac{1}{\sqrt{2}}\langle \boldsymbol{e}_2 \otimes \boldsymbol{b}'_{i,a}, \boldsymbol{b}''_0\rangle \\ &= \frac{1}{2}\|\boldsymbol{b}_{i,a}\|^2 + \frac{1}{2}\|\boldsymbol{b}'_{i,a}\|^2 \\ &= \|\boldsymbol{b}''_{i,a}\|^2,\end{aligned}$$

Similarly, $\|\boldsymbol{b}''_0\|^2 = \frac{1}{2}\|\boldsymbol{b}_0\|^2 + \frac{1}{2}\|\boldsymbol{b}'_0\|^2 = 1$. Consider $\mu_C = \frac{1}{2}\mu_C + \frac{1}{2}\mu'_C$ for every $C \in \mathtt{supp}(\mathcal{C})$. We have

$$\begin{aligned}\langle \boldsymbol{b}''_{i,a}, \boldsymbol{b}''_{j,b}\rangle &= \langle \frac{1}{\sqrt{2}}(\boldsymbol{e}_1 \otimes \boldsymbol{b}_{i,a}) + \frac{1}{\sqrt{2}}(\boldsymbol{e}_2 \otimes \boldsymbol{b}'_{i,a}), \frac{1}{\sqrt{2}}(\boldsymbol{e}_1 \otimes \boldsymbol{b}_{j,b}) + \frac{1}{\sqrt{2}}(\boldsymbol{e}_2 \otimes \boldsymbol{b}'_{j,b})\rangle \\ &= \frac{1}{2}\langle \boldsymbol{b}_{i,a}, \boldsymbol{b}_{j,b}\rangle + \frac{1}{2}\langle \boldsymbol{b}'_{i,a}, \boldsymbol{b}'_{j,b}\rangle \\ &= \Pr_{x\sim\mu''_C}[x_i = a, x_j = b]\end{aligned}$$

Therefore, the vectors $\boldsymbol{V}'' = \{\boldsymbol{b}''_{i,a}\} \cup \{\boldsymbol{b}''_0\}$ along with the local distributions $\boldsymbol{\mu}''$ satisfy all the SDP constraints. Furthermore, we have $\mathtt{supp}(\mu''_C) = \mathtt{supp}(\mu_C) \cup \mathtt{supp}(\mu'_C)$ for every $C \in \mathtt{supp}(\mathcal{C})$. Finally, the SDP value of the solution $(\boldsymbol{V}'', \boldsymbol{\mu})$ is 1 as μ_C is supported on the set of satisfying assignments to C for every $C \in \mathtt{supp}(\mathcal{C})$.

□

Lemma 7.9. *Fix a satisfiable instance $\Upsilon(\mathcal{V}, \mathcal{C})$ of a Max-$\mathcal{P}$-CSP and a SDP formulation SDP(Υ) such that every satisfiable assignment to Υ is a valid integral solution to SDP(Υ).*

There is a polynomial-time algorithm that returns a SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ with value 1 such that for every satisfiable assignment $\boldsymbol{\alpha}$ to the instance Υ, and for every $C \in \mathtt{supp}(\mathcal{C})$, $\mu_C(\alpha|_{\mathcal{V}(C)}) > 0$.

Proof. To prove this lemma, we will combine different SDP solutions using Claim 7.8. For every constraint C and every $d = (a_1, a_2, a_3) \in \Sigma^3$ such that $C(d) = 1$, we augment the relaxation SDP(Υ) with a constraint $\mu_C(d) = 1$; call this new relaxation SDP$_{C,d}(\Upsilon)$. Let $(\boldsymbol{V}_{C,d}, \boldsymbol{\mu}_{C,d})$ be any arbitrary solution to SDP$_{C,d}(\Upsilon)$. We then combine all the solutions $(\boldsymbol{V}_{C,d}, \boldsymbol{\mu}_{C,d})$ with value 1 iteratively using Claim 7.8 to get a solution $(\boldsymbol{V}, \boldsymbol{\mu})$. This solution satisfies the required guarantee. To see this, fix any satisfiable assignment $\boldsymbol{\alpha}$ to the instance Υ and fix any $C \in \mathtt{supp}(\mathcal{C})$. Since, $\mu_C(\alpha|_{\mathcal{V}(C)}) = 1$, the solution $(\boldsymbol{V}_{C,\alpha|_{\mathcal{V}(C)}}, \boldsymbol{\mu}_{C,\alpha|_{\mathcal{V}(C)}})$ has value 1 and hence $\mu_C(\alpha|_{\mathcal{V}(C)}) > 0$ using Claim 7.8. Furthermore, Claim 7.8 guarantees that the value of the solution $(\boldsymbol{V}, \boldsymbol{\mu})$ is 1, proving this lemma. □

While Algorithm 2 makes sure that the SDP solution and the GE System are compatible towards applying the invariance principle in the analysis of our dictatorship test, it is important to show that the modified SDP has all the satisfying assignments as valid integral solutions. The following claim shows precisely this.

Claim 7.10. *If $\boldsymbol{x} \in \Sigma^N$ is a satisfying assignment to the instance Υ, then it remains a valid integral solution to the SDP $\mathcal{S}$ that we get after running Algorithm 2.*

Proof. We will show that the procedure maintains the following two invariants: 1) every satisfying assignment from $\boldsymbol{x} \in [q]^N$ is also a valid satisfying assignment (after interpreting $\boldsymbol{x}_v$ with $r^v_{x_v}$) to the GE System after the modification, and 2) Every satisfying assignment from $\boldsymbol{x} \in [q]^N$ is a valid integral solution to SDP. These invariants are clearly satisfied at the beginning.

When we run ModifySDP, the modified SDP has all the satisfying assignments to the instance Υ as valid integral solutions. This follows as $\mathcal{T}$ has the embedding of every satisfying assignment to start with, and we only excluded assignments from the local distributions that are not implied by the GE System.

Next, when we modify the GE System, we add equations, and these equations do not exclude any embedding from the support of μ_C for any $C \in \mathsf{supp}(\mathcal{C})$. As the solution from step 3 is guaranteed to keep $\alpha|_{\mathcal{V}(C)}$ for every satisfying assignment α and C, every satisfying assignment from $[q]^N$ is also a valid satisfying assignment to the GE System after the modification.

□

Algorithm 2 Massaging the SDP and the GE System towards applying the mixed invariance principle

1: **Input:** An instance $\Upsilon(\mathcal{V}, \mathcal{C})$ of Max-$\mathcal{P}$-CSP.
2: Let SDP(Υ) be the basic semidefinite program from Figure 2.
3: Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be the SDP solution to SDP(Υ) guaranteed by Lemma 7.9.
4: **if** $\mathrm{val}(\boldsymbol{V}, \boldsymbol{\mu}) \neq 1$ **then**
5:     Abort. ▷ Υ is not satisfiable
6: **end if**
7: **if** for a constraint $C \in \mathsf{supp}(\mathcal{C})$, $\mathsf{supp}(\mu_C)$ is either pairwise disconnected or has a $\mathbb{Z}$-embedding **then**
8:     Abort. ▷ Υ is not satisfiable or C is not MILDLY-SYMMETRIC
9: **else**
10:     Set up a GE System using the SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ and solve it.
11:     Let $\mathcal{T} \subseteq \prod_{v=1}^{N} H_v^\star$ be the set of solutions to the GE System.
12:     **if** $\exists\, (a, b, c) \in \mathsf{supp}(\mu_C)$ such that $(r_a^{s_1}, r_b^{s_2}, r_c^{s_3}) \notin \mathcal{T}|_{\mathcal{V}(C)}$ where $\mathcal{V}(C) = (s_1, s_2, s_3)$ **then**
13:         Run ModifySDP below.
14:     **else**
15:         Run ModifyGESystem below.
16:     **end if**
17:     **if** the SDP $\mathcal{S}$ or the GE System is modified **then**
18:         Repeat from step 3 above
19:     **else**
20:         Return $\mathcal{S}$ and the GE System
21:     **end if**
22: **end if**

Algorithm 3 ModifySDP

1: **for** every $(a, b, c) \in \mathsf{supp}(\mu_C)$ such that $(r_a^{s_1}, r_b^{s_2}, r_c^{s_3}) \notin \mathcal{T}|_{\mathcal{V}(C)}$ where $\mathcal{V}(C) = (s_1, s_2, s_3)$ **do**
2:     Augment SDP(Υ) by adding the constraint $\mu_C(a, b, c) = 0$.
3: **end for**

We now have the following important lemma that shows that applying Algorithm 2 on a CSP instance where every predicate is MILDLY-SYMMETRIC give returns an SDP solution and a system of linear equations that are consistent with respect to each other.

Algorithm 4 ModifyGESystem

for every $C \in \mathsf{supp}(\mathcal{C})$ **do**
  Let $\mathcal{V}(C) = (s_1, s_2, s_3)$, and $H^\star_C$ be as defined in the proof of Claim 7.6.
  Let $\mathsf{Hspan}(\mathsf{supp}(\mu_C))$ be the subgroup of $H^\star_C \times H^\star_C \times H^\star_C$, again defined in the proof of Claim 7.6.
  Let $\mathsf{Hspan}(\mathcal{T}|_{\mathcal{V}(C)}) := \{(\boldsymbol{x}_{s_1}|_{H^\star_C}, \boldsymbol{x}_{s_2}|_{H^\star_C}, \boldsymbol{x}_{s_3}|_{H^\star_C}) \mid \boldsymbol{x} \in \mathcal{T}\}$.
  if $\mathsf{Hspan}(\mathsf{supp}(\mu_C)) \subsetneq \mathsf{Hspan}(\mathcal{T}|_{\mathcal{V}(C)})$ **then**
    Let $(\chi_1, \chi_2, \chi_3) \in \widehat{H^\star_C \times H^\star_C \times H^\star_C}$ be the character that evaluates to the constant 1 on $\mathsf{Hspan}(\mathsf{supp}(\mu_C))$ and not constant on any subgroup containing it.
    Add the equation $y_{s_1}^{\chi_1} \cdot y_{s_2}^{\chi_2} \cdot y_{s_3}^{\chi_3} = 1$ to the $\mathsf{GE\ System}$.
  end if
end for

Lemma 7.11. *Applying Algorithm 2 on a satisfiable CSP instance $\Upsilon(\mathcal{V}, \mathcal{C})$ where every predicate $C \in \mathcal{C}$ is* MILDLY-SYMMETRIC *will return an SDP solution $(\boldsymbol{V}^\star, \boldsymbol{\mu}^\star)$ and a system of linear equations* $\mathsf{GE\ System}^\star$ *that are consistent with each other. More formally, For every constraint $C \in \mathtt{supp}(\mathcal{C})$ where $\mathcal{V}(C) = \{s_1, s_2, s_3\}$, and for every cyclic embedding functions P, Q and R, $PQR \equiv 1$ under the local distribution from the SDP solution $(\boldsymbol{V}^\star, \boldsymbol{\mu}^\star)$ iff $PQR \equiv 1$ under the solutions to the* $\mathsf{GE\ System}^\star$.

Proof. We rule out that the algorithm will never reach the Abort steps (Steps 5. and 8.). Using Claim 7.6 and Claim 7.7, eventually, the algorithm will return an SDP solution and a $\mathsf{GE\ System}$ that are consistent with each other.

Using Claim 7.10, the algorithm will never reach Step 5. Fix a satisfying assignment $\boldsymbol{\alpha} \in \Sigma^N$ to the instance. Using Lemma 7.9, for every $C \in \mathtt{supp}(\mathcal{C})$, the local distribution μ_C in the SDP solution has in its support the set $Z_{\boldsymbol{\alpha},C} := \{(\tau_i(\sigma_1), \tau_i(\sigma_2), \tau_i(\sigma_3)) \mid i \in [\ell]\} \subseteq \Sigma^3$ for $\sigma = \boldsymbol{\alpha}|_C$. Because every predicate is MILDLY-SYMMETRIC , the set $Z_{\boldsymbol{\alpha},C}$ (and hence $\mathtt{supp}(\mu_C)$) is pairwise connected and does not have a $(\mathbb{Z}, +)$-embedding. Hence, the algorithm will never reach Step 8. ☐

Finally, we compute the running time of Algorithm 2

Lemma 7.12. *Algorithm 2 on a satisfiable CSP instance $\Upsilon(\mathcal{V}, \mathcal{C})$ where every predicate $C \in \mathcal{C}$ is* MILDLY-SYMMETRIC *runs in time $poly(|\mathcal{V}|, |\Sigma|)$.*

Proof. First, the procedure ModifySDP runs in time $poly(|\mathcal{V}|, |\Sigma|)$. As every modification to the SDP program strictly reduces the support of one of the local distributions, there can be at most $poly(|\mathcal{V}|, |\Sigma|)$ calls to the ModifySDP procedure through the execution of Algorithm 2.

In the procedure ModifyGESystem, for a constraint C, if there are cyclic embedding functions P, Q and R such that $PQR \equiv 1$ under the local distribution from the SDP solution $(\boldsymbol{V}^\star, \boldsymbol{\mu}^\star)$ but $PQR \not\equiv 1$ under the solutions to the $\mathsf{GE\ System}^\star$, then the procedure runs in time $poly(|\mathcal{V}|, |\Sigma|)$ to modify the $\mathsf{GE\ System}$ to make sure $PQR \equiv 1$ under the solutions to the $\mathsf{GE\ System}$. The number of cyclic embedding functions is $O_{|\Sigma|}(1)$, and therefore, every call to ModifyGESystem takes at most $poly(|\mathcal{V}|, |\Sigma|)$ time.

Therefore, the overall running time of Algorithm 2 is $poly(|\mathcal{V}|, |\Sigma|)$ ☐

7.4 Hybrid Approximation Algorithm

In this section, we give our hybrid algorithm for satisfiable instances of Max-$\mathcal{P}$-CSP where $\mathcal{P}$ is a collection of MILDLY-SYMMETRIC predicates.

Hybrid Algorithm ($\mathcal{ALG}$): the algorithm is given below.

Algorithm 5 Hybrid Approximation algorithm for MILDLY-SYMMETRIC predicates

Input: An instance $\Upsilon(\mathcal{V}, \mathcal{C})$ of Max-$\mathcal{P}$-CSP where $\mathcal{P}$ is a collection of 3-ary MILDLY-SYMMETRIC predicates.
Perform the operations mentioned in Algorithm 2.
if the procedure fails to output an SDP solution $(\boldsymbol{V}^\star, \boldsymbol{\mu}^\star)$ and a system of linear equations GE System* that are consistent with each other **then**
    Reject.
else
    Accept.
end if

It can be checked easily that given Lemma 7.12, the algorithm runs in polynomial time. We define the following quantity $\alpha_{\mathcal{P}}^{\mathcal{ALG}}$, and by definition, the approximation guarantee of the hybrid algorithm is $\alpha_{\mathcal{P}}^{\mathcal{ALG}}$.

$$\alpha_{\mathcal{P}}^{\mathcal{ALG}} = \inf_{\beta} \begin{cases} \exists \text{ an instance } \Upsilon \text{ of Max-}\mathcal{P}\text{-CSP such that} \\ 1. \text{ OPT}(\Upsilon) \leqslant \beta, \\ 2. \text{ SDP value} = 1, \\ 3. \text{ The hybrid algorithm accepts.} \end{cases}$$

Theorem 7.1. *For any collection of* MILDLY-SYMMETRIC *predicates $\mathcal{P}$, the algorithm $\mathcal{ALG}$ for Max-$\mathcal{P}$-CSP distinguishes between the following two cases.*

1. *The input instance is satisfiable.*
2. *The input instance has value at most $\alpha_{\mathcal{P}}^{\mathcal{ALG}}$.*

Proof. If the instance is satisfiable, then by Lemma 7.11, $\mathcal{ALG}$ accepts. If the instance has value strictly less than $\alpha_{\mathcal{ALG}}$, then by definition, $\mathcal{ALG}$ rejects such instances. □

In the next section, we study a dictatorship test for Max-$\mathcal{P}$-CSP whose soundness matches with the approximation guarantee of the hybrid algorithm. In other words, we design a test where the test accepts according to the predicate $\mathcal{P}$, has perfect completeness, and has soundness $\alpha_{\mathcal{P}}^{\mathcal{ALG}} + \varepsilon$, for every constant $\varepsilon > 0$.

7.5 Dictatorship Test

We fix an $\Upsilon = (\mathcal{V}, \mathcal{C})$ of a Max-$\mathcal{P}$-CSP such that the algorithm $\mathcal{ALG}$ accepts Υ. Let Σ be the alphabet of the CSP and we identify Σ with the set $[q] = \{1, 2, \ldots, q\}$ where $q = |\Sigma|$. Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be a solution for the SDP relaxation of Υ.

Given a function $F : \Sigma^R \to \Sigma$, in our dictatorship test $\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}$, we will sample three queries according to the distribution $\mu_C^{\otimes R}$ for $C \sim \mathcal{C}$. For the test $\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}$, there is no single natural choice of probability measure μ^n on Σ^n using which we can define the function F to be 'far-from-dictator' required (as in Definition 6.6) for the application of our mixed invariance principle. Therefore, we need to define quasirandomness of a function appropriately, and we do so in the next section.

7.6 The Notion of Quasirandom Functions

For each variable $s \in \mathcal{V}$ in the original Max-$\mathcal{P}$-CSP instance $\Upsilon = (\mathcal{V}, \mathcal{C})$, there is a corresponding probability space $\Omega_s = (\Sigma, \mu_s)$. Therefore, we will define the relative notion of "quasirandom with respect to $(\boldsymbol{V}, \boldsymbol{\mu})$". Roughly speaking, we shall call a function "quasirandom" if all its "influences" are low under nearly all of the probability distributions corresponding to variables $s_i \in \mathcal{V}$.

Definition 7.13. *(Quasirandom function w.r.t. $(\boldsymbol{V}, \boldsymbol{\mu})$) A function $F : \Sigma^n \to \Sigma$ is called a (d, τ)-quasirandom w.r.t. $(\boldsymbol{V}, \boldsymbol{\mu})$ if the following property holds. Consider the functions $(F_{s,1}, F_{s,2}, \dots, F_{s,q})$, where for every $j \in [q]$, $F_{s,j} : ([q]^n, \mu_s^{\otimes n}) \to \mathbb{R}$ is defined as $F_{s,j}(\boldsymbol{x}) := \mathbf{1}_{F(\boldsymbol{x})=j}$. Suppose there exist product functions $\mathcal{P}_{s,j}$ with $\mathsf{rk}\mathcal{P}_{s,j}) \geqslant T$ such that functions $\tilde{F}_{s,j}$ of the form*

$$\tilde{F}_{s,j}(x) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s,j})} P(x) \cdot L_P(x)$$

satisfy:

1. *$\|F_{s,j} - \tilde{F}_{s,j}\|_2 \leqslant \varepsilon$.*
2. *$\mathsf{deg}(L_P)$ and $\|L_P\|_2$ are both $d = O_{T,\alpha,q,\varepsilon}(1)$.*
3. *Furthermore, for every $C \in \mathsf{supp}(\mathcal{C})$ where $\mathcal{V}(C) = (s_1, s_2, s_3)$, $j, k, \ell \in [q]$, and for any $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1,j})$, $Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2,k})$ and $R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3,\ell})$, either $P(x)Q(y)R(z) \equiv 1$ in the support of μ_C^n, or $\Delta_{\mathsf{symbolic}}(PQR, 1) \geqslant T'$.*

Then, F is called (d, τ)-quasirandom w.r.t. $(\boldsymbol{V}, \boldsymbol{\mu})$ if for every $s \in \mathcal{V}$, $j \in [q]$ and $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s,j})$, for every $i \in [n]$ it holds that $I_i[L_P] \leqslant \tau$.

Remark 7.14. *A standard notion of (d, τ)-quasirandomness that gets used in proving the hardness results from dictatorship tests has the conditions $I_i[F_{s,j}^{\leqslant d}] \leqslant \tau$ for all $j \in [q]$ and $i \in [n]$, where $F_{s,j}^{\leqslant d}$ is a low-degree truncation of $F_{s,j}$ up to degree d. The reason behind ignoring the high-degree component of $F_{s,j}$ in the hardness reduction is that the underlying distribution is connected and hence the high-degree components have a negligible contribution. In our case, $F_{s,j}$ has high-degree components $P(x)$, and we cannot ignore such components in hardness reductions for satisfiable Max-$\mathcal{P}$-CSPs as the underlying distribution cannot be made connected. Thus, a natural extension of quasirandomness is to enforce the small influence conditions on the low-degree components L_P in $F_{s,j}$s. At this point, we do not know how to use this notion of quasirandomness to convert a dictatorship test to a (conditional) hardness result. We leave this as an open question for future research.*

One may wonder if it is even possible to find the functions $\tilde{F}_{s,j}$ such that all the above (except for the quasirandomness property) properties hold for small enough ε and large T and T'. Lemma 5.12 already shows the existence of product functions $\mathcal{P}_{s,j}$ for $F_{s,j}$ such that properties 1. and 2. above hold. The following lemma states that for any collection of functions $(f_1, f_2, \dots, f_s)$, we can find a collection of product functions such that the above three properties hold simultaneously.

Lemma 7.15. *For all $\alpha > 0$, $m \in \mathbb{N}$, $T \geqslant 1$, and $\xi > 0$ there exist $r \in \mathbb{N}$ and $\varepsilon_0 > 0$ such that the following holds for any decay function $w\colon [0,1] \to [0,1]$. Let Σ be an alphabet of size at most m, let $\mu_{jk\ell}$ be pairwise consistent distributions over Σ^3 in which the probability of each atom is at least α for all $1 \leqslant j, k, \ell \leqslant s$. Then there exists an Abelian group G whose size depends only on $|\Sigma|$, $\sigma\colon \Sigma \to G$, such that for any function $f_j\colon (\Sigma^n, \mu_j^n) \to \mathbb{C}$, for $1 \leqslant j \leqslant s$, there are $r' \leqslant r$, $\varepsilon \geqslant \varepsilon_0$ and a collections $\mathcal{P}_j$ of cyclic embedding functions of size at most r' such that:*

1. *We have* $\mathsf{rk}(\mathcal{P}_j) \geqslant \frac{1}{w(\varepsilon)}$.

2. *For all* $\varepsilon' \in (w(\varepsilon), \varepsilon)$ *we have that* $\|f_j - \mathrm{T}_{\mathcal{P}_j, 1-\varepsilon'} f_j\|_{\nu_j, \alpha} \leqslant \xi$, *where* ν_j *is the marginal distribution* $\mu_{jk\ell}$ *on* j.

3. *for any* $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_j)$, $Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_k)$ *and* $R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_\ell)$, *either* $P(x)Q(y)R(z) \equiv 1$ *in the support of* $\mu_{jk\ell}^n$, *or* $\Delta_{\mathsf{symbolic}}(PQR, 1) \geqslant T$.

Proof. Fix a decay function $w' : [0,1] \to [0,1]$. For given m, α and sufficiently small ξ, we first apply Lemma 5.12 to the functions f_js individually to get a collection of product functions $\mathcal{P}_j$ for f_j such that $\mathsf{rk}(\mathcal{P}_j) \geqslant T$ and for all $\varepsilon' \in (w(\varepsilon), \varepsilon)$, $\|f_j - T_{\mathcal{P}_j, 1-\varepsilon'} f_j\|_{\mu_j, \alpha} \leqslant \xi$, for a sufficiently small decay function w compared to w'. We will update the collection $\mathcal{P}_j$ iteratively to arrive at a final collection $\mathcal{P}_j'$ of size at most $|\mathcal{P}_j|$ that will satisfy all the conditions from the lemma with the decay function w', $\varepsilon_0' = \Omega_{r,s}(\xi^2/T)$, $\xi' = 2\xi$.

1. Set $t = 1$ and $U_1 = [n]$.

2. Construct a subset $A_t \subseteq U_t$ as follows:

   (a) Set $A_t = \emptyset$.

   (b) For every j, k, l and $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_j)$, $Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_k)$, and $R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_\ell)$ such that $P(x)Q(y)R(z) \not\equiv 1$ in the support of $\mu_{jk\ell}^n$, and $\Delta_{\mathsf{symbolic}}(PQR, 1) < T$,

      i. Add $i \in [n]$ to A_t for which $P_i Q_i R_i \not\equiv 1$ in the support of $\mu_{jk\ell}$.

      ii. Change $P' \in \mathcal{P}_j$ to $\prod_{i \in U_t \setminus A_t} P_i'(x)$, and similarly do the same for every $j \in [s]$.

      iii. Set $U_{t+1} = U_t \setminus A_t$.

3. Increment t by 1 and repeat step 2 unless for every j, k, l and $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_j)$, $Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_k)$, and $R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_\ell)$ either $P(x)Q(y)R(z) \equiv 1$ in the support of $\mu_{jk\ell}^n$, or $\Delta_{\mathsf{symbolic}}(PQR, 1) \geqslant T$,

We start with a simple observation. If once triple P, Q, R is responsible for adding a coordinate to A_t, then it will never be responsible for adding a coordinate in later iterations. This follows as we change the P, Q, R so that $P(x)Q(y)R(z) \equiv 1$. The consequence of this is that the procedure stops after a finite number of iterations. Let $t^\star$ be the last iteration. Clearly, $t^\star \leqslant \prod_j \prod_{P \in \mathcal{P}_j} \mathsf{ord}(P)$, as the RHS is the number of distinct triples from the collections $\{\mathsf{spn}_{\mathbb{N}}(\mathcal{P}_j)\}_{j \in [q]}$. Thus $t^\star = O_{r,s}(1)$, where $|\mathcal{P}_j| \leqslant r$ for every j. Furthermore, at any given iteration, $|A_t| \leqslant \prod_j \prod_{P \in \mathcal{P}_j} \mathsf{ord}(P) \cdot T = O_{r,s}(T)$ as every triple contributes at most T to the set A_t. If we let $A := \cup_{t=1}^{t^\star} A_t$, then $|A| = O_{r,s}(T)$.

At the end of the procedure, we have that a product function $P(x) = \prod_{i=1}^n P_i(x)$ is changed to $P'(x) = \prod_{i \in [n] \setminus A} P_i(x)$. Let the final collection of these product functions be $\{\mathcal{P}_j'\}_{j \in [s]}$. Since the procedure ends, we have that for any $P' \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_j')$, $Q' \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_k')$ and $R' \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_\ell')$, either $P'(x)Q'(y)R'(z) \equiv 1$ in the support of $\mu_{jk\ell}^n$, or $\Delta_{\mathsf{symbolic}}(P'Q'R', 1) \geqslant T$. Furthermore, as we changed $|A|$ many coordinates to 1 in each of the product functions, we have $\mathsf{rk}(\mathcal{P}_j) \geqslant \frac{1}{w(\varepsilon)} - |A| \geqslant \frac{1}{w'(\varepsilon)}$. We now show that the second property holds for these updated collections.

We have the following claim which is similar to Fact 5.3, except we do the analysis in one shot by observing that the coupling works for the collections $\mathcal{P}_j$ and $\mathcal{P}_j'$ directly.

Claim 7.16. *For every* $\varepsilon > 0$, *for any* 1*-bounded function* $f_j \colon (\Sigma^n, \nu_j^n) \to \mathbb{C}$,

$$\|\mathrm{T}_{\mathcal{P}_j, 1-\varepsilon} f_j - \mathrm{T}_{\mathcal{P}_j', 1-\varepsilon} f_j\|_{\nu_j, 2} \leqslant 2\sqrt{|A|\varepsilon}.$$

Proof. We first show that there is a coupling of (x, y, y') such that (x, y) is distributed according to $(x, \mathrm{T}_{\mathcal{P}_j,1-\varepsilon}x)$, (x, y') is distributed according to $(x, \mathrm{T}_{\mathcal{P}'_j,1-\varepsilon}x)$ and $\Pr[y \neq y'] \leqslant |A|\varepsilon$.

Note that A is the set of all the coordinates in which $P \in \mathcal{P}_j$ and $P' \in \mathcal{P}'_j$ differ. Sample $x \sim \nu$, then $I \subseteq_{1-\varepsilon} [n]$ and then $y \sim \nu_j$ conditioned on $y_I = x_I$ and $P(y) = P(x)$ for all $P \in \mathcal{P}_j$; note that y is distributed according to $\mathrm{T}_{\mathcal{P}_j,1-\varepsilon}x$. Let $y' \sim \mathrm{T}_{\mathcal{P}'_j,1-\varepsilon}x$. Note that if $I \cap A = \emptyset$, then $P'(y) = P'(x)$. Thus, letting E be the event that $I \cap A = \emptyset$, we have that $\Pr[E] \geqslant 1 - |A|\varepsilon$ and that the distributions of $y \mid E$ and $y' \mid E$ are identical. Thus, the statistical distance between y and y' is at most $\Pr[E] \leqslant |A|\varepsilon$. It follows that there is a coupling between y and y' such that $\Pr[y \neq y'] \leqslant |A|\varepsilon$.

Fix the coupling (x, y, y') so that we may write

$$\|\mathrm{T}_{\mathcal{P}'_j,1-\varepsilon}f_j - \mathrm{T}_{\mathcal{P}_j,1-\varepsilon}f_j\|_{\nu_j,2}^2 = \mathop{\mathbb{E}}_{x}\left[\left|\mathop{\mathbb{E}}_{y,y'}\left[f_j(y) - f_j(y')\right]\right|^2\right] \leqslant 4\mathop{\mathbb{E}}_{x}\left[1_{y\neq y'}\right] \leqslant 4|A|\varepsilon. \qquad \square$$

There exists $\varepsilon'_0 \geqslant \frac{\xi^2}{4|A|} = \Omega_{r,s}(\frac{\xi^2}{T})$, and $\varepsilon \geqslant \varepsilon'_0$, such that for all $\varepsilon' \in (w'(\varepsilon), \varepsilon)$, we have

$$\begin{aligned}\|f_j - \mathrm{T}_{\mathcal{P}'_j,1-\varepsilon'}f_j\|_{\nu_j,\alpha} &\leqslant \|f_j - \mathrm{T}_{\mathcal{P}_j,1-\varepsilon'}f_j\|_{\nu_j,\alpha} + \|\mathrm{T}_{\mathcal{P}'_j,1-\varepsilon'}f_j - \mathrm{T}_{\mathcal{P}_j,1-\varepsilon'}f_j\|_{\nu_j,\alpha} \\ &\leqslant \|f_j - \mathrm{T}_{\mathcal{P}_j,1-\varepsilon'}f_j\|_{\nu_j,\alpha} + \|\mathrm{T}_{\mathcal{P}'_j,1-\varepsilon'}f_j - \mathrm{T}_{\mathcal{P}_j,1-\varepsilon'}f_j\|_2 \\ &\leqslant \xi + 2\sqrt{|A|\varepsilon'} \\ &\leqslant 2\xi.\end{aligned}$$

$\square$

7.6.1 The Dictatorship Test for MILDLY-SYMMETRIC Predicates

We are now ready to state the dictatorship test and prove its completeness and soundness. Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be a solution for the SDP relaxation of Υ and $\mathcal{S} \subseteq G^{\mathcal{V}}_{\mathsf{master}}$, where $G^{\mathcal{V}}_{\mathsf{master}} := \prod_{v\in\mathcal{V}} H^\star_v$, be the subspace of the set of satisfying assignments of the instance $\Upsilon(\mathcal{V}, \mathcal{C})$ coming from the $\mathsf{GE\ System}$ that satisfy the conditions from Lemma 7.11. Let $\alpha > 0$ be a lower bound on the non-zero probabilities from the distributions μ_C, and we treat α as a constant as the instance size is fixed.

In Figure 3, we give the dictatorship test $\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}$ for functions $F : \Sigma^R \to \Sigma$. Note that Lemma 7.11 guarantees that $(\boldsymbol{V}, \boldsymbol{\mu})$ and $\mathcal{S} \subseteq G^{\mathcal{V}}_{\mathsf{master}}$ are consistent with respect to every P, Q, R coming from the decompositions of $F_{s_1,j}, F_{s_2,k}, F_{s_3,\ell}$, for every $C \in \mathtt{supp}(\mathcal{C})$ and $1 \leqslant j, k, \ell \leqslant q$ such that $\mathcal{V}(C) = (s_1, s_2, s_3)$ and we will use this fact in the analysis of our dictatorship test.

Completeness: a function $F : \Sigma^R \to \Sigma$ is called a dictator function if $F(\boldsymbol{z}) = \boldsymbol{z}^{(i)}$ for some $i \in [R]$. The completeness of the test is defined as follows,

$$\text{Completeness}(\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}) = \min_{\substack{i\in[R],\\ F \text{ is the } i^{th} \text{ dictator}}} \Pr[F \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}].$$

As the distribution μ_C is supported on the satisfying assignment of the constraint C, the test passes with probability 1 when F is a dictator function. Therefore, $\text{Completeness}(\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}) = 1$.

Soundness: the soundness of the test is the maximum probability with which it accepts quasirandom functions. More formally, define the soundness of the test as:

1. Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be a solution for the basic SDP relaxation of Υ that satisfy the conditions from Lemma 7.11.
2. Sample a payoff $C \sim \mathcal{C}$. Let $\mathcal{V}(C) = \{s_1, s_2, s_3\}$.
3. Sample $\boldsymbol{z}_C = \{\boldsymbol{z}_{s_1}, \boldsymbol{z}_{s_2}, \boldsymbol{z}_{s_3}\}$ from the product distribution $\mu_C^{\otimes R}$, i.e., independently for each $i \in [R]$, $(\boldsymbol{z}_{s_1}^{(i)}, \boldsymbol{z}_{s_2}^{(i)}, \boldsymbol{z}_{s_3}^{(i)}) \sim \mu_C$.
4. Query the function values $F(\boldsymbol{z}_{s_1}), F(\boldsymbol{z}_{s_2}), F(\boldsymbol{z}_{s_3})$.
5. Accept iff $C(F(\boldsymbol{z}_{s_1}), F(\boldsymbol{z}_{s_2}), F(\boldsymbol{z}_{s_3})) = 1$.

Figure 3: SDP integrality gap to a dictatorship test $\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}$.

$$\text{Soundness}_{(d,\tau)}(\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}) = \sup_{\substack{F:\Sigma^R \to \Sigma \\ F \text{ is } (d,\tau)-\text{quasirandom w.r.t.}(\boldsymbol{V},\boldsymbol{\mu})}} \Pr[F \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}].$$

We now state our main theorem regarding the soundness of the dictatorship test $\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}$.

Theorem 7.2. *Fix any collection of* MILDLY-SYMMETRIC *predicates $\mathcal{P}$. Given an instance $\Upsilon = (\mathcal{V}, \mathcal{C})$ of a Max-$\mathcal{P}$-CSP such that the algorithm $\mathcal{ALG}$ accepts Υ, the test $\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}$ has completeness 1 and soundness*

$$\textit{Soundness}_{(d,\tau)}(\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}) = \mathsf{OPT}(\Upsilon) + o_{d,\tau}(1),$$

where $\mathsf{OPT}(\Upsilon)$ is the optimal value of the instance Υ and $o_{d,\tau}(1) \to 0$ as $\tau \to 0$.

The following corollary follows from the above theorem and the approximation guarantee of our hybrid algorithm.

Corollary 7.17. *Fix any collection of* MILDLY-SYMMETRIC *predicates $\mathcal{P}$. For every $\varepsilon > 0$, there exists a dictator vs. quasirandom test with completeness 1 and soundness $\alpha_{\mathcal{P}}^{\mathcal{ALG}} + \varepsilon$, and the accepting criterion of the test is from the set of predicates $\mathcal{P}$.*

Theorem 7.1 along with Corollary 7.17 prove our main Theorem 1.1. We prove Theorem 7.2 in the rest of the section. We first set up a few notations in Section 7.6.2 that will be used in the analysis of the soundness of the test.

7.6.2 Functions on Product Spaces

Let (Ω, μ) be a probability space with $|\Omega| = q$ and μ has full support on Ω. Define the inner product between two functions $f, g : \Omega \to \mathbb{R}$ on this space as follows: $\langle f, g \rangle = \mathbb{E}_{x \sim \mu}[f(x)g(x)]$.

Definition 7.18. *An orthonormal ensemble consists of a basis of real orthonormal random variables $\mathcal{L} = \{\ell_0 \equiv \mathbf{1}, \ell_1, \ldots, \ell_{q-1}\}$, where $\mathbf{1}$ is the constant 1 function.*

Henceforth, we will sometimes refer to orthonormal ensembles as just ensembles. For an ensemble $\mathcal{L} = \{\ell_0 \equiv \mathbf{1}, \ell_1, \ldots, \ell_{q-1}\}$ of random variables, we will use $\mathcal{L}^R$ to denote the ensemble obtained by taking R independent copies of $\mathcal{L}$. Further $\mathcal{L}^{(i)} = \{\ell_0^{(i)}, \ell_1^{(i)}, \ldots, \ell_{q-1}^{(i)}\}$ will denote the i^{th} independent copy of $\mathcal{L}$.

Fix an ensemble $\mathcal{L} = \{\ell_0 \equiv \mathbf{1}, \ell_1, \ldots, \ell_{q-1}\}$ that forms a basis for $L^2(\Omega)$. Given such a basis for $L^2(\Omega)$, it induces a basis for the space $L^2(\Omega^R)$, given by the random variables

$$\left\{ \ell_{\boldsymbol{\sigma}} := \prod_{i=1}^{R} \ell_{\sigma_i}^{(i)} \;\middle|\; \boldsymbol{\sigma} \in \{0, 1, \ldots, q-1\}^R \right\}.$$

Therefore, any function $\mathcal{F} : \Omega^R \to \mathbb{R}$ has a multilinear expansion

$$\mathcal{F}(\boldsymbol{z}) = \sum_{\boldsymbol{\sigma} \in \{0,1,\ldots,q-1\}^R} \hat{\mathcal{F}}(\boldsymbol{\sigma}) \ell_{\boldsymbol{\sigma}}(\boldsymbol{z}),$$

where $\ell_{\boldsymbol{\sigma}}(\boldsymbol{z}) = \prod_{i=1}^{R} \ell_{\sigma_i}(z_i)$.

Definition 7.19. *A multi-index $\boldsymbol{\sigma}$ is a vector $(\sigma_1, \sigma_2, \ldots . \sigma_R) \in \{0, 1, \ldots, q-1\}^R$ and the degree of $\boldsymbol{\sigma}$ is denoted by $|\boldsymbol{\sigma}|$ which is equal to $|\boldsymbol{\sigma}| = |\{i \in [R] \mid \sigma_i \neq 0\}|$. Given a set of indeterminates $\boldsymbol{\mathcal{X}} = \{x_j^{(i)} | j \in \{0, 1, \ldots, q-1\}, i \in [R]\}$ and a multi-index $\boldsymbol{\sigma}$, define the monomial $x_{\boldsymbol{\sigma}}$ as*

$$x_{\boldsymbol{\sigma}} = \prod_{i=1}^{R} x_{\sigma_i}^{(i)}.$$

The degree of the monomial is given by $|\boldsymbol{\sigma}|$. A multilinear polynomial over such indeterminates is given by

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{\sigma} \in \{0,1,\ldots,q-1\}^R} \hat{F}_{\boldsymbol{\sigma}} x_{\boldsymbol{\sigma}}.$$

Given any function $\mathcal{F} : \Omega^R \to \mathbb{R}$, with the multilinear expansion $\mathcal{F}(\boldsymbol{z}) = \sum_{\sigma \in \{0,1,\ldots,q-1\}^R} \hat{\mathcal{F}}(\sigma) \ell_\sigma(\boldsymbol{z})$ with respect to the orthonormal ensemble $\mathcal{L} = \{\ell_0 \equiv \mathbf{1}, \ell_1, \ldots, \ell_{q-1}\}$, we define a corresponding *formal polynomial* in the indeterminates $\boldsymbol{\mathcal{X}} = \{x_j^{(i)} | j \in \{0, 1, \ldots, q-1\}, i \in [R]\}$, as follows:

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{\sigma}} \hat{\mathcal{F}}(\boldsymbol{\sigma}) x_{\boldsymbol{\sigma}}.$$

We will always use the symbol $\mathcal{F}$ to denote a real-valued function on a product probability space Ω^R. Further $F(\boldsymbol{x})$ will denote the formal multilinear polynomial corresponding to $\mathcal{F}$. Hence, $F(\mathcal{L}^R)$ is a random variable obtained by substituting the random variables $\mathcal{L}^R$ in place of $\boldsymbol{x}$. For instance, the following equation holds in this notation:

$$\mathop{\mathbb{E}}_{\boldsymbol{z} \in \Omega^R}[\mathcal{F}(\boldsymbol{z})] = \mathbb{E}[F(\mathcal{L}^R)].$$

Vector Valued Functions. We will always use the symbol $\boldsymbol{\mathcal{F}} = (\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_q)$ to denote a vector-valued function on a product probability space Ω^R. Further, $\boldsymbol{F}(\boldsymbol{x}) = (F_1, F_2, \ldots, F_q)$ will denote the formal multilinear polynomial corresponding to $\boldsymbol{\mathcal{F}}$.

7.6.3 Soundness Analysis of the Dictatorship Test

We now analyze the soundness of the test. Towards this, we assume that F is quasirandom according to Definition 7.13 and we fix a decomposition of F that guarantees the quasirandomness of F throughout the analysis. For each $s \in \mathcal{V}$, let $\Omega_s = (\Sigma, \mu_s)$ be a probability space with atoms in Σ where the probability of $a \in \Sigma$ is $\|\boldsymbol{b}_{s,a}\|_2^2$.

We will fix an arbitrary mapping from Σ to $\{1, 2, \dots, q\}$, denoted by $\varsigma : \Sigma \to \{1, 2, \dots, q\}$. The domain of the payoff $C : \Sigma^3 \to \{0, 1\}$ can be extended from Σ^3 to $\blacktriangle_q^3$. To see this, by the abuse of notation, first define a Δ_q-representation of a payoff $C : \Sigma^3 \to \{0, 1\}$ as $C : \Delta_q^3 \to \{0, 1\}$ where

$$C(\boldsymbol{e}_{a_1}, \boldsymbol{e}_{a_2}, \boldsymbol{e}_{a_3}) = C(\varsigma^{-1}(a_1), \varsigma^{-1}(a_2), \varsigma^{-1}(a_3)), \text{ for all } (a_1, a_2, a_3) \in \{1, 2, \dots, q\}^3.$$

The function C can be extended to the domain $\blacktriangle_q^3$ by its multi-linear extension. Again, by abusing the notation, define the extension C as:

$$C(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = \sum_{\sigma \in \Sigma^3} C(\sigma) \prod_{i=1}^{3} x_{i,\varsigma(\sigma_i)}, \text{ for all } \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \in \blacktriangle_q. \tag{37}$$

Extending C to $\mathbb{R}^{3q}$: We will extend the payoff function C further to a real valued function on $(\mathbb{R}^q)^3$, by plugging the real values in the expansion of C given in the Equation (37). This extension of C is smooth in the following sense:

1. All the partial derivatives of C up to order 3 are uniformly bounded by $C_0(q)$.
2. C is a Lipschitz function with Lipschitz constant $C_0(q)$, i.e.,$\forall \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}, \{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3\} \in (\mathbb{R}^q)^3$,

$$|C(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) - C(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)| \leqslant C_0(q) \sum_{i=1}^{3} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|_2.$$

Local and Global Ensembles. Fix a given SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ with value 1. We define the following local and global orthonormal ensembles of random variables for every $s \in \mathcal{V}$ as follows.

- **Local Integral Ensembles $\mathcal{L}$:** The Local Integral Ensemble $\mathcal{L} = \{\boldsymbol{\ell}_s \mid s \in \mathcal{V}\}$ for a variable $s \in \mathcal{V}$, $\boldsymbol{\ell}_s = \{\ell_{s,0} \equiv \mathbf{1}, \ell_{s,1}, \dots, \ell_{s,q-1}\}$ is a set of random variables that are orthonormal ensembles for the space Ω_s.

We also define the following global ensembles of random variables:

- **Global Gaussian Ensembles $\mathcal{G}$:** The Global Gaussian Ensembles $\mathcal{G} = \{\boldsymbol{g}_s \mid s \in \mathcal{V}\}$ are generated by setting $\boldsymbol{g}_s = \{g_{s,0} \equiv \mathbf{1}, g_{s,1}, \dots, g_{s,q-1}\}$ where

$$g_{s,j} = \sum_{\omega \in \Sigma} \ell_{s,j}(\omega) \langle \boldsymbol{b}_{s,\omega}, \boldsymbol{\zeta} \rangle, \quad \forall j \in \{1, \dots, q-1\},$$

  and $\boldsymbol{\zeta}$ is a normal Gaussian random vector of appropriate dimension.

The following lemma states that the local integral ensemble and the global Gaussian ensemble have matching first and second moments. We need this to apply the invariance principle in our analysis below.

Lemma 7.20. *([7]) For every $s \in \mathcal{V}$, $\boldsymbol{g}_s$ is an orthonormal ensemble w.r.t. the space Ω_s. Also, for any payoff $C \in \mathcal{C}$, the global ensembles $\mathcal{G}$ match the following moments of the local integral ensembles $\mathcal{L}$:*

$$\mathop{\mathbb{E}}_{\boldsymbol{\zeta}}[g_{s,j}.g_{s',j'}] = \mathop{\mathbb{E}}_{(\omega,\omega')\sim\mu_C|(s,s')}[\ell_{s,j}(\omega).\ell_{s',j'}(\omega')] \quad \forall j,j' \in \{1,\ldots,q-1\}, s,s' \in \mathcal{V}(P'),$$

where $\mu_C|(s,s')$ is the marginal distribution of μ_C on the coordinates of s, s'.

Soundness analysis. The acceptance probability of the test for a given function $\boldsymbol{\mathcal{F}}$ is given by:

$$\Pr[\boldsymbol{\mathcal{F}} \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}] = \mathop{\mathbb{E}}_{C\sim\mathcal{C}} \mathop{\mathbb{E}}_{\boldsymbol{z}_C\sim\mu_C^{\otimes R}}[C(\boldsymbol{\mathcal{F}}_{s_1}(\boldsymbol{z}_{s_1}), \boldsymbol{\mathcal{F}}_{s_2}(\boldsymbol{z}_{s_2}), \boldsymbol{\mathcal{F}}_{s_3}(\boldsymbol{z}_{s_3}))], \tag{38}$$

where $\mathcal{V}(C) = (s_2, s_2, s_3)$. Consider the following expression.

$$C(\boldsymbol{\mathcal{F}}_{s_1}(\boldsymbol{z}_{s_1}), \boldsymbol{\mathcal{F}}_{s_2}(\boldsymbol{z}_{s_2}), \boldsymbol{\mathcal{F}}_{s_3}(\boldsymbol{z}_{s_3})) = \sum_{\sigma\in\Sigma^3} C(\sigma) \prod_{j=1}^{3} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}). \tag{39}$$

We now apply Theorem 6.9, to the expression

$$\mathop{\mathbb{E}}_{\boldsymbol{z}_C\sim\mu_C^{\otimes R}} \left[\prod_{j=1}^{3} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j})\right].$$

The invariant distribution we wish to move to is slightly different from the setting of Theorem 6.9. Specifically, we wish to consider the following two distributions:

1. The distribution $\mathcal{D}_{\mathcal{S}}$: sample R assignments $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \ldots, \boldsymbol{\alpha}^{(R)}$ from the set of satisfying assignment $\mathcal{S} \subseteq G_{\mathsf{master}}^{\mathcal{V}}$ to the $\mathsf{GE\ System}$ independently and uniformly at random, consider $(a_i, b_i, c_i) := (\alpha_{s_1}^{(i)}, \alpha_{s_2}^{(i)}, \alpha_{s_3}^{(i)}) \in H_{s_1}^{\star} \times H_{s_2}^{\star} \times H_{s_3}^{\star}$ and output

$$(\chi_P(\boldsymbol{a}), \chi_Q(\boldsymbol{b}), \chi_R(\boldsymbol{c}))_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1,\sigma_1}), Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2,\sigma_2}), R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3,\sigma_3})}.$$

2. The distribution $\mathcal{D}_{\mu_C}$: sample $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \sim \mu_C^{\otimes R}$ and output

$$(\chi_P(\sigma(\boldsymbol{x})), \chi_Q(\gamma(\boldsymbol{y})), \chi_R(\phi(\boldsymbol{z})))_{P\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1,\sigma_1}), Q\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2,\sigma_2}), R\in\mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3,\sigma_3})}.$$

The following lemma asserts that the distributions $\mathcal{D}_H$ and $\mathcal{D}_\mu$ are close in statistical distance.

Lemma 7.21. *Suppose the sizes of each one of $\mathcal{P}_{s_1,\sigma_1}, \mathcal{P}_{s_2,\sigma_2}, \mathcal{P}_{s_3,\sigma_3}$ is at most r, $|G_{\mathsf{master}}| \leqslant m$ and that for any $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1,\sigma_1}))$, $Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2,\sigma_2}))$ and $R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3,\sigma_3}))$ it holds that either $PQR \equiv 1$ or else $P_iQ_iR_i \not\equiv 1$ for at least T' of the coordinates $i \in [R]$. Then*

$$\mathsf{SD}(\mathcal{D}_{\mu_C}, \mathcal{D}_{\mathcal{S}}) \lesssim_{m,r} (1 - \Omega_{m,\alpha}(1))^{T'}.$$

where α is the minimum non-zero probability of an item from $\mathtt{supp}(\mu)$.

Proof. The proof of the lemma is similar to the proof of Lemma 6.12 and we only sketch the proof. Here, we crucially use the fact that the set of solutions $\mathcal{S}$ to the $\textsf{GE System}$ and the SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ are consistent to each other as stated in Lemma 7.11. Consider S in the support of one of the distribution. For any P, Q, R, consider their contribution to $\mathcal{D}_{\mu_C}(S)$ and $\mathcal{D}_{\mathcal{S}}(S)$ defined in Lemma 6.12, but for the two distributions $\mathcal{D}_{\mu_C}$ and $\mathcal{D}_{\mathcal{S}}$. For P, Q, R such that $PQR \equiv 1$, the two contributions are the same. Else, by assumption $P_iQ_iR_i \not\equiv 1$ for at least T' coordinates, and each one of the expectations is at most $(1 - \Omega_{m,\alpha}(1))^{T'}$ in absolute value. □

Now, define the functions $F_{s,j}^{\mathsf{dec}} \colon (H_s^\star)^R \times \mathbb{R}^{(q-1)R} \to \mathbb{C}$ in the following way. Let $\mathcal{P}_{s,j}$ be the collection of product functions associated with $\tilde{F}_{s,j}$ from the decomposition of $F : \Sigma^R \to \Sigma$ satisfying the quasirandomness property with respect to the Definition 7.13. Write each $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s,j})$ as $P = \prod_{i=1}^{R} \chi_{P,i}(\sigma(x_i))$. Thus,

$$\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s,j})} L_P(\boldsymbol{y}) \cdot \prod_{i=1}^{R} \chi_{i,P}(\sigma(x_i)). \tag{40}$$

Interpreting the first input over the group $(H_s^\star)^R$, we define $\mathcal{F}_{s,j}^{\mathsf{dec}} \colon (H_s^\star)^R \times \Sigma^R \to \mathbb{C}$ by replacing $\sigma(x_i)$ by the group element input. Namely,

$$\mathcal{F}_{s,j}^{\mathsf{dec}}(\boldsymbol{a}, \boldsymbol{y}) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s,j})} L_P(\boldsymbol{y}) \cdot \prod_{i=1}^{R} \chi_{i,P}(a_i). \tag{41}$$

Next, we replace the low-degree polynomials with their corresponding multilinear extensions to get the functions $F_{s,j}^{\mathsf{dec}} \colon (H_s^\star)^R \times \mathbb{R}^{(q-1)R} \to \mathbb{C}$.

$$F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}, \boldsymbol{\ell}) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s,j})} L_P(\boldsymbol{\ell}) \cdot \prod_{i=1}^{R} \chi_{i,P}(a_i).$$

Finally, the function $\widetilde{F_{s,j}^{\mathsf{dec}}} \colon (H_s^\star)^R \times \mathbb{R}^{(q-1)R} \to \mathbb{R}$ is defined by taking only real part of the output of $F_{s,j}^{\mathsf{dec}}$ and truncating it using the function trunc.

$$\widetilde{F_{s,j}^{\mathsf{dec}}}(\boldsymbol{a}, \boldsymbol{\ell}) = \mathsf{trunc}(\mathfrak{Re}(\tilde{F}_{s,j}^{\mathsf{dec}}(\boldsymbol{a}, \boldsymbol{\ell}))).$$

Applying Theorem 6.9 but with respect to the distributions $(\mathcal{D}_{\mathcal{S}}, \mathcal{D}_{\mu_C})$, and with the function $\Psi(a, b, c) := abc$ for $a, b, c \in \mathbb{C}$, we get

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{z}_C \sim \mu_C^{\otimes R}} \left[\prod_{j=1}^{3} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \right] - \mathop{\mathbb{E}}_{\substack{\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \ldots, \boldsymbol{\alpha}^{(R)} \sim \mathcal{S}, \\ a_{s,i} = \alpha_s^{(i)}, \\ \mathcal{G}^R}} \left[\prod_{j=1}^{3} \widetilde{F_{s_j,\sigma_j}^{\mathsf{dec}}}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j}) \right] \right| \leqslant 2^{-\Omega_{q,\alpha,\varepsilon}(T')} + \eta(\tau).$$

Combining the above with Equations (38) and (39), we get

$$\Pr[\boldsymbol{\mathcal{F}} \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}] \leqslant \mathop{\mathbb{E}}_{C\sim\mathcal{C}} \mathop{\mathbb{E}}_{\substack{\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)},\ldots,\boldsymbol{\alpha}^{(R)}\sim\mathcal{S},\\ a_{s,i}=\alpha_s^{(i)},\\ \mathcal{G}^R}} [C(\widetilde{\boldsymbol{F}_{s_1}^{\mathsf{dec}}}(\boldsymbol{a}_{s_1},\boldsymbol{g}_{s_1}), \widetilde{\boldsymbol{F}_{s_2}^{\mathsf{dec}}}(\boldsymbol{a}_{s_2},\boldsymbol{g}_{s_2}), \widetilde{\boldsymbol{F}_{s_3}^{\mathsf{dec}}}(\boldsymbol{a}_{s_3},\boldsymbol{g}_{s_3}))] + 2^{-\Omega_{q,\alpha,\varepsilon}(T')} + \eta(\tau) \tag{42}$$

Note that $\widetilde{\boldsymbol{F}_s^{\mathsf{dec}}}(\boldsymbol{a},\boldsymbol{g}) \in \mathbb{R}^q$. We now analyze the loss that happens because of scaling, as given in Step 4 of the rounding scheme. Let $\boldsymbol{F}_s^{\mathsf{dec}}(\boldsymbol{a},\boldsymbol{g})^\star \in \blacktriangle_q$ be the vector obtained by scaling it using the function $\mathbf{Scale}$ defined in Step 4 of the rounding scheme $\mathbf{Round}_{\boldsymbol{\mathcal{F}}}$ below.

The expected value of the solution returned by the rounding scheme $\mathbf{Round}_{\boldsymbol{\mathcal{F}}}$ is given by the following expression.

$$\mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V},\boldsymbol{\mu}) = \mathop{\mathbb{E}}_{C\sim\mathcal{C}} \mathop{\mathbb{E}}_{\substack{\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)},\ldots,\boldsymbol{\alpha}^{(R)}\sim\mathcal{S},\\ a_{s,i}=\alpha_s^{(i)},\\ \mathcal{G}^R}} \left[C(\boldsymbol{F}_{s_1}^{\mathsf{dec}}(\boldsymbol{a}_{s_1},\boldsymbol{g}_{s_1})^\star, \boldsymbol{F}_{s_2}^{\mathsf{dec}}(\boldsymbol{a}_{s_2},\boldsymbol{g}_{s_2})^\star, \boldsymbol{F}_{s_3}^{\mathsf{dec}}(\boldsymbol{a}_{s_3},\boldsymbol{g}_{s_3})^\star)\right] \tag{43}$$

Let $\delta = o_{\tau,T'}(1)$ where $\delta \to 0$ as $\tau \to 0$ and $T' \to \infty$. Fix a constraint C and a variable $s \in \mathcal{V}(C)$. Let E_s be the event that $\sum_j \mathfrak{Re}(F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}_s,\boldsymbol{g}_s)) \in [1-\delta, 1+\delta]$. Note that for all s and $\boldsymbol{z} \in \Sigma^R$, we have $\sum_j \mathcal{F}_{s,j}(\boldsymbol{z}) = 1$. The next claim shows that with high probability over $\boldsymbol{a}_s, \boldsymbol{g}_s$ sampled according to the rounding scheme, $\sum_j \mathfrak{Re}(F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}_s,\boldsymbol{g}_s)) \in [1-\delta, 1+\delta]$.

Claim 7.22. *For every $s \in \mathcal{V}$, $\Pr[E_s] \geqslant 1 - O_q(\delta)$.*

Proof. Define $\mathcal{F}_{\boldsymbol{a}}(\boldsymbol{z}) = \sum_j \mathcal{F}_{s,j}^{\mathsf{dec}}(\boldsymbol{a},\boldsymbol{z})$ where $\mathcal{F}_{s,j}^{\mathsf{dec}}$ is defined in Equation 41. Let $F_{\boldsymbol{a}}(\boldsymbol{\ell}) := F_{s,j}^{\mathsf{dec}}(\boldsymbol{a},\boldsymbol{\ell})$ be the multilinear extension of $\mathcal{F}_{\boldsymbol{a}}$. By definition, all the influences of $F_{\boldsymbol{a}}$ are bounded by $O_q(\tau)$.

Consider the function $\zeta(x) = |1-x|^2$. Applying the invariance principle from Theorem 3.9, we get

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{\ell}_s\sim\mathcal{L}|_s}[\zeta(F_{\boldsymbol{a}}(\boldsymbol{\ell}_s))] - \mathop{\mathbb{E}}_{\boldsymbol{g}_s}[\zeta(F_{\boldsymbol{a}}(\boldsymbol{g}_s))] \right| \leqslant o_\tau(1). \tag{44}$$

Now, we have

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell}_s\sim\mathcal{L}|_s}[\zeta(F_{\boldsymbol{a}}(\boldsymbol{\ell}_s))] = \mathop{\mathbb{E}}_{\boldsymbol{x}_s''\sim\mu_s^{\otimes R}}[\zeta(\mathcal{F}_{\boldsymbol{a}}(\boldsymbol{x}_s''))] = \mathop{\mathbb{E}}_{\boldsymbol{x}_s''\sim\mu_s^{\otimes R}}\left[\zeta\left(\sum_j \mathcal{F}_{s,j}^{\mathsf{dec}}(\boldsymbol{a},\boldsymbol{x}_s'')\right)\right].$$

We now fix a coupling $(\boldsymbol{x}_s', \boldsymbol{x}_s'', \boldsymbol{x}_s)$ between $\mu_s^{\otimes R} \times \mu_s^{\otimes R}$ and $\mu_s^{\otimes R}$ given by Lemma 6.5. Let $\boldsymbol{a}_s$ denote the sample $(\alpha_s^{(1)}, \alpha_s^{(2)}, \ldots, \alpha_s^{(R)})$ corresponding to the variable s according to the rounding scheme. Using Lemma 7.21, we switch $\boldsymbol{a}_s$ to a distribution $\sigma(\boldsymbol{x}')$

$$\mathop{\mathbb{E}}_{\boldsymbol{a}_s} \mathop{\mathbb{E}}_{\boldsymbol{x}_s''\sim\mu_s^{\otimes R}}\left[\zeta\left(\sum_j \mathcal{F}_{s,j}^{\mathsf{dec}}(\boldsymbol{a},\boldsymbol{x}_s'')\right)\right] \leqslant \mathop{\mathbb{E}}_{(\boldsymbol{x}_s',\boldsymbol{x}_s'',\boldsymbol{x}_s)}\left[\zeta\left(\sum_j \tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(\boldsymbol{x}_s',\boldsymbol{x}_s'')\right)\right] + 2^{-\Omega_{q,\alpha,\varepsilon}(T')}, \tag{45}$$

where $\tilde{\mathcal{F}}_{s,j}$ is the function defined in Equation(40). Using the fact that $\zeta(1+b) = |b|^2$ for every $b \in \mathbb{C}$, we get

$$\mathop{\mathbb{E}}_{(\boldsymbol{x}'_s,\boldsymbol{x}''_s,\boldsymbol{x}_s)}\left[\zeta\left(\sum_j \mathcal{F}^{\mathsf{dec}}_{s,j}(\boldsymbol{x}'_s,\boldsymbol{x}''_s)\right)\right] = \mathop{\mathbb{E}}_{(\boldsymbol{x}'_s,\boldsymbol{x}''_s,\boldsymbol{x}_s)}\left[\zeta\left(\sum_j \mathcal{F}^{\mathsf{dec}}_{s,j}(\boldsymbol{x}'_s,\boldsymbol{x}''_s) + \sum_j \mathcal{F}_{s,j}(\boldsymbol{x}_s) - \sum_j \mathcal{F}_{s,j}(\boldsymbol{x}_s)\right)\right]$$

$$= \mathop{\mathbb{E}}_{(\boldsymbol{x}'_s,\boldsymbol{x}''_s,\boldsymbol{x}_s)}\left[\left|\sum_j \mathcal{F}^{\mathsf{dec}}_{s,j}(\boldsymbol{x}'_s,\boldsymbol{x}''_s) - \sum_j \mathcal{F}_{s,j}(\boldsymbol{x}_s)\right|^2\right]$$

$$\leqslant \sum_j \mathop{\mathbb{E}}_{(\boldsymbol{x}'_s,\boldsymbol{x}''_s,\boldsymbol{x}_s)}\left[\left|\mathcal{F}^{\mathsf{dec}}_{s,j}(\boldsymbol{x}'_s,\boldsymbol{x}''_s - \mathcal{F}_{s,j}(\boldsymbol{x}_s)\right|^2\right]$$

Using Lemma 6.5 we have for every $1 \leqslant j \leqslant q$,

$$\mathop{\mathbb{E}}_{(\boldsymbol{x}'_s,\boldsymbol{x}''_s,\boldsymbol{x}_s)}\left[\left|\mathcal{F}^{\mathsf{dec}}_{s,j}(\boldsymbol{x}'_s,\boldsymbol{x}''_s) - \mathcal{F}_{s,j}(\boldsymbol{x}_s)\right|^2\right] \leqslant 2^{-\Omega_{q,\alpha,\varepsilon}(T')}$$

Therefore, we get

$$\mathop{\mathbb{E}}_{(\boldsymbol{x}'_s,\boldsymbol{x}''_s,\boldsymbol{x}_s)}\left[\zeta\left(\sum_j \mathcal{F}^{\mathsf{dec}}_{s,j}(\boldsymbol{x}'_s,\boldsymbol{x}''_s)\right)\right] \leqslant 2^{-\Omega_{q,\alpha,\varepsilon}(T')}. \tag{46}$$

Using Equations (44), (45), and (46), we have

$$\mathop{\mathbb{E}}_{\boldsymbol{a}_s,\boldsymbol{g}_s}\left[\zeta\left(\sum_j F^{\mathsf{dec}}_{s,j}(\boldsymbol{a}_s,\boldsymbol{g}_s)\right)\right] \leqslant 2^{-\Omega_{q,\alpha,\varepsilon}(T')} + o_\tau(1).$$

Setting $\delta' \approx 2^{-\Omega_{q,\alpha,\varepsilon}(T')} + o_\tau(1)$, we get that with probability at least $1-\sqrt{\delta'}$, $\sum_j \mathfrak{Re}(F^{\mathsf{dec}}_{s,j}(\boldsymbol{a}_s,\boldsymbol{g}_s)) \in [1-\sqrt{\delta'}, 1+\sqrt{\delta'}]$. This finished the proof of the claim. □

Using Equation(43), the above claim, and the facts that $\boldsymbol{F}^{\mathsf{dec}}_s(\boldsymbol{a}_s,\boldsymbol{g}_s)^\star$ is a 1-bounded function and the function C is upper bounded by $O_q(1)$ on 1-bounded inputs, we have

$$\begin{aligned}&\mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V},\boldsymbol{\mu})\\ &\geqslant \mathop{\mathbb{E}}_{C\sim\mathcal{C}} \mathop{\mathbb{E}}_{\substack{\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)},\ldots,\boldsymbol{\alpha}^{(R)}\sim\mathcal{S},\\ a_{s,i}=\alpha_s^{(i)},\\ \mathcal{G}^R}}\left[C(\boldsymbol{F}^{\mathsf{dec}}_{s_1}(\boldsymbol{a}_{s_1},\boldsymbol{g}_{s_1})^\star, \boldsymbol{F}^{\mathsf{dec}}_{s_2}(\boldsymbol{a}_{s_2},\boldsymbol{g}_{s_2})^\star, \boldsymbol{F}^{\mathsf{dec}}_{s_3}(\boldsymbol{a}_{s_3},\boldsymbol{g}_{s_3})^\star) \mid \wedge_{j=1}^3 E_{s_j}\right] - O_q(\delta).\end{aligned} \tag{47}$$

We now remove the truncation and scaling, and see the effect of this on the above expression. Towards this, we have a following simple claim.

Claim 7.23. *Let $\boldsymbol{a} \in \mathbb{R}^q$ such that $\sum_i a_i \in [1-\delta, 1+\delta]$, and let $\boldsymbol{a}^\star$ be the vector that we get after scaling the vector $\tilde{\boldsymbol{a}} := \mathsf{trunc}(\boldsymbol{a})$, then*

$$\sum_i (\tilde{a}_i - a_i^\star)^2 \leqslant q \sum_i (a_i - \tilde{a}_i)^2 + O_q(\delta).$$

Proof.

$$\sum_i (\tilde{a}_i - a_i^\star)^2 = \sum_i \left(\tilde{a}_i - \frac{\tilde{a}_i}{\sum_j \tilde{a}_j^2}\right)^2 = \left(\sum_i \tilde{a}_i - 1\right)^2 \sum_i \frac{\tilde{a}_i^2}{(\sum_j \tilde{a}_j)^2} \leqslant \left(\sum_i \tilde{a}_i - 1\right)^2$$

Now,

$$\left(\sum_i \tilde{a}_i - 1\right)^2 \leqslant \left(\sum_i \tilde{a}_i - a_i\right)^2 + O_q(\delta) \leqslant q\sum_i (\tilde{a}_i - a_i)^2 + O_q(\delta),$$

where the last inequality is the Cauchy-Schwarz inequality. □

Claim 7.24. *For every constraint C on (s_1, s_2, s_3), if $(\boldsymbol{a}_{s_1}, \boldsymbol{a}_{s_2}, \boldsymbol{a}_{s_3})$ and $(\boldsymbol{g}_{s_1}, \boldsymbol{g}_{s_2}, \boldsymbol{g}_{s_3})$ are distributed according to the distribution from Equation (47), then*

$$\left| \begin{array}{l} \mathbb{E}\left[C(\boldsymbol{F}^{\mathsf{dec}}_{s_1}(\boldsymbol{a}_{s_1}, \boldsymbol{g}_{s_1})^\star, \boldsymbol{F}^{\mathsf{dec}}_{s_2}(\boldsymbol{a}_{s_2}, \boldsymbol{g}_{s_2})^\star, \boldsymbol{F}^{\mathsf{dec}}_{s_3}(\boldsymbol{a}_{s_3}, \boldsymbol{g}_{s_3})^\star) \mid \wedge_{j=1}^3 E_{s_j}\right] - \\ \quad \mathbb{E}\left[C(\widetilde{\boldsymbol{F}^{\mathsf{dec}}_{s_1}}(\boldsymbol{a}_{s_1}, \boldsymbol{g}_{s_1}), \widetilde{\boldsymbol{F}^{\mathsf{dec}}_{s_2}}(\boldsymbol{a}_{s_2}, \boldsymbol{g}_{s_2}), \widetilde{\boldsymbol{F}^{\mathsf{dec}}_{s_3}}(\boldsymbol{a}_{s_3}, \boldsymbol{g}_{s_3})) \mid \wedge_{j=1}^3 E_{s_j}\right] \end{array} \right| \leqslant o_\tau(1) + O_q(\delta).$$

Proof. Using the fact that C is a Lipschitz function with Lipschitz constant $C_0(q)$, we get that the LHS is upper bounded by

$$C_0(q) \cdot \sum_{j=1}^3 \mathbb{E}\left[\|\boldsymbol{F}^{\mathsf{dec}}_{s_j}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})^\star - \widetilde{\boldsymbol{F}^{\mathsf{dec}}_{s_j}}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})\|_2 \mid \wedge_{j=1}^3 E_{s_j}\right],$$

which by Claim 7.23 is upper bounded by

$$O_q(1) \cdot \sum_{j=1}^3 \mathbb{E}\left[\|\mathfrak{Re}(\boldsymbol{F}^{\mathsf{dec}}_{s_j}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})) - \widetilde{\boldsymbol{F}^{\mathsf{dec}}_{s_j}}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})\|_2 \mid \wedge_{j=1}^3 E_{s_j}\right] + O_q(\delta),$$

where we use the fact that conditioned on the event E_{s_j}, $\sum_\sigma \mathfrak{Re}(F^{\mathsf{dec}}_{s_j,\sigma}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})) \in [1-\delta, 1+\delta]$. For a non-negative random variable, $\mathbb{E}[X \mid E] \leqslant \frac{\mathbb{E}[X]}{\Pr[E]}$, and hence,

$$\begin{aligned} &\mathbb{E}\left[\|\mathfrak{Re}(\boldsymbol{F}^{\mathsf{dec}}_{s_j}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})) - \widetilde{\boldsymbol{F}^{\mathsf{dec}}_{s_j}}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})\|_2 \mid \wedge_{j=1}^3 E_{s_j}\right] \\ &\qquad \leqslant \frac{1}{\Pr[\wedge_{j=1}^3 E_{s_j}]} \mathbb{E}\left[\|\mathfrak{Re}(\boldsymbol{F}^{\mathsf{dec}}_{s_j}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})) - \widetilde{\boldsymbol{F}^{\mathsf{dec}}_{s_j}}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})\|_2\right] \\ &\qquad \leqslant 2\,\mathbb{E}\left[\|\mathfrak{Re}(\boldsymbol{F}^{\mathsf{dec}}_{s_j}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})) - \widetilde{\boldsymbol{F}^{\mathsf{dec}}_{s_j}}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})\|_2\right], \end{aligned}$$

using the fact that $\Pr[\wedge_{j=1}^3 E_{s_j}] = 1 - O_q(\delta) \leqslant 1/2$. Therefore, the LHS from the claim is upper bounded by

$$O_q(1) \cdot \sum_{j=1}^3 \mathbb{E}\left[\|\mathfrak{Re}(\boldsymbol{F}^{\mathsf{dec}}_{s_j}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})) - \widetilde{\boldsymbol{F}^{\mathsf{dec}}_{s_j}}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})\|_2\right] + O_q(\delta) \tag{48}$$

Now, for a given variable $s \in \mathcal{V}$, by the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}\left[\|\mathfrak{Re}(\boldsymbol{F}_s^{\mathsf{dec}}(\boldsymbol{a}_s, \boldsymbol{g}_s)) - \widetilde{\boldsymbol{F}_s^{\mathsf{dec}}}(\boldsymbol{a}_s, \boldsymbol{g}_s)\|_2\right]^2 &\leqslant \mathbb{E}\left[\|\mathfrak{Re}(\boldsymbol{F}_s^{\mathsf{dec}}(\boldsymbol{a}_s, \boldsymbol{g}_s)) - \widetilde{\boldsymbol{F}_s^{\mathsf{dec}}}(\boldsymbol{a}_s, \boldsymbol{g}_s)\|_2^2\right] \\ &= \mathbb{E}\left[\sum_{j=1}^{q}(\mathfrak{Re}(F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}_s, \boldsymbol{g}_s)) - \widetilde{F_{s,j}^{\mathsf{dec}}}(\boldsymbol{a}_s, \boldsymbol{g}_s))^2\right] \\ &= \sum_{j=1}^{q}\mathbb{E}\left[(\mathfrak{Re}(F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}_s, \boldsymbol{g}_s)) - \widetilde{F_{s,j}^{\mathsf{dec}}}(\boldsymbol{a}_s, \boldsymbol{g}_s))^2\right] \\ &= \sum_{j=1}^{q}\mathbb{E}\left[\boldsymbol{\gamma}\left(F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}_s, \boldsymbol{g}_s)\right)\right], \end{aligned} \tag{49}$$

where $\boldsymbol{\gamma}(x) := (\mathfrak{Re}(x) - \mathsf{trunc}(\mathfrak{Re}(x))^2$. Applying the invariance principle from Theorem 6.9 with the smooth function $\Psi = \boldsymbol{\gamma}$ for functions $f = \mathcal{F}_{s,j}$, $g \equiv 1$, and $h \equiv 1$, we get

$$\left|\mathbb{E}\left[\boldsymbol{\gamma}\left(F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}_s, \boldsymbol{g}_s)\right)\right] - \mathop{\mathbb{E}}_{\boldsymbol{z}_s \sim \mu_s^{\otimes R}}\left[\boldsymbol{\gamma}\left(\mathcal{F}_{s,j}(\boldsymbol{z}_s)\right)\right]\right| \leqslant 2^{-\Omega_{q,\alpha,\varepsilon}(T')} + o_\tau(1).$$

As $\gamma\left(\mathcal{F}_{s,j}(\boldsymbol{z}_s)\right) = 0$, we have

$$\mathbb{E}\left[\boldsymbol{\gamma}\left(F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}_s, \boldsymbol{g}_s)\right)\right] \leqslant 2^{-\Omega_{q,\alpha,\varepsilon}(T')} + o_\tau(1). \tag{50}$$

Combining Equations (48), (49), and (50), we get that the LHS from the claim is upper bounded by $2^{-\Omega_{q,\alpha,\varepsilon}(T')} + o_\tau(1) + O_q(\delta)$ as claimed. □

Using the above claim and Equation(47), we get

$$\begin{aligned}&\mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V}, \boldsymbol{\mu}) \\ &\geqslant \mathop{\mathbb{E}}_{C\sim\mathcal{C}} \mathop{\mathbb{E}}_{\substack{\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)},\dots,\boldsymbol{\alpha}^{(R)}\sim\mathcal{S}, \\ a_{s,i}=\alpha_s^{(i)}, \\ \mathcal{G}^R}}\left[C(\widetilde{\boldsymbol{F}_{s_1}^{\mathsf{dec}}}(\boldsymbol{a}_{s_1}, \boldsymbol{g}_{s_1}), \widetilde{\boldsymbol{F}_{s_2}^{\mathsf{dec}}}(\boldsymbol{a}_{s_2}, \boldsymbol{g}_{s_2}), \widetilde{\boldsymbol{F}_{s_3}^{\mathsf{dec}}}(\boldsymbol{a}_{s_3}, \boldsymbol{g}_{s_3})) \mid \wedge_{j=1}^{3} E_{s_j}\right] - O_q(\delta).\end{aligned}$$

We can now remove the conditioning using the fact that $\widetilde{\boldsymbol{F}_s^{\mathsf{dec}}}$ is 1-bounded and $\Pr[E] \geqslant 1 - O_q(\delta)$ to get.

$$\mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V}, \boldsymbol{\mu}) \geqslant \mathop{\mathbb{E}}_{C\sim\mathcal{C}} \mathop{\mathbb{E}}_{\substack{\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)},\dots,\boldsymbol{\alpha}^{(R)}\sim\mathcal{S}, \\ a_{s,i}=\alpha_s^{(i)}, \\ \mathcal{G}^R}}\left[C(\widetilde{\boldsymbol{F}_{s_1}^{\mathsf{dec}}}(\boldsymbol{a}_{s_1}, \boldsymbol{g}_{s_1}), \widetilde{\boldsymbol{F}_{s_2}^{\mathsf{dec}}}(\boldsymbol{a}_{s_2}, \boldsymbol{g}_{s_2}), \widetilde{\boldsymbol{F}_{s_3}^{\mathsf{dec}}}(\boldsymbol{a}_{s_3}, \boldsymbol{g}_{s_3}))\right] - O_q(\delta). \tag{51}$$

We already had the following bound from Equation(42)

$$\Pr[\boldsymbol{\mathcal{F}} \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}] \leqslant \mathop{\mathbb{E}}_{C\sim\mathcal{C}} \mathop{\mathbb{E}}_{\substack{\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)},\ldots,\boldsymbol{\alpha}^{(R)}\sim\mathcal{S},\\ a_{s,i}=\alpha_s^{(i)},\\ \mathcal{G}^R}} [C(\widetilde{\boldsymbol{F}_{s_1}^{\mathsf{dec}}}(\boldsymbol{a}_{s_1},\boldsymbol{g}_{s_1}),\widetilde{\boldsymbol{F}_{s_2}^{\mathsf{dec}}}(\boldsymbol{a}_{s_2},\boldsymbol{g}_{s_2}),\widetilde{\boldsymbol{F}_{s_3}^{\mathsf{dec}}}(\boldsymbol{c},\boldsymbol{g}_{s_3}))]$$
$$+ 2^{-\Omega_{q,\alpha,\varepsilon}(T')} + \eta(\tau)$$

Combining the above two inequalities, we get

$$\Pr[\boldsymbol{\mathcal{F}} \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}] \leqslant \mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V},\boldsymbol{\mu}) + 2^{-\Omega_{q,\alpha,\varepsilon}(T')} + \eta'(\tau)$$

where $\eta'(\tau) \to 0$ as $\tau \to 0$. As the integral value of the instance Υ is at most $\mathsf{OPT}(\Upsilon)$, we have $\mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V},\boldsymbol{\mu}) \leqslant \mathsf{OPT}(\Upsilon)$ and hence

$$\Pr[\boldsymbol{\mathcal{F}} \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}] \leqslant \mathsf{OPT}(\Upsilon) + 2^{-\Omega_{q,\alpha,\varepsilon}(T')} + \eta'(\tau),$$

and this finishes the proof of Theorem 7.2. □

Setup: For each $s \in \mathcal{V}$, the probability space $\Omega_s = (\Sigma, \mu_s)$ consists of atoms in Σ with the distribution $\mu_s(a) = \|b_{s,a}\|^2$. Let $\boldsymbol{\mathcal{F}}_s$ denote the function obtained by interpreting the function $\boldsymbol{\mathcal{F}} : \Sigma^R \to \blacktriangle_q$ as a function over Ω_s^R.
Input: An instance $\Upsilon = (\mathcal{V}, \mathcal{C})$ of a Max-$\mathcal{P}$-CSP such that the algorithm $\mathcal{ALG}$ accepts Υ. Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be a solution for the basic SDP relaxation of Υ and $\mathcal{S} \subseteq G^{\mathcal{V}}_{\mathsf{master}}$ be the subspace of the set of satisfying assignments of the instance $\Upsilon(\mathcal{V}, \mathcal{C})$ over $G^{\mathcal{V}}_{\mathsf{master}} := \prod_{v \in \mathcal{V}} H^\star_v$ that satisfy the conditions from Lemma 7.11.

Rounding Scheme:
Step I: Sample R Gaussian vectors $\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}, \dots, \boldsymbol{\zeta}^{(R)}$ with the same dimension as $\boldsymbol{V}$.

Step II: For each $s \in \mathcal{V}$, do the following:

1. For each $j \in [R]$, let $g^{(j)}_{s,0} \equiv \mathbf{1}$ and for $c \in \{1, \dots, q-1\}$, set
$$g^{(j)}_{s,c} = \sum_{\omega \in \Sigma} \ell_{s,c}(\omega) \langle \boldsymbol{b}_{s,\omega}, \boldsymbol{\zeta}^{(j)} \rangle.$$
Let $\boldsymbol{g}^{(j)}_s = (g^{(j)}_{s,0} \equiv \mathbf{1}, g^{(j)}_{s,1}, \dots, g^{(j)}_{s,q-1})$ and $\boldsymbol{g}_s = (\boldsymbol{g}^{(1)}_s, \boldsymbol{g}^{(2)}_s, \dots, \boldsymbol{g}^{(R)}_s)$.

2. Sample R uniformly random assignments $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(R)}$ from the set of satisfying assignments to the instance Υ over $G^{\mathcal{V}}_{\mathsf{master}}$. Let $\boldsymbol{a}_s = (\boldsymbol{\alpha}^{(1)}_s, \boldsymbol{\alpha}^{(2)}_s, \dots, \boldsymbol{\alpha}^{(R)}_s)$.

3. Evaluate the polynomial $\boldsymbol{F}^{\mathsf{dec}}_s$ with $(\boldsymbol{a}_s, \boldsymbol{g}_s)$ as inputs to obtain $\boldsymbol{p}_s \in \mathbb{C}^q$, and let, $\tilde{\boldsymbol{p}}_s = \mathsf{trunc}(\Re\mathfrak{e}(\boldsymbol{p}_s))$ where
$$\mathsf{trunc}(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leqslant x \leqslant 1, \\ 1 & \text{if } x > 1, \end{cases}$$

4. Round $\boldsymbol{p}_s$ to $\boldsymbol{p}^\star_s$.
$$\boldsymbol{p}^\star_s = \mathbf{Scale}((\boldsymbol{p}_s)_1, (\boldsymbol{p}_s)_2, \dots, (\boldsymbol{p}_s)_q),$$
where
$$\mathbf{Scale}(x_1, x_2, \dots, x_q) = \begin{cases} \frac{1}{\sum_i x_i}(x_1, x_2, \dots, x_q) & \text{if } \sum_i x_i \neq 0, \\ (1, 0, 0, \dots, 0) & \text{if } \sum_i x_i = 0. \end{cases}$$

5. Assign the variable $s \in \mathcal{V}$ a value $a \in \Sigma$ with probability $(\boldsymbol{p}^\star_s)_{\varsigma^{-1}(a)}$.

Step III: Output the assignment from Step II.

Figure 4: Rounding Scheme $\mathbf{Round}_{\boldsymbol{\mathcal{F}}}$.

References

[1] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, May 1998. (Preliminary version

in *33rd FOCS*, 1992).

[2] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, January 1998. (Preliminary version in *33rd FOCS*, 1992).

[3] Libor Barto and Marcin Kozik. Constraint satisfaction problems of bounded width. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 595–603, 2009.

[4] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8(12):269–289, 2012.

[5] Vitaly Bergelson, Terence Tao, and Tamar Ziegler. An inverse theorem for the uniformity seminorms associated with the action of. *Geometric and Functional Analysis*, 19(6):1539–1596, 2010.

[6] Amey Bhangale and Subhash Khot. Optimal inapproximability of satisfiable k-lin over non-abelian groups. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1615–1628. ACM, 2021.

[7] Amey Bhangale, Subhash Khot, and Dor Minzer. On Approximability of Satisfiable *k*-CSPs: I. In *54th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2022, Rome, Italy, June 20 - 24, 2022*, pages 976–988, 2022.

[8] Amey Bhangale, Subhash Khot, and Dor Minzer. Effective bounds for restricted 3-arithmetic progressions in $\mathbb{F}_p^n$. *arXiv preprint arXiv:2308.06600*, 2023.

[9] Amey Bhangale, Subhash Khot, and Dor Minzer. On Approximability of Satisfiable *k*-CSPs: II. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 632–642, 2023.

[10] Amey Bhangale, Subhash Khot, and Dor Minzer. On Approximability of Satisfiable *k*-CSPs: III. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 643–655, 2023.

[11] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable k-CSPs: IV. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 1423–1434. ACM, 2024.

[12] Joshua Brakensiek and Venkatesan Guruswami. New Hardness Results for Graph and Hypergraph Colorings. In *31st Conference on Computational Complexity (CCC 2016)*, volume 50 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 14:1–14:27, Dagstuhl, Germany, 2016. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[13] Joshua Brakensiek and Venkatesan Guruswami. An algorithmic blend of LPs and ring equations for promise CSPs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 436–455. SIAM, 2019.

[14] Joshua Brakensiek, Venkatesan Guruswami, Marcin Wrochna, and Stanislav Živný. The power of the combined basic linear programming and affine relaxation for promise constraint satisfaction problems. *SIAM Journal on Computing*, 49(6):1232–1248, 2020.

[15] Joshua Brakensiek, Neng Huang, Aaron Potechin, and Uri Zwick. On the mysteries of MAX NAE-SAT. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 484–503. SIAM, 2021.

[16] Mark Braverman, Subhash Khot, and Dor Minzer. On Rich 2-to-1 Games. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, pages 27:1–27:20, 2021.

[17] Andrei A. Bulatov. A dichotomy theorem for nonuniform CSPs. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 319–330, 2017.

[18] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. *ACM Trans. Algorithms*, 5(3), jul 2009.

[19] Lorenzo Ciardo and Stanislav Živný. Clap: A new algorithm for promise CSPs. *SIAM Journal on Computing*, 52(1):1–37, 2023.

[20] Lorenzo Ciardo and Stanislav Živný. Semidefinite programming and linear equations vs. homomorphism problems. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 1935–1943, New York, NY, USA, 2024. Association for Computing Machinery.

[21] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, STOC '71, page 151–158, New York, NY, USA, 1971. Association for Computing Machinery.

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

[23] Tomás Feder and Moshe Y. Vardi. The computational structure of monotone monadic snp and constraint satisfaction: A study through datalog and group theory. *SIAM Journal on Computing*, 28(1):57–104, 1998.

[24] Uriel Feige, Shafi Goldwasser, Laszlo Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM (JACM)*, 43(2):268–292, 1996.

[25] Harry Furstenberg and Yitzhak Katznelson. A density version of the Hales-Jewett theorem for k= 3. In *Annals of Discrete Mathematics*, volume 43, pages 227–241. Elsevier, 1989.

[26] Hillel Furstenberg and Yitzhak Katznelson. An ergodic Szemerédi theorem for commuting transformations. *Journal d'Analyse Mathématique*, 34(1):275–291, 1978.

[27] Hillel Furstenberg and Yitzhak Katznelson. A density version of the Hales-Jewett theorem. *Journal d'Analyse Mathématique*, 57(1):64–119, 1991.

[28] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995.

[29] Ben Green. Montreal lecture notes on quadratic fourier analysis. *arXiv preprint math/0604089*, 2006.

[30] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613–622, 2001.

[31] Venkatesan Guruswami and Euiwoong Lee. Strong inapproximability results on balanced rainbow-colorable hypergraphs. *Combinatorica*, 38(3):547–599, 2018.

[32] Alfred W Hales and Robert I Jewett. Regularity and positional games. In *Classic Papers in Combinatorics*, pages 320–327. Springer, 2009.

[33] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.

[34] Johan Håstad. On the NP-hardness of max-not-2. *SIAM Journal on Computing*, 43(1):179–193, 2014.

[35] David Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, mar 1998.

[36] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th STOC*, pages 767–775. ACM, 2002.

[37] Subhash Khot and Rishi Saket. A 3-query non-adaptive PCP with perfect completeness. In *21st Annual IEEE Conference on Computational Complexity (CCC'06)*, pages 11–pp. IEEE, 2006.

[38] 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*, STOC '14, page 634–643, New York, NY, USA, 2014. Association for Computing Machinery.

[39] Richard E Ladner. On the structure of polynomial time reducibility. *Journal of the ACM (JACM)*, 22(1):155–171, 1975.

[40] Leonid Anatolevich Levin. Universal sequential search problems. *Problemy peredachi informatsii*, 9(3):115–116, 1973.

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

[42] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010.

[43] Ryan O'Donnell and Yi Wu. 3-bit dictator testing: 1 vs. 5/8. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 365–374. SIAM, 2009.

[44] Ryan O'Donnell and Yi Wu. Conditional hardness for satisfiable 3-CSPs. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, page 493–502, New York, NY, USA, 2009. Association for Computing Machinery.

[45] DHJ Polymath. A new proof of the density Hales-Jewett theorem. *Annals of Mathematics*, pages 1283–1327, 2012.

[46] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the fortieth annual ACM symposium on Theory of computing (STOC)*, pages 245–254, 2008.

[47] Prasad Raghavendra and David Steurer. How to round any CSP. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 586–594, 2009.

[48] Thomas J. Schaefer. The complexity of satisfiability problems. In *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing*, STOC '78, page 216–226, New York, NY, USA, 1978. Association for Computing Machinery.

[49] Grant Schoenebeck. Linear level lasserre lower bounds for certain k-CSPs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 593–602. IEEE, 2008.

[50] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Analysis & PDE*, 3(1):1–20, 2010.

[51] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *Annals of Combinatorics*, 16(1):121–188, 2012.

[52] Dmitriy Zhuk. A proof of the CSP dichotomy conjecture. *J. ACM*, 67(5), August 2020.

[53] Uri Zwick. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *SODA*, volume 98, pages 201–210, 1998.

[54] Uri Zwick. Outward rotations: a tool for rounding solutions of semidefinite programming relaxations, with applications to max cut and other problems. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, STOC '99, page 679–687, New York, NY, USA, 1999. Association for Computing Machinery.

A Omitted Proofs

Let $p \geqslant 5$ be a prime, take $\Sigma = \mathbb{F}_p$ and consider the predicate $P\colon \Sigma^3 \to \{0,1\}$ defined as $P(x,y,z) = 1_{x,y,z \text{ are distinct}}$.

Claim A.1. *The collection* $\{P\}$ *is* MILDLY-SYMMETRIC .

Proof. For each $a \in \mathbb{F}_p \setminus \{0\}$ and $b \in \mathbb{F}_p$ define the map $\tau_{a,b}\colon \Sigma \to \Sigma$ by $\tau_{a,b}(u) = au + b$. It is clear that each one of these maps preserves the satisfying assignments of P, so it remains to check the second condition in Definition 1.3.

Let (x,y,z) be any satisfying assignment. We can find $a \neq 0$ and b such that $ax+b=0$ and $ay+b=1$, so we may assume without loss of generality that $(x,y,z) = (0,1,z)$ for some $z \neq 0,1$; this is justified as the orbits of $(0,1,az+b)$ and (x,y,z) under $\{\tau_{a',b'}\}_{a'\neq 0,b'}$ are the same, and $az+b$ is some element in $\mathbb{F}_p$ not equal to $0,1$. Take $z' = ax+b$, and to simplify notation we omit z' and simply call it z (the original x,y,z will not be used henceforth in the argumnet).

Let $A = \{\,(b, a+b, az+b) \mid a \neq 0, b \in \mathbb{F}_p\}$ be the orbit of $(0,1,z)$ under $\{\tau_{a,b}\}_{a\neq 0,b}$, and suppose that σ, γ, ϕ is an embedding of A into $(\mathbb{Z}, +)$. By applying an affine shift to all embeddings we may assume that $\gamma(0) = \phi(0) = 0$, and by the definition of embeddings we get that

$$\sigma(b) + \gamma(a+b) + \phi(az+b) = 0 \qquad \forall a \neq 0, b. \tag{52}$$

Taking $b = -a$ and using $\gamma(0) = 0$ we get that $\sigma(-a) + \phi(a(z-1)) = 0$ for $a \neq 0$, so $\phi(a(z-1)) = -\sigma(-a)$. Hence we get that

$$-\phi(-b(z-1)) + \gamma(a+b) + \phi(az+b) = 0 \qquad \forall a,b \neq 0.$$

Taking $b = -az$ and using $\phi(0) = 0$ we get $-\phi(az(z-1)) + \gamma(-a(z-1)) = 0$, meaning that $\gamma(y) = \phi(-zy)$ for all $y \neq 0$. Equality also holds for $y = 0$ (as both values are 0) and we conclude that

$$-\phi(-b(z-1)) + \phi(-z(a+b)) + \phi(az+b) = 0 \qquad \forall a, b \neq 0.$$

Note that the image of $(-b(z-1), -z(a+b))$ under $a, b \neq 0$ consists of (α, β) such that $\alpha \neq 0$ and $(z-1)\beta - z\alpha \neq 0$, and we get that for any such α, β it holds that $\phi(\alpha) = \phi(\beta) + \phi(\alpha - \beta)$. Note that this equality trivially holds for $\beta = 0$ (as $\phi(0) = 0$), so we conclude that $\phi(\alpha) = \phi(\beta) + \phi(\alpha - \beta)$ holds whenever $(z-1)\beta - z\alpha \neq 0$. We now use the idea of local self correction to argue that $\phi(\alpha) = \phi(\beta) + \phi(\alpha - \beta)$ in fact holds for all $\alpha, \beta \in \mathbb{F}_p$.

Take any α, β satisfying $(z-1)\beta - z\alpha = 0$, and choose $\alpha', \beta' \in \mathbb{F}_p$ uniformly and independently. With probability at least $1 - \frac{4}{p} > 0$ we have that $(z-1)\beta' - z\alpha' \neq 0$, $(z-1)(\beta - \beta') - z(\alpha - \alpha') \neq 0$, $(z-1)\beta' - z\beta \neq 0$, $(z-1)(\alpha' - \beta') - z(\alpha - \beta) \neq 0$ and $(z-1)\alpha' - z\alpha \neq 0$ all hold,[13] and we fix $\alpha', \beta' > 0$ satisfying all of these inequalities. Adding up the constraints we get from the first two conditions we get that

$$\phi(\alpha') + \phi(\alpha - \alpha') = (\phi(\beta') + \phi(\alpha' - \beta')) + (\phi(\beta - \beta') + \phi(\alpha - \beta - \alpha' + \beta'))$$

Using the third and foruth conditions we have $\phi(\beta') + \phi(\beta - \beta') = \phi(\beta)$ and $\phi(\alpha' - \beta') + \phi(\alpha - \beta - \alpha' + \beta') = \phi(\alpha - \beta)$ so that the right hand side simplifies to $\phi(\beta) + \phi(\alpha - \beta)$. Using the fifth condition the left hand side simplifies to $\phi(\alpha)$, altogether giving that $\phi(\alpha) = \phi(\beta) + \phi(\alpha - \beta)$.

We conclude that $\phi(\alpha) = \phi(\beta) + \phi(\alpha - \beta)$ for all $\alpha, \beta \in \mathbb{F}_p$. Thus, we get that $\phi(x) = x\phi(1)$ for all $x \in \mathbb{F}_p$ and also that $\phi(1) = \phi(2 \cdot (p-1)/2 + 2) = 2\phi((p-1)/2) + \phi(2) = (p+1)\phi(1)$, hence $\phi(1) = 0$. This implies that $\phi \equiv 0$, and using $\gamma(y) = \phi(-zy)$ that holds for $y \neq 0$ and $\gamma(0) = 0$ it follows that $\gamma \equiv 0$. Plugging this into (52) gives that $\sigma \equiv 0$, and the proof is concluded. $\square$

[13]Note that we have 5 conditions, so a naive application of the union bound only gives a lower bound of $1 - \frac{5}{p}$ on the probability that all of these events hold. However, since $(z-1)\beta - z\alpha = 0$, the first two conditions can be seen to be equivalent, so the result of the union bound improves to $1 - \frac{4}{p}$.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il