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 Gaussian 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 collection 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 algorithm

is a polynomial-time algorithm that, given an instance Υ, finds an assignment satisfying at least $\alpha \cdot \mathrm{OPT}(\Upsilon)$ fraction of the constraints, where $\mathrm{OPT}(\Upsilon)$ 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, 41], the PCP theorem [1, 2, 24], the development of semidefinite programming algorithms [28, 36, 55, 56, 48, 18], the Unique Games Conjecture and its consequences [37, 48], the dichotomy theorem [23, 17, 54] 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 [50], 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 [40], 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, 54], who proved that, indeed, for any family of predicates $\mathcal{P}$, 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 $\mathrm{OPT}(\Upsilon) \geqslant 1 - \varepsilon$ 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 [34] states that for all $\varepsilon > 0$, given an instance Υ of the Max-3-Lin problem promised to have $\mathrm{OPT}(\Upsilon) \geqslant 1 - \varepsilon$, 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 [34] 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, known as the Unique Games Conjecture (UGC) [37]. Assuming UGC, Raghavendra [48] 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 $\mathrm{OPT}(\Upsilon) \geqslant 1-\varepsilon$, 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 Approximating 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 [37]. 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 collections $\mathcal{P}$ 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. There are even fewer collections $\mathcal{P}$ for which we know the value of $\alpha_{\mathcal{P}}$. An example for such problem is the 3-SAT problem, for which Håstad [34] 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] [45, 46, 38], for which the optimal threshold of $\alpha_{NTW} = 5/8$ was shown by Håstad [35]. 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 progress on 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, 52, 53]. 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 $k = 3$, and it can be used to study the density of restricted 3-AP free sets [8], a result which was

---

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

previously unknown. As the difficulty in the study of Gowers' uniformity norms sharply increase when k is large, it stands to reason that 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 [48] as well as the dichotomy theorem [17, 54], 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 relatively elementary Hales-Jewett theorem [33] and the more challenging density Hales-Jewett theorem [25, 27, 47].

1.3.2 Approximation Algorithms vs Hardness Reductions

To characterize the right approximation threshold for Max-$\mathcal{P}$-CSP, one has to work both on the algorithmic front as well as on the hardness front, and make them meet. In this section we discuss the way it is manifested in almost satisfiable CSPs, and the difference between that and our setting 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 there is $i \in [n]$ such that $f(x) = x_i$ for all x. A dictatorship test is a procedure that queries f at a few (correlated) locations randomly, and based on these values 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.,

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

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 the 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 [48], the class of quasirandom functions consists of low-degree functions with no influential coordinates [42]. To utilize these functions algorithmically, the powerful invariance principle of [43] 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 [43] 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 Abelian embeddings, also referred to as linear embeddings.

Definition 1.2. *Let $\Sigma_1, \ldots, \Sigma_k$ 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 $i = 1, \ldots, k$ form an Abelian embedding of A if*

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

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 Abelianly embeddable if $\mathsf{supp}(\mu)$ is.

---

[3]Often times one requires a plausible complexity-theoretic assumption, such as the Unique-Games Conjecture [37] or the Rich-2-to-1 Games Conjecture [16].

5

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 [32] 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 $\mathcal{P} \subseteq \{P \colon \Sigma^3 \to \{0, 1\}\}$ is called* MILDLY-SYMMETRIC *if there are actions $\tau_1, \tau_2, \ldots, \tau_\ell \colon \Sigma \to \Sigma$ 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 soundness $\alpha_\mathcal{P} + \varepsilon$.*

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 [48] 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.

7

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 $x_1, \ldots, x_n$ 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, 51]. 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 $\{1, 2, \ldots, n\}$, 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 $\displaystyle\mathop{\mathbb{E}}_{c \in \mathcal{C}} \mathop{\mathbb{E}}_{x \sim \mu_c} [c(x)]$ subject to (1) $\langle \boldsymbol{b}_{i,a}, \boldsymbol{b}_{j,b} \rangle = \displaystyle\Pr_{x \sim \mu_c}[x_i = a, x_j = b]$ $\qquad c \in \mathcal{C}, \quad i, j \in \mathcal{V}(c), \quad a, b \in \Sigma,$ (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 $(a_1, a_2, a_3) \in \text{supp}(\mu_c)$. 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 7.12.

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

10

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:

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

$$\geqslant (1-\varepsilon) \underset{(c,P)}{\mathbb{E}}\left[\underset{(y_{1,j},y_{2,j},\ldots,y_{k,j})\sim\mu_c}{\mathbb{E}}[P(y_{1,j}, y_{2,j},\ldots, y_{k,j})]\right]$$

$$= (1-\varepsilon)\cdot 1,$$

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:

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

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's argument [48] 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 $(y_{1,j}, y_{2,j}, \ldots, y_{k,j})$ has full support, a result of [42] 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 [43] 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

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

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 $(y_1, \ldots, y_k) \sim \mu^R$ 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 [42]. 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 $\chi \in \widehat{G}^R$, 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

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

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.

12

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 [43] 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

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

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

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

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

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

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 algorithm presented in the beginning 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 the result should hold for the more general class of pairwise 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 [49] and Khot, Tulsiani, and Worah [39] 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

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

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

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

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 invariance principle, we decouple the inputs to the characters-part of $\tilde{f}_1$ and the low-degree parts of $\tilde{f}_1$ and define

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

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

---

[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 jth 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_P with their corresponding multi-linear expansions, we have*

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

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, 47], 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 outline the proof of Theorem 1.2. The formal proof appears in Sections 5 and 6.

2.1 List Decoding

Consider an expectation of the form (8) and suppose it is non-negligible in absolute value. We already know, using results of Mossel [42], 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 $\sigma, \gamma, \phi \colon \Sigma \to G_{\text{master}}$ of μ into G_{master}, such that if f_1, f_2, f_3 are $O(1)$-bounded functions satisfying

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

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 $\chi \in \widehat{G}^n_{\text{master}}$ 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 $\mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f_1(x) f_2(y) f_3(z)]$ 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, which is not true in our case. 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 [43, 42] 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 ε, then sample $y \sim \nu$ conditioned on $y_{\bar{I}} = x_{\bar{I}}$ and $P_i(y) = P_i(x)$ for all $i = 1, \ldots, r$ (see Definition 4.1). The function $\mathrm{T}_{\mathcal{P}, 1-\varepsilon} f_1$ is then defined by

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

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 does not 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.

16

2.2.1 The Regularity Lemma

With the new noise operator $T_{\mathcal{P},1-\varepsilon}$ 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 $\tilde{f}_1 = T_{\mathcal{P},1-\varepsilon}f$ 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 $\tilde{f}_1 = T_{\mathcal{P},1-\varepsilon}f$ 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 some 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 $T_{\{P'\},1-\varepsilon}(f - T_{\{P\},1-\varepsilon}f)$ from $f - T_{\{P\},1-\varepsilon}f$ to get the function

$$\tilde{f}_1 = f - T_{\{P\},1-\varepsilon}f - T_{\{P'\},1-\varepsilon}(f - T_{\{P\},1-\varepsilon}f),$$

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 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 $\tilde{f}_1 = T_{\mathcal{P},1-\varepsilon}f$ 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 $T_{\mathcal{P},1-\varepsilon}f_1$ 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

$$T_{\mathcal{P},1-\varepsilon}f(x) = \mathop{\mathbb{E}}_{I \subseteq_\varepsilon [n]} \left[\frac{1}{A_I(x)} \mathbb{E}_{y\sim\nu} \left[f(y)1_{P_i(y)=P_i(x)\ \forall i} \mid y_I = x_I \right] \right],$$

where $A_I(x) = \Pr_{y\sim\nu} [P_i(y) = P_i(x)\ \forall i \mid y_{\bar{I}} = x_{\bar{I}}]$. Morally, we show that on most inputs, both $1/A_I(x)$ and the expectation above are close to functions of the form of the right-hand side of (10), except that the function L_P is not a low-degree function but instead only depends on the coordinates from $\bar{I}$. The expectation over I then turns these L_P into low-degree functions (see Lemma 3.4).

---

[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 $T_{\mathcal{P},1-\varepsilon}f_1$, $T_{\mathcal{Q},1-\varepsilon}f_2$ and $T_{\mathcal{R},1-\varepsilon}f_3$ 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

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

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 $\mathcal{P}$ has a high rank, 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 high-rank, we mean that besides the all 1 function, 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.

The intuition for 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 rank 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 the functions $P(X)$ and the functions $L_P(X)$ is almost independent of each other.

2.4 Deducing the Invariance Principle

In conclusion, we have shown that the values of the noised function $T_{\mathcal{P},1-\varepsilon}f_1$ can be coupled with the values of decoupled function from (11) in a way that is close in L_2 distance. Intuitively, that ought to imply that

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

While this is morally the case, there are some issues one has to address. The first issue is of purely technical nature: the output of the decoupled functions is not necessarily $O(1)$, so one should replace these values by some sort of truncation. The second issue is that above, we actually need a multi-variate version of our mixed invariance principle. While we show how to deduce it such result (in a relatively straightforward manner) from the uni-variate version the mixed invariance, it requires the collections of products functions we have produced for each one of f_1, f_2 and f_3 to be aligned in a certain manner, and we defer the reader to Section 6.3 for more details.

With the above result, 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 by $I \subseteq_\rho [n]$ a random subset of $[n]$ in which we include each $i \in [n]$ independently with probability ε.

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*

$$V_{=d}(\Sigma^n, \nu^n) = V_{\leqslant d}(\Sigma^n) \cap V_{\leqslant d-1}(\Sigma^n)^\perp.$$

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

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

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

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

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.

Definition 3.3. *The level d-weight of a function f is defined as $W^{=d}[f] = \|f^{=d}\|_2^2$, and by Praseval's equality we have that $W^{=d}[f] = \sum_{|S|=d} \|f^{=S}\|_2^2$. We also denote $W^{\geqslant d}[f] = \sum_{i=d}^{n} W^{=i}[f]$.*

We will need the following lemma:

Lemma 3.4. *Suppose that we have functions $f_I \colon \Sigma^n \to \mathbb{C}$, where each f_I is an $\bar{I}$-junta, and define $f(x) = \mathbb{E}_{I \subseteq_\varepsilon [n]}[f_I(x)]$.*

1. *If each f_I is 1-bounded, then f is 1-bounded.*

2. *If $\|f_I\|_2 \leqslant 1$ for each I, then for each $d \geqslant 1$ we have that $W^{\geqslant d}[f] \leqslant (1 - \varepsilon)^d$.*

Proof. The first item is clear by the triangle inequality. For the second item, note that for each $S \subseteq [n]$ we have that $f^{=S}(x) = \mathbb{E}_{I \subseteq_\varepsilon [n]}\left[f_I^{=S}(x)\right]$. As f_I is an $\bar{I}$-junta, we have that $f_I^{=S} \equiv 0$ unless $S \subseteq \bar{I}$, and so $f^{=S}(x) = \mathbb{E}_{I \subseteq_\varepsilon [n]}\left[f_I^{=S}(x) 1_{S \subseteq \bar{I}}\right]$. It follows from the Cauchy-Schwarz inequality that

$$\|f^{=S}\|_2^2 = \left| \mathbb{E}_x \left[\mathbb{E}_{I \subseteq_\varepsilon [n]} \left[f_I^{=S}(x) 1_{S \subseteq \bar{I}} \right] \right] \right|^2 \leqslant \mathbb{E}_x \left[\mathbb{E}_{I \subseteq_\varepsilon [n]} \left[\left| f_I^{=S}(x) \right|^2 \right] \mathbb{E}_{I \subseteq_\varepsilon [n]} \left[1_{S \subseteq \bar{I}} \right] \right]$$

$$= (1 - \varepsilon)^{|S|} \mathbb{E}_{I \subseteq_\varepsilon [n]} \left[\|f_I^{=S}\|_2^2 \right].$$

Thus,

$$W^{\geqslant d}[f] = \sum_{|S| \geqslant d} \|f^{=S}\|_2^2 \leqslant \sum_{|S| \geqslant d} (1 - \varepsilon)^{|S|} \mathbb{E}_{I \subseteq_\varepsilon [n]} \left[\|f_I^{=S}\|_2^2 \right] \leqslant (1 - \varepsilon)^d \mathbb{E}_{I \subseteq_\varepsilon [n]} \left[\sum_{|S| \geqslant d} \|f_I^{=S}\|_2^2 \right],$$

which by Parseval is at most $(1 - \varepsilon)^d \mathbb{E}_{I \subseteq_\varepsilon [n]}\left[\|f_I\|_2^2\right] \leqslant (1 - \varepsilon)^d$. $\qquad\square$

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 $f(x) = \sum_{\vec{\alpha} \in [m]^n} \widehat{f}(\vec{\alpha}) v_{\vec{\alpha}}(x)$ 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 the functions $v_1, \ldots, v_{m-1}$ are not unique. Nevertheless, it will be useful for in the presentation of the invariance principle [43] in Section 3.2.

3.1.4 Hypercontractivity

Definition 3.5. *For $q \geqslant 2$ and $f \colon (\Sigma^n, \nu^n) \to \mathbb{C}$, we define the q-norm of f as $\|f\|_q = (\mathbb{E}_{x \sim \nu^n} [|f(x)|^q])^{1/q}$.*

Theorem 3.6. *[[44, Chapter 10]] 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 3.7. *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 $\mathrm{T}_\rho f(x) = \mathbb{E}_{y \sim \mathrm{T}_\rho x} [f(y)]$. We will also define an operator T_I for each $I \subseteq [n]$, as follows:

Definition 3.8. *For a subset $I \subseteq [n]$, and for $x \in \Sigma^n$, we define the distribution of $\mathrm{T}_I x$ as: to sample $y \sim \mathrm{T}_I x$, set $y_i = x_i$ for $i \notin I$, and $y_i \sim \nu$ independently for each $i \in I$. Abusing notation, we will also think of T_I as an operator $\mathrm{T}_I \colon L_2(\Sigma^n, \nu^n) \to L_2(\Sigma^n, \nu^n)$ defined by $\mathrm{T}_I f(x) = \mathbb{E}_{y \sim \mathrm{T}_I x} [f(y)]$.*

3.1.6 Influences and Low-degree Influences

Next, we define the notions of influences and low-degree influences.

Definition 3.9. *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*

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

Definition 3.10. *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 $I_i^{\leqslant d}[f] = I_i[f^{\leqslant d}]$.*

3.2 The Invariance Principle

In this section, we present the invariance principle of [43], 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 $\mathcal{X}$. 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 [42] 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

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

where the coefficients m_T are complex numbers and $a_{i,j}$ are thought of as complex-valued inputs.

Definition 3.11. *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 3.12. *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 $\varrho\colon \mathbb{C}^k \to \mathbb{R}$ by $\varrho(a_1, \ldots, a_k) = \sqrt{\sum_{i=1}^{k} |\mathsf{trunc}(a_i) - a_i|^2}$. Then*

$$\left| \mathbb{E}\left[\varrho(M(\mathcal{X}^n))\right] - \mathbb{E}\left[\varrho(M(\mathcal{G}^n))\right] \right| \leqslant \varepsilon.$$

Last, we need the following elementary fact.

Fact 3.13. *Consider the function $\varrho\colon \mathbb{C} \to [0, \infty)$ defined as $\varrho(a) = |\mathsf{trunc}(a) - a|$.*

1. trunc *is a 2-Lipschitz function.*

2. ϱ *is 3-Lipschitz function.*

Proof. Let $a, b \in \mathbb{C}$; we show that $|\varrho(a) - \varrho(b)| \leqslant 2 |a - b|$. 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
$$|\varrho(a) - \varrho(b)| = |\varrho(b)| = |b| - 1 \leqslant |b| - |a| \leqslant |b - a|,$$
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, first by the triangle inequality $|\varrho(a) - \varrho(b)| \leqslant |\mathsf{trunc}(a) - a - (\mathsf{trunc}(b) - b)| \leqslant |\mathsf{trunc}(a) - \mathsf{trunc}(b)| + |a - b|$, and we next analyze trunc. We have
$$|\mathsf{trunc}(a) - \mathsf{trunc}(b)| = \frac{|a\,|b| - b\,|a||}{|a|\,|b|} \leqslant \frac{|a\,|b| - a\,|a|| + |a\,|a| - b\,|a||}{|a|\,|b|} = \frac{||b| - |a||}{|b|} + \frac{||a - b||}{|b|}$$

which is at most $2|a - b|$, and the proof is concluded. $\qquad\square$

3.3 The μ-norm and the CSP Stability Result

Definition 3.14. *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*
$$\|f\|_\mu = \sup_{\substack{g\colon \Gamma^n \to \mathbb{C} \\ h\colon \Phi^n \to \mathbb{C} \\ \text{1-bounded}}} \left| \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)g(y)h(z)] \right|.$$

For general distributions μ, $\|\circ\|_\mu$ 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 distributions 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 3.15. *For alphabets Σ, Γ and Φ, a distribution ν over Σ and a parameter $\alpha > 0$, define the collections*
$$M_\nu = \left\{ \mu \mid \text{pairwise connected distribution over } \Sigma \times \Gamma \times \Phi \text{ with no } (\mathbb{Z}, +)\text{-embedding}, \mu_x = \nu \right\},$$
$$M_\alpha = \left\{ \mu \mid \mu(x, y, z) \geqslant \alpha \ \forall (x, y, z) \in \mathsf{supp}(\mu) \right\},$$
and $M_{\nu,\alpha} = M_\nu \cap M_\alpha$.

Definition 3.16. *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 3.17. *The results below apply to the more general notion of semi-norm $\|\circ\|_{M,\nu}$ with suitable adaptations. 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 $\|\circ\|_{\nu,\alpha}$.*

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 3.18. *For all $m \in \mathbb{N}$ there is an Abelian group 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. Then there exists an embedding (σ, γ, ϕ) of $\mathsf{supp}(\mu)$ into G, such that if $f \colon \Sigma^n \to \mathbb{C}$ is a 1-bounded function with $\|f\|_\mu \geqslant \varepsilon$, 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*

$$\left| \langle f, L \cdot \chi \circ \sigma^{\otimes n} \rangle \right| \geqslant \delta.$$

Throughout, a function of the type $\chi \circ \sigma^{\otimes n}$ will be referred to as a character function or a character embedding function, and we will often simply write such a function as $\chi \circ \sigma$.

3.4 Product Functions

In this section we introduce some useful notions about embedding functions arising in Theorem 3.18.

Definition 3.19. *Let Σ be an alphabet, G be a group and $\sigma \colon \Sigma \to G$ be a map. The class of product functions $\mathcal{P}(\Sigma, G, \sigma)$ is defined as the collection of functions of the form $P \colon \Sigma^n \to \mathbb{C}$ for which there are $u_1, \ldots, u_n \in \hat{G}$ and $\theta \in \mathbb{C}$ of a root of unity of order at most $|G|$ such that for all $x \in \Sigma^n$, $P(x) = \theta \prod_{i=1}^n u_i(\sigma(x_i))$. In the case that $\theta = 1$ we also refer to P as a character product function.*

We will often refer to functions as in Definition 3.19 as product functions, often without specifying the group G or the map σ. The product function we deal with will often in fact be character product functions.

Definition 3.20. *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 3.21. *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 $u_1, \ldots, u_n, v_1, \ldots, v_n \in \hat{G}$, for which $u_i = v_i$ for all but k of the indices $i \in [n]$, and $\theta, \theta' \in \mathbb{C}$ of absolute value 1 such that*

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

Definition 3.22. *For a set $\mathcal{P} = \{P_1, \ldots, P_r \colon \Sigma^n \to \mathbb{C}\} \subseteq P(\Sigma, G, \sigma)$, we define*

$$\mathsf{spn}_{\mathbb{N}}(\mathcal{P}) = \left\{ \prod_{i=1}^r P_i^{\alpha_i} \;\middle|\; \alpha_i \in \mathbb{N}, 0 \leqslant \alpha_i \leqslant |G| \; \forall i \right\}.$$

Definition 3.23. *The rank of $\mathcal{P} \subseteq \mathcal{P}(\Sigma, G, \sigma)$ is defined as $\mathsf{rk}(\mathcal{P}) = \min_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}), P \neq 1} \Delta_{\mathsf{symbolic}}(P, 1)$.*

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

Lemma 3.24. *Suppose that $P\colon \Sigma^n \to \mathbb{C}$ is a product function from $\mathcal{P}(\Sigma, G, \sigma)$ such that $\Delta_{\mathsf{symbolic}}(P, 1) \geqslant 2$, suppose that $\mathcal{P}(\Sigma, G, \sigma)$ is τ-separated and let ν be a distribution whose support is Σ and in which the probability of each atom is at least α. Then for all $c > 0$ and $\xi \in [c, 1]$ we have that*

$$\|\mathrm{T}_{1-\xi}P\|_2 \leqslant 2^{-\Omega_{\alpha,\tau,c}(\Delta_{\mathsf{symbolic}}(P,1))},$$

where the 2-norm and the operator $\mathrm{T}_{1-\xi}$ are with respect to ν^n.

Proof. By definition, we may write $P(x) = \prod_{i=1}^{n} u_i(\sigma(x_i))$ where for at least $k := \Delta_{\mathsf{symbolic}}(P, 1)$ many i's we have that $u_i \circ \sigma$ is not constantly 1. By minimality, it follows that for at least $k - 1$ of i's we have that $u_i \circ \sigma$ is not constant, so by separatedness it has variance at least $\Omega_{\alpha,\tau}(1)$. For each such i get that $\|\mathrm{T}_{1-\xi}u_i \circ \sigma\|_2 \leqslant 1 - \Omega_{\alpha,\tau}(\xi) \leqslant 1 - \Omega_{\alpha,\tau,\xi}(1)$, and the result follows as $\|\mathrm{T}_{1-\xi}P\|_2 = \prod_{i=1}^{n} \|\mathrm{T}_{1-\xi}u_i \circ \sigma\|_2$. $\quad\square$

Observe that if $\mathsf{rk}(\mathcal{P})$ is large, then the above lemma (when $\xi = 1$) implies that for every $1 \neq P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$, the average of P is negligible.

4 The Noise Operator and Its Properties

In this section we introduce a variant of the standard noise operator over product spaces that will be crucial for our applications.

Definition 4.1. *Let Σ be a finite alphabet, let ν be a distribution over Σ, let $\mathcal{P} = \{P_1, \ldots, P_r \colon \Sigma^n \to \mathbb{C}\}$ be a collection of functions and let $I \subseteq [n]$. For each $x \in \Sigma^n$ we define the distribution $\mathrm{T}_{\nu,\mathcal{P},I}x$ as:*

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

2. *Output y.*

For $\varepsilon > 0$, we define the distribution of $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}x$ by sampling $I \subseteq_{\varepsilon} [n]$, and then outputting $y \sim \mathrm{T}_{\nu,\mathcal{P},I}x$.

As usual, we will associated with $\mathrm{T}_{\nu,\mathcal{P},I}, \mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$ averaging operators over functions, which by abuse of notation we denote as $\mathrm{T}_{\nu,\mathcal{P},I}, \mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}\colon L_2(\Sigma^n, \nu^{\otimes n}) \to L_2(\Sigma^n, \nu^{\otimes n})$. We define

$$\mathrm{T}_{\nu,\mathcal{P},I}f(x) = \underset{y \sim \mathrm{T}_{\nu,\mathcal{P},I}x}{\mathbb{E}}[f(y)], \qquad \mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}f(x) = \underset{I \subseteq_{\varepsilon}[n]}{\mathbb{E}}\left[\underset{y \sim \mathrm{T}_{\nu,\mathcal{P},I}x}{\mathbb{E}}[f(y)]\right] = \underset{I \subseteq_{\varepsilon}[n]}{\mathbb{E}}[\mathrm{T}_{\nu,\mathcal{P},I}f(x)].$$

We will use noise operators as a replacement for harsh truncations: as we shall see in L_2 it acts a truncation-like yet operator, which in addition preserves the boundedness of functions. The distribution ν will sometimes be omitted from the notation and will be clear from context.

4.1 Some Basic Properties of the Noise Operator

The goal of this section is to prove a few basic properties of the noise operator $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$.

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

We first prove 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.[11]

Fact 4.2. *For all $I \subseteq [n]$ and $\mathcal{P}$, $\nu^{\otimes n}$ is a stationary distribution of $\mathrm{T}_{\nu,\mathcal{P},I}$. Consequently $\nu^{\otimes n}$ is a stationary distribution of $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$.*

Proof. Fix I, write $\mathcal{P} = \{P_1, \ldots, P_r\}$ and fix $w \in \Sigma^n$. We calculate the probability that sampling $x \sim \nu^{\otimes n}$ and $y \sim \mathrm{T}_{\nu,\mathcal{P},I}x$ we have $y = w$. Denoting $a_i = P_i(w)$, we have

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

Let A be the set of $x \in \Sigma^n$ such that $P_i(x) = a_i$ for all i and $x_I = w_I$, so that the first term is equal to $\nu^{\otimes n}(A)$. By definition of $\mathrm{T}_{\nu,\mathcal{P},I}$, y is distributed according to $y \sim \nu^{\otimes n}$ conditioned on $y \in A$, so the second term is equal to $\nu^{\otimes n}(w)/\nu^{\otimes n}(A)$. Together we get $\Pr_{x,y}[y = w] = \nu^{\otimes n}(w)$. $\qquad\square$

4.1.2 Relating Different Noise Operators

We next 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 4.3. *Let $\mathcal{P}$ be a collection of product functions, let $\varepsilon > 0$, let $P' \colon \Sigma^n \to \mathbb{C}$ be a product function 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 $f \colon \Sigma^n \to \mathbb{C}$, $\|\mathrm{T}_{\nu,\mathcal{P}\cup\{P'\},1-\varepsilon}f - \mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}f\|_2 \leqslant 2\sqrt{k\varepsilon}$.*

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. We note that if $K \subseteq \bar{I}$, then $\mathrm{T}_{\nu,\mathcal{P}\cup\{P'\},I} = \mathrm{T}_{\nu,\mathcal{P},I}$, and as $\Pr_{I \subseteq_\varepsilon [n]}[K \subseteq \bar{I}] \geqslant 1 - k\varepsilon$, we get that for each x the statistical distance between $\mathrm{T}_{\nu,\mathcal{P}\cup\{P'\},1-\varepsilon}x$ and $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}x$ is at most $k\varepsilon$, and the first item follows. For the second item, fix the coupling (x, y, y') so that we may write

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

4.1.3 Nearly Preserving Low Degree Functions

We would like to argue that $\mathrm{T}_{\mathcal{P},1-\varepsilon}$ roughly preserves low-degree functions. Toward this end, for each $I \subseteq [n]$ we introduce the function

$$A_{\mathcal{P},I}(x) = \Pr_{x' \sim \mathrm{T}_I x}\left[P_i(x') = P_i(x) \,\forall i\right].$$

We have the following fact:

---

[11] We remark that whenever the set $\mathcal{P}$ is non-empty, the Markov chain $\mathrm{T}_{\nu,\mathcal{P},1-\varepsilon}$ is disconnected, so it has multiple stationary distributions. This will not be important for us.

Fact 4.4. *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 $I \subseteq [n]$ and $\tau > 0$ we have that*

$$\Pr_{x \sim \nu^{\otimes n}} [A_{\mathcal{P},I}(x) \leqslant \tau] \leqslant O_{r,m}(\tau).$$

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

$$B_{\vec{a}} = \{ x \in \Sigma^n \mid P_i(x) = a_i \; \forall i = 1, \ldots, r \}, \qquad B'_{\vec{a}} = \{ x \in B_{\vec{a}} \mid A_{\mathcal{P},I}(x) \leqslant \tau \}.$$

Note that for all $\vec{a}$,

$$\Pr_{x \sim \nu^{\otimes n}, y \sim \mathrm{T}_I x} [x, y \in B'_{\vec{a}}] = \langle 1_{B'_{\vec{a}}}, \mathrm{T}_I 1_{B'_{\vec{a}}} \rangle = \|\mathrm{T}_I 1_{B'_{\vec{a}}}\|_2^2 \geqslant \|\mathrm{T}_I 1_{B'_{\vec{a}}}\|_1^2 = \nu^{\otimes n}(B'_{\vec{a}})^2.$$

In the second transition we used the fact that T_I is self adjoint and $\mathrm{T}_I^2 = \mathrm{T}_I$. On the other hand,

$$\Pr_{x \sim \nu^{\otimes n}, y \sim \mathrm{T}_I x} [x, y \in B'_{\vec{a}}] = \nu^{\otimes n}(B'_{\vec{a}}) \cdot \mathbb{E}_{x \sim \nu^{\otimes n}} [A_{\mathcal{P},I}(x) \mid x \in B'_{\vec{a}}] \leqslant \nu^{\otimes n}(B'_{\vec{a}})\tau.$$

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

$$\Pr_{x \sim \nu^{\otimes n}} [A_{\mathcal{P},I}(x) \leqslant \tau] \leqslant \nu^{\otimes n}(X_{\mathsf{bad}}) \lesssim_{r,m} \tau. \qquad \square$$

Claim 4.5. *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 $\mathcal{P} \subseteq \mathcal{P}(\Sigma, G, \sigma)$ be a collection of size at most r, size of G at most m, and $\sigma \colon \Sigma \to G$ is some map. 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

$$\mathbb{E}_{x \sim \nu^{\otimes n}} \left[\left| \mathbb{E}_{y \sim \mathrm{T}_{\mathcal{P},1-\varepsilon} x} [L(x) - L(y)] \right|^2 \right] \leqslant \mathbb{E}_{x \sim \nu^{\otimes n}, I \subseteq_\varepsilon [n]} \left[\mathbb{E}_{y \sim \mathrm{T}_{\mathcal{P},I} x} \left[|L(x) - L(y)|^2 \right] \right]. \tag{13}$$

Let $\tau > 0$ be a parameter to be chosen. We break right-hand side of (13) to $(I) + (II)$ where:

$$(I) = \mathbb{E}_{\substack{x \sim \nu^{\otimes n} \\ I \subseteq_\varepsilon [n]}} \left[\mathbb{E}_{y \sim \mathrm{T}_{\mathcal{P},I} x} \left[|L(x) - L(y)|^2 \right] 1_{A_{\mathcal{P},I}(x) < \tau} \right],$$

$$(II) = \mathbb{E}_{\substack{x \sim \nu^{\otimes n} \\ I \subseteq_\varepsilon [n]}} \left[\mathbb{E}_{y \sim \mathrm{T}_{\mathcal{P},I} x} \left[|L(x) - L(y)|^2 \right] 1_{A_{\mathcal{P},I}(x) \geqslant \tau} \right].$$

27

For (I), using Cauchy-Schwarz twice we have

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

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

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

where we used Fact 4.2 in the last inequality. By hypercontractivity $\|L\|_4^2 \lesssim_{\alpha,d} \|L\|_2^2 = 1$, and combining with Fact 4.4 we conclude that $(I) \lesssim_{\alpha,m,d} \sqrt{\tau}$. For (II), we have that

$$(II) = \mathop{\mathbb{E}}_{\substack{x \sim \nu^{\otimes n} \\ I \subseteq_{\varepsilon} [n]}} \left[\frac{\mathbb{E}_{y \sim \mathrm{T}_I x} \left[|L(y) - L(x)|^2 \, \mathbf{1}_{P_i(y) = P_i(x) \, \forall i} \right]}{\mathcal{A}_{\mathcal{P},I}(x)} \mathbf{1}_{A_{\mathcal{P},I}(x) \geqslant \tau} \right],$$

which is at most

$$\frac{1}{\tau} \mathop{\mathbb{E}}_{\substack{x \sim \nu^{\otimes n} \\ I \subseteq_{\varepsilon} [n]}} \left[\mathop{\mathbb{E}}_{y \sim \mathrm{T}_I x} \left[|L(y) - L(x)|^2 \right] \right] = \frac{1}{\tau} \left(2 - 2\langle L, \mathrm{T}_{1-\varepsilon} L \rangle \right) \leqslant \frac{2}{\tau} \left(1 - (1 - \varepsilon)^d \right),$$

which is at most $\frac{2d}{\tau} \varepsilon$. Combining, we get (13) $\lesssim_{\alpha,m,d} \sqrt{\tau} + \frac{\varepsilon}{\tau}$, and choosing $\tau = \varepsilon^{2/3}$ finishes the proof. $\square$

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

Note that any function $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$ is an eigenvector of $T_{\mathcal{P},1-\varepsilon}$ of eigenvalue 1. Thus, by Claim 4.5 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 by a linear combination of such functions. This statement, which is Lemma 4.9 below, will be used in 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 $A_{\mathcal{P},I}(x)$. In Claim 4.4 we proved that its values are almost always positive numbers bounded away from 0. In the next fact, we show that $A_{\mathcal{P},I}(x)$ can be written as a linear combination of functions of the form $P \cdot L$ for bounded functions L which are $\bar{I}$-juntas.

Fact 4.6. *For all $r, m \in \mathbb{N}$, $I \subseteq [n]$ and any collection of r product functions $\mathcal{P} \subseteq \mathcal{P}(\Sigma, G, \sigma)$ with $|\Sigma| = m$ and $|G| = O_m(1)$, we may write*

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

where for every P the function $\tilde{L}_{I,P}$ is a $\bar{I}$-junta $O_{m,r}(1)$ bounded.

Proof. Let $\mathcal{P} = \{P_1, P_2, \ldots, P_r\}$. Note that as all of the functions P_i get the values of characters over an Abelian group of size at most $O_m(1)$, we have that in the set $\mathcal{I} = \bigcup_i \text{Image}(P_i)$ 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 $z_1 = z_2 \in \mathcal{I}$, $Q(z_1, z_2) = 0$ if $z_1, z_2 \in \mathcal{I}$ are distinct.[12] Using Q, we may write

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

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

$$A_{\mathcal{P},I}(x) = \sum_{j_1,\ldots,j_r,k_1,\ldots,k_r=0}^{d} a_{j_1,k_1} \cdots a_{j_r,k_r} P_1(x)^{k_1} \cdots P_r(x)^{k_r} \underset{y \sim \mathrm{T}_I x}{\mathbb{E}} \left[P_1(y)^{j_1} \cdots P_r(y)^{j_r} \right]$$

$$= \sum_{k_1,\ldots,k_r=0}^{d} P_1(x)^{k_1} \cdots P_r(x)^{k_r} \mathrm{T}_I \left(\sum_{j_1,\ldots,j_r=1}^{d} a_{j_1,k_1} \cdots a_{j_r,k_r} P_1^{j_1} \cdots P_r^{j_r} \right)(x).$$

$\square$

We will also want to show that the function $1/A_{\mathcal{P},I}(x)$ is close to being a linear combination of functions of the type $P \cdot L_I$, but some care is required for x such that $A_{\mathcal{P},I}(x)$ is close to 0. The following fact asserts that such approximation holds when we ignore this type of inputs.

Fact 4.7. *Consider the setup from Fact 4.6. For all $m, r \in \mathbb{N}$, $\tau, \xi > 0$ and $I \subseteq [n]$, there is $A_I'': \Sigma^n \to \mathbb{C}$ with the following properties:*

1. *A_I'' may be written as $A_I''(x) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})} P \cdot \tilde{L}_{I,P}$ where for all P, $\tilde{L}_{I,P}$ is a $\bar{I}$-junta which is $O_{m,\xi,\tau,r}(1)$-bounded.*

2. *$\mathbb{E}_x \left[\left| \frac{1}{A_{\mathcal{P},I}(x)} - A_I''(x) \right|^2 1_{A_{\mathcal{P},I}(x) \geqslant \tau} \right] \leqslant \xi.$*

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

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

where $|\text{err}(x)| \leqslant e^{-\tau T}$. We take $T = \log(2/\xi)/\tau$ and define $A_I''(x) = \sum_{k=0}^{T} (1 - A_{\mathcal{P},I}(x))^k$; then the second item is clear. For the first item, we use the formula for $A_{\mathcal{P},I}$ from Fact 4.6, expand A_I'' and re-group terms to get that A_I'' may be written as a linear combination of functions of the form $P \cdot \tilde{L}_{I,P,1} \cdots \tilde{L}_{I,P,T+1}$ where $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})$ and each $\tilde{L}_{I,P,j}$ is a $\bar{I}$-junta which is $O_{m,r}(1)$-bounded. In particular, we may write

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

where $\tilde{L}_{I,P}$ is a $\bar{I}$-junta which is $O_{m,T,r}(1) = O_{m,\xi,\tau,r}(1)$-bounded.

$\square$

---

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

We next give a similar approximate formula for $\mathrm{T}_{\nu,\mathcal{P},I}f$.

Lemma 4.8. *Fix $m, r \in \mathbb{N}$, $\alpha, \xi > 0$, suppose that Σ has size at most m, that 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 1-bounded function $f \colon \Sigma^n \to \mathbb{C}$ and $I \subseteq [n]$, there exists a function $f_I' \colon \Sigma^n \to \mathbb{C}$ such that:*

1. *The function f_I' approximates $\mathrm{T}_{\mathcal{P},I}f \colon \|\mathrm{T}_{\mathcal{P},I}f - f_I'\|_2 \leqslant \xi$.*

2. *The function f_I' can be written as $f_I'(x) = \sum\limits_{P \in \mathsf{spn}_\mathbb{N}(\mathcal{P})} P(x) \cdot L_{I,P}(x)$, where for all P, $L_{I,P}$ is a $\bar{I}$-junta and is $O_{m,r,\alpha,\xi}(1)$-bounded.*

Proof. Take the parameters
$$0 < \zeta \ll \tau \ll r^{-1}, m^{-1}, \alpha, \xi \leqslant 1.$$
Let $\tau' \in [\tau, \tau + \sqrt{\zeta}]$ be a parameter to be determined, define $F_1(x) = \mathrm{T}_{\mathcal{P},I}f(x)1_{A_{\mathcal{P},I}(x) \geqslant \tau'}$ and note that

$$\|\mathrm{T}_{\mathcal{P},I}f - F_1\|_2 = \|\mathrm{T}_{\mathcal{P},I}f(x)1_{A_{\mathcal{P},I}(x) \leqslant \tau}\|_2 \leqslant \|1_{A_{\mathcal{P},I}(x) \leqslant \tau'}\|_2 \lesssim_{m,r} \sqrt{\tau'} \leqslant \tau^{1/8}, \tag{14}$$

where we used Fact 4.4. Next, expanding the definition we get that

$$\mathrm{T}_{\mathcal{P},I}f(x) = \frac{1}{A_{\mathcal{P},I}(x)} \underbrace{\mathbb{E}_{x' \sim \mathrm{T}_I x} \left[f(x')1_{P_i(x')=P_i(x)\ \forall i} \right]}_{(I)}.$$

Take $A_I''(x)$ from Fact 4.7 with the parameter ξ therein being τ in the current setting. Then

$$\left\| A_I'' 1_{A_{\mathcal{P},I} \geqslant \tau'} - \frac{1}{A_{\mathcal{P},I}} 1_{A_{\mathcal{P},I} \geqslant \tau'} \right\|_2 \leqslant \tau.$$

Thus, as $(I)(x)$ is a 1-bounded function of x we conclude that for $F_2(x) = (I)(x)A_I''(x)1_{A_{\mathcal{P},I}(x) \geqslant \tau'}$

$$\|F_1 - F_2\|_2 \leqslant \tau. \tag{15}$$

Next, note that $\mathbb{E}_{x \sim \nu^{\otimes n}}\left[|I(x)|^2 |A_I''(x)|^2 \right] \leqslant O_\tau(1)$, therefore there exists $1 \leqslant j \leqslant 1/\sqrt{\zeta}$ such that

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

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

$$\|F_2 - F_3\|_2 \leqslant \sqrt{\mathbb{E}_x \left[(I)(x)^2 A_I''(x)^2 1_{A_{\mathcal{P},I}(x) \in [\tau + j\zeta, \tau + (j+1)\zeta]} \right]} \leqslant \zeta^{1/4}. \tag{16}$$

At this point, our approximating function F_3 almost fits the form as needed for f_I', but we still need to study the structure of $(I)(x)$ and of $h(A_{\mathcal{P},I}(x))$ further. For $h(A_{\mathcal{P},I}(x))$, by Weierstrass 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 $F_4(x) = (I)(x)A''(x)W(A_{\mathcal{P},I}(x))$ and get that

$$\|F_3 - F_4\|_2 \leqslant \zeta \sqrt{\mathbb{E}_x \left[(I)(x)^2 A_I''(x)^2 \right]} \leqslant \zeta \sqrt{\mathbb{E}_x \left[A_I''(x)^2 \right]} \leqslant O_\tau(\zeta) \leqslant \sqrt{\zeta}. \tag{17}$$

The Structure of (I). As in the proof of Fact 4.6, 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

$$(I)(x) = \mathop{\mathbb{E}}_{x' \sim \mathrm{T}_I x} \left[f(x') \prod_{i=1}^r Q(P_i(x'), P_i(x)) \right].$$

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

$$(I)(x) = \sum_{j_1, \dots, j_r = 0}^d P_1(x)^{j_1} \cdots P_r(x)^{j_r} \mathop{\mathbb{E}}_{x' \sim \mathrm{T}_I x} \left[\sum_{i_1, \dots, i_r = 0}^d a_{i_1, j_1} \cdots a_{i_r, j_r} f(x') P_1(x')^{i_1} \cdots P_r(x')^{i_r} \right]$$

$$= \sum_{j_1, \dots, j_r = 0}^d P_1(x)^{j_1} \cdots P_r(x)^{j_r} F_{j_1, \dots, j_r}(x),$$

where

$$F_{j_1, \dots, j_r}(x) = \mathrm{T}_I \left(\sum_{i_1, \dots, i_r = 0}^d a_{i_1, j_1} \cdots a_{i_r, j_r} f P_1^{i_1} \cdots P_r^{i_r} \right)(x)$$

$$= \mathop{\mathbb{E}}_{x' \sim \mathrm{T}_I x} \left[\sum_{i_1, \dots, i_r = 0}^d a_{i_1, j_1} \cdots a_{i_r, j_r} f(x') P_1^{i_1}(x') \cdots P_r^{i_r}(x') \right].$$

Note each $F_{j_1, \dots, j_r}$ is $O_{m,r}(1)$-bounded by the triangle inequality, and is also a $\bar{I}$-junta.

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

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

Plugging in the formula for $A_{\mathcal{P}, I}(x)$ from Fact 4.6, we get that $W(A_{\mathcal{P}, I}(x))$ may be written as a linear combination of at most $O_q(1)$ may terms each of the form $P \cdot \tilde{L}_{I,P,1}(x) \cdots \tilde{L}_{I,P,q}(x)$, where for all i, $\tilde{L}_{I,P,i}$ is a $\bar{I}$ junta which is $O_{m,r}(1)$-bounded. Re-arranging gives

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

where $\tilde{L}_{I,P}$ is a $\bar{I}$-junta which is $O_\zeta(1)$-bounded. By choice of A_I'' from Fact 4.7, we may write

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

where $\tilde{L}_{I,P}'$ is a $\bar{I}$-junta which is $O_\tau(1)$-bounded.

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

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

where each $\tilde{L}''_P$ is a linear combination of at most $O_{m,r}(1)$ products of at most 3 of the functions $F_{j_1,\ldots,j_r}$, $\tilde{L}_{I,Q'}$ and $\tilde{L}'_{I,Q''}$. Thus, $\tilde{L}''_P$ is a $\bar{I}$-junta which is $O_\zeta(1)$-bounded. Combining (14), (15), (16) and (17) gives the second item of the lemma. $\square$

Lemma 4.9. *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 any 1-bounded function $f \colon \Sigma^n \to \mathbb{C}$, 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*

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

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

Proof. For each $I \subseteq [n]$, apply Lemma 4.8 to get f'_I as there such that $\|\mathrm{T}_{\mathcal{P},I}f - f'_I\| \leqslant \xi/2$. Define $f'(x) = \mathbb{E}_{I \subseteq_\varepsilon [n]}[f'_I(x)]$. First, by the triangle inequality we have that

$$\|\mathrm{T}_{\mathcal{P},1-\varepsilon}f - f'\|_2 \leqslant \underset{I \subseteq_\varepsilon [n]}{\mathbb{E}} \left[\|\mathrm{T}_{\mathcal{P},I}f - f'_I\|_2\right] \leqslant \frac{\xi}{2}. \tag{18}$$

Next, write $f'_I(x) = \sum_{P \in \mathsf{spn}_\mathbb{N}(\mathcal{P})} P(x) \cdot L_{I,P}(x)$ as in Lemma 4.8. We get that

$$f'(x) = \sum_{P \in \mathsf{spn}_\mathbb{N}(\mathcal{P})} P(x) \underset{I \subseteq_\varepsilon [n]}{\mathbb{E}}[L_{I,P}(x)].$$

Defining $L_P(x) = \mathbb{E}_{I \subseteq_\varepsilon [n]}[L_{I,P}(x)]$, we have by Lemma 3.4 that $W^{\geqslant D}[L_P] \leqslant (1 - \varepsilon)^D \cdot C$, where $C = O_{m,r,\alpha,\xi}(1)$ is from Lemma 4.8. Taking $\tilde{L}_P = L_P^{\leqslant D}$ and $f''(x) = \sum_{P \in \mathsf{spn}_\mathbb{N}(\mathcal{P})} P(x)\tilde{L}_P(x)$, we get that

$$\|f' - f''\|_2 \leqslant \sum_{P \in \mathsf{spn}_\mathbb{N}(\mathcal{P})} \|L_P - \tilde{L}_P\|_2 \leqslant O_{m,r}((1 - \varepsilon)^{D/2}\sqrt{C}) \leqslant \frac{\xi}{2}$$

for sufficiently large D. Combining this with (18) we get that $\|\mathrm{T}_{\mathcal{P},1-\varepsilon}f - f''\|_2 \leqslant \xi$. As $\|\tilde{L}_P\|_2 \leqslant \|L_P\|_2 \leqslant C$, the statement follows. $\square$

5 The Regularity Lemma

In this section we prove a regularity lemma for the semi-norm $\| \circ \|_{\nu,\alpha}$.

5.1 A Basic Version of the Regularity Lemma

We begin with a basic version of our regularity lemma that already contains all of the essential ideas.

Lemma 5.1. *For all $\alpha > 0$, $m \in \mathbb{N}$ there is an Abelian group G of size $O_m(1)$ such that for all $\xi > 0$ there exist $\varepsilon > 0$ 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 $f : \Sigma^n \to \mathbb{C}$ be a 1-bounded function. Then there exist $\sigma : \Sigma \to G$ and a collection $\mathcal{P} \subseteq \mathcal{P}(\Sigma, G, \sigma)$ 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 5.1.

We look at all $\mu \in M_{\nu, \alpha}$ and apply Theorem 3.18 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 3.18 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 3.18 and σ to be the concatenation of all of σ's from the applications of Theorem 3.18. We fix these parameters, set $R = \lceil \frac{100}{\delta^2} \rceil$ and further use the following parameters:

$$0 < d_R^{-1} \ll \zeta_R \ll \varepsilon_R \ll \ldots \ll d_1^{-1} \ll \zeta_1 \ll \varepsilon_1 \ll d_0^{-1}, \delta \ll \alpha, \xi, m^{-1} \leqslant 1. \tag{19}$$

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 $L_i : \Sigma^n \to \mathbb{C}$ of 2-norm 1, we do the following:

1. Define $\mathcal{P}_{i+1} = \mathcal{P}_i \cup \{P_i\}$, $g_{i+1} = \mathrm{T}_{\mathcal{P}_{i+1}, 1-\varepsilon_{i+1}} f$ 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 5.1. Once it is established, the proof is concluded quickly.

Claim 5.2. *For all $i \leqslant R$ we have that $\|g_{i+1}\|_2^2 \geqslant i\delta^2/10$.*

Proof. Fix i, and for ease of notation denote $\mathrm{T}_i = \mathrm{T}_{\mathcal{P}_i, 1-\varepsilon_i}$. For $j = 1, \ldots, i+1$ we have that $|\langle f_j, P_j L_j \rangle| \geqslant \delta$, and we pick a complex number θ_j of absolute value 1 such that $\langle f_j, \theta_j P_j L_j \rangle \geqslant \delta$. 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

$$\langle f, \theta_j (I - \mathrm{T}_j) P_j L_j \rangle \geqslant \delta \qquad \forall j = 1, \ldots, i+1. \tag{20}$$

We next inspect that quantity $(I) = \langle \mathrm{T}_{i+2} f, \sum_{j=1}^{i+1} \theta_j (I - \mathrm{T}_j) P_j L_j \rangle$, 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 $(I) = \langle f, \sum_{j=1}^{i+1} \theta_j \mathrm{T}_{i+2} (I - \mathrm{T}_j) P_j L_j \rangle$, and we next argue that $\mathrm{T}_{i+2} (I - \mathrm{T}_j) P_j L_j$ is very close to $(I - \mathrm{T}_j) P_j L_j$ for each j. Clearly,

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

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 4.5 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 $\|\mathrm{T}_{i+2}P_jL_j - P_jL_j\|_2 \leqslant \varepsilon_{i+2}^{1/6}$. For the second term, namely for $\mathrm{T}_{i+2}\mathrm{T}_jP_jL_j$, first apply Lemma 4.9 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 4.5 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 into (I) gives

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

where in the last inequality we used (20). 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

$$\|\sum_{j=1}^{i+1}\theta_j(I - \mathrm{T}_j)P_jL_j\|_2^2 \leqslant \sum_j \|(I - \mathrm{T}_j)P_jL_j\|_2^2 + 2\sum_{j'<j}\left|\langle(I - \mathrm{T}_j)P_jL_j, (I - \mathrm{T}_{j'})P_{j'}L_{j'}\rangle\right|. \quad (21)$$

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 (21) gives that

$$(21) \leqslant 2(i+1) + (i+1)^2\varepsilon_1^{1/6} \leqslant 2(i+2),$$

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 $\|\mathrm{T}_{i+2}f\|_2^2 \geqslant i\delta^2/10$. $\qquad\square$

Using Claim 5.2, 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 3.18 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. $\qquad\square$

5.2 A Version of the Regularity Lemma Allowing Noise Modification

In this section, we state and prove a variant of Lemma 5.1 in which we have flexibility in picking the noise parameter ε. For the standard noise operator $T_{1-\varepsilon}$, if we have that $T_{1-\varepsilon}f$ is close to f in 2-norm, then $T_{1-\varepsilon'}f$ would be close to f for any $0 \leqslant \varepsilon' \leqslant \varepsilon$. With this analogy, one expects that the first item in Lemma 5.1 to hold not only for ε but rather for any $0 < \varepsilon' < \varepsilon$. While we do not know how to make such an argument go through, we show a different argument that essentially achieves this extra flexibility.

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 form $w(\varepsilon) = 2^{-1/\varepsilon}$ or a more rapidly decaying function.

Lemma 5.3. *For all $\alpha > 0$, $m \in \mathbb{N}$ there is an Abelian group G of size $O_m(1)$ such that for all $\xi > 0$ there exist $\varepsilon > 0$ and $r \in \mathbb{N}$ such that for any decay function $w\colon [0,1] \to [0,1]$ there is ε_0 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 exist $\sigma\colon \Sigma \to G$, a collection $\mathcal{P} \subseteq \mathcal{P}(\Sigma, G, \sigma)$ of size at most r and $\varepsilon \geqslant \varepsilon_0$ such that*

$$\|f - T_{\mathcal{P},1-\varepsilon'}f\|_{\nu,\alpha} \leqslant \xi$$

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

Proof. We run the same process as in the proof of Lemma 5.1. Once the process terminates, say at step i, we know that $(I - T_{\mathcal{P}_i,1-\varepsilon})f$ 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 3.18 it follows that $\|(I - T_{\mathcal{P}_i,1-\varepsilon})f\|_{\nu,\alpha} \leqslant \xi$ for all $\varepsilon \in (w(\varepsilon_i), \varepsilon_i)$. Otherwise, we may find $\varepsilon_i' \in (w(\varepsilon_i), \varepsilon_i)$ such that $(I - T_{\mathcal{P}_i,1-\varepsilon_i'})f$ 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 5.1, we conclude that this modified process also terminates within R steps, and the proof is concluded. $\square$

The analysis of our dictatorship test in Section 7.5 requires an analogous statement in the setting of multiple functions. More precisely, in this setting we have a collection of functions $f_{j,\ell}$ for $j = 1, \ldots, L$ and $\ell = 1, \ldots, q$, where for each fixed j the domain of $f_{j,\ell}$ is the same. We show that we can find collection of product functions and noise rates for all functions simultaneously, and furthermore that for each j, the collections for $f_{j,1}, f_{j,2}, \ldots, f_{j,q}$ are identical.

Lemma 5.4. *For all $\alpha > 0$, $L, q, m \in \mathbb{N}$ there is an Abelian group G of size $O_m(1)$ such that for all $\xi > 0$ there exist $\varepsilon > 0$ and $r \in \mathbb{N}$ such that for any decay function $w\colon [0,1] \to [0,1]$ there is ε_0 such that the following holds. Let $\Sigma_1, \ldots, \Sigma_L$ be alphabets of size at most m, let ν_j be a distribution over Σ_j in which the probability of each atom is at least α and let $f_{j,\ell}\colon (\Sigma_j^n, \nu_j^n) \to \mathbb{C}$ be a 1-bounded function. Then there exist $\sigma_j\colon \Sigma_j \to G$, a collection $\mathcal{P}_j \subseteq \mathcal{P}(\Sigma_j, G, \sigma_i)$ of size at most r and $\varepsilon \geqslant \varepsilon_0$ such that*

$$\|f_{j,\ell} - T_{\mathcal{P}_j,1-\varepsilon'}f_{j,\ell}\|_{\nu_j,\alpha} \leqslant \xi$$

for all $j = 1, \ldots, L$, $\ell = 1, \ldots, q$ and $w(\varepsilon) \leqslant \varepsilon' \leqslant \varepsilon$.

Proof. We run the following process. We start with $g_{j,\ell}^{(0)} = \mathbb{E}[f_{j,\ell}]$, $f_{j,\ell}^{(0)} = f_{j,\ell} - g_{j,\ell}^{(0)}$, $\mathcal{P}_j^{(0)} = \emptyset$ and $i_j = 0$ for every j and ℓ. If $f_{j,\ell}$ has correlation at least δ with a function of the form $P_{i_j}^{(j)} \cdot L_{i_j}^{(j)}$ for $P_{i_j}^{(j)} = \chi \circ \sigma_j$ and $L_{i_j}^{(j)}\colon \Sigma_j^n \to \mathbb{C}$ of 2-norm 1, we do the following:

1. Define $\mathcal{P}_j^{(i_j+1)} = \mathcal{P}_j^{(i_j)} \cup \{P_{i_j}^{(j)}\}$.

2. Define $g_{j,\ell}^{(i_j+1)} = \mathrm{T}_{\mathcal{P}_j^{(i_j+1)}, 1-\varepsilon_{i_j+1}} f$ and $f_{j,\ell}^{(i_j+1)} = f - g_{j,\ell}^{(i_j+1)}$.

3. Increase i_j by 1 and repeat.

Similar to the proof of Claim 5.2, we can show that $\sum_\ell \|g_{j,\ell}^{i_j+1}\|_2^2 \geqslant i_j \delta^2/2$, and we continue the argument as in the proof of Lemma 5.4 to establish that the process stops after performing at most qR steps for each j. When the process stops, we have $\|f_{j,\ell}^{(i_j)}\|_{\nu_j,\alpha} \leqslant \xi$ for all j and ℓ. $\qquad\square$

5.3 A Multi-Variate Regularity Lemma with High Rank

We finish this section off by presenting a strengthening of Lemma 5.4 in the setting of multi-variate distributions and multiple functions. We will use the following convenient notation:

Definition 5.5. *For a product function $P = P_1 \cdots P_n\colon \Sigma^n \to \mathbb{C}$ and a subset $I \subseteq [n]$, we denote by $P|_I\colon \Sigma^n \to \mathbb{C}$ defined as $P|_I(x) = \prod\limits_{i\in I} P_i(x_i)$.*

Lemma 5.6. *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, Γ, Φ be finite alphabets of size at most m and let μ be a distribution over $\Sigma \times \Gamma \times \Phi$ in which the probability of each atom is at least α. Then there exists a group G whose size depends only on m, such that for all 1-bounded functions $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$ and $h\colon \Phi^n \to \mathbb{C}$ there are collections $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ of embedding functions with $|\mathcal{P}|, |\mathcal{Q}|, |\mathcal{R}| \leqslant r$ and $\varepsilon \geqslant \varepsilon_0$ such that:*

1. *$\|f - \mathrm{T}_{\mathcal{P},1-\varepsilon'} f\|_{\mu_x,\alpha} \leqslant \xi$, $\|g - \mathrm{T}_{\mathcal{Q},1-\varepsilon'} g\|_{\mu_y,\alpha} \leqslant \xi$, and $\|h - \mathrm{T}_{\mathcal{R},1-\varepsilon'} h\|_{\mu_z,\alpha} \leqslant \xi$ for all $\varepsilon' \in (w(\varepsilon), \varepsilon)$.*

2. *Considering the collection $\mathcal{D} = \{D_P, D_Q, D_R \mid P \in \mathcal{P}, Q \in \mathcal{Q}, R \in \mathcal{R}\}$ of product functions over $\mathrm{supp}(\mu)$ where $D_P(x,y,z) = P(x)$, $D_Q(x,y,z) = Q(y)$ and $D_R(x,y,z) = R(z)$, we have that $\mathrm{rk}(\mathcal{D}) \geqslant 1/w(\varepsilon)$.*

Proof. Fix α, m, ξ, take r from Lemma 5.4 and let $w'\colon [0,1] \to [0,1]$ be a decay function to be determined (depending only on α, m, ξ, r and the decay function w). Applying Lemma 5.4 on f, g, h gives collections $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ and ε_0 satisfying the first item with w' in place of w, and we next explain how to get the other two items. Let $\varepsilon \geqslant \varepsilon_0$.

Note that the size of $S = \mathrm{spn}_{\mathbb{N}}(\mathcal{D})$ is $A = O_{m,\alpha,r,\xi}(1)$. For each $i = 1, \ldots, 6A - 1$ define $w_{i+1}(a) = 0.5w_i(w(a))$ and $w_1(a) = 0.5w(a)$. Then we pick the decay function $w' = w_{6A}$. For $i = 1, \ldots, 2A - 1$ consider the interval $I_i = [w_{3i}(\varepsilon)^{-1}, w_{3i+3}(\varepsilon)^{-1})$. Then we may find i such that $\Delta_{\mathsf{symbolic}}(D, 1) \notin I_i$ for all $D \in S$. Define $S_{\mathsf{bad}} \subseteq S$ as the collection of $D \in S$ such that $\Delta_{\mathsf{symbolic}}(D, 1) \leqslant w_{3i}(\varepsilon)^{-1}$, and let

$$J = \{i \in [n] \mid \exists D = D_1 \cdots D_n \in S_{\mathsf{bad}}, D_i \not\equiv 1\}.$$

Then $|J| \leqslant A \cdot w_{3i}(\varepsilon)^{-1}$. Define the collections

$$\mathcal{P}' = \{P|_{\bar{J}} \mid P \in \mathcal{P}\}, \quad \mathcal{Q}' = \{Q|_{\bar{J}} \mid Q \in \mathcal{Q}\}, \quad \mathcal{R}' = \{R|_{\bar{J}} \mid R \in \mathcal{R}\}, \quad \mathcal{D}' = \{D|_{\bar{J}} \mid D \in \mathcal{D}\}.$$

We note that each one of $\mathcal{P}', \mathcal{Q}', \mathcal{R}', \mathcal{D}'$ has rank

$$w_{3i+3}(\varepsilon)^{-1} - |J| \, w_{3i}(\varepsilon)^{-1} \geqslant \frac{1}{2} w_{3i+3}(\varepsilon)^{-1} \geqslant \frac{1}{w(w_{3i+2}(\varepsilon))},$$

so the second item holds. The next claim establishes the first item, thereby finishing the proof:

Claim 5.7. *For $\varepsilon' \in (w_{3i+2}(\varepsilon), w_{3i+1}(\varepsilon))$ we have $\|f - \mathrm{T}_{\mathcal{P}',1-\varepsilon'}f\|_{\mu_x,\alpha} \leqslant 2\eta$, $\|g - \mathrm{T}_{\mathcal{Q}',1-\varepsilon'}g\|_{\mu_y,\alpha} \leqslant 2\eta$ and $\|h - \mathrm{T}_{\mathcal{R}',1-\varepsilon'}h\|_{\mu_z,\alpha} \leqslant 2\eta$.*

Proof. We prove the inequality for f, and the other two follow in the same way. Recall that by choice of $\mathcal{P}$ we have $\|f - \mathrm{T}_{\mathcal{P},1-\varepsilon'}f\|_{\mu_x,\alpha} \leqslant \eta$, and we next argue that $\|\mathrm{T}_{\mathcal{P},1-\varepsilon'}f - \mathrm{T}_{\mathcal{P}',1-\varepsilon'}f\|_2 \leqslant 2|\mathcal{P}|\sqrt{|J|\varepsilon'}$. Indeed, write $\mathcal{P} = \{P_1,\ldots,P_r\}$, and for $0 \leqslant i \leqslant r$ define $\mathcal{P}(i) = \{P_1|_{\bar{J}},\ldots,P_i|_{\bar{J}}, P_{i+1},\ldots,P_r\}$. Then by the triangle inequality for each i, we have that

$$\|\mathrm{T}_{\mathcal{P}(i),1-\varepsilon'}f - \mathrm{T}_{\mathcal{P}(i+1),1-\varepsilon'}f\|_2 \leqslant \|\mathrm{T}_{\mathcal{P}(i),1-\varepsilon'}f - \mathrm{T}_{\mathcal{P}(i)\cup\{P_{i+1}|_{\bar{J}}\},1-\varepsilon'}f\|_2$$
$$+ \|\mathrm{T}_{\mathcal{P}(i)\cup\{P_{i+1}|_{\bar{J}}\},1-\varepsilon'}f - \mathrm{T}_{\mathcal{P}(i+1),1-\varepsilon'}f\|_2. \qquad (22)$$

Note that the first expression on the right-hand side is at most $\sqrt{\varepsilon'|J|}$ by Fact 4.3. As for the second expression, we note that $\mathcal{P}(i) \cup \{P_{i+1}|_{\bar{J}}\} = \mathcal{P}(i+1) \cup \{P_{i+1}\}$, so it is also at most $\sqrt{\varepsilon'|J|}$ by Fact 4.3. Summing up (22) over i and using the triangle inequality gives that $\|\mathrm{T}_{\mathcal{P},1-\varepsilon'}f - \mathrm{T}_{\mathcal{P}',1-\varepsilon'}f\|_2 \leqslant 2|\mathcal{P}|\sqrt{|J|\varepsilon'} \leqslant \eta$. Thus, $\|\mathrm{T}_{\mathcal{P},1-\varepsilon'}f - \mathrm{T}_{\mathcal{P}',1-\varepsilon'}f\|_{\mu_x,\alpha} \leqslant \eta$, and the result follows from the triangle inequality. $\qquad\square$

$\square$

6 Moving to the Mixed Space

In this section we prove Theorem 1.2, and we first show how to associate with a given product space (Σ^n, ν^n) a mixed space $((\mathbb{R}^{|\Sigma|-1})^n \times \Sigma^n, \gamma^n \times \nu^n)$, where γ is a jointed distribution over $|\Sigma| - 1$ correlated Gaussian random variables. We then show a multi-variate version of this association.

6.1 Univariate Decoupling

Suppose we have a 1-bounded function $f\colon (\Sigma^n, \nu) \to \mathbb{C}$ and we found a collection $\mathcal{P}$ satisfying the first and third item in Lemma 5.6. Namely:

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. $\mathrm{rk}(\mathcal{P}) \geqslant M$ where $M \gg 1/\eta, 1/\varepsilon, 1/\alpha, m$.

For simplicity of notation we denote $f' = \mathrm{T}_{\mathcal{P},1-\varepsilon}f$. By Lemma 4.9 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 $\|f' - \tilde{f}\|_2 \leqslant \zeta$, $\deg(L_P) \leqslant d = O_{m,\alpha,\eta,\zeta,\varepsilon}(1)$ and $\|L_P\|_2 \leqslant O_{m,\alpha,\eta,\zeta,\varepsilon}(1)$. We 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 $\tilde{f}_{\mathsf{decoupled}}(x, x')$ when x, x' are sampled from ν^n independently is close to the distribution of $\tilde{f}(X)$ when $X \sim \nu^n$. To see that, inspect the formula of $\tilde{f}(X)$, and note that the values of the low-degree functions L_P are mostly determined after we expose $1 - \delta/d$ of the coordinates of X. On the other hand, since $\mathcal{P}$ has a high rank, conditioning on exposing these coordinates of X, the distribution of $(P(X))_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P})}$ is still mostly the same. 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, hence it is close to what happens in $\tilde{f}_{\mathsf{decoupled}}$.

To prove this formally we start with a few auxiliary facts. The first of which asserts that given a high rank collection of product functions $\mathcal{P} = \{P_1, \ldots, P_r\}$, we have that for each $a_i \in \mathsf{Image}(P_i)$, the probability over $x \sim \nu^{\otimes n}$ that $P_i(x) = a_i$ for all i is either essentially 0, or else it is bounded away from 0.

Fact 6.1. *For all $m, r \in \mathbb{N}$, for sufficiently large $M \in \mathbb{N}$, if $\mathcal{P} = \{P_1, \ldots, P_r \colon \Sigma^n \to \mathbb{C}\} \subseteq \mathcal{P}(\Sigma, G, \sigma)$ has $|G|, |\Sigma| \leqslant m$ and $\mathsf{rk}(\mathcal{P}) \geqslant M$, and ν is a distribution whose support is Σ and in which the probability of each atom is at least α, then for all $a_i \in \mathsf{Image}(P_i)$, one of the following holds:*

1. $\Pr_{x \sim \nu^n} [P_i(x) = a_i \ \forall i] \leqslant 2^{-\Omega_{m, r, \alpha}(M)}$.

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

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

$$C \mathbb{E}_x \left[\prod_i \prod_{b_i \in \mathsf{Image}(P_i) \setminus \{a_i\}} (b_i - P_i(x)) \right],$$

where $C = \prod\limits_i \prod\limits_{b_i \in \mathsf{Image}(P_i) \setminus \{a_i\}} \frac{1}{b_i - a_i}$. To compute the expectation, we expand it out and get that

$$\mathbb{E}_x \left[\prod_i \prod_{b_i \in \mathsf{Image}(P_i) \setminus \{a_i\}} (b_i - P_i(x)) \right] = \sum_{\alpha_1, \ldots, \alpha_r} C(\alpha_1, \ldots, \alpha_r) \mathbb{E}_x \left[\prod_{i=1}^r P_i(x)^{\alpha_i} \right],$$

where each one of the coefficients $C(\alpha_1, \ldots, \alpha_r)$ is a signed sum of products of elements in the images of the functions P_i. Defining

$$\mathcal{A} = \left\{ (\alpha_1, \ldots, \alpha_r) \mid \prod_{i=1}^r P_i(x)^{\alpha_i} \not\equiv 1 \right\}, \qquad C' = \sum_{(\alpha_1, \ldots, \alpha_r) \notin \mathcal{A}} C(\alpha_1, \ldots, \alpha_r),$$

we get that $\Pr_{x \sim \nu^n} [P_i(x) = a_i \ \forall i] = C \cdot C' + C \sum\limits_{(\alpha_1, \ldots, \alpha_r) \notin \mathcal{A}} C(\alpha_1, \ldots, \alpha_r) \mathbb{E}_x \left[\prod_{i=1}^r P_i(x)^{\alpha_i} \right]$. Note that by the triangle inequality,

$$\left| C \sum_{(\alpha_1, \ldots, \alpha_r) \notin \mathcal{A}} C(\alpha_1, \ldots, \alpha_r) \mathbb{E}_x \left[\prod_{i=1}^r P_i(x)^{\alpha_i} \right] \right| \lesssim_{m, r} \max_{(\alpha_1, \ldots, \alpha_r) \notin \mathcal{A}} \left| \mathbb{E}_x \left[\prod_{i=1}^r P_i(x)^{\alpha_i} \right] \right| \lesssim_{m, r} 2^{-\Omega_{m, r, \alpha}(M)},$$

where the last inequality is by Lemma 3.24. As M is sufficiently large, we conclude that

$$\left| C \cdot C' \right| - 2^{-\Omega_{m, r}(M)} \leqslant \Pr_{x \sim \nu^n} [P_i(x) = a_i \ \forall i] \leqslant \left| C \cdot C' \right| + 2^{-\Omega_{m, r}(M)}.$$

38

Because the image of each P_i is a root of unity of order at most $|G| \leqslant m$, we conclude that $|C| \geqslant \Omega_{m,r}(1)$. As C' is a signed sum of at most r^r products, each product consisting of elements from images of the functions P_i, we get that either $C' = 0$ or else $|C'| \geqslant \Omega_{m,r}(1)$. In conclusion, it is either the case that $|C \cdot C'| = 0$, in which the first item holds, and otherwise $|C \cdot C'| \geqslant \Omega_{m,r}(1)$, in which case the second item holds. $\qquad\square$

Next, suppose that $\mathcal{P} = \{P_1, \ldots, P_r\}$ product functions, and for each i take $a_i \in \mathsf{Image}(P_i)$ and consider the set $S = \{x \in \Sigma^n \mid P_i(x) = a_i \; \forall i = 1, \ldots, r\}$. The next result shows that the function $\mathrm{T}_{1-\kappa} 1_S$ is close to a constant function.

Fact 6.2. *For all $r, m \in \mathbb{N}$, $\xi > 0$ the following holds for sufficiently large $M \in \mathbb{N}$. Suppose $\mathcal{P} = \{P_1, \ldots, P_r \colon \Sigma^n \to \mathbb{C}\} \subseteq \mathcal{P}(\Sigma, G, \sigma)$ where $|G|, |\Sigma| \leqslant m$ has $\mathsf{rk}(\mathcal{P}) \geqslant M$ and let $a_i \in \mathsf{Image}(P_i)$ for $i = 1, \ldots, r$. Then for*
$$S = \{x \in \Sigma^n \mid P_i(x) = a_i \; \forall i = 1, \ldots, r\}$$
we have $\|\mathrm{T}_{1-\kappa} 1_S - \nu^n(S)\|_2 \leqslant 2^{-\Omega_{m,\kappa,r}(M)}$.

Proof. Expanding 1_S as in the proof of Fact 6.1, we see that $1_S = A + \sum\limits_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}), P \neq 1} C(P) P$ where $C(p)$ are constants satisfying $|C(P)| = O_{m,r}(1)$. Taking expectation of both sides as in Fact 6.1 gives $\nu^n(S) = A + 2^{-\Omega_{m,r}(M)}$, so $|\nu^n(S) - A| \leqslant 2^{-\Omega_{m,r}(M)}$. We get from the triangle inequality that

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

where we used Lemma 3.24. $\qquad\square$

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 $r, m \in \mathbb{N}$, $\alpha, \kappa, \xi > 0$, let $M \in \mathbb{N}$ be sufficiently large and let $\mathcal{P} = \{P_1, \ldots, P_r \colon \Sigma^n \to \mathbb{C}\} \subseteq \mathcal{P}(\Sigma, G, \sigma)$ be a collection of product functions satisfying $|G|, |\Sigma| \leqslant m$ and $\mathsf{rk}(\mathcal{P}) \geqslant M$. Let ν be a distribution whose support is Σ in which the probability of each atom is at least α, and consider the joint distribution of (x, x') and X sampled as:*

1. *Sample $X \sim \nu^n$.*

2. *independently sample $x \sim \mathrm{T}_{\mathcal{P},0} X$ and $x' \sim \mathrm{T}_{1-\kappa} X$.*

Then then distribution of (x, x') is $2^{-\Omega_{r,m,\kappa}(M)}$ close to $\nu^n \times \nu^n$ in statistical distance.

Proof. Fix $w, w' \in \Sigma^n$ and define $q_{w,w'} = \mathrm{Pr}_{X,x,x'}[x = w, x' = w']$, $q_{w'|w} = \mathrm{Pr}_{X,x,x'}[x' = w' \mid x = w]$. By Fact 4.2 the marginal distribution of x is ν^n and so $q_{w,w'} = \nu^n(w) q_{w'|w}$. For $a_i \in \mathsf{Image}(P_i)$ define

$$S_{a_1, \ldots, a_r} = \{u \in \Sigma^n \mid P_i(u) = a_i \; \forall i\},$$

and let $S_w = S_{P_1(w),...,P_r(w)}$. By Fact 6.1, for each $\vec{a} = (a_1,\ldots,a_r)$ we either have that $\nu^n(S_{\vec{a}}) \leqslant 2^{-\Omega_{m,r,\alpha}(M)}$ and otherwise $\nu^n(S_{\vec{a}}) \geqslant \Omega_{m,r}(1)$, and we say w is bad if $\nu^n(S_w) \leqslant 2^{-\Omega_{m,r,\alpha}(M)}$. Note that

$$\sum_{\substack{w,w' \\ w \text{ is bad}}} q_{w,w'} = \sum_{w \text{ is bad}} \nu^n(w) = \sum_{\substack{\vec{a} \\ \nu^n(S_{\vec{a}})\leqslant 2^{-\Omega_{m,r,\alpha}(M)}}} \nu^n(S_{\vec{a}}) \leqslant 2^{-\Omega_{m,r,\alpha}(M)}, \tag{23}$$

where in the last inequality we used the fact that the total number of $\vec{a}$ is $O_{m,r}(1)$. Similarly,

$$\sum_{\substack{w,w' \\ w \text{ is bad}}} \nu^n(w)\nu^n(w') = \sum_{w \text{ is bad}} \nu^n(w) \leqslant 2^{-\Omega_{m,r,\alpha}(M)}, \tag{24}$$

We next analyze good w. Note that conditioned on $x = w$, the distribution of X is ν conditioned on S_w:

$$\Pr_{X,x,x'}[X = u \mid x = w] = \nu^n(u)\frac{1_{S_w}(u)}{\nu^n(S_w)}.$$

Thus,

$$q_{w'|w} = \sum_u \nu^n(u)\frac{1_{S_w}(u)}{\nu^n(S_w)}\Pr_{u'\sim T_{1-\kappa}u}[u' = w'] = \sum_u \nu^n(u)\frac{1_{S_w}(u)}{\nu^n(S_w)}T_{1-\kappa}1_{w'}(u),$$

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

$$q_{w'|w} = \langle\frac{1_{S_w}}{\nu^n(S_w)}, T_{1-\kappa}1_{w'}\rangle = \langle T_{1-\kappa}\frac{1_{S_w}}{\nu^n(S_w)}, 1_{w'}\rangle = \nu^n(w')T_{1-\kappa}\frac{1_{S_w}}{\nu(S_w)}(w'),$$

and so $q_{w,w'} = \nu^n(w)\nu^n(w')T_{1-\kappa}\frac{1_{S_w}}{\nu^n(S_w)}(w')$. Computing, we get that

$$\sum_{\substack{w,w' \\ w \text{ not bad}}} |q_{w,w'} - \nu^n(w)\nu^n(w')| = \sum_{\substack{w,w' \\ w \text{ not bad}}} \nu^n(w)\nu^n(w')\left|T_{1-\kappa}\frac{1_{s_w}}{\nu^n(S_w)}(w') - 1\right|$$

$$\lesssim_{m,r} \sum_{w,w'} \nu^n(w)\nu^n(w')\left|T_{1-\kappa}1_{S_w}(w') - \nu^n(S_w)\right|$$

$$= \mathop{\mathbb{E}}_{w\sim\nu^n}[\|T_{1-\kappa}1_{S_w} - \nu^n(S_w)\|_1],$$

which is at most $2^{-\Omega_{m,r,\alpha}(M)}$ by Fact 6.2. Combining this with (23) and (24) gives the lemma. $\qquad\square$

We can now state and prove the relation between the functions $\tilde{f}$ and $\tilde{f}_{\text{decoupled}}$.

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

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

Proof. Writing $\tilde{f}(x) = \sum_{P\in\text{spn}_\mathbb{N}(\mathcal{P})} P(x)L_P(x)$ where $\deg(L_P) \leqslant d$, we get that the left-hand side in the lemma is equal to (as $P(x) = P(X)$ for $x \sim T_{\mathcal{P},0}X$):

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

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

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

6.2 Multi-Variate Decoupling

In this section we generalize the results of the previous section to the multi-variate setting. In the multi-variate setting we have alphabets Σ, Γ, Φ each of size m and a distribution μ over $\Sigma \times \Gamma \times \Phi$. The distribution μ will not be fully supported over $\Sigma \times \Gamma \times \Phi$, and when we discuss product functions over $\Sigma \times \Gamma \times \Phi$ we always mean that their domain is only $\mathsf{supp}(\mu)$. The following result is the multi-variate analog of Lemma 6.3:

Lemma 6.5. *Let Σ, Γ, Φ be alphabets of size at most m, let μ be a distribution over $\Sigma \times \Gamma \times \Phi$ in which the probability of each atom is at least α, and let $\mathcal{D} \subseteq \mathcal{P}(\mathsf{supp}(\mu), G, \sigma)$ be a collection of product functions with $|G| \leqslant m$ and $\mathsf{rk}(\mathcal{D}) \geqslant M$. Consider the coupling $((x,y,z),(x',y',z'),(X,Y,Z))$ defined as follows:*

1. *Sample $(X,Y,Z) \sim \mu^{\otimes n}$.*

2. *Sample $(x,y,z) \sim \mathrm{T}_{\mathcal{D},0}(X,Y,Z)$.*

3. *Sample $(x',y',z') \sim \mathrm{T}_{1-\kappa}(X,Y,Z)$.*

Then then distribution of $((x',y',z'),(x,y,z))$ is $2^{-\Omega_{r,m,\kappa}(M)}$ close to $\mu^n \times \mu^n$ in statistical distance.

Proof. Immediate from Lemma 6.3. $\qquad \square$

We next state a version of Lemma 6.4 for the coupling in Lemma 6.5. For that, given a functions 1-bounded functions $f \colon (\Sigma^n, \mu_x^n) \to \mathbb{C}$, $g \colon (\Gamma^n, \mu_y^n) \to \mathbb{C}$ and $h \colon (\Phi^n, \mu_z^n) \to \mathbb{C}$ we find collections of product functions $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ as in Lemma 5.6, and then set up the functions $\tilde{f}, \tilde{f}_{\mathsf{decoupled}}$ and analogously $\tilde{g}, \tilde{g}_{\mathsf{decoupled}}$ and $\tilde{h}, \tilde{h}_{\mathsf{decoupled}}$ as in Section 6.1.

Lemma 6.6. *In the above setting, consider the joint distribution from Lemma 6.5. Then*

$$\mathbb{E}\left[\left| \tilde{f}_{\mathsf{decoupled}}(x,x') - \tilde{f}(X) \right|^2 \right], \mathbb{E}\left[\left| \tilde{g}_{\mathsf{decoupled}}(y,y') - \tilde{g}(Y) \right|^2 \right], \mathbb{E}\left[\left| \tilde{h}_{\mathsf{decoupled}}(z,z') - \tilde{h}(Z) \right|^2 \right] \lesssim_{d,m,r} \kappa.$$

Proof. Identical to the proof of Lemma 6.4, as the marginal distribution on (x, x', X) in the coupling from Lemma 6.5 is identical to the distribution on (x, x', X) in the coupling from Lemma 6.3. $\qquad \square$

One technical issue with the coupling is that it only guarantees closeness in statistical distance. To remedy that, we have the following result:

Lemma 6.7. *In the setting of Lemma 6.5 there is a coupling $((X',Y',Z'),(X'',Y'',Z''))$, (X,Y,Z) between $\mu^n \times \mu^n$ and μ^n such that*

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

and similarly for g and h.

Proof. Let $((x, y, z), (x', y', z'))$ and (X, Y, Z) be a coupling as in Lemma 6.5. As the statistical distance between (x, x') and $\nu^n \times \nu^n$ is at most $2^{-\Omega_{r,m,\kappa}(M)}$, we may couple $((x, y, z), (x', y', z'))$ with $((X', Y', Z'), (X'', Y'', Z''))$ distributed according to $\mu^n \times \mu^n$ such that

$$\Pr\left[(x, y, z, x', y', z') \neq (X', Y', Z', X'', Y'', Z'')\right] \leqslant 2^{-\Omega_{r,m,\kappa}(M)}.$$

We prove that the coupling $((X', Y', Z'), (X'', Y'', Z''))$ and (X, Y, Z) satisfies the lemma. Note that the left-hand side of the lemma is at most

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

The first expectation above is at most $O_{r,m,d}(\kappa)$ by Lemma 6.6. For the second expectation, let E be the event that $(x, y, z, x', y', z') \neq (X', Y', Z', X'', Y'', Z'')$. Then by Cauchy-Schwarz

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

$$\lesssim \mathbb{E}\left[\left|\tilde{f}_{\mathsf{decoupled}}(x, x')\right|^4 + \left|\tilde{f}_{\mathsf{decoupled}}(X', X'')\right|^4\right]\Pr[E].$$

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

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

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

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

6.3 The Mixed Invariance Principle

In this section, we prove the mixed invariance principle. The notion of shifted low-degree influences will be important for us:

Definition 6.8. *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

$$\left|\Psi(a, b, c) - \Psi(a', b', c')\right| \lesssim \sqrt{|a - a'|^2 + |b - b'|^2 + |c - c'|^2} \tag{25}$$

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 $h \colon (\Phi^n, \nu_z^n) \to \mathbb{C}$, collections $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ of product

functions, we define $f', \tilde{f}, \tilde{f}_{\text{decoupled}}$ as in Section 6.1 and similarly $g', \tilde{g}, \tilde{g}_{\text{decoupled}}$ and $h', \tilde{h}, \tilde{h}_{\text{decoupled}}$. 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

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

where $\text{trunc} \colon \mathbb{C} \to \mathbb{C}$ is as in Section 3.2. We are now ready to prove Theorem 1.2, formally stated below:

Theorem 6.9. *Let $m \in \mathbb{N}$, $\alpha > 0$, let Σ, Γ, Φ be alphabets of size at most m, and let μ be a distribution over $\Sigma \times \Gamma \times \Phi$ with no $\mathbb{Z}$-embeddings in which the probability of each atom is at least α. Then for every $\xi > 0$ there exist M and $\tau > 0$ such that the following holds. 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, consider collections $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ and the functions F, G, H defined as above, and suppose that*

1. $\tilde{f}, \tilde{g}, \tilde{h}$ *have τ-small shifted low-degree influences.*

2. *The collection $\mathcal{D}$ as defined in Lemma 5.6 satisfies that $\text{rk}(\mathcal{D}) \geqslant M$.*

Then for a smooth $\Psi \colon \mathbb{C}^3 \to \mathbb{C}$ satisfying (25) with bounded third order derivatives we have

$$\left| \mathop{\mathbb{E}}_{(x,y,z) \sim \mu^n} [\Psi(f(x), g(y), h(z))] - \mathop{\mathbb{E}}_{\substack{(x,y,z) \sim \mu^n \\ (G_x, G_y, G_z) \sim \mathcal{G}^n}} [\Psi(F(x, G_x), G(y, G_y), H(z, G_z))] \right| \leqslant \xi. \qquad (26)$$

Proof. We take the parameters:

$$0 < \tau, M^{-1} \ll \kappa \ll d^{-1} \ll \eta, r^{-1} \ll m^{-1}, \alpha, \xi < 1. \qquad (27)$$

Here, m, α, ξ are as in the statement, r is an upper bound on the size of the collections $\mathcal{P}, \mathcal{Q}, \mathcal{R}$. The parameter η is the L_2-closeness parameter between $f', \tilde{f}$ and similarly $g', \tilde{g}$ and the $\|\|_{\mu_x, \alpha}$-closeness parameter between f, f' and similarly g, g' and h, h'. The parameter d is the degree of the functions L_P, L_Q, L_R, κ is the parameter for the coupling, M is a lower bound on the rank of $\mathcal{D}$ and τ is a bound on the influences. It can be seen that the parameters can be taken as (27) using a suitable decay function in Lemma 5.6.

First, as $\|f - f'\|_{\mu_x, \alpha} \leqslant \eta$, $\|g - g'\|_{\mu_y, \alpha} \leqslant \eta$ and $\|h - h'\|_{\mu_z, \alpha} \leqslant \eta$, we get that

$$\left| \mathop{\mathbb{E}}_{(x,y,z) \sim \mu^n} [\Psi(f(x), g(y), h(z))] - \mathop{\mathbb{E}}_{(x,y,z) \sim \mu^n} [\Psi(f'(x), g'(y), h'(z))] \right| \lesssim \eta. \qquad (28)$$

By the smoothness of Ψ it follows that

$$\left| \mathop{\mathbb{E}}_{(x,y,z) \sim \mu^n} [\Psi(f'(x), g'(y), h'(z))] - \mathop{\mathbb{E}}_{(x,y,z) \sim \mu^n} [\Psi(\tilde{f}(x), \tilde{g}(y), \tilde{h}(z))] \right|$$
$$\lesssim \mathop{\mathbb{E}}_{(x,y,z) \sim \mu^n} \left[\sqrt{\left| f'(x) - \tilde{f}(x) \right|^2 + |g'(y) - \tilde{g}(y)|^2 + \left| h'(z) - \tilde{h}(z) \right|^2} \right]$$
$$\lesssim \eta. \qquad (29)$$

In the last transition, we first use Cauchy-Schwarz, and then the fact that f' and $\tilde{f}$ are η-close in ℓ_2 distance to get that $\mathbb{E}_{(x,y,z)\sim\mu^n}\left[\left|f'(x) - \tilde{f}(x)\right|^2\right] \leqslant \eta$. Similarly $g', \tilde{g}$ and $h', \tilde{h}$. Using the coupling from Lemma 6.7 and the smoothness of Ψ it follows that:

$$
\left| \underset{(X,Y,Z)}{\mathbb{E}} \left[\Psi(\tilde{f}(X), \tilde{g}(Y), \tilde{h}(Z)) \right] \right.
$$

$$
\left. - \underset{\substack{(X,Y,Z)\\(X',Y',Z')\\(Z'',Y'',Z'')}}{\mathbb{E}} \left[\Psi(\tilde{f}_{\text{decoupled}}(X', X''), \tilde{g}_{\text{decoupled}}(Y', Y''), \tilde{h}_{\text{decoupled}}(Z', Z'')) \right] \right|
$$

$$
\lesssim \underset{\substack{(X,Y,Z)\\(X',Y',Z')\\(X'',Y'',Z'')}}{\mathbb{E}} \left[\sqrt{|\Delta_1(X, X', X'')|^2 + |\Delta_2(Y, Y', Y'')|^2 + |\Delta_3(Z, Z', Z'')|^2} \right]. \tag{30}
$$

Here, $\Delta_1(X, X', X'') = \tilde{f}(X) - \tilde{f}_{\text{decoupled}}(X', X'')$ and Δ_2, Δ_3 are defined analogously for g and h. By Cauchy-Schwarz and Lemma 6.7, we get that (30) $\lesssim_{d,m,r} \kappa$, and so (30) $\leqslant \eta$. Denote

$$
(I) = \underset{\substack{(x,y,z)\sim\mu^n\\(x',y',z')\sim\mu^n}}{\mathbb{E}} \left[\Psi(\tilde{f}_{\text{decoupled}}(x', x), \tilde{g}_{\text{decoupled}}(y', y), \tilde{h}_{\text{decoupled}}(z', z)) \right],
$$

so that the second expectation on the left hand side of (30) is equal to (I). Fix x, y, z in (I), and note that the 2-norm of $\tilde{f}_{\text{decoupled}}(x', x)$ over the choice of x' is at most $O_{m,r,\eta,d}(1)$ and the influences are at most $O(\tau)$. We apply Theorem 3.12 and average over $(x, y, z) \sim \mu^n$, to get that, provided that τ is small enough

$$
\left| (I) - \underset{\substack{(x',y',z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}}{\mathbb{E}} \left[\Psi(\tilde{f}_{\text{decoupled}}(x', G_x), \tilde{g}_{\text{decoupled}}(y', G_y), \tilde{h}_{\text{decoupled}}(z', G_z)) \right] \right| \leqslant \eta. \tag{31}
$$

Next, by the smoothness of Ψ it follows that

$$
\left| \underset{\substack{(x',y',z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}}{\mathbb{E}} \left[\Psi(\tilde{f}_{\text{decoupled}}(x', G_x), \tilde{g}_{\text{decoupled}}(y', G_y), \tilde{h}_{\text{decoupled}}(z', G_z)) \right] \right.
$$

$$
\left. - \underset{\substack{(x',y',z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}}{\mathbb{E}} \left[\Psi(F(x', G_x), G(y', G_y), H(z', G_z)) \right] \right| \leqslant E, \tag{32}
$$

where

$$
E = \underset{\substack{(x',y',z')\sim\mu^n\\(G_x,G_y,G_z)\sim\mathcal{G}^n}}{\mathbb{E}} \left[\sqrt{\varrho(\tilde{f}_{\text{decoupled}}(x', G_x)) + \varrho(\tilde{g}_{\text{decoupled}}(y', G_y)) + \varrho(\tilde{h}_{\text{decoupled}}(z', G_z))} \right],
$$

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

Claim 6.10. $E \lesssim \sqrt{\eta}$.

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

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

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

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 3.12, provided that τ is small enough

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

By Fact 3.13 the function ϱ is $O(1)$-Lipschitz, and so

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

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

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

Applying Cauchy-Schwarz and Lemma 6.6 gives that $E_1 \lesssim \sqrt{\eta}$, as desired. □

Combining all of the inequalities (28), (29), (30), (31) and (32) and Claim 6.10 proves (26). □

7   Proof of Theorem 1.1

Following Raghavendra [48] (see also [7]), a typical way to design a dictatorship for an optimization problem $\mathcal{P}$ is based on considering its SDP relaxation and solving it. On satisfiable instances, an SDP solution gives to each constraint a local distribution on its satisfying assignments, which then can be used in constructing the queries of the dictatorship test. We note that even if the SDP value is 1, there is no guarantee that a given local distribution is fully supported on $P^{-1}(1)$. In particular, this local distribution may not be pairwise connectedness and could admit $\mathbb{Z}$-embeddings even if the predicate P does not. In this case we cannot apply the analytical lemma from [11], hence we cannot analyze the soundness of the dictatorship test.

We circumvent this issue by demonstrating the use of our mixed invariance principle on MILDLY-SYMMETRIC predicates. These predicates have sufficient symmetry to guarantee that any SDP solution with value 1 can be converted into another SDP solution with value 1 in which the support of all local distributions is pairwise connectedness and has no $\mathbb{Z}$-embeddings.

Notations: fix a collection $\mathcal{P}$ of MILDLY-SYMMETRIC predicates. An instance of Max-$\mathcal{P}$-CSP, $\Upsilon = (\mathcal{V}, \mathcal{C})$, consists of a variable set $\mathcal{V}$ which take values from Σ, and a distribution $\mathcal{C}$ on the constraint set. We 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 $P_C : \Sigma^3 \to \{0, 1\}$ from $\mathcal{P}$. An assignment (x, y, z) to the tuple $\mathcal{V}(C)$ satisfies the constraint C iff $(x, y, z) \in P_C^{-1}(1)$. For ease of notation, we often use C to refer to both the constraint and the underlying predicate P_C, and simply write $C(x, y, z) = 1$ or $(x, y, z) \in C^{-1}(1)$ if $P_C(x, y, z) = 1$. Given a set T, the set of all the distributions on T is denoted by $\blacktriangle(T)$.

7.1 The SDP Program

The basic semidefinite programming relaxation of an instance $\Upsilon = (\mathcal{V}, \mathcal{C})$ is given in Figure 2. The SDP formulation consists of vectors $\{\boldsymbol{b}_{i,a}\}_{i \in \mathcal{V}, a \in \Sigma}$, distributions $\{\mu_C\}_{C \in \mathtt{supp}(\mathcal{C})}$ over local assignments (i.e., on $\Sigma^{\mathcal{V}(C)}$) and a unit vector $\boldsymbol{b}_0$. We denote by $\mathbf{val}(\boldsymbol{V}, \boldsymbol{\mu})$ 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 $\mathrm{poly}(n, \log(1/\eta))$ (see, for instance [31]). We will ignore this issue of approximation and assume that the SDP can be solved optimally in polynomial time, and in Remark 7.13 we explain the modifications necessary to accommodate for this.

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

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

During the execution of our algorithm, we will modify the SDP by imposing additional conditions on the local distributions. A key feature of these modification will be that they preserve valid *integral solutions*:

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

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

along with $\boldsymbol{\mu}$ where for every $C \in \mathtt{supp}(\mathcal{C})$, $\mu_C(\alpha|_{\mathcal{V}(C)}) = 1$ 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.

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, we 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 $C \in \mathtt{supp}(\mathcal{C})$. Towards showing the optimality of our algorithm, we require that for every $C \in \mathtt{supp}(\mathcal{C})$, the support of μ_C is pairwise connected and has no $\mathbb{Z}$-embedding. This condition is easy to achieve if the predicates in the instance are MILDLY-SYMMETRIC and we will show this towards the end. For the following discussion, we assume that the SDP solution satisfies this condition.

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 $v \in \mathcal{V}$ 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 $[q] = \{1, 2, \ldots, q\}$ 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. For every constraint C such that $v \in \mathcal{V}(C)$, and every embedding $(\sigma_1, \sigma_2, \sigma_3, g)$ of $\mathrm{supp}(\mu_C)$ into an Abelian 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. Similarly to Step 2 we add all the linear combinations of the exponents of the rows.

This completes the informal description of the algorithm for computing M_v.

In the formal algorithm below, we also keep track of the group elements that are in the images of the embeddings. This is stored in the variable $\{r_\ell^v\}_{\ell \in \Sigma}$, where r_ℓ^v is the tuple of group elements that ℓ is mapped to under the various embeddings of v. The group G_{master}^v is the product of all the groups associated with the embeddings of the constraints C with $v \in \mathcal{V}(C)$. In a sense, this is a large group which encompasses all of the Abelian embeddings v may participate in. We denote by $H_v^\star$ a subgroup of G_{master}^v generated by $\{r_\ell^v\}_{\ell \in \Sigma}$. In a sense $H_v^\star$ is the sub-group generated by feasible values of v in the system of equations.

In the algorithm below, in steps 4 to 10 we find all Abelian embeddings of local distributions μ_C that a variable $v \in \mathcal{V}$ participates in. Formally speaking, these are embeddings and characters over the group G, but it will be convenient to think of them as characters of the larger group G_{master}^v. This is done by identifying a character $\chi \in \hat{G}$ in step 7 with the character $(1, \ldots, 1, \chi, 1, \ldots, 1) \in \widehat{G_{\mathrm{master}}^v}$.

Claim 7.3. *The columns of the matrix M_v contain evaluations of all the characters of G_{master}^v on a subgroup generated by $\{r_1^v, r_2^v, \ldots, r_q^v\}$.*

Proof. In steps 14-18, we add columns that correspond to every other character from $\widehat{G_{\mathrm{master}}^v}$ evaluated at $\{r_\ell^v\}_{\ell \in [q]}$. This follows as the character $\vec{\chi} := (\chi_1, \chi_2, \ldots, \chi_t) \in \widehat{G_{\mathrm{master}}^v}$ is, by definition, $\vec{\chi}(\boldsymbol{a}) = \prod_{i=1}^{t} \chi_i(a_i)$, and step 16 is precisely adding such columns.

Finally, in lines 20-24, we are adding rows that correspond to the evaluations of all the characters from $\widehat{G_{\mathrm{master}}^v}$ on every element of G_{master}^v generated by $\{r_1^v, r_2^v, \ldots, r_q^v\}$. This follows as $\vec{\chi} \in \widehat{G_{\mathrm{master}}^v}$ is a group homomorphism $\vec{\chi} : G_{\mathrm{master}}^v \to \mathbb{C}$, i.e., $\vec{\chi}(\boldsymbol{a}_1 + \boldsymbol{a}_2) = \vec{\chi}(\boldsymbol{a}_1) \cdot \vec{\chi}(\boldsymbol{a}_2)$ for every $\boldsymbol{a}_1, \boldsymbol{a}_2 \in G_{\mathrm{master}}^v$. $\qquad \square$

The variables in GE System associated to $v \in \mathcal{V}$ are $\{y_v^{\vec{\chi}}\}$ where $\vec{\chi} \in \widehat{G_{\mathrm{master}}^v}$. It would be convenient to think of the variable set $\{y_v^{\vec{\chi}}\}$ taking values that correspond to a row of the matrix M_v.

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 embedding $((\sigma_1, \sigma_2, \sigma_3), g)$ of $\mathsf{supp}(\mu_{C_i})$ 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 $\ell \in [q]$. $\triangleright r_\ell^v \in H$.
10:      **end for**
11: **end for**
12: Set $G_{\mathsf{master}}^v \leftarrow H$.
13: Let L' be the number of columns in M_v
14: **for** every subset $S \subseteq [L']$ **do** $\triangleright$ Adding more columns
15:      **if** $\circ_{i \in S} M_v[.][i]$ is not present as a column in M_v ($M_v[.][i]$ is the ith column of M_v and $\circ$ is pointwise multiplication) **then**
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** $\triangleright$ 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_{master}^v.
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**

7.2.2 Adding Equations to the GE System

Our next step is to set up a system of linear equations. Ideally, we would have liked to assign a group element from $\{r_1^v, r_2^v, \ldots, r_q^v\} \subseteq H_v^\star$ to a variable v (since they directly correspond to elements from Σ). However, we do not know how to find such an assignment by an efficient algorithm, let alone by a set of linear equations. We thus relax this requirement and settle for an assignment from $H_v^\star$. To do so we add to the linear system GE System the following two types of equations:

1. **Valid character constraints.** These equations enforce that the variables $\{y_v^{\vec{\chi}}\}_{\vec{\chi} \in \widehat{G_{\mathsf{master}}^v}}$ correspond to a row of the matrix M_v, and hence, the vector assignment corresponds to a group element from $H_v^\star$. 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 $\vec{\chi}, \vec{\chi}', \vec{\chi}''$ 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 $H_v^\star$ may be a proper subgroup of G_{master}^v, 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 $y_v^{\vec{\chi}} = c$.

48

2. **Valid satisfying assignments constraints.** These equations ensure that any solution to GE System is consistent with the image of the support of local distributions under various embeddings into Abelian groups. 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 the embedding $((\sigma_1, \sigma_2, \sigma_3), g)$ 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, stating that a valid integral solution to the SDP gives a solution to the GE System.

Observation 7.4. *Suppose $\boldsymbol{x} \in [q]^N$ 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 $\{y_v^{\vec{\chi}}\}$ 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 $\mathcal{T}$ to the GE System as a subset of $\prod_{v \in \mathcal{V}} H_v^{\star}$.

7.3 Getting Consistent SDP and GE System Solutions

Jumping ahead, in the soundness analysis of our dictatorship test we are interested in computing the expectation of the form

$$\mathbb{E}_{C(s_1, s_2, s_3) \sim \mathcal{C}} \left[\mathbb{E}_{(x,y,z) \sim \mu_C^R} \left[F_{s_1}(\sigma_1(x), x) \cdot F_{s_2}(\sigma_2(y), y) \cdot F_{s_3}(\sigma_3(z), z) \right] \right]$$

for functions for functions $F_s : (H_s^{\star})^R \times \Sigma^R \to \mathbb{C}$, where the first input goes to the product functions from the decomposition of F_{s_i}, and the second input goes to the low-degree functions. During the rounding procedure, we want to replace the local distribution $(\sigma_1(x), \sigma_2(y), \sigma_3(z))$ 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 show the optimality of the algorithm[13] is that when we make this switch, the embedding functions coming from the functions F_{s_i} should satisfy the following property.

- ◆ For every constraints C over (s_1, s_2, s_3) and tuple of embedding 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 μ_C iff $PQR \equiv 1$ under the distribution from GE System.

To guarantee this condition we define below a condition called *consistent solutions*, which can be achieved via our hybrid algorithm and is sufficient to guarantee the above condition.

For a variable $s \in \mathcal{V}$, let σ_s be the map $\sigma_s : \Sigma \to G_{\mathsf{master}}^s$ given by $\sigma_s(a) = r_a^s$, where r_a^s is the group element from Algorithm 1 associated with the variable s and $a \in \Sigma$. We now define the span of the local distribution μ_C under the embeddings $(\sigma_{s_1}, \sigma_{s_2}, \sigma_{s_3})$.

---

[13] Optimality with respect to the soundness of the dictatorship test.

Definition 7.5. *For a constraint $C \in \mathcal{C}$ with $\mathcal{V}(C) = (s_1, s_2, s_3)$ and the embedding maps σ_{s_i} to $G_{\text{master}}^{s_i}$ for $1 \leqslant i \leqslant 3$, define* $\text{span}(\text{supp}(\mu_C))$ *to be the subgroup of* $G_{\text{master}}^{s_1} \times G_{\text{master}}^{s_2} \times G_{\text{master}}^{s_3}$ *generated by* $\{(\sigma_{s_1}(a), \sigma_{s_2}(b), \sigma_{s_3}(c)) \mid (a, b, c) \in \text{supp}(\mu_C)\}$.

We are now ready to define the notion of consistent solutions.

Definition 7.6 (Consistent solutions). *The solution to the semidefinite program $(\boldsymbol{V}, \boldsymbol{\mu})$ is* consistent *with the solution space $\mathcal{T}$ to the* GE System *if the following condition holds: for every constraint $C \in \mathcal{C}$ with $\mathcal{V}(C) = (s_1, s_2, s_3)$, we have*

$$\text{span}(\text{supp}(\mu_C)) = \mathcal{T}|_C,$$

where $\mathcal{T}|_C := \{(a_{s_1}, a_{s_2}, a_{s_3}) \mid \boldsymbol{a} \in \mathcal{T}\}$.

The utility of consistent solutions is given by the following lemma, which is crucial in the soundness analysis of the dictatorship test to achieve the condition ♦ mentioned above.

Lemma 7.7. *Fix a constraint $C \in \text{supp}(\mathcal{C})$, let $\mathcal{V}(C) = (s_1, s_2, s_3)$ and let $\sigma_{s_1} : \Sigma \to G_{\text{master}}^{s_1}$, $\sigma_{s_2} : \Sigma \to G_{\text{master}}^{s_2}$ and $\sigma_{s_3} : \Sigma \to G_{\text{master}}^{s_3}$ be as above. For $\chi \in \widehat{G_{\text{master}}^{s_1}}$, $\chi' \in \widehat{G_{\text{master}}^{s_2}}$ and $\chi'' \in \widehat{G_{\text{master}}^{s_3}}$, consider the functions $P(x) = \chi(\sigma_{s_1}(x))$, $Q(y) = \chi'(\sigma_{s_2}(y))$ and $R(z) = \chi''(\sigma_{s_3}(z))$ and the functions $P'(a) = \chi(a)$, $Q'(b) = \chi'(b)$ and $R'(c) = \chi''(c)$. If the SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ is consistent with the set of solutions $\mathcal{T}$ to the* GE System, *then $PQR \equiv 1$ under $\text{supp}(\mu_C)$ iff $P'Q'R' \equiv 1$ in the support of $\mathcal{T}|_C$.*

Proof. To prove the lemma in one direction, suppose $PQR \equiv 1$ in the support of $\text{supp}(\mu_C)$. Take any $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \in \mathcal{T}|_C$. Since $\text{span}(\text{supp}(\mu_C)) = \mathcal{T}|_C$ there are $(x_p, y_p, z_p) \in \text{supp}(\mu_C)$ such that $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \sum_{p=1}^{L}(\sigma_{s_1}(x_p), \sigma_{s_2}(y_p), \sigma_{s_3}(z_p))$. It follows from the multiplicativity of characters that

$$P'(\boldsymbol{a})Q'(\boldsymbol{b})R'(\boldsymbol{c}) = \prod_{p=1}^{L} P'(\sigma_{s_1}(x_p)) \cdot \prod_{p=1}^{L} Q'(\sigma_{s_2}(y_p)) \prod_{p=1}^{L} R'(\sigma_{s_3}(z_p)) = \prod_{p=1}^{L} P(x_p) \cdot Q(y_p) \cdot R(z_p),$$

which is equal to 1 as $PQR \equiv 1$ in the support of μ_C and $(x_p, y_p, z_p) \in \text{supp}(\mu_C)$ for $p = 1, \ldots, L$.
The other direction of the lemma is proved in the same way. $\qquad\square$

We next explain how to achieve consistent solutions. Towards this end, we iteratively modify our SDP solution, as well as the GE System, as stated in Algorithm 4. This algorithm consists of two sub-routines, which we present next:

1. `ModifyGESystem`: In this procedure, if $\text{span}(\text{supp}(\mu_C)) \subsetneq \mathcal{T}|_C$, we modify the GE System to reduce $\mathcal{T}|_C$.

2. `ModifySDP`: In this procedure, if $\text{span}(\text{supp}(\mu_C)) \supsetneq \mathcal{T}|_C$, we modify the SDP formulation to reduce $\text{span}(\text{supp}(\mu_C))$.

The following claim shows the correctness of the procedure `ModifyGESystem` given in Algorithm 2.

Claim 7.8. *Fix a constraint $C \in \mathcal{C}$ with $\mathcal{V}(C) = (s_1, s_2, s_3)$. If $\text{span}(\text{supp}(\mu_C)) \subsetneq \mathcal{T}|_C$, then the* GE System, *as modified in Algorithm 2, satisfies $\text{span}(\text{supp}(\mu_C)) \supseteq \mathcal{T}|_C$.*

Algorithm 2 `ModifyGESystem`

1: **for** every $C \in \mathsf{supp}(\mathcal{C})$ **do**
2:     Let $\mathcal{V}(C) = (s_1, s_2, s_3)$.
3:     **if** $\mathsf{span}(\mathsf{supp}(\mu_C)) \subsetneq \mathcal{T}|_C$ **then**
4:         For all $(\chi_1, \chi_2, \chi_3) \in \widehat{G^{s_1}_{\mathsf{master}}} \times \widehat{G^{s_2}_{\mathsf{master}}} \times \widehat{G^{s_3}_{\mathsf{master}}}$ that evaluates to the constant 1 on $\mathsf{span}(\mathsf{supp}(\mu_C))$, add the equation $y^{\chi_1}_{s_1} \cdot y^{\chi_2}_{s_2} \cdot y^{\chi_3}_{s_3} = 1$ to the GE System.
5:     **end if**
6: **end for**

Proof. Fix a constraint $C \in \mathsf{supp}(\mathcal{C})$ and let $\mathcal{V}(C) = (s_1, s_2, s_3)$. Consider the collection of all the characters $\{(\chi^i_1, \chi^i_2, \chi^i_3)\}_i$ of the group $G^{s_1}_{\mathsf{master}} \times G^{s_2}_{\mathsf{master}} \times G^{s_3}_{\mathsf{master}}$ that evaluate to the constant 1 on the subgroup $\mathsf{span}(\mathsf{supp}(\mu_C))$. Using these characters, we are adding the following equations to the GE System: $y^{\chi^i_1}_{s_1} \cdot y^{\chi^i_2}_{s_2} \cdot y^{\chi^i_3}_{s_3} = 1$ for all i. These equations enforce that $\{(a_{s_1}, a_{s_2}, a_{s_3}) \mid a \in \mathcal{T}\}$ is a subset of $\mathsf{span}(\mathsf{supp}(\mu_C))$ as required. $\qquad\square$

The above claim fixes the problem in one direction towards getting consistent solutions. We also require that $\mathsf{span}(\mathsf{supp}(\mu_C)) \subseteq \mathcal{T}|_C$. This is done in Algorithm 3 below by simply excluding the assignments from a local distribution not implied by the GE System.

Algorithm 3 `ModifySDP`

1: **for** every $(a, b, c) \in \mathsf{supp}(\mu_C)$ such that $(r^{s_1}_a, r^{s_2}_b, r^{s_3}_c) \notin \mathcal{T}|_C$ 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**

The following claim shows the correctness of the procedure `ModifySDP` given in Algorithm 3.

Claim 7.9. *Fix a constraint $C \in \mathcal{C}$ with $\mathcal{V}(C) = (s_1, s_2, s_3)$. Suppose $\mathsf{span}(\mathsf{supp}(\mu_C)) \supsetneq \mathcal{T}|_C$. Then the SDP, as modified in Algorithm 3, satisfies $\mathsf{span}(\mathsf{supp}(\mu_C)) \subseteq \mathcal{T}|_C$.*

Proof. Suppose $\mathsf{span}(\mathsf{supp}(\mu_C)) \supsetneq \mathcal{T}|_C$, This implies that there exists $(a, b, c) \in \mathsf{supp}(\mu_C)$ such that $(r^{s_1}_a, r^{s_2}_b, r^{s_3}_c) \notin \mathcal{T}|_C$. Any solution to the modified SDP from Algorithm 3 ensures that $\mu_C(a, b, c) = 0$ for such tuples. Therefore, any SDP solution to the modified SDP satisfies $\mathsf{span}(\mathsf{supp}(\mu_C)) \subseteq \mathcal{T}|_C$ for every constraint C. $\qquad\square$

Thus, we run the procedures modifyGESystem and modifySDP towards achieving consistent solutions as defined in Definition 7.6. Note that the procedures depend on the initial SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$, and as discussed, we will want to preserve all satisfying assignments in it. In more detail, we need to ensure that, after running the subroutines on a satisfiable instance, the final SDP and the GE System still have all the satisfying assignments preserved. We formally define this below for the SDP solution.

Definition 7.10 (Solution preserving all the integral solutions). *We say an SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ preserves all the valid integral solutions, if for every satisfying assignment $\boldsymbol{\alpha} \in \Sigma^n$ to the instance and any constraint $C \in \mathsf{supp}(\mathcal{C})$, $\alpha|_{\mathcal{V}(C)} \in \mathsf{supp}(\mu_C)$.*

Algorithm 4 stated below calls the two subroutines iteratively in a specific manner so that subsequent SDP formulations preserve all valid integral solutions.

The following section shows that step 3 of the algorithm can be done in polynomial time. In Section 7.3.2, we prove that the algorithm returns consistent solutions in polynomial time.

Algorithm 4 Massaging the SDP solution and the GE System consistent

1: **Input:** An instance $\Upsilon(\mathcal{V}, \mathcal{C})$ of Max-$\mathcal{P}$-CSP.
2: Let $\text{SDP}(\Upsilon)$ be the basic semidefinite program from Figure 2.
3: Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be the SDP solution to $\text{SDP}(\Upsilon)$ preserving all the integral solutions (we explain how to achieve this in Lemma 7.12).
4: **if** $\text{val}(\boldsymbol{V}, \boldsymbol{\mu}) \neq 1$ **then**
5:      Abort. $\triangleright$ Υ is not satisfiable
6: **end if**
7: **if** there is $C \in \text{supp}(\mathcal{C})$ such that $\text{supp}(\mu_C)$ is either pairwise disconnected or has a $\mathbb{Z}$-embedding **then**
8:      Abort. $\triangleright$ 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** $\exists\, (a, b, c) \in \text{supp}(\mu_C)$ such that $(r_a^{s_1}, r_b^{s_2}, r_c^{s_3}) \notin \mathcal{T}|_C$ where $\mathcal{V}(C) = (s_1, s_2, s_3)$ **then**
13:          Run `ModifySDP`.
14:      **else**
15:          Run `ModifyGESystem`.
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**

7.3.1 Preserving All of the Integral Solutions

To make sure that we do not exclude any satisfying assignment from subsequent SDP formulations during the execution of Algorithm 4, we will ensure that the SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ in step 3 preserves all integral solutions. Lemma 7.12 below states that such a solution $(\boldsymbol{V}, \boldsymbol{\mu})$ can always be found in polynomial time. To state and prove the lemma, we need the following claim.

Claim 7.11. *Suppose* $(\boldsymbol{V}, \boldsymbol{\mu})$ *and* $(\boldsymbol{V}', \boldsymbol{\mu}')$ *are two SDP solutions with value* 1. *Then there is a SDP solution* $(\boldsymbol{V}'', \boldsymbol{\mu}'')$ *with value* 1 *such that for every constraint* $C \in \text{supp}(\mathcal{C})$, $\text{supp}(\mu_C'') = \text{supp}(\mu_C) \cup \text{supp}(\mu_C')$.

Proof. The proof is based on a standard tensorization argument. Suppose that 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 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

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

$$= \frac{1}{2}\|\boldsymbol{b}_{i,a}\|^2 + \frac{1}{2}\|\boldsymbol{b}_{i,a}'\|^2$$

$$= \|\boldsymbol{b}_{i,a}''\|^2,$$

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

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

$$= \frac{1}{2}\langle \boldsymbol{b}_{i,a}, \boldsymbol{b}_{j,b} \rangle + \frac{1}{2}\langle \boldsymbol{b}_{i,a}', \boldsymbol{b}_{j,b}' \rangle$$

$$= \Pr_{x \sim \mu_C''}[x_i = a, x_j = b]$$

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 $C \in \mathtt{supp}(\mathcal{C})$. $\qquad \square$

Lemma 7.12. *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 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 $C \in \mathtt{supp}(\mathcal{C})$, $\mu_C(\alpha|_{\mathcal{V}(C)}) > 0$.

Proof. To prove this lemma, we will combine different SDP solutions using Claim 7.11. 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 $\mathtt{SDP}_{C,d}(\Upsilon)$. Let $(\boldsymbol{V}_{C,d}, \boldsymbol{\mu}_{C,d})$ be any arbitrary solution to $\mathtt{SDP}_{C,d}(\Upsilon)$. We then combine all the solutions $(\boldsymbol{V}_{C,d}, \boldsymbol{\mu}_{C,d})$ with value 1 iteratively using Claim 7.11 to get a solution $(\boldsymbol{V}, \boldsymbol{\mu})$. Note that each application of Claim 7.11 blows up the number of coordinates of the vectors by a multiplicative factor of 2. However, since the SDP solution consists of $qn + 1$ vectors, once the number of coordinates goes beyond $qn + 1$, we can find a linear transformation that maps every vector to a vector with only the first $qn + 1$ coordinates nonzero, while preserving all the inner products. We keep doing this dimension reduction after each application of Claim 7.11, and hence the overall process takes polynomial time. We now show that the final solution satisfies the required guarantee. To see this, fix any satisfiable assignment $\boldsymbol{\alpha}$ to the instance Υ and fix any $C \in \mathtt{supp}(\mathcal{C})$. As the solution $(\boldsymbol{V}_{C,\alpha|_{\mathcal{V}(C)}}, \boldsymbol{\mu}_{C,\alpha|_{\mathcal{V}(C)}})$ has $\mu_{C,\alpha|_{\mathcal{V}(C)}}(\alpha|_{\mathcal{V}(C)}) = 1$, we have $\mu_C(\alpha|_{\mathcal{V}(C)}) > 0$ using Claim 7.11. Furthermore, Claim 7.11 guarantees that the value of the solution $(\boldsymbol{V}, \boldsymbol{\mu})$ is 1, proving this lemma. $\qquad \square$

Remark 7.13. *As mentioned at the beginning of this section, for every $\eta > 0$, the SDP can be solved up to an additive accuracy of η in time $\mathtt{poly}(n, \log(1/\eta))$ (see, for instance, [31]). However, Claim 7.8 and Claim 7.9 rely on solving the SDP program exactly as they rely on the exact support of μ_C. One way to get around this issue is as follows. The solution constructed by Lemma 7.12 has the property that each valid entry where we want $\mu_C(\alpha|_{\mathcal{V}(C)}) > 0$ can indeed be forced to have $\mu_C(\alpha|_{\mathcal{V}(C)}) \geqslant 2^{-n^c}$ for a fixed constant c depending on the number of constraints and the number of satisfying assignments to a constraint. This is because each iteration reduces the original weight in the distribution by $\frac{1}{2}$, and there will be at most n^c*

iterations. Thus, if we solve the SDP with accuracy 2^{-n^C} for $C \gg c$ in time $\mathsf{poly}(n)$, then we can treat every weight $\mu_C(\alpha|_{\mathcal{V}(C)}) \leqslant 2^{-n^C}$ as zero towards applying Claim 7.8 and Claim 7.9. In this way, we will end up preserving the local views of all the satisfying assignments in the SDP solution.

The following claim shows that at each iteration in Algorithm 4, the modified SDP has all the satisfying assignments as valid integral solutions.

Claim 7.14. *If $\boldsymbol{\alpha} \in \Sigma^N$ is a satisfying assignment to instance Υ, then it remains a valid integral solution to the SDP $\mathcal{S}$ that we get after running Algorithm 4.*

Proof. We will show that the procedure maintains the following two invariants: 1) every satisfying assignment from $\boldsymbol{\alpha} \in [q]^N$ is also a valid satisfying assignment (after interpreting α_s with $r_{\alpha_s}^s$) to the GE System after the modification, and 2) every satisfying assignment from $\boldsymbol{\alpha} \in [q]^N$ 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, as the solution from step 3 is guaranteed to keep $\alpha|_{\mathcal{V}(C)}$ for every satisfying assignment $\boldsymbol{\alpha}$ and C, every satisfying assignment $\boldsymbol{\alpha} \in [q]^N$ (after interpreting α_s with $r_{\alpha_s}^s$) is also a valid satisfying assignment to the GE System after the modification. To see this, the elements $(r_{\alpha_{s_1}}^{s_1}, r_{\alpha_{s_2}}^{s_2}, r_{\alpha_{s_3}}^{s_3})$ satisfy all the additional equations of the from $y_{s_1}^{\chi_1} \cdot y_{s_2}^{\chi_2} \cdot y_{s_3}^{\chi_3} = 1$ added to the GE System related to the constraint C for every C with $\mathcal{V}(C) = (s_1, s_2, s_3)$. Hence, $(r_{\alpha_s}^s)_{s \in \mathcal{V}}$ is in the support of $\mathcal{T}$ even after modification of the GE System. $\qquad \square$

7.3.2 Getting Consistent Solutions to the SDP and the GE System

We now have the following important lemma that shows that applying Algorithm 4 on a CSP instance where every predicate is MILDLY-SYMMETRIC returns an SDP solution and a system of linear equations that are consistent with respect to each other.

Lemma 7.15. *Applying Algorithm 4 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 GE System* that are consistent with each other. Furthermore, for every constraint $C \in \mathsf{supp}(\mathcal{C})$, the $\mathsf{supp}(\mu_C^\star)$ is pairwise connected and does not have a $(\mathbb{Z}, +)$-embedding.*

Proof. We will rule out that the algorithm will never reach the Abort steps (steps 5 and 8). Lemma 7.16 below shows that the algorithm terminates in polynomial time. Using Claim 7.8 and Claim 7.9, eventually, the algorithm will return an SDP solution and a GE System that are consistent with each other.

Using Claim 7.14, the algorithm will never reach step 5. Next, we rule out that the algorithm will never reach step 8. Fix a satisfying assignment $\boldsymbol{\alpha} \in \Sigma^N$ to the instance. Consider the maps $\{\tau_i : \Sigma \to \Sigma \mid 1 \leqslant i \leqslant \ell\}$ from the Definition 1.3 of the predicates being MILDLY-SYMMETRIC. Using Property 1. from the Definition 1.3, we have that for all $1 \leqslant i \leqslant \ell$, $\boldsymbol{\alpha}^{(i)} \in \Sigma^N$, where $\alpha_j^{(i)} = \tau_i(\alpha_j)$ for $1 \leqslant j \leqslant N$, is also a satisfying assignment to the instance. Using this and Lemma 7.12, for every $C \in \mathsf{supp}(\mathcal{C})$, the local distribution $\mu_C^\star$ 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 $\sigma = \boldsymbol{\alpha}|_C$. Now using Property 2. from the Definition 1.3 of the predicates being MILDLY-SYMMETRIC, the set $Z_{\boldsymbol{\alpha}, C}$ (and hence $\mathsf{supp}(\mu_C^\star)$) is pairwise connected and does not have a $(\mathbb{Z}, +)$-embedding. Hence, the algorithm will never reach step 8. $\qquad \square$

Finally, we compute the running time of Algorithm 4

Lemma 7.16. *Algorithm 4 on a satisfiable CSP instance* $\Upsilon(\mathcal{V}, \mathcal{C})$ *where every predicate* $C \in \mathcal{C}$ *is* MILDLY-SYMMETRIC *runs in time* $O_{|\Sigma|}(1) \cdot |\mathcal{V}|^{O(1)}$.

Proof. First, the procedure ModifySDP runs in time $O_{|\Sigma|}(1) \cdot |\mathcal{V}|^{O(1)}$. 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 4.

In the procedure ModifyGESystem, for a constraint C such that $\mathsf{span}(\mathsf{supp}(\mu_C)) \subsetneq \mathcal{T}|_C$, then the GE System, as modified in Algorithm 2, satisfies $\mathsf{span}(\mathsf{supp}(\mu_C)) \supseteq \mathcal{T}|_C$. The number of embedding functions is $O_{|\Sigma|}(1)$, and therefore, every call to ModifyGESystem takes at most $O_{|\Sigma|}(1) \cdot |\mathcal{V}|^{O(1)}$ time.

Therefore, the overall running time of Algorithm 4 is $O_{|\Sigma|}(1) \cdot |\mathcal{V}|^{O(1)}$ □

7.4 The 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
---
1: **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.
2: Perform the operations mentioned in Algorithm 4.
3: **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**
4:     Reject.
5: **else**
6:     Accept.
7: **end if**

It can be checked easily that Lemma 7.16 implies 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 $\alpha_{\mathcal{P}}^{\mathcal{ALG}}$.

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

Theorem 7.17. *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 7.15, $\mathcal{ALG}$ accepts. If the instance has value strictly less than $\alpha_{\mathcal{ALG}}$, then by definition, $\mathcal{ALG}$ rejects such instances. □

In Section 7.5, 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 predicates $\mathcal{P}$, has perfect completeness, and has soundness $\alpha_{\mathcal{P}}^{\mathcal{ALG}} + \varepsilon$, for every constant $\varepsilon > 0$.

7.5 The Dictatorship Test

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 identify it with the set $[q] = \{1, 2, \ldots, q\}$ where $q = |\Sigma|$. Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be a solution for the final 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 μ_C^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 6.8) 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.

7.5.1 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", after removing the high-degree contributions, are low under all of the probability distributions corresponding to variables $s_i \in \mathcal{V}$. We make this formal in the following definition.

Definition 7.18. *(Quasirandom function w.r.t. $(\boldsymbol{V}, \boldsymbol{\mu})$) Fix a function $F : \Sigma^n \to \Sigma$. Let α be the minimum probability of the atoms in the SDP solution. Consider the functions $(F_{s,1}, F_{s,2}, \ldots, F_{s,q})$, where for every $j \in [q]$, $F_{s,j} : ([q]^n, \mu_s^{\otimes n}) \to \mathbb{R}$ is defined as $F_{s,j}(\boldsymbol{x}) := \mathbf{1}_{F(\boldsymbol{x})=j}$. Fix the group G_{master}^s for each $s \in \mathcal{V}$ from the hybrid algorithm. Suppose that for each s we have a collection of product functions $\mathcal{P}_s$ over G_{master}^s such that there are functions $F'_{s,j}$ of the form $F'_{s,j}(x) = \sum\limits_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_s)} P(x) \cdot L_{P,s,j}(x)$ satisfying:*

1. *$\|F_{s,j} - F'_{s,j}\|_{\mathcal{D}_s,\alpha} \leqslant \varepsilon$, where $\mathcal{D}_s = \{\mu_C \mid s \in \mathcal{V}(C)\}$.*

2. *$|\mathcal{P}_s|$, $\deg(L_{P,s,j})$ and $\|L_{P,s,j}\|_2$ are all at most $d = O_{\alpha,q,\varepsilon}(1)$.*

3. *For every $C \in \mathsf{supp}(\mathcal{C})$ with $\mathcal{V}(C) = (s_1, s_2, s_3)$, considering the collection of product functions $\mathcal{D}_C = \{D_P, D_Q, D_R \mid P \in \mathcal{P}_{s_1}, Q \in \mathcal{P}_{s_2}, R \in \mathcal{P}_{s_3}\}$ over $\mathsf{supp}(\mu_C)$ where $D_P(x,y,z) = P(x)$, $D_Q(x,y,z) = Q(y)$ and $D_R(x,y,z) = R(z)$, we have that $\mathsf{rk}(\mathcal{D}_C) \geqslant T$.*

Then, F is called (d, τ)-quasirandom w.r.t. $(\boldsymbol{V}, \boldsymbol{\mu})$ if for every $\varepsilon > 0$, there is arbitrary large enough $T \gg \varepsilon^{-1}, d, q, \alpha^{-1}$ such that every $s \in \mathcal{V}$, $j \in [q]$ the above properties hold and furthermore for every $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_s)$ and $j \in [q]$, $\max_{i \in [n]} I_i[L_{P,s,j}] \leqslant \tau$.

Remark 7.19. *The standard notion of (d, τ)-quasirandomness that is often used in hardness of approximation is that $I_i[F_{s,j}^{\leqslant d}] \leqslant \tau$ for all $j \in [q]$ and $i \in [n]$, where $F_{s,j}^{\leqslant d}$ is the truncation of $F_{s,j}$ up to degree d. In particular, the high-degree component of $F_{s,j}$ is often ignored since it can often be shown not to contribute much. In our case, since we care about perfect completeness, the high degree part of $F_{s,j}$ cannot be simply ignored. Thus, the above notion of quasirandomness can be seen as a natural analog of the usual one, where we enforce small influence conditions on all the low-degree components $L_{P,s,j}$ 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, and we leave this as an open challenge for future research.*

One may wonder if it is even possible to find the functions $F'_{s,j}$ such that all the above (except for the low influence property) properties hold for small enough ε and large T. Lemma 5.4 and Lemma 5.6 already show the existence of product functions $\mathcal{P}_s$ for individual $F_{s,j}$ satisfying the above conditions in some setting. The following lemma generalizes that argument to our current setting.

Lemma 7.20. *For all $\alpha > 0$, $m, M \in \mathbb{N}$ 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, and let $\mathcal{M}$ be a family of subsets of $\{1, 2, \ldots, M\}$ of size 3. Let $\{\mu_{(s_1,s_2,s_3)} \mid (s_1, s_2, s_3) \in \mathcal{M}\}$ be distributions over Σ^3 in which the probability of each atom is at least α and the distribution $\mu_{(s_1,s_2,s_3)}$ is pairwise connected and has no $\mathbb{Z}$ embedding for all $(s_1, s_2, s_3) \in \mathcal{M}$. Suppose the marginals of $\mu_{(s_1,s_2,s_3)}$ are μ_{s_1}, μ_{s_2}, μ_{s_3} on the three coordinates, respectively. Then there are Abelian groups G^s whose sizes depend only on $|\Sigma|$, such that for any 1-bounded functions $F_{s,j} \colon (\Sigma^n, \mu_s^n) \to \mathbb{C}$, for $1 \leqslant s \leqslant M$ and $1 \leqslant j \leqslant |\Sigma|$, there is $\varepsilon \geqslant \varepsilon_0$ and a collections $\mathcal{P}_s$ of product functions of size at most r over the group G^s such that:*

1. *For all $\varepsilon' \in (w(\varepsilon), \varepsilon)$ we have that $\|F_{s,j} - T_{\mathcal{P}_s, 1-\varepsilon'} F_{s,j}\|_{\mathcal{D}_s, \alpha} \leqslant \xi$, where $\mathcal{D}_s = \{\mu_{(s_1,s_2,s_3)} \mid s_1 = s, s_2 = s \text{ or } s_3 = s\}$.*

2. *For every $(s_1, s_2, s_3) \in \mathcal{M}$, considering the collection of product functions*

$$\mathcal{D}_{(s_1,s_2,s_3)} = \{D_P, D_Q, D_R \mid P \in \mathcal{P}_{s_1}, Q \in \mathcal{P}_{s_2}, R \in \mathcal{P}_{s_3}\}$$

   over $\mathrm{supp}(\mu_{(s_1,s_2,s_3)})$ where $D_P(x, y, z) = P(x)$, $D_Q(x, y, z) = Q(y)$ and $D_R(x, y, z) = R(z)$, we have that $\mathrm{rk}(\mathcal{D}_{(s_1,s_2,s_3)}) \geqslant \frac{1}{w(\varepsilon)}$.

Furthermore, if the collection of distributions $\{\mu_{(s_1,s_2,s_3)} \mid (s_1, s_2, s_3) \in \mathcal{M}\}$ are coming from the valid SDP solution from Lemma 7.15, then the groups G^s can be taken to be the groups G^s_{master} from the hybrid algorithm.

Proof. Fix a decay function $w' \colon [0,1] \to [0,1]$ depending on w (to be determined). We first apply Lemma 5.4 to get a collection of product functions $\mathcal{P}_s$ such that for all s, j and $\varepsilon' \in (w'(\varepsilon), \varepsilon)$,

$$\|F_{s,j} - T_{\mathcal{P}_j, 1-\varepsilon'} F_{s,j}\|_{\mathcal{D}_s, \alpha} \leqslant \xi.$$

Inspecting the proof of Lemma 5.4, we see that because we apply it with the collection $\mathcal{D}_s$, the collection $\mathcal{P}_s$ consists of product functions over G^s_{master}. Let $r = \max_s |\mathcal{P}_s|$, and note that $r = O_{\alpha,m,M,\xi}(1)$. Let ε_0 be from Lemma 5.4 and let $\varepsilon \geqslant \varepsilon_0$.

Next, we modify collection $\mathcal{P}_s$ as in Lemma 5.6 to achieve the second item of the lemma. Denote $A = \sum_{(s_1,s_2,s_3)\in\mathcal{M}} \left|\mathsf{spn}_{\mathbb{N}}(\mathcal{D}_{(s_1,s_2,s_3)})\right|$, and note that $A = O_{\alpha,m,M,\xi}(1)$. Define the decay functions $w_1(z) = 0.5w(z)$ and $w_{i+1}(z) = 0.5w(w_i(z))$ for $i \geqslant 1$. Choose $w'(z) = w_{6A}(z)$ and inspect the intervals $I_i = (w_{3i}(\varepsilon)^{-1}, w_{3(i+1)}(\varepsilon)^{-1})$ for $i = 1, \ldots, 2A - 1$. By the pigeonhole principle, we may find i such that for each $(s_1, s_2, s_3) \in \mathcal{M}$ and $D \in \mathsf{spn}_{\mathbb{N}}(\mathcal{D}_{s_1,s_2,s_3})$ we have that $\Delta_{\mathsf{symbolic}}(D, 1) \notin I_i$. We take

$$J = \left\{\ell \in [n] \mid \exists (s_1, s_2, s_3) \in \mathcal{M}, D \in \mathsf{spn}_{\mathbb{N}}(\mathcal{D}_{s_1,s_2,s_3}) \text{ with } \Delta_{\mathsf{symbolic}}(D, 1) \leqslant w_{3i}(\varepsilon)^{-1}, D_\ell \neq 1\right\}$$

and set $\mathcal{P}'_s = \{P|_{\bar{J}} \mid P \in \mathcal{P}_s\}$ and $\mathcal{D}'_{s_1,s_2,s_3}$ accordingly. Then for all $(s_1, s_2, s_3) \in \mathcal{M}$ and $D \in \mathsf{spn}_{\mathbb{N}}(\mathcal{D}'_{s_1,s_2,s_3})$ we either have $D \equiv 1$ or

$$\Delta_{\mathsf{symbolic}}(D, 1) \geqslant w_{3(i+1)}(\varepsilon)^{-1} - |J| \geqslant w_{3(i+1)}(\varepsilon)^{-1} - w_{3i}(\varepsilon)^{-1} A \geqslant \frac{1}{w(w_{3i+2}(\varepsilon))},$$

so the second item holds. Applying the argument in the proof of Claim 5.7 gives the first item. $\square$

Thus, Lemma 7.20 shows that given any function $f : \Sigma^n \to \Sigma^n$, we can find decompositions of f guaranteeing properties 1., 2. and 3. listed in Definition 7.18. Next, we will design a dictatorship test with perfect completeness, such that for a given f, if any such decompositions happen to satisfy the quasirandom property (namely, it additionally satisfies the low-influence property), the test passes with probability at most the optimal value of the instance. If for any such decompositions of f, some L_P has an influential variable, then we hope such a structure can be used in the (conditional) hardness reductions, specifically in the *decoding* procedure of the soundness analysis of a reduction.

7.5.2 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 $\mathcal{T} \subseteq \prod_{v \in \mathcal{V}} H_v^\star$ be the subspace of the set of satisfying assignments of the instance $\Upsilon(\mathcal{V}, \mathcal{C})$ coming from the GE System that satisfy the conditions from Lemma 7.15. Let $\alpha > 0$ be a lower bound on the non-zero probabilities from the distributions μ_C; 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 $F : \Sigma^R \to \Sigma$ for large enough R compared to $|\mathcal{V}|, \alpha^{-1}, |\Sigma|$.

---

1. Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be a solution for the basic SDP relaxation of Υ that satisfy the conditions from Lemma 7.15.

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 μ_C^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 $\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}$.

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 follows.

$$\text{Soundness}_{(d,\tau)}(\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}) = \sup_{\substack{F : \Sigma^R \to \Sigma \\ F \text{ is } (d,\tau)\text{-quasirandom w.r.t.}(\boldsymbol{V},\boldsymbol{\mu})}} \Pr[F \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}].$$

We now state our main theorem regarding the soundness of the dictatorship test $\mathbf{Dict}_{V,\mu}$.

Theorem 7.21. *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 $\mathbf{Dict}_{V,\mu}$ has completeness 1 and soundness*

$$\text{Soundness}_{(d,\tau)}(\mathbf{Dict}_{V,\mu}) = \mathsf{OPT}(\Upsilon) + o_\tau(1),$$

where $\mathsf{OPT}(\Upsilon)$ is the optimal value of the instance Υ and $o_\tau(1) \to 0$ as $\tau \to 0$.

The following corollary follows from the above theorem and the approximation guarantee of our hybrid algorithm.

Corollary 7.22. *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 7.17 along with Corollary 7.22 prove our main Theorem 1.1. We prove Theorem 7.21 in the rest of the section. We first set up a few notations in Section 7.5.3 that will be used in the analysis of the soundness of the test.

7.5.3 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 $f, g : \Omega \to \mathbb{C}$ on this space as follows: $\langle f, g \rangle = \mathbb{E}_{x \sim \mu}[f(x)\overline{g(x)}]$.

Definition 7.23. *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 ith 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 7.24. *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 $\boldsymbol{\mathcal{X}} = \{x_j^{(i)} \mid j \in \{0, 1, \ldots, q-1\}, i \in [R]\}$ 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*

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{\sigma} \in \{0,1,\ldots,q-1\}^R} \hat{F}_{\boldsymbol{\sigma}} x_{\boldsymbol{\sigma}}.$$

59

Given any function $\mathcal{F} : \Omega^R \to \mathbb{C}$, with the multilinear expansion $\mathcal{F}(z) = \sum_{\sigma \in \{0,1,\dots,q-1\}^R} \hat{\mathcal{F}}(\sigma) \ell_\sigma(z)$ with respect to the orthonormal ensemble $\mathcal{L} = \{\ell_0 \equiv \mathbf{1}, \ell_1, \dots, \ell_{q-1}\}$, we define a corresponding *formal polynomial* in the indeterminates $\boldsymbol{\mathcal{X}} = \{x_j^{(i)} | j \in \{0, 1, \dots, q-1\}, i \in [R]\}$, as follows:

$$F(\boldsymbol{x}) = \sum_{\sigma} \hat{\mathcal{F}}(\boldsymbol{\sigma}) x_{\boldsymbol{\sigma}}.$$

We will always use the symbol $\mathcal{F}$ to denote a 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:

$$\mathop{\mathbb{E}}_{z \in \Omega^R} [\mathcal{F}(z)] = \mathbb{E}[F(\mathcal{L}^R)].$$

Vector Valued Functions. We will always use the symbol $\boldsymbol{\mathcal{F}} = (\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_q)$ to denote a vector-valued function on a product probability space Ω^R. Further $\boldsymbol{F}(\boldsymbol{x}) = (F_1, F_2, \dots, F_q)$ will denote the formal multilinear polynomial corresponding to $\boldsymbol{\mathcal{F}}$.

7.5.4 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 7.18 and we fix decompositions of F that guarantee 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 $\mathbb{C}^{3q}$: We will extend the payoff function C further to a real-valued function on $(\mathbb{C}^q)^3$, by constructing a smooth extension of C. This smooth extension of C satisfies the following properties, controlled by a constant $C_0(q)$ that depends only on q.

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.,$\forall \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}, \{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3\} \in (\mathbb{C}^q)^3$,

$$|C(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) - C(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)| \leqslant C_0(q) \sum_{i=1}^3 \|\boldsymbol{x}_i - \boldsymbol{y}_i\|_2.$$

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 $s \in \mathcal{V}$, $\boldsymbol{\ell}_s = \{\ell_{s,0} \equiv \mathbf{1}, \ell_{s,1}, \ldots, \ell_{s,q-1}\}$, is a set of random variables that are orthonormal ensembles for the space Ω_s.

- **Global Gaussian Ensembles $\mathcal{G}$:** The Global Gaussian Ensembles $\mathcal{G} = \{\boldsymbol{g}_s \mid s \in \mathcal{V}\}$ are generated by setting $\boldsymbol{g}_s = \{g_{s,0} \equiv \mathbf{1}, g_{s,1}, \ldots, g_{s,q-1}\}$ where

$$g_{s,j} = \sum_{\omega \in \Sigma} \ell_{s,j}(\omega)\langle \boldsymbol{b}_{s,\omega}, \boldsymbol{\zeta}\rangle, \quad \forall j \in \{1, \ldots, q-1\},$$

and $\boldsymbol{\zeta}$ is a random normal Gaussian 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 7.25. *([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 $\mathcal{L}$:*

$$\mathbb{E}_{\boldsymbol{\zeta}}[g_{s,j} \cdot g_{s',j'}] = \mathbb{E}_{(\omega,\omega')\sim\mu_C|(s,s')}[\ell_{s,j}(\omega) \cdot \ell_{s',j'}(\omega')] \quad \forall j, j' \in \{1, \ldots, q-1\}, s, s' \in \mathcal{V}(P'),$$

where $\mu_C|(s,s')$ is the marginal distribution of μ_C on the coordinates of s, s'.

Functions used in the soundness analysis. In order to analyze the above expectation, we define functions related to $\mathcal{F}_s$ for $s \in \mathcal{V}$ analogously to Section 6.3. For small enough $\xi > 0$, let $\varepsilon > 0$ be arbitrarily small constant compared to ξ and take a parameter T such that $T \gg 1/\xi, 1/\varepsilon, 1/\alpha, |\Sigma|$. Let $\mathcal{P}_s$ be the collection of product functions associated with $F'_{s,j}$ from the decomposition of $F : \Sigma^R \to \Sigma$ satisfying the quasirandomness property as in Definition 7.18 with the above parameters where τ is much smaller than the parameters ξ, ε. Write each $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_s)$ as $P = \prod_{i=1}^{R} \chi_{P,i}(\sigma_s(x_i))$. Here, $\sigma_s : \Sigma \to H_s^\star$ is the embedding into the group $H_s^\star$ with respect to the set of distributions $M_{\mu_s} = \{\mu_C \mid s \in \mathcal{V}(C)\}$. With these collections of product functions for each $s \in \mathcal{V}$, we first define the noisy version of the functions $\mathcal{F}_{s,j}$ for $j \in [q]$ as:

$$\mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(\boldsymbol{z}) = \mathbb{E}_{I \subseteq_\varepsilon [n]}\left[\mathbb{E}_{\boldsymbol{y} \sim \mathrm{T}_{\mu_s,\mathcal{P}_s,I}\boldsymbol{z}} [\mathcal{F}_{s,j}(\boldsymbol{y})] \right], \tag{38}$$

where the operator $\mathrm{T}_{\mu_s,\mathcal{P}_s,I}$ is defined in Definition 4.1. Using this, define the following decomposition which is close to the above function in the ℓ_2 norm as proved in Lemma 4.9,

$$\mathcal{F}'_{s,j}(\boldsymbol{z}) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_s)} L_{P,s,j}(\boldsymbol{z}) \cdot \prod_{i=1}^{R} \chi_{P,i}(\sigma_s(z_i)). \tag{39}$$

Here $\deg(L_{P,s,j}) \leqslant d$ and $\|L_{P,s,j}\|_2 \leqslant d$, where d depends on $\xi, \varepsilon, \alpha, |\Sigma|$. These are the same decompositions used to define the quasirandomness of the function F in Definition 7.18. Again, by the (d, τ) quasirandomness condition, we have for every $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_s)$ and $j \in [q]$, $\max_{i \in [n]} I_i[L_{P,s,j}] \leqslant \tau$. Next,

we define a decoupled version of the above function where we provide separate inputs for the low-degree part and the product function part (we recall that by Lemma 6.4 there exists a coupling between $\mu_s^R \times \mu_s^R$ and μ_s^R under which these two functions are close to each):

$$\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_s)} L_{P,s,j}(\boldsymbol{y}) \cdot \prod_{i=1}^{R} \chi_{P,i}(\sigma_s(x_i)). \tag{40}$$

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 $\sigma_s(x_i)$ with the group element input. Namely,

$$\mathcal{F}_{s,j}^{\mathsf{dec}}(\boldsymbol{a}, \boldsymbol{y}) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s,j})} L_{P,s,j}(\boldsymbol{y}) \cdot \prod_{i=1}^{R} \chi_{P,i}(a_i). \tag{41}$$

Towards applying the invariance principle, we replace the low-degree polynomials with their corresponding multilinear extensions to get the functions to get the following function $F_{s,j}^{\mathsf{dec}} \colon (H_s^\star)^R \times \mathbb{R}^{(q-1)R} \to \mathbb{C}$.

$$F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}, \boldsymbol{\ell}) = \sum_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s,j})} L_{P,s,j}(\boldsymbol{\ell}) \cdot \prod_{i=1}^{R} \chi_{P,i}(a_i).$$

We will also need to work with the analogous function $F_{s,j}^{\mathsf{dec}} \colon \Sigma^R \times \mathbb{R}^{(q-1)R} \to \mathbb{C}$, where in the above definition a_i is replaced with $\sigma(x_i)$, and we use the same notation to refer to this function as the distinction will be clear from the inputs.

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 $\mathsf{trunc}_{[0,1]}$ defined in Figure 4.

$$\widetilde{F}_{s,j}^{\mathsf{dec}}(\boldsymbol{a}, \boldsymbol{\ell}) = \mathsf{trunc}_{[0,1]}(\mathfrak{Re}(F_{s,j}^{\mathsf{dec}}(\boldsymbol{a}, \boldsymbol{\ell}))).$$

We also use the notation $\sim$ on any function with the same definition as above. This completes the description of the various functions that we will encounter in the soundness analysis below.

Soundness analysis. The acceptance probability of the test for a given function $\mathcal{F}$ is given by:

$$\Pr[\mathcal{F} \text{ passes } \mathbf{Dict}_{V,\mu}] = \underset{C \sim \mathcal{C}}{\mathbb{E}} \underbrace{\underset{\boldsymbol{z}_C \sim \mu_C^R}{\mathbb{E}} [C(\mathcal{F}_{s_1}(\boldsymbol{z}_{s_1}), \mathcal{F}_{s_2}(\boldsymbol{z}_{s_2}), \mathcal{F}_{s_3}(\boldsymbol{z}_{s_3}))]}_{(I)}, \tag{42}$$

where $\mathcal{V}(C) = (s_2, s_2, s_3)$ and C is the smooth extension of the payoff function. Note that for all s and $\boldsymbol{z} \in \Sigma^R$, we have $\sum_j \mathcal{F}_{s,j}(\boldsymbol{z}) = 1$. Furthermore, we have the same property for the noisy functions as given by the following claim.

Claim 7.26. *For all s and $\boldsymbol{z} \in \Sigma^R$, we have $\sum_j \mathrm{T}_{\mathcal{P}_s, 1-\varepsilon} \mathcal{F}_{s,j}(\boldsymbol{z}) = 1$.*

Proof.

$$\sum_j \mathrm{T}_{\mathcal{P}_s, 1-\varepsilon} \mathcal{F}_{s,j}(\boldsymbol{z}) = \sum_j \underset{I \subseteq_\varepsilon [n]}{\mathbb{E}} \left[\underset{\boldsymbol{y} \sim \mathrm{T}_{\mu_s, \mathcal{P}, I} \boldsymbol{z}}{\mathbb{E}} [\mathcal{F}_{s,j}(\boldsymbol{y})] \right] = \underset{I \subseteq_\varepsilon [n]}{\mathbb{E}} \left[\underset{\boldsymbol{y} \sim \mathrm{T}_{\mu_s, \mathcal{P}_s, I} \boldsymbol{z}}{\mathbb{E}} \left[\sum_j \mathcal{F}_{s,j}(\boldsymbol{y}) \right] \right] = 1. \quad \square$$

We now apply Theorem 6.9 to the expression (I), and we use C as Ψ therein. We remark that in the current setting we use $\mathsf{trunc}_{[0,1]}$ instead of trunc as in the proof of Theorem 6.9, but the argument goes through in the same way. The only difference is that in the argument in Claim 6.10 instead of using the fact that $f' = \mathrm{T}_{\mathcal{P},1-\varepsilon} f$ is 1-bounded (as there), we use the fact that $\mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}$ is $[0,1]$-valued. Therefore, we conclude the following

$$\left| (I) - \underbrace{\mathop{\mathbb{E}}_{\substack{\boldsymbol{a}_C \sim \tilde{\mu}_C^R \\ \mathcal{G}^R}} \left[C(\widetilde{\boldsymbol{F}_{s_1}^{\mathsf{dec}}}(\boldsymbol{a}_{s_1}, \boldsymbol{g}_{s_1}), \widetilde{\boldsymbol{F}_{s_2}^{\mathsf{dec}}}(\boldsymbol{a}_{s_2}, \boldsymbol{g}_{s_2}), \widetilde{\boldsymbol{F}_{s_3}^{\mathsf{dec}}}(\boldsymbol{a}_{s_3}, \boldsymbol{g}_{s_3})) \right]}_{(II)} \right| \lesssim \xi, \tag{43}$$

where $a_C = (\boldsymbol{a}_{s_1}, \boldsymbol{a}_{s_2}, \boldsymbol{a}_{s_3}) \sim \tilde{\mu}_C^R$ is the following distribution on $(H_{s_1}^\star \times H_{s_2}^\star \times H_{s_3}^\star)^R$ given by first sampling $\boldsymbol{z}_C = (\boldsymbol{z}_{s_1}, \boldsymbol{z}_{s_2}, \boldsymbol{z}_{s_3}) \sim \mu_C^R$ and then applying the embedding maps $\sigma_s : \Sigma \to H_s^\star$ coordinatewise to z_{s_j}.

At this point, we would like to make a couple of observations. In the expectation given in (II), the distribution $\mathcal{G}^R$ can be sampled globally using the vectors from the SDP solution. However, the distribution $\tilde{\mu}_C^R$ is specific to the constraint C. We do not know how to sample the strings $\{\boldsymbol{a}_s\}_{s \in \mathcal{V}}$ globally that marginally match with the distributions $\tilde{\mu}_C^R$ for every constraint C. To remedy this, we next show that we can move from the distribution $\tilde{\mu}_C^R$ to a distribution that samples a random set of R solutions to the GE System and outputs the restriction of these solutions to the variables s_1, s_2, s_3. In order to show the closeness of these two distributions, we use the property that the product functions that we choose satisfy the high-rank property and the SPD solution is consistent with the GE System, guaranteed by Lemma 7.15.

Specifically, consider the following two distributions:

1. The distribution $\mathcal{D}_\mathcal{T}$: sample R assignments $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \ldots, \boldsymbol{\alpha}^{(R)}$ from the set of satisfying assignment $\mathcal{T} \subseteq G_{\mathsf{master}}^\mathcal{V}$ to the 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 $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$.

2. The distribution $\mathcal{D}_{\mu_C}$: sample $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \sim \mu_C^R$ and output $(\sigma_{s_1}(\boldsymbol{x}), \sigma_{s_2}(\boldsymbol{y}), \sigma_{s_3}(\boldsymbol{z}))$.

For any distribution $\mathcal{D}$ on $(H_{s_1}^\star \times H_{s_2}^\star \times H_{s_3}^\star)^R$, consider the distribution $\chi(\mathcal{D})$ defined as follows: Sample $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \sim \mathcal{D}$ and output

$$(\chi_P(\boldsymbol{a}), \chi_Q(\boldsymbol{b}), \chi_R(\boldsymbol{c}))_{P \in \mathsf{spn}_\mathbb{N}(\mathcal{P}_{s_1}), Q \in \mathsf{spn}_\mathbb{N}(\mathcal{P}_{s_2}), R \in \mathsf{spn}_\mathbb{N}(\mathcal{P}_{s_3})}.$$

Note that although each coordinate $i \in [R]$ of the output of the distribution $\chi(\mathcal{D})$ is a vector in $\mathbb{C}^D$ for some $D = O_{|\Sigma|}(1)$, the support of the distribution for each $i \in [R]$ is bounded by $O_{|\Sigma|}(1)$. The following lemma (Lemma 7.29) asserts that the distributions $\chi(\mathcal{D}_{\mu_C})$, and $\chi(\mathcal{D}_\mathcal{T})$ are close in statistical distance. Here, we crucially use the fact that the set of solutions $\mathcal{T}$ to the GE System and the SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$ are consistent with each other as stated in Lemma 7.15. Before we state and prove the lemma, we need the following two facts.

Fact 7.27. *Suppose $P \in \mathsf{spn}_\mathbb{N}(\mathcal{P}_{s_1})$, $Q \in \mathsf{spn}_\mathbb{N}(\mathcal{P}_{s_2})$, $R \in \mathsf{spn}_\mathbb{N}(\mathcal{P}_{s_3})$ are such that $(P_i Q_i R_i) \not\equiv 1$ for at least T coordinates under the distribution μ_C. Then letting α be the minimum probability of an atom in μ_C, we have that*

$$\left| \mathop{\mathbb{E}}_{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \sim \mu_C^R} [\chi_P(\sigma_{s_1}(\boldsymbol{x})) \chi_Q(\sigma_{s_2}(\boldsymbol{y})) \chi_R(\sigma_{s_3}(\boldsymbol{z}))] \right| \leqslant (1 - \Omega_{\alpha, |\Sigma|}(1))^T.$$

Proof. Let $\mathcal{J}$ be the set of i's such that $(P_i Q_i R_i) \not\equiv 1$ under μ_C. Then the left-hand side is equal to

$$\prod_{i \in \mathcal{J}} \left| \underset{(x_i, y_i, z_i) \sim \mu_C}{\mathbb{E}} [\chi_{P_i}(\sigma_{s_1}(x_i)) \chi_{Q_i}(\sigma_{s_2}(y_i)) \chi_{R_i}(\sigma_{s_3}(z_i))] \right|,$$

and it suffices to upper-bound each one of the terms by $1 - \Omega_{\alpha, |\Sigma|}(1)$. Fix $i \in \mathcal{J}$, and note that the values that $P_i Q_i R_i$ may receive are discrete, all have absolute value 1 and they are $\Omega_{|\Sigma|}(1)$ far apart in absolute value. Thus, there is a fixed constant $c_{\alpha, |\Sigma|} > 0$ such that for $i \in \mathcal{J}$, we have

$$\left| \underset{(x_i, y_i, z_i) \sim \mu_C}{\mathbb{E}} [\chi_{P_i}(\sigma_{s_1}(x_i)) \chi_{Q_i}(\sigma_{s_2}(y_i)) \chi_{R_i}(\sigma_{s_3}(z_i))] \right| \leqslant 1 - c_{\alpha, |\Sigma|},$$

and the fact follows. □

A similar statement is also true if we replace the distribution with $\mathcal{D}_{\mathcal{T}}$.

Fact 7.28. *Suppose $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1})$, $Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2})$, $R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3})$ are such that $(P_i Q_i R_i) \not\equiv 1$ for at least T coordinates under the distribution μ_C. If the SDP solution and the solutions to the* GE System *are consistent, we have that*

$$\left| \underset{(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \sim \mathcal{D}_{\mathcal{T}}}{\mathbb{E}} [\chi_P(\boldsymbol{a}) \chi_Q(\boldsymbol{b}) \chi_R(\boldsymbol{c})] \right| \leqslant (1 - \Omega_{|\Sigma|}(1))^T.$$

Proof. Again, for the set $\mathcal{J}$ as before, we have

$$\left| \underset{(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \sim \mathcal{D}_{\mathcal{T}}}{\mathbb{E}} [\chi_P(\boldsymbol{a}) \chi_Q(\boldsymbol{b}) \chi_R(\boldsymbol{c})] \right| = \prod_{i \in \mathcal{J}} \left| \underset{(a_i, b_i, c_i) \sim \mathcal{D}_{\mathcal{T}}^{(i)}}{\mathbb{E}} [\chi_{P_i}(a_i) \chi_{Q_i}(b_i) \chi_{R_i}(c_i)] \right|,$$

where we let $\mathcal{D}_{\mathcal{T}}^{(i)}$ be the tuple on (s_1, s_2, s_3) corresponding to the assignment $\boldsymbol{\alpha}^{(i)}$. Here, we used the fact that the SDP solution and the solutions to the GE System are consistent, and hence, for $i \notin \mathcal{J}$, $P_i Q_i R_i \equiv 1$ under $\mathcal{D}_{\mathcal{T}}^{(i)}$ by Lemma 7.7. It suffices to upper-bound each one of the terms in the above product by $1 - \Omega_{|\Sigma|}(1)$. Fix $i \in \mathcal{J}$. We know that $(P_i Q_i R_i) \not\equiv 1$ under the distribution μ_C. As the SDP solution and the solutions to the GE System are consistent, again using Lemma 7.7, we have $\chi_{P_i}(a) \chi_{Q_i}(b) \chi_{R_i}(c) \neq \chi_{P_i}(a') \chi_{Q_i}(b') \chi_{R_i}(c')$ for some $(a, b, c) \neq (a', b', c')$ in the support of $\mathcal{D}_{\mathcal{T}}^{(i)}$. The probability that this (a, b, c) is sampled according to $\mathcal{D}_{\mathcal{T}}^{(i)}$ is at least $\Omega_{|\Sigma|}(1)$.

Similar to the above proof, the distinct values of $\chi_{P_i} \chi_{Q_i} \chi_{R_i}$ are $\Omega_{|\Sigma|}(1)$ far apart in absolute value. Thus, there is a fixed constant $c_{|\Sigma|} > 0$ such that for $i \in \mathcal{J}$, we have

$$\left| \underset{(a_i, b_i, c_i) \sim \mathcal{D}_{\mathcal{T}}^{(i)}}{\mathbb{E}} [\chi_{P_i}(a_i) \chi_{Q_i}(b_i) \chi_{R_i}(c_i)] \right| \leqslant 1 - c_{|\Sigma|},$$

and the fact follows. □

We now prove the lemma.

Lemma 7.29. *Fix $C \in \mathcal{C}$. Suppose the sizes of each one of $\mathcal{P}_{s_1}, \mathcal{P}_{s_2}, \mathcal{P}_{s_3}$ is at most r, and that the size of each one of $H_{s_1}^{\star}, H_{s_2}^{\star}, H_{s_3}^{\star}$ is at most m. Suppose further that for any $P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1})$, $Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2})$ and $R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3})$, it holds that either $PQR \equiv 1$ under μ_C^R or else $P_i Q_i R_i \not\equiv 1$ under μ_C for at least T of the coordinates $i \in [R]$. Then*

$$\mathsf{SD}(\chi(\mathcal{D}_{\mu_C}), \chi(\mathcal{D}_{\mathcal{T}})) \lesssim_{m,r} (1 - \Omega_{m,\alpha}(1))^T.$$

Proof. Take any tuple $S = (a_P, b_Q, c_R)_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1}), Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2}), R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3})}$ in either one of the supports. Consider their contribution to the two distributions.

$$\chi(\mathcal{D}_{\mathcal{T}})(S) = \mathop{\mathbb{E}}_{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) \sim \mathcal{D}_{\mathcal{T}}} \left[\prod_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1})} 1_{\chi_P(a) = a_P} \prod_{Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2})} 1_{\chi_Q(b) = b_Q} \prod_{R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3})} 1_{\chi_R(c) = c_R} \right]$$

and

$$\chi(\mathcal{D}_{\mu_C})(S) = \mathop{\mathbb{E}}_{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) \sim \mathcal{D}_{\mu_C}} \left[\prod_{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1})} 1_{\chi_P(a) = a_P} \prod_{Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2})} 1_{\chi_Q(b) = b_Q} \prod_{R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3})} 1_{\chi_R(c) = c_R} \right].$$

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 $B(P, Q, R)$ that are at most $O_{m,r}(1)$ in absolute value such that

$$\chi(\mathcal{D}_{\mathcal{T}})(S) = \sum_{\substack{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1}) \\ Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2}) \\ R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3})}} B(P, Q, R) \mathop{\mathbb{E}}_{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) \sim \mathcal{D}_{\mathcal{T}}} [\chi_P(a) \chi_Q(b) \chi_R(c)]$$

and

$$\chi(\mathcal{D}_{\mu_C})(S) = \sum_{\substack{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1}) \\ Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2}) \\ R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3})}} B(P, Q, R) \mathop{\mathbb{E}}_{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}) \sim \mathcal{D}_{\mu_C}} [\chi_P(a) \chi_Q(b) \chi_R(c)].$$

Fix P, Q, R and consider their contribution to $\chi(\mathcal{D}_{\mathcal{T}})(S)$ and $\chi(\mathcal{D}_{\mu_C})(S)$. For P, Q, R such that $PQR \equiv 1$ in the support of μ^R, the two contributions are the same. This follows from Lemma 7.7. Else, by assumption $P_i Q_i R_i \not\equiv 1$ under μ_C for at least T coordinates, and by Fact 7.27 the second expectations is at most $(1 - \Omega_{m,\alpha}(1))^T$ in absolute value. Since the SDP solution is consistent with the GE System, using Fact 7.28, the same holds for the first expectation for the terms corresponding to such $P_i Q_i R_i$. It follows that

$$|\chi(\mathcal{D}_{\mathcal{T}})(S) - \chi(\mathcal{D}_{\mu_C})(S)| \leqslant \sum_{\substack{P \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_1}) \\ Q \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_2}) \\ R \in \mathsf{spn}_{\mathbb{N}}(\mathcal{P}_{s_3})}} |B(P, Q, R)| (1 - \Omega_{m,\alpha}(1))^T \lesssim_{m,r} (1 - \Omega_{m,\alpha}(1))^T.$$

Therefore, we get that $\mathsf{SD}(\chi(\mathcal{D}_{\mu_C}), \chi(\mathcal{D}_{\mathcal{T}})) \lesssim_{m,r} \sum_S (1 - \Omega_{m,\alpha}(1))^T \lesssim_{m,r} (1 - \Omega_{m,\alpha}(1))^T.$ $\qquad\square$

Now, consider the following expectation where both the inputs to the function $\widetilde{\boldsymbol{F}_{s,\sigma}^{\mathsf{dec}}}$ are sampled using global distributions.

$$(III) = \mathop{\mathbb{E}}_{\mathcal{D}_{\mathcal{T}},\mathcal{G}^R}\left[C(\widetilde{\boldsymbol{F}_{s_1}^{\text{dec}}}(\boldsymbol{a}_{s_1},\boldsymbol{g}_{s_1}),\widetilde{\boldsymbol{F}_{s_2}^{\text{dec}}}(\boldsymbol{a}_{s_2},\boldsymbol{g}_{s_2}),\widetilde{\boldsymbol{F}_{s_3}^{\text{dec}}}(\boldsymbol{a}_{s_3},\boldsymbol{g}_{s_3}))\right].$$

As the quantities inside the expectations from (II) and (III) are bounded by $O_q(1)$, using the above lemma, we have

$$|(II) - (III)| \lesssim_q \mathsf{SD}(\chi(\mathcal{D}_{\mu_C}),\chi(\mathcal{D}_{\mathcal{T}})) \lesssim_{m,r} (1 - \Omega_{m,\alpha}(1))^T \lesssim \xi, \tag{44}$$

by choosing an appropriately large value of T.

Combining Equations (42), (43)and (44), we get

$$\left|\Pr[\mathcal{F} \text{ passes } \mathbf{Dict}_{V,\mu}] - \mathop{\mathbb{E}}_{C\sim\mathcal{C}}\left[\mathop{\mathbb{E}}_{(\mathcal{D}_{\mathcal{T}},\mathcal{G}^R)}\left[C(\widetilde{\boldsymbol{F}_{s_1}^{\text{dec}}}(\boldsymbol{a}_{s_1},\boldsymbol{g}_{s_1}),\widetilde{\boldsymbol{F}_{s_2}^{\text{dec}}}(\boldsymbol{a}_{s_2},\boldsymbol{g}_{s_2}),\widetilde{\boldsymbol{F}_{s_3}^{\text{dec}}}(\boldsymbol{a}_{s_3},\boldsymbol{g}_{s_3}))\right]\right]\right| \lesssim \xi. \tag{45}$$

The second expectation in the above expression almost looks like the approximation guarantee of our rounding algorithm given in Figure 4. The only difference is that we scale the vector $\widetilde{\boldsymbol{F}_s^{\text{dec}}}(\boldsymbol{a},\boldsymbol{g}) \in \mathbb{R}^q$ in step 4. We now analyze the loss that occurs because of this scaling. Let $\boldsymbol{F}_s^{\text{dec}}(\boldsymbol{a},\boldsymbol{g})^\star \in \blacktriangle_q$ be the vector obtained by scaling it using the function $\mathbf{Scale}$ defined in step 4 of the rounding scheme $\mathbf{Round}_{\mathcal{F}}$ below. The expected value of the solution returned by the rounding scheme $\mathbf{Round}_{\mathcal{F}}$ is given by:

$$\mathbf{Round}_{\mathcal{F}}(V,\boldsymbol{\mu}) = \mathop{\mathbb{E}}_{C\sim\mathcal{C}}\left[\mathop{\mathbb{E}}_{(\mathcal{D}_{\mathcal{T}},\mathcal{G}^R)}\left[C(\boldsymbol{F}_{s_1}^{\text{dec}}(\boldsymbol{a}_{s_1},\boldsymbol{g}_{s_1})^\star,\boldsymbol{F}_{s_2}^{\text{dec}}(\boldsymbol{a}_{s_2},\boldsymbol{g}_{s_2})^\star,\boldsymbol{F}_{s_3}^{\text{dec}}(\boldsymbol{a}_{s_3},\boldsymbol{g}_{s_3})^\star)\right]\right]. \tag{46}$$

Let $\delta = o_\xi(1)$. Here, the notation $o_\xi(1)$ means that the expression goes to 0 as ξ goes to 0. Fix a constraint C and a variable $s \in \mathcal{V}(C)$. Let $(\boldsymbol{z}_s,\boldsymbol{z}_s',\boldsymbol{Z}_s)$ to be distributed according to the coupled distribution from Lemma 6.3 for κ such that $T^{-1} \ll \kappa \ll d^{-1},\xi$. Let E_s be the event that $\sum_j \mathfrak{Re}(\tilde{\mathcal{F}}_{s,j}^{\text{dec}}(\boldsymbol{z}_s,\boldsymbol{z}_s')) \in [1-\delta,1+\delta]$. The next claim shows that the event E_s occurs with high probability over $(\boldsymbol{z}_s,\boldsymbol{z}_s')$. We will use this later to show that the effect of truncation and scaling has little effect on the expression (46).

Claim 7.30. *For every $s \in \mathcal{V}$, $\Pr[E_s] \geqslant 1 - \delta$.*

Proof. Let $1 \gg \delta' > 0$ to be fixed later. Define $\varrho'(x)$ where $\varrho'(x) = |\mathfrak{Re}(x)|$ if $\mathfrak{Re}(x) \in [-\delta'^2,\delta'^2]$ and 1 otherwise. Consider the function $\zeta(b) = |1 - b|$ such that the function. Let $(\boldsymbol{z}_s,\boldsymbol{z}_s',\boldsymbol{Z}_s)$ to be distributed according the coupled distribution from Lemma 6.3. As $\zeta(1+b) = |b|$ for every $b \in \mathbb{C}$ we get

$$\mathop{\mathbb{E}}_{(\boldsymbol{z}_s,\boldsymbol{z}_s',\boldsymbol{Z}_s)}\left[\varrho' \circ \zeta\left(\sum_j \tilde{\mathcal{F}}_{s,j}^{\text{dec}}(\boldsymbol{z}_s,\boldsymbol{z}_s')\right)\right]$$

$$= \mathop{\mathbb{E}}_{(\boldsymbol{z}_s,\boldsymbol{z}_s',\boldsymbol{z}_s)}\left[\varrho' \circ \zeta\left(\sum_j \tilde{\mathcal{F}}_{s,j}^{\text{dec}}(\boldsymbol{z}_s,\boldsymbol{z}_s') + \sum_j \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(\boldsymbol{Z}_s) - \sum_j \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(\boldsymbol{Z}_s)\right)\right]$$

$$= \mathop{\mathbb{E}}_{(\boldsymbol{z}_s,\boldsymbol{z}_s',\boldsymbol{z}_s)}\left[\varrho'\left(\left|\sum_j \tilde{\mathcal{F}}_{s,j}^{\text{dec}}(\boldsymbol{z}_s,\boldsymbol{z}_s') - \sum_j \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(\boldsymbol{Z}_s)\right|\right)\right]. \qquad \text{(Using Claim 7.26)}$$

66

Now, using the fact that $\varrho'(|x|) \leqslant \delta'^{-2}|x|$ for all x, we have

$$\mathop{\mathbb{E}}_{(z_s,z'_s,Z_s)}\left[\varrho' \circ \zeta\left(\sum_j \tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s)\right)\right] \leqslant \delta'^{-2} \mathop{\mathbb{E}}_{(z_s,z'_s,Z_s)}\left[\left|\sum_j \tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s) - \sum_j \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(Z_s)\right|\right]$$

$$\leqslant \delta'^{-2} \mathop{\mathbb{E}}_{(z_s,z'_s,Z_s)}\left[\left|\sum_j \tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s) - \sum_j \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(Z_s)\right|^2\right]^{1/2}$$

$$\lesssim_q \delta'^{-2} \sum_j \mathop{\mathbb{E}}_{(z_s,z'_s,Z_s)}\left[\left|\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s) - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(Z_s)\right|^2\right]^{1/2},$$

where the second line is the Cauchy-Schwarz inequality. Now,

$$\mathop{\mathbb{E}}_{(z_s,z'_s,Z_s)}\left[\left|\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s) - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(Z_s)\right|^2\right]$$

$$= \sum_{j=1}^{q} \mathbb{E}\left[\left|\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s) - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(Z_s)\right|_2^2\right]$$

$$= \sum_{j=1}^{q} \mathbb{E}\left[\left|\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s) - \mathcal{F}'_{s,j}(Z_s) + \mathcal{F}'_{s,j}(Z_s) - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(Z_s)\right|_2^2\right]$$

$$\lesssim \sum_{j=1}^{q} \mathbb{E}\left[\left|\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s) - \mathcal{F}'_{s,j}(Z_s)\right|^2\right] + \mathbb{E}\left[\left|\mathcal{F}'_{s,j}(Z_s) - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(Z_s)\right|_2^2\right].$$

Using Lemma 4.9, the second expectation is upper bounded by ξ and using Lemma 6.6, the first expectation is upper bounded by $\lesssim_{d,q} \kappa \ll \xi$. Thus, we have for every $1 \leqslant j \leqslant q$,

$$\mathop{\mathbb{E}}_{(z_s,z'_s,Z_s)}\left[\left|\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s) - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(Z_s)\right|^2\right] = o_\xi(1).$$

Therefore, we get

$$\mathop{\mathbb{E}}_{(z_s,z'_s,Z_s)}\left[\varrho' \circ \zeta\left(\sum_j \tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s)\right)\right] \lesssim_q \frac{o_\xi(1)}{\delta'^2}. \tag{47}$$

For some setting of $\delta \approx \delta'^{-2}o_\xi(1)$, we get that with probability at least $1 - \delta$, $\sum_j \mathfrak{Re}(\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(z_s,z'_s)) \in [1 - \delta, 1 + \delta]$. This finishes the proof of the claim. $\square$

We now remove the truncation and scaling, and see the effect of this on the expression from (46). Towards this, we have the following simple claim.

Claim 7.31. *Let $a \in \mathbb{R}^q$ such that $\sum_i a_i \in [1 - \delta, 1 + \delta]$, and let $a^\star$ be the vector that we get after scaling the vector $\tilde{a} := \mathsf{trunc}_{[0,1]}(a)$ (as defined in Figure 4), then*

$$\sum_i (\tilde{a}_i - a_i^\star)^2 \leqslant q \sum_i (a_i - \tilde{a}_i)^2 + O_q(\delta).$$

Proof. First, note that $\sum_i \tilde{a}_i > 0$ if $\sum_i a_i \in [1-\delta, 1+\delta]$. We have

$$\sum_i (\tilde{a}_i - a_i^\star)^2 = \sum_i \left(\tilde{a}_i - \frac{\tilde{a}_i}{\sum_j \tilde{a}_j} \right)^2 = \left(\sum_i \tilde{a}_i - 1 \right)^2 \sum_i \frac{\tilde{a}_i^2}{(\sum_j \tilde{a}_j)^2} \leqslant \left(\sum_i \tilde{a}_i - 1 \right)^2 ,$$

where we used $(\sum_j \tilde{a}_j)^2 \geqslant \sum_j \tilde{a}_j^2$ which holds as $\tilde{a}_j$ are non-negative. 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. $\qquad\square$

Claim 7.32. *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 (46), then*

$$\left| \begin{array}{l} \mathbb{E}\left[C(\boldsymbol{F}_{s_1}^{\mathsf{dec}}(\boldsymbol{a}_{s_1}, \boldsymbol{g}_{s_1})^\star, \boldsymbol{F}_{s_2}^{\mathsf{dec}}(\boldsymbol{a}_{s_2}, \boldsymbol{g}_{s_2})^\star, \boldsymbol{F}_{s_3}^{\mathsf{dec}}(\boldsymbol{a}_{s_3}, \boldsymbol{g}_{s_3})^\star) \right] - \\ \mathbb{E}\left[C(\widetilde{\boldsymbol{F}_{s_1}^{\mathsf{dec}}}(\boldsymbol{a}_{s_1}, \boldsymbol{g}_{s_1}), \widetilde{\boldsymbol{F}_{s_2}^{\mathsf{dec}}}(\boldsymbol{a}_{s_2}, \boldsymbol{g}_{s_2}), \widetilde{\boldsymbol{F}_{s_3}^{\mathsf{dec}}}(\boldsymbol{a}_{s_3}, \boldsymbol{g}_{s_3})) \right] \end{array} \right| \leqslant o_\tau(1) + o_\xi(1) + O_q(\delta).$$

Proof. Throughout this proof, we will use the following simple fact related to adding and removing conditioning on the event happening with high probability. If X is a non negative random variable which is c-bounded and E be the event that $X \in [1-\delta, 1+\delta]$ with $\Pr[E] \geqslant 1 - \eta \geqslant 1/2$, then $\mathbb{E}[X] \leqslant \mathbb{E}[X \mid E] + \eta \cdot c$. Also, $\mathbb{E}[X \mid E] \leqslant 2 \cdot \mathbb{E}[X]$.

We now begin with the 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

$$C_0(q) \cdot \sum_{j=1}^3 \mathbb{E}\left[\| \boldsymbol{F}_{s_j}^{\mathsf{dec}}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j})^\star - \widetilde{\boldsymbol{F}_{s_j}^{\mathsf{dec}}}(\boldsymbol{a}_{s_j}, \boldsymbol{g}_{s_j}) \|_2 \right] .$$

Using Lemma 7.29, the fact that the expectation is bounded by $O_q(1)$, we get that the LHS can be upper bounded by

$$C_0(q) \cdot \sum_{j=1}^3 \underbrace{\mathbb{E}\left[\| \boldsymbol{F}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{g}_{s_j})^\star - \widetilde{\boldsymbol{F}_{s_j}^{\mathsf{dec}}}(\boldsymbol{z}_{s_j}, \boldsymbol{g}_{s_j}) \|_2 \right]}_{(I)} + o_\xi(1).$$

For each fixing of $\boldsymbol{z}_{s_j}$, the corresponding function has all the influences bounded by $\lesssim_q \tau$. Applying the invariance principle from Theorem 3.12, we get

$$\mathbb{E}\left[\| \boldsymbol{F}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{g}_{s_j})^\star - \widetilde{\boldsymbol{F}_{s_j}^{\mathsf{dec}}}(\boldsymbol{z}_{s_j}, \boldsymbol{g}_{s_j}) \|_2 \right] \leqslant \mathbb{E}\left[\| \tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})^\star - \widetilde{\tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}) \|_2 \right] + o_\tau(1)$$

Overall, the LHS from the claim is upper bounded by

$$C_0(q) \cdot \sum_{j=1}^3 \mathbb{E}\left[\| \tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})^\star - \widetilde{\tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}) \|_2 \right] + o_\xi(1) + o_\tau(1).$$

In this expression, $(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})$ are distributed according to the distribution $\mu_{s_j}^R \times \mu_{s_j}^R$. Consider $(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}, \boldsymbol{Z}_{s_j})$ the distributed from Lemma 6.3 with the same κ as required for Claim 7.30. As $T \gg d, \kappa^{-1}$, using

Lemma 6.3, the statistical distance between these two distributions on $(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})$ can be upper bounded by $o_\xi(1)$. The quantity inside the expectation is $O_q(1)$ bounded, and hence, we can move to the coupled distribution by incurring an error of $o_\xi(1)$. Thus, we assume this distribution on $(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})$ now onwards. Hence, the above quantity is upper bounded by

$$C_0(q) \cdot \sum_{j=1}^{3} \mathbb{E}_{(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}, \boldsymbol{Z}_{s_j})} \left[\| \tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})^\star - \widetilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}) \|_2 \right] + o_\xi(1) + o_\tau(1),$$

It follows from Claim 7.30 that $\Pr[E_{s_j}] \geqslant 1 - \delta$. Hence, the above quantity is upper bounded by

$$O_q(1) \cdot \sum_{j=1}^{3} \mathbb{E} \left[\| \tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})^\star - \widetilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}) \|_2 \mid E_{s_j} \right] + o_\xi(1) + o_\tau(1) + O_q(\delta).$$

Using Claim 7.31, the above quantity (and hence the LHS from the claim) is upper bounded by

$$O_q(1) \cdot \sum_{j=1}^{3} \mathbb{E} \left[\| \mathfrak{R}(\tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})) - \widetilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}) \|_2 \mid E_{s_j} \right] + o_\xi(1) + o_\tau(1) + O_q(\delta).$$

We have:

$$\mathbb{E} \left[\| \mathfrak{Re}(\tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})) - \widetilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}) \|_2 \mid E_{s_j} \right]$$

$$\leqslant \frac{1}{\Pr[E_{s_j}]} \mathbb{E} \left[\| \mathfrak{Re}(\tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})) - \widetilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}) \|_2 \right]$$

$$\leqslant 2 \mathbb{E} \left[\| \mathfrak{Re}(\tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})) - \widetilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}) \|_2 \right],$$

using the fact that $\Pr[E_{s_j}] = 1 - \delta \geqslant 1/2$. Therefore, the LHS from the claim is upper bounded by

$$O_q(1) \cdot \sum_{j=1}^{3} \mathbb{E} \left[\| \mathfrak{Re}(\tilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j})) - \widetilde{\boldsymbol{\mathcal{F}}}_{s_j}^{\mathsf{dec}}(\boldsymbol{z}_{s_j}, \boldsymbol{z}'_{s_j}) \|_2 \right] + o_\xi(1) + o_\tau(1) + O_q(\delta). \tag{48}$$

Now, for a given variable $s \in \mathcal{V}$, define

$$\Theta_s := \mathbb{E} \left[\| \mathfrak{Re}(\tilde{\boldsymbol{\mathcal{F}}}_s^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}'_s)) - \widetilde{\boldsymbol{\mathcal{F}}}_s^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}'_s) \|_2 \right] = \mathbb{E} \left[\varrho(\tilde{\boldsymbol{\mathcal{F}}}_s^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}'_s)) \right], \tag{49}$$

where $\varrho(a_1, \ldots, a_q) = \sqrt{\sum_{i=1}^{q} \left| \mathsf{trunc}_{[0,1]}(\mathfrak{Re}(a_i)) - \mathfrak{Re}(a_i) \right|^2}$. Using the fact that ϱ is $O(1)$ Lipschitz,

$$\mathbb{E} \left[\varrho \left(\tilde{\boldsymbol{\mathcal{F}}}_s^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}'_s) \right) \right] \lesssim \mathbb{E} \left[\varrho \left(\mathrm{T}_{\mathcal{P}_s, 1-\varepsilon} \boldsymbol{\mathcal{F}}_s(\boldsymbol{Z}_s) \right) + \| \tilde{\boldsymbol{\mathcal{F}}}_s^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}'_s) - \mathrm{T}_{\mathcal{P}_s, 1-\varepsilon} \boldsymbol{\mathcal{F}}_s(\boldsymbol{Z}_s) \|_2 \right].$$

As $\mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j} \in [0,1]$ for $1 \leqslant j \leqslant q$, $\varrho\left(\mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_s(\boldsymbol{z}_s)\right) = 0$. Therefore, we have

$$
\begin{aligned}
\Theta_s^2 &\lesssim \mathbb{E}\left[\|\tilde{\boldsymbol{\mathcal{F}}}_s^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}_s') - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\boldsymbol{\mathcal{F}}_s(\boldsymbol{Z}_s)\|_2\right]^2 \\
&\leqslant \mathbb{E}\left[\|\tilde{\boldsymbol{\mathcal{F}}}_s^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}_s') - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\boldsymbol{\mathcal{F}}_s(\boldsymbol{Z}_s)\|_2^2\right] \\
&= \sum_{j=1}^{q} \mathbb{E}\left[\left|\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}_s') - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(\boldsymbol{Z}_s)\right|_2^2\right] \\
&= \sum_{j=1}^{q} \mathbb{E}\left[\left|\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}_s') - \mathcal{F}_{s,j}'(\boldsymbol{Z}_s) + \mathcal{F}_{s,j}'(\boldsymbol{Z}_s) - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(\boldsymbol{Z}_s)\right|_2^2\right] \\
&\lesssim \sum_{j=1}^{q} \mathbb{E}\left[\left|\tilde{\mathcal{F}}_{s,j}^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}_s') - \mathcal{F}_{s,j}'(\boldsymbol{Z}_s)\right|^2\right] + \mathbb{E}\left[\left|\mathcal{F}_{s,j}'(\boldsymbol{Z}_s) - \mathrm{T}_{\mathcal{P}_s,1-\varepsilon}\mathcal{F}_{s,j}(\boldsymbol{Z}_s)\right|_2^2\right].
\end{aligned}
$$

Using Lemma 4.9 and Lemma 6.4 we can bound the first and second expectations from each summand by $o_\xi(1)$ respectively to get that $\Theta_s \lesssim o_\xi(1)$. Therefore,

$$
\mathbb{E}\left[\varrho(\tilde{\boldsymbol{\mathcal{F}}}_s^{\mathsf{dec}}(\boldsymbol{z}_s, \boldsymbol{z}_s'))\right] \leqslant o_\xi(1). \tag{50}
$$

Combining Equations (48), (49), and (50), we get that the LHS from the claim is upper bounded by $o_\tau(1) + o_\xi(1) + O_q(\delta)$ as claimed. $\qquad\square$

Using the above claim and Equation(46), and the fact that $\tau \ll \xi$, we get

$$
\mathbf{Round}_{\mathcal{F}}(\boldsymbol{V}, \boldsymbol{\mu}) \geqslant \mathop{\mathbb{E}}_{C \sim \mathcal{C}}\left[\mathop{\mathbb{E}}_{(\mathcal{D}_{\mathcal{T}}, \mathcal{G}^R)}\left[C(\widetilde{\boldsymbol{F}_{s_1}^{\mathsf{dec}}}(\boldsymbol{a}_{s_1}, \boldsymbol{g}_{s_1}), \widetilde{\boldsymbol{F}_{s_2}^{\mathsf{dec}}}(\boldsymbol{a}_{s_2}, \boldsymbol{g}_{s_2}), \widetilde{\boldsymbol{F}_{s_3}^{\mathsf{dec}}}(\boldsymbol{a}_{s_3}, \boldsymbol{g}_{s_3}))\right]\right] - o_\xi(1). \tag{51}
$$

Combining with (45) gives

$$
\Pr[\mathcal{F} \text{ passes } \mathbf{Dict}_{V,\mu}] \leqslant \mathbf{Round}_{\mathcal{F}}(\boldsymbol{V}, \boldsymbol{\mu}) + o_\xi(1).
$$

As the integral value of the instance Υ is at most $\mathsf{OPT}(\Upsilon)$, we have $\mathbf{Round}_{\mathcal{F}}(\boldsymbol{V}, \boldsymbol{\mu}) \leqslant \mathsf{OPT}(\Upsilon)$ it follows that

$$
\Pr[\mathcal{F} \text{ passes } \mathbf{Dict}_{V,\mu}] \leqslant \mathsf{OPT}(\Upsilon) + o_\xi(1),
$$

finishing the proof of Theorem 7.21. $\qquad\square$

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 $\mathcal{F}_s$ denote the function obtained by interpreting the function $\mathcal{F} : \Sigma^R \to \blacktriangle_q$ 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 $\mathcal{T} \subseteq G_{\mathsf{master}}^{\mathcal{V}}$ be the subspace of the set of satisfying assignments of the instance $\Upsilon(\mathcal{V}, \mathcal{C})$ over $G_{\mathsf{master}}^{\mathcal{V}} := \prod_{v \in \mathcal{V}} H_v^\star$ that satisfy the conditions from Lemma 7.15.

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 $g_{s,0}^{(j)} \equiv 1$ and for $c \in \{1, \dots, q-1\}$, set

$$g_{s,c}^{(j)} = \sum_{\omega \in \Sigma} \ell_{s,c}(\omega) \langle \boldsymbol{b}_{s,\omega}, \boldsymbol{\zeta}^{(j)} \rangle.$$

   Let $\boldsymbol{g}_s^{(j)} = (g_{s,0}^{(j)} \equiv 1, g_{s,1}^{(j)}, \dots, g_{s,q-1}^{(j)})$ and $\boldsymbol{g}_s = (\boldsymbol{g}_s^{(1)}, \boldsymbol{g}_s^{(2)}, \dots, \boldsymbol{g}_s^{(R)})$.

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 $G_{\mathsf{master}}^{\mathcal{V}}$. Let $\boldsymbol{a}_s = (\boldsymbol{\alpha}_s^{(1)}, \boldsymbol{\alpha}_s^{(2)}, \dots, \boldsymbol{\alpha}_s^{(R)})$.

3. Evaluate the polynomial $\boldsymbol{F}_s^{\mathsf{dec}}$ with $(\boldsymbol{a}_s, \boldsymbol{g}_s)$ as inputs to obtain $\boldsymbol{p}_s \in \mathbb{C}^q$, and let $\tilde{\boldsymbol{p}}_s \in \mathbb{C}^q$ be such that $(\tilde{\boldsymbol{p}}_s)_j = \mathsf{trunc}_{[0,1]}(\mathfrak{Re}((\boldsymbol{p}_s)_j))$ where

$$\mathsf{trunc}_{[0,1]}(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leqslant x \leqslant 1, \\ 1 & \text{if } x > 1, \end{cases}$$

4. Round $\tilde{\boldsymbol{p}}_s$ to $\boldsymbol{p}_s^\star$.

$$\boldsymbol{p}_s^\star = \mathbf{Scale}((\tilde{\boldsymbol{p}}_s)_1, (\tilde{\boldsymbol{p}}_s)_2, \dots, (\tilde{\boldsymbol{p}}_s)_q),$$

   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 $(\boldsymbol{p}_s^\star)_{\varsigma^{-1}(a)}$.

Step III: Output the assignment from Step II.

Figure 4: Rounding Scheme $\mathbf{Round}_{\mathcal{F}}$.

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] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

[32] Venkatesan Guruswami and Euiwoong Lee. Strong inapproximability results on balanced rainbow-colorable hypergraphs. *Combinatorica*, 38(3):547–599, 2018.

[33] Alfred W Hales and Robert I Jewett. Regularity and positional games. In *Classic Papers in Combinatorics*, pages 320–327. Springer, 2009.

[34] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.

[35] Johan Håstad. On the NP-hardness of max-not-2. *SIAM Journal on Computing*, 43(1):179–193, 2014.

[36] David Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, mar 1998.

[37] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th STOC*, pages 767–775. ACM, 2002.

[38] 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.

[39] 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.

[40] Richard E Ladner. On the structure of polynomial time reducibility. *Journal of the ACM (JACM)*, 22(1):155–171, 1975.

[41] Leonid Anatolevich Levin. Universal sequential search problems. *Problemy peredachi informatsii*, 9(3):115–116, 1973.

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

[43] 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.

[44] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[45] 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.

[46] 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.

[47] DHJ Polymath. A new proof of the density Hales-Jewett theorem. *Annals of Mathematics*, pages 1283–1327, 2012.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] Dmitriy Zhuk. A proof of the CSP dichotomy conjecture. *J. ACM*, 67(5), August 2020.

[55] Uri Zwick. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *SODA*, volume 98, pages 201–210, 1998.

[56] 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 $0, 1$.

Let $A = \{ (b, a + b, az + b) \mid a \neq 0, b \in \mathbb{F}_p \}$ be the orbit of $(0, 1, z)$ under $\{\tau_{a,b}\}_{a \neq 0, b}$ for $z \neq 0, 1$, 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 $\phi(\alpha) = \phi(\beta) + \phi(\alpha - \beta)$. 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 $(z-1)\beta - z\alpha = 0$. If $\beta = 0$ the equality $\phi(\alpha) = \phi(\beta) + \phi(\alpha - \beta)$ is clear, so assume otherwise. It follows that $\alpha = \frac{z-1}{z}\beta$, so $\alpha \neq 0, \beta$. Choose $\alpha' \in \mathbb{F}_p \setminus \{\alpha, 0, \frac{z}{z-1}\alpha\}$ arbitrarily and then $\beta' \in \mathbb{F}_p$ uniformly. With probability at least $1 - \frac{3}{p} > 0$ we have that $(z-1)\beta' - z\alpha' \neq 0$, $(z-1)(\beta - \beta') - z(\alpha - \alpha') \neq 0$, $(z-1)\beta' - z\beta \neq 0$, $(z-1)(\alpha' - \beta') - z(\alpha - \beta) \neq 0$.[14] Also, by choice of α' we have $(z-1)\alpha' - z\alpha \neq 0$ and $\alpha' \neq 0, \alpha, \beta, \beta \neq 0$ and $\alpha \neq 0$. Fix α', β' 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 fourth 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. $\qquad \square$

---

[14]Note that we have 4 conditions, so a naive application of the union bound only gives a lower bound of $1 - \frac{4}{p}$ 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 $1 - \frac{3}{p}$.