Electronic Colloquium on Computational Complexity, Report No. 61 (2022)

On Approximability of Satisfiable k-CSPs: I

Amey Bhangale* Subhash Khot† Dor Minzer‡

Abstract

We consider the P-CSP problem for 3-ary predicates P on satisfiable instances. We show that under certain conditions on P and a $(1, s)$ *integrality gap* instance of the P-CSP problem, it can be translated into a dictatorship vs. quasirandomness test with perfect completeness and soundness $s + \varepsilon$, for every constant $\varepsilon > 0$. Compared to Ragahvendra's result [Rag08], we do not lose perfect completeness. This is particularly interesting as this test implies new hardness results on satisfiable constraint satisfaction problems, assuming the Rich 2-to-1 Games Conjecture by Braverman, Khot, and Minzer [BKM21]. Our result can be seen as the first step of a potentially long-term challenging program of characterizing optimal inapproximability of every satisfiable k-ary CSP.

At the heart of the reduction is our main analytical lemma for a class of 3-ary predicates, which is a generalization of a lemma by Mossel [Mos10]. The lemma and a further generalization of it that we propose as a hypothesis may be of independent interest.

1 Introduction

Constraint satisfaction problems (CSPs) are some of the most fundamental problems in computer science. Given a predicate $P : \Sigma^k \to \{0, 1\}$, for some alphabet Σ, a P-CSP instance consists of a set of variables $x_1, x_2, \ldots, x_n$ and a collection of *local* constraints $C_1, C_2, \ldots, C_m$. Each constraint is of the type $P(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$. The constraints might involve *literals* instead of just the variables.[1] An algorithmic task is to decide if there exists an assignment to the variables that satisfies all the constraints. In a related problem, called the Max-P-CSP problem, the task is to find an assignment to the variables that satisfies the maximum fraction of the constraints. An α-approximation algorithm is a polynomial-time algorithm which always returns an assignment that satisfies at least $\alpha \cdot \textsc{Opt}$ fraction of the constraints, where $\textsc{Opt}$ is the value of the optimum assignment. The focus of the current work is on approximability of fully satisfiable instances.

A systematic study of the complexity of solving CSPs was started by Schaefer in 1978 [Sch78] who showed that for every P over a 2-element alphabet, the problem of checking satisfiability of a P-CSP is either in **P** or is **NP**-complete. A famous Dichotomy Conjecture of Feder and Vardi [FV98], which was resolved recently in huge breakthroughs by Bulatov and Zhuk independently [Bul17, Zhu20], states that for every P, checking satisfiability of a P-CSP is either in **P** or is **NP**-complete.

However, when it comes to designing optimal approximation algorithms for Max-P-CSP on fully satisfiable instances, the question is wide open. The PCP Theorem [FGL+96, ALM+98, AS98] proved in the early 90s shows that it is **NP**-hard to approximate many P-CSPs within a constant factor $\alpha < 1$. This was vastly improved in a seminal result by Håstad [Hås01] for certain CSPs. Håstad showed that for many CSPs, it is **NP**-hard to do better than the approximation factor achieved by a random assignment. More specifically, he showed that 3SAT cannot be approximated better than $\frac{7}{8} + \varepsilon$ for any constant $\varepsilon > 0$ in polynomial time

*Department of Computer Science and Engineering, University of California, Riverside. Email: ameyrb@ucr.edu.

†Department of Computer Science, Courant Institute of Mathematical Sciences, New York University. Supported by the NSF Award CCF-1422159, NSF CCF award 2130816, the Simons Collaboration on Algorithms and Geometry, and the Simons Investigator Award. Email: khot@cs.nyu.edu .

‡Department of Mathematics, Massachusetts Institute of Technology, Cambridge, USA. Supported by a Sloan Research Fellowship. Email: dminzer@mit.edu.

[1] See Definition 2.7 and Remark 2.8 for more details.

ISSN 1433-8092

unless $\mathbf{P} = \mathbf{NP}$. Note that if we select a random assignment, then it satisfies $\frac{7}{8}$-fraction of the clauses in expectation. The result proved in [Hås01] is stronger than what is stated – even if we know that a given instance is *fully satisfiable*, i.e., there exists an assignment that satisfies all the clauses, it is $\mathbf{NP}$-hard to come up with an assignment that satisfies more than $(\frac{7}{8} + \varepsilon)$-fraction of the clauses for any constant $\varepsilon > 0$.

Håstad also showed that it is $\mathbf{NP}$-hard to find an assignment to a given 3LIN instance[2] that satisfies more that $(\frac{1}{2} + \varepsilon)$-fraction of the constraints, even if we are guaranteed that there exists an assignment that satisfies $(1 - \varepsilon)$-fraction of the constraints. This is interesting because unlike 3SAT, we can in fact find an assignment that satisfies all the constraints of a given 3LIN instance, if there exists one, in polynomial time. Thus, knowing that a given instance of P-CSP is fully satisfiable, in principle, can be used to design better approximation algorithms for Max-P-CSPs. In this paper, we study the inapproximability of fully satisfiable instances. On the other hand, as we will explain next, if the instance is almost satisfiable, then by Raghavendra's work [Rag08], we know the precise approximation threshold for every P-CSP and the optimal algorithm is given by semi-definite programming.

In order to gain better understanding of complexity of approximation algorithms for various optimization problems, Khot [Kho02] in 2002 proposed the Unique Games Conjecture (UGC). Since then, for various optimization problems, we now know the precise approximation factor that one can achieve in polynomial time assuming the UGC. Max-Cut is one of the simplest CSPs in which the constraints are of the type $x \oplus y = 1$. Goemans and Williamson [GW95] gave a α_{GW}-approximation algorithm for Max-Cut problem where $\alpha_{GW} \approx 0.878$. Surprisingly, [KKMO07] showed that the approximation algorithm by Goemans and Williamson is tight assuming the UGC. Their hardness result relied on the 'Majority is Stablest' theorem which was proved in [MOO05].

For general CSPs, Austrin and Mossel [AM09] gave a very simple sufficient criterion for a predicate P to be *approximation resistant*. A predicate P is called approximation resistant if it is $\mathbf{NP}$-hard (or UG-hard) to achieve an approximation algorithm better than the random assignment algorithm. 3SAT and 3LIN predicates described above are examples of approximation resistant predicates. Austrin and Mossel showed that if there exists a distribution supported only on the satisfying assignments in P, which is balanced and pairwise independent, then P is approximation resistant assuming the UGC.

The Max-Cut hardness result was beautifully generalized to *all* constraint satisfaction problems by Raghavendra [Rag08]. More specifically, he showed that for any P-CSP problem, if there exists a (c, s) basic SDP integrality gap instance[3], then it is UG-hard to find an assignment that satisfies $(s + \varepsilon)$ fraction of the constraints, even if the given instance is $(c - \varepsilon)$-satisfiable, for every constant $\varepsilon > 0$. For all $c \in (0, 1]$, let $s(c)$ be the infimum value such that there exists an $(c, s(c))$ integrality gap instance. By definition, the SDP relaxation promises $s(c)$ satisfying assignment on every c-satisfiable instance. Raghavendra gives the rounding algorithm that actually finds the $s(c)$-satisfying assignment. Thus, Raghavendra's result gives a complete answer to the complexity of approximating Max-P-CSP assuming the UGC. However, it does not imply hardness on instances that are fully satisfiable. This is because in translating the integrality gap parameters (c, s) to hardness parameter, there is always a loss of some small constant $\varepsilon > 0$ in the completeness parameter.

The most important building-block in Raghavendra's result (and also in many prior works) is the dictatorship test. A function $f : \Sigma^n \to \Sigma$ is called a dictator function if it depends only on one variable. A dictatorship test is a procedure which queries f at a few (correlated) locations randomly and based on the function values at these locations, it decides if f is a dictator function or *far* from any dictator function (also referred to as *quasirandom* functions). We briefly describe the notion of being *far* from dictator functions here. Influence of a coordinate i in a function f is the probability that for a random input $(x_1, x_2, \dots, x_n)$, f changes its value if we change the i^{th} coordinate. Note that dictator functions have one coordinate whose influence is 1. A function is called far from dictator functions if for every coordinate i, the influence of the coordinate i in f is small.

There are three important properties of the test which are useful in getting hardness of approximation result for Max-P-CSP. The first one is the *completeness parameter* c — this is the probability that the test accepts any dictator function. The second property is the *soundness parameter* s — this is the probability with which

[2] This CSP is over the Boolean domain and constraints are of the type $x_{i_1} \oplus x_{i_2} \oplus x_{i_3} = 1/0$.

[3] See Definition 2.16 for the formal definition.

the test accepts far from dictator functions. The third property is the decision predicate that the test uses in accepting or rejecting the function f. If the decision predicate is P and the test has completeness c and soundness s, then such a test can be translated into a UG-hardness result for Max-P-CSP with completeness $(c-\varepsilon)$ and soundness $(s+\varepsilon)$, for any constant $\varepsilon > 0$. In other words, it is UG-hard to find an assignment that satisfies $(s+\varepsilon)$ fraction of the constraints, even if the given instance is $(c-\varepsilon)$-satisfiable, for every constant $\varepsilon > 0$.

Raghavendra proved his result by designing a dictatorship test starting with a (c, s) integrality gap instance for Max-P-CSP such that the test has completeness $(c-\varepsilon)$ and soundness $(s+\varepsilon)$, for any constant $\varepsilon > 0$. Therefore, his test loses in the completeness parameter and hence cannot be used in proving hardness result on satisfiable instances. Note that even if the completeness parameter of the test is c, because of the conjectured hardness of Unique Games, one still loses small constant ε in the completeness parameter of the final UG-hardness result.[4] In order to save this loss, Braverman, Khot, and Minzer [BKM21] proposed a Rich 2-to-1 Games Conjecture and if we use this instead of Unique Games, then there is no loss in the completeness parameter. Therefore, it is important that we do not lose anything in the completeness parameter when designing the dictatorship test.

In this work, we initiate a systematic study of completely characterizing the precise approximability of every k-ary CSP on satisfiable instances (recognizing, of course, that the prior works have obtained such a characterization for specific predicates, e.g., 3SAT). In order to answer this challenging question, it was necessary first to understand the complexity of checking satisfiability of CSP which is the famous Dichotomy Conjecture. Now that this conjecture is resolved, we can embark on the study of approximability of satisfiable CSPs.

As with the case with 3SAT and 3LIN, a predicate being linear makes a big difference on the complexity of the CSP. Addressing this issue of linearity is also a challenging aspect in the proof of the Dichotomy Conjecture. In this work, we take the first step by considering special class of non-linear predicates. We show how to convert any $(1, s)$-integrality gap instance of a 3-ary CSP to a dictatorship test with completeness 1 and soundness $s + \varepsilon$, for any constant $\varepsilon > 0$. For our conclusion to hold, we need a few additional properties from the predicate as well as from the integrality gap instance that we describe next.

- *Predicates not satisfying any linear embedding:* Given a predicate $P\colon \Sigma^3 \to \{0, 1\}$, it is said to satisfy a linear equation if there exists an Abelian group $(G, +)$ and 3 embeddings $\sigma : \Sigma \to G$, $\phi : \Sigma \to G$ and $\gamma : \Sigma \to G$ such that the following hold: At least one of the embeddings is non-constant and for every tuple $(x, y, z) \in P^{-1}(1)$, $\sigma(x) + \phi(y) + \gamma(z) = 0$ where 0 is the identity element of G.
- *Semi-rich predicates:* A predicate $P\colon \Sigma^3 \to \{0, 1\}$ is called semi-rich if for each $(x, y) \in \Sigma \times \Sigma$, there exists a $z \in \Sigma$ such that $(x, y, z) \in P^{-1}(1)$. Also, for every $(x, z) \in \Sigma \times \Sigma$, there exists a $y \in \Sigma$ such that $(x, y, z) \in P^{-1}(1)$.
- *SDP solution that is semi-rich and that is not linearly embeddable:* An SDP solution for a given P-CSP instance consists of a local distribution for each constraint. We say the SDP solution is semi-rich and is not linearly embeddable if the support of every local distribution is semi-rich and is not linearly embeddable in any Abelian group (See Definitions 2.1, 2.2 and 2.3).

We now state our main theorem.

Theorem 1.1. *Let $P\colon \Sigma^3 \to \{0, 1\}$ be any predicate that satisfies the following conditions. (1) P does not satisfy any linear embedding, (2) P is a semi-rich predicate, and (3) there exists an instance of P-CSP that has a $(1, s)$-integrality gap for the basic SDP relaxation and an optimal SDP solution is semi-rich and is not linearly embeddable. Then for every $\varepsilon > 0$, there is a dictatorship test for P that has perfect completeness and soundness $s + \varepsilon$.*

We do not believe that the semi-rich condition is really needed in the theorem, but this is what we could show currently.

[4]Unique Games can hard only on almost satisfiable instances. Therefore, any hardness from Unique Games loses perfect completeness.

In order to focus on designing new dictatorship tests and a new way to analyze the tests, in this work we will not discuss in detail the application of this towards getting the conditional **NP**-hardness results. A recent work by Braverman, Khot, Lifshitz and Minzer [BKLM22] gives some supporting evidence towards dictatorship tests implying hardness results for the predicates satisfying conditions from Theorem 1.1. Note that, the first condition from the hypothesis is *necessary* towards getting the hardness result for the predicate with a gap $(1, s+\varepsilon)$. This can be seen from the Max-3LIN problem on an Abelian group G.[5] This predicate has a linear embedding as well as there exists an instance with a SDP integrality gap of $(1, \frac{1}{|G|}+\varepsilon)$, for every constant $\varepsilon > 0$. However, if the instance is satisfiable, then one can find the satisfying assignment in polynomial time using Gaussian elimination.

It might be instructive to consider a couple of examples of predicates that satisfy the first two conditions.

1. Linear equations over a quasirandom group: Fix any group $(G, \cdot)$ such that any non-trivial irreducible representation of G has dimension greater than 1.[6] Consider the predicate $P_G : G^3 \to \{0,1\}$ where $P_G^{-1}(1) = \{(x,y,z) \mid x \cdot y \cdot z = 1_G\}$, where 1_G is the identity element. The fact that G does not have any non-trivial representation of dimension 1 implies that P does not satisfy any linear embedding. Also, it is easily observed that the predicate is semi-rich.

2. Arithmetic progression over a quasirandom group: For a similar group as above, consider a predicate $P_{AP} : G^3 \to \{0,1\}$ where $P_{AP}^{-1}(1) = \{(x, x\cdot g, x\cdot g^2) \mid x, g \in G\}$. It can be shown that this predicate does not satisfy any linear embedding. To see that P_{AP} is semi-rich, we need to permute the coordinates. Note that permuting the coordinates of a predicate does not really change the complexity of the corresponding CSP problem. By the change of variables $x \cdot g = h$ we can write $P_{AP}^{-1}(1) = \{(h \cdot g^{-1}, h, h \cdot g) \mid h, g \in G\}$. We can permute the coordinates to get the following predicate $\tilde{P}_{AP}^{-1}(1) = \{(h, h\cdot g^{-1}, h \cdot g) \mid h, g \in G\}$. Now, it is easily observed that the predicate is semi-rich.

Remark 1.2. *A dictatorship test with optimal parameters (in fact, the optimal* **NP***-hardness result for satisfiable instances) for the predicate P_G was shown by Bhangale and Khot [BK21]. Our main theorem gives new results for the predicate P_{AP} (and many more). The predicate P_{AP} is fundamentally different from P_G as it does not support any pairwise-independent distribution, whereas P_G does.*

1.1 Related Work

Many hardness results on satisfiable CSPs are known for specific CSPs. In this section, we state these results. Here, $\varepsilon > 0$ is an arbitrary small constant. Håstad [Hås01], in his seminal result, showed that for every $k \geqslant 3$, k-SAT is **NP**-hard to approximate within a factor of $1 - 1/2^k + \varepsilon$, even if the instance is satisfiable. Håstad and Khot [HK05] proved that Boolean CSPs on k variables are **NP**-hard to approximate within a ratio $\frac{2^{O(k^{1/2})}}{2^k}$. For every prime p, they also showed the hardness result for CSPs over an alphabet of size p, where the hardness factor is $\frac{p^{O(k^{1/2})}}{p^k}$. Huang [Hua14] improved the result for Boolean CSPs to the factor $\frac{2^{\tilde{O}(k^{1/3})}}{2^k}$. Brakensiek and Guruswami [BG21] formulated a problem called the 'V Label Cover' to improve these results on satisfiable k-ary CSPs. Towards this, assuming the hardness of the V Label Cover, they showed that there is an absolute constant c_0 such that for $k \geqslant 3$, given a satisfiable instance of Boolean k-CSP, it is hard to find an assignment satisfying more than $c_0 k^2/2^k$ fraction of the constraints. These results are non-trivial only for large values of k.

Towards getting an improved hardness result for Boolean satisfiable 3-CSPs, Håstad [Hå14] showed that the predicate NTW[7] is **NP**-hard to approximate within a factor of $5/8 + \varepsilon$. For larger alphabet, Engebretsen and Holmerin [EH05] showed that 3-ary CSPs over an alphabet of size q is **NP**-hard to approximate within a factor of $\frac{1}{q} + \frac{1}{q^2} + \varepsilon$. Tang [Tan09] showed a conditional result with the hardness factor $\frac{1}{q} + \frac{1}{q^2} - \frac{1}{q^3} + \varepsilon$.[8] Very recently, the first two authors [BK21] improved these results for 3-ary CSPs where where they showed

[5] Here, the predicate is $\{(x,y,z) \mid x+y+z=0\}$ where 0 is the identity element in G.

[6] See Section 3.5 for the definition of irreducible representations.

[7] The satisfying assignments for the NTW (Not-TWo-ones) predicate are all 3 bit strings such that the number of 1s in them is not *two*.

[8] The theorem in [EH05] holds for every $q \geqslant 3$, and the theorem in [Tan09] holds for every $q \geqslant 4$.

that it is **NP**-hard to approximate satisfiable 3-ary CSPs over an alphabet of size q to within a factor of $\frac{1}{q} + \varepsilon$, for infinitely many q.

1.2 Techniques

For a given predicate P, we are interested in finding the maximum α_P such that (1) there exists an approximation algorithm that satisfies at least α_P fraction of the constraints on satisfiable instances, and (2) for all $\varepsilon > 0$, it is hard to find $(\alpha_P + \varepsilon)$-satisfying assignments on satisfiable instances.

In order to answer the above question, the starting point is the Dichotomy Theorem which gives a full characterization of predicates for which the corresponding CSP is **NP**-complete or is in **P** (i.e., deciding if $\alpha_P = 1$ or $\alpha_P < 1$). The characterization is based on whether a certain non-trivial polymorphism exists for a given predicate. For a given predicate $P : \Sigma^k \to \{0, 1\}$, a function $f : \Sigma^n \to \Sigma$ is called a polymorphism if for every $k \times n$ matrix constructed by letting every column to be an arbitrary satisfying assignment to P and letting $x_1, x_2, \ldots, x_k \in \Sigma^n$ be the rows of the matrix, it is the case that $(f(x_1), f(x_2), \ldots, f(x_k))$ is also a satisfying assignment to P. It is easy to see that a dictator function, i.e., $f(x) = x_i$ for some $1 \leqslant i \leqslant n$ is always a polymorphism, and any other polymorphism is called a non-trivial polymorphism. The Dichotomy Theorem states that for a predicate P, checking satisfiability of P-CSP is in **P** if there exists a non-trivial polymorphism; otherwise, it is **NP**-complete (ignoring some subtle issues).

Dictatorship test. Similar to polymorphisms, dictatorship tests form the back-bone of proving hardness of approximating Max-P-CSPs. Here we formally define the dictatorship test for a given predicate.

Definition 1.3. *A dictatorship test for a predicate $P : \Sigma^k \to \{0, 1\}$ can query a function $f : \Sigma^n \to \Sigma$. The test picks a random $k \times n$ matrix by letting every column to be a random satisfying assignment to P (i.e., in $P^{-1}(1)$, with some fixed distribution μ on $P^{-1}(1)$) and letting $x_1, x_2, \ldots, x_k \in \Sigma^n$ be the rows of the matrix. The test accepts if $(f(x_1), f(x_2), \ldots, f(x_k))$ is also a satisfying assignment to P.*

Here again, it is obvious that if f is a dictator function, then the test accepts with probability 1. If solving P-CSP is **NP**-complete then it has no non-trivial polymorphisms according to the Dichotomy Theorem. Therefore, the question here is to determine the maximum probability the test accepts a function f if f is *far* from being a dictator function. If such a test exists where the maximum probability of acceptance for far from dictator functions is at most α_P, then using the Rich 2-to-1 Conjecture of Braverman, Khot, and Minzer [BKM21], one gets an **NP**-hardness of approximating P-CSP on satisfiable instances to within a factor of α_P.

We now describe the dictatorship test that we design for a large class of predicates. The starting point is an instance ϕ of P-CSP and let the value (i.e., maximum fraction of the constraints that can be satisfied by an assignment) of this instance be s. The distribution μ in the test depends on the SDP solution for ϕ and we only consider instances whose SDP value is 1. The SDP solution consists of vectors as well as local distribution for each constraint. Since the SDP value is 1, all these local distributions are supported on the satisfying assignments to P. Let μ_i be the local distribution corresponding to the i^{th} constraint of the instance. The test is as follows. Here $\varepsilon > 0$ is a small constant independent of n.

Given $f : \Sigma^n \to \Sigma$,

1. Select a constraint from ϕ according to the weights of the constraints. Let i be the selected constraint.
2. Construct a $k \times n$ matrix by setting each column of the matrix independently according to the following distribution: sample the column using μ_i.
3. Check if $P(f(x_1), f(x_2), \ldots, f(x_k) = 1$.

If f is a dictator function, then the test accepts with probability 1. This follows because for every i, the distribution μ_i is supported on the satsifying assignments to P and therefore every column of the matrix is from $P^{-1}(1)$. A challenging task is to compute the acceptance probability when f is far from dictator functions.

This test is a slight modification of Raghavendra's test [Rag08]. In his test, in Step (2) with probability ε, a random sample is chosen from Σ^k. This uniform *noise* has an effect of killing all the *high-degree* monomials of f and hence the analysis boils down to only considering the low-degree functions. At this point, one can apply the invariance principle for low-degree functions from Mossel [Mos10] and can replace the inputs with correlated Gaussians. Finally, the expression involving the Gaussians is interpreted as a rounding algorithm that rounds the SDP solution to an integral solution and the value is upper-bounded by the integral value of the instance which is s. Thus, the soundness of the test essentially matches the value of the integral solution. However, because of the uniform noise, the dictator functions will no longer pass the test with probability 1 and hence this test will not give hardness results on satisfiable instances.

Coming back to our test, we cannot add uniform noise as we want to maintain the completeness of the test to be 1. However, this introduces a few challenges in the analysis of the test. The main challenge is to show that the local distribution is enough to kill the high-degree part of f. This in general is not true. Specifically, if the predicate satisfies a linear equation, then this distribution is not enough to kill the high-degree part (see the counterexample in Remark 1.7). This is where we need the predicate (and the local distributions) to *not* satisfy any linear equation. In this case, we use our main analytical lemma, that we will discuss later, to show that the high-degree part of f contributes little to the test acceptance probability. However, we additionally need the predicate and the SDP solution to be semi-rich.

Finally, similar to Raghavendra's analysis, we use the low degree-part of f in the rounding algorithm and relate the performance of the algorithm to the test acceptance probability. This shows that if f is far from dictator function, then the acceptance probability of the test is upper bounded by the value of the assignment returned by the rounding procedure, which is always upper-bounded by s.

Main analytical lemma. Analyzing the acceptance probability of the test is a challenging task in general. One begins by thinking of the function f as a real valued function, e.g. as an indicator of the event that it takes a specific symbol in Σ as its value. Skipping some details, one needs to analyze expectations of the form

$$\mathop{\mathbf{E}}_{x_1,x_2,\ldots,x_k\sim\mu^{\otimes n}}\left[\prod_{i=1}^{k} f(x_i)\right],$$

here $x_1, x_2, \ldots, x_k$ are distributed as discussed in Definition 1.3. As the low-degree part of f corresponds to the SDPs from the algorithmic side, in order to prove our main theorem, we need to show that when f is a high-degree function, then this expectation is small. Our main analytical lemma shows that this is indeed the case. Following is the informal statement of the lemma (for a formal statement, see Lemma 2.6).

Lemma 1.4 (Informal)**.** *Let P be any 3-ary predicate that is semi-rich and does not satisfy any linear embedding. Let μ be any distribution that is fully supported on $P^{-1}(1)$. Then for any high-degree bounded function f, we have*

$$\left|\mathop{\mathbf{E}}_{x_1,x_2,x_3\sim\mu^{\otimes n}}[f(x_1)f(x_2)f(x_3)]\right| \leqslant \delta,$$

where $\delta \to 0$ as the degree of the function increases.

We note that a high-degree function has $\mathbf{E}[f] \approx 0$. This lemma is proved in Section 3 and it is evident that the proof of this lemma is quite involved. We believe that the semi-rich condition is not needed for the conclusion to hold. Generalizing the lemma for k-ary predicates and proving it without the semi-rich condition is a fascinating analytical question for future work.

The lemma is a generalization of Lemma 6.2 by Mossel [Mos10]. That lemma states that if the distribution μ is *connected* then the expectation is small. The connectedness condition can be stated as follows: For every pair of assignments (a, b, c) and (a', b', c') in $P^{-1}(1)$, there is a way to convert the first assignment to the second by replacing only once coordinate at a time such that every intermediate triple is in $P^{-1}(1)$. The predicate P_G that was mentioned earlier where $P_G^{-1}(1) = \{(x, y, z) \mid x \cdot y \cdot z = 1_G\}$ for some non-Abelian group does not satisfy the connectedness condition, as changing one coordinate from any satisfying assignment gives a triple which is outside of $P_G^{-1}(1)$. This predicate, however, does not satisfy any linear embedding if G does not have any non-trivial representation of dimension 1. P_G is also semi-rich and hence we can apply the above analytical lemma for P_G.

The proof of the above lemma for the predicate P_G is implicit in the work of Bhangale and Khot [BK21]. Given this fact, our high-level strategy to prove the lemma is as follows. We modify the underlying distribution μ and the predicate P so that the modified predicate can be viewed as a set of equations over some non-Abelian group. We do this by carefully adding more satisfying triplets to the predicate. During the modifications, we maintain the properties of the original predicate (i.e., semi-richness and not having any linear embedding) as well as make sure that the expectation does not change by much. Since the original predicate does not satisfy any linear embedding, the group must be non-Abelian and also lacks any non-trivial representation of dimension 1. Therefore, the final expectation must be small. This shows that the earlier expectation is also small.

1.3 Conclusion and Future Work

Our work leaves open many interesting problems. One obvious open problem is to extend our main theorem for other class of predicates. We could prove our analytical lemma for 3-ary semi-rich predicates. However, we believe that this semi-richness condition is not necessary for the conclusion to hold. One obvious open question is to extend our main theorem to other 3-ary predicates that are not semi-rich. More ambitiously, we put forth the following hypothesis for general k-ary predicates. One can naturally extend the definition of 3-ary predicates not satisfying any linear equation to k-ary predicates as follows.

Definition 1.5. *Let $P : \Sigma^k \to \{0, 1\}$ be any k-ary predicate such that the support on each coordinate is full. We say P satisfies a linear embedding if there exists an Abelian group $(G, +)$ and mappings $\sigma_i : \Sigma \to G$ such that*

- $\sum_{i=1}^{k} \sigma_i(x_i) = 0$ *for every $(x_1, x_2, \ldots, x_k) \in P^{-1}(1)$, where 0 is the identity element of G.*
- *one of the mappings $\{\sigma_i\}_{i=1}^{k}$ is non-constant.*

Otherwise, we say P does not satisfy any linear embedding.

With this definition, we propose the following hypothesis.

Hypothesis 1.6 (Informal)**.** *Let P be any k-ary predicate that does not satisfy any linear embedding. Let μ be any distribution that is fully supported on $P^{-1}(1)$. Given k functions $f_1, f_2, \ldots, f_k : \Sigma^n \to [-1, 1]$, such that one of the f_is is a high-degree function, then we have*

$$\left| \mathop{\mathbf{E}}_{x_1, x_2, \ldots, x_k \sim \mu^{\otimes n}} [f_1(x_1) f_2(x_2) \cdots f_k(x_k)] \right| \leqslant \delta,$$

where $\delta \to 0$ as the degree of the function increases.

Remark 1.7. *We note that if the predicate satisfies a linear equation, then the conclusion does not hold. To see this, suppose P satisfies a linear equation over an Abelian group G given by the embeddings $\{\sigma_i\}_{i=1}^{k}$. Let χ be any non-trivial character of G and define $f_i(x_i) = \prod_{j=1}^{n} \chi(\sigma_i((x_i)_j))$. Now,*

$$f_1(x_1) f_2(x_2) \cdots f_k(x_k) = \prod_{i=1}^{k} \prod_{j=1}^{n} \chi(\sigma_i((x_i)_j)) = \prod_{j=1}^{n} \prod_{i=1}^{k} \chi(\sigma_i((x_i)_j)) = \prod_{j=1}^{n} \chi\left(\sum_{i=1}^{k} \sigma_i((x_i)_j) \right),$$

where the last equality is because of the multiplicativity of the character χ. For every j, we have $\sum_{i=1}^{k} \sigma_i((x_i)_j) = 0$ and hence the product becomes 1 as $\chi(0) = 1$. Moreover, for large n, since one of the embeddings is non-constant, one of the f_is is a high-degree function.

With a positive answer to the hypothesis, we may be able to make progress on predicates P that do not satisfy any linear equation. On the other hand, if P does satisfy a certain linear equation, then a hybrid algorithm that solves the SDP as well as the system of linear equations might give an optimal algorithm for satisfiable CSPs. We leave these as open problems for future work.

1.4 Organization

In Section 2 we state our main dictatorship test for 3-ary CSPs satisfying conditions from Theorem 1.1. We start with preliminaries in Section 2.1 where we define constraint satisfaction problems, functions on product spaces and state the invariance principle. We also state our main analytical lemma that we use in our dictatorship test analysis in this section. In Section 2.2 we define the basic SDP relaxation for a Max-P-CSP. In Section 2.3, we state our dictatorship test and prove the completeness and soundness analysis of the test. We prove our main analytical lemma in Section 3. The proof consists of a series of steps which we prove in the subsequent subsections.

2 From Integrality Gap to Dictatorship Test

In this section, we show that if a P-CSP instance has a $(1, s)$ integrality gap for the basic SDP relaxation, then there is a dictatorship test with completeness 1 and soundness $s + \varepsilon$ for any $\varepsilon > 0$, if the predicate and the SDP solution satisfy certain conditions.

2.1 Preliminaries

The focus of this paper is on special type of predicates that do not satisfy any linear equation and that are semi-rich. We define these two properties next. We define these properties for more general predicates having different alphabet for each location, although in our dictatorship test we only consider predicates of the type $P : \Sigma^3 \to \{0, 1\}$.

Definition 2.1. *Let Σ, Φ, Γ be finite alphabets. Let $H \subseteq \Sigma \times \Phi \times \Gamma$ and $\Sigma' \subseteq \Sigma, \Phi' \subseteq \Phi$ and $\Gamma' \subseteq \Gamma$ be the subsets on which H is supported. We say H can be linearly embedded in an Abelian group if there is an Abelian group $(G, +)$ and maps $\sigma \colon \Sigma' \to G$, $\phi \colon \Phi' \to G$, $\gamma \colon \Gamma' \to G$ such that*

1. *for all $(x, y, z) \in H$ it holds that $\sigma(x) + \phi(y) + \gamma(z) = 0$;*
2. *at least one of σ, ϕ, γ is non-constant.*

Otherwise, we say H cannot be embedded linearly into an Abelian group, or simply H does not satisfy any linear equation.

Definition 2.2. *Let Σ, Φ, Γ be finite alphabets. Let $H \subseteq \Sigma \times \Phi \times \Gamma$ and $\Sigma' \subseteq \Sigma, \Phi' \subseteq \Phi$ and $\Gamma' \subseteq \Gamma$ be the subsets on which H is supported. We say H is* semi-rich *if the following two properties hold.*

1. *For all $(x, y) \in \Sigma' \times \Phi'$, there exists $z \in \Gamma'$ such that $(x, y, z) \in H$.*
2. *For all $(x, z) \in \Sigma' \times \Gamma'$, there exists $y \in \Phi'$ such that $(x, y, z) \in H$.*

Note that in the definition of semi-rich one of the three coordinates is special. However, the location of the special coordinate does not matter as we can permute the coordinates and study the modified subset instead.

We now define the predicates that have these two properties.

Definition 2.3. *A predicate $P \colon \Sigma \times \Phi \times \Gamma \to \{0, 1\}$ is said to be linearly embedded into an Abelian group if and only if $P^{-1}(1)$ can be linearly embedded in an Abelian group. P is called semi-rich if $P^{-1}(1)$ is semi-rich.*

Let (Ω, μ) be a probability space. Define the inner product on this space by $\langle f, g\rangle_\mu := \mathbf{E}_{x\in\mu}[f(x)g(x)]$. We will use the notation $\|f\|_{p;\mu} := \mathbf{E}_{x\in\mu}[|f(x)|^p]^{1/p}$ to denote the p^{th} norm of f. In order to state our main analytical lemma, we need the following definition of the noise operator.

Definition 2.4. *Let Φ be a finite alphabet, and ν be a measure on Φ. For a parameter $\rho \in [0, 1]$, we define the ρ-correlated distribution with respect to ν as follows. For any $y \in \Phi$, the distribution of inputs that are ρ-correlated with y is denoted by $\mathbf{y}' \sim \mathrm{T}_\rho y$ and is defined by taking $\mathbf{y}' = y$ with probability ρ, and otherwise sampling $\mathbf{y}' \sim \nu$.*

As is often the case, we also view T_ρ as an operator on functions, mapping $L^2(\Phi,\nu)$ to $L^2(\Phi,\nu)$ by

$$(\mathrm{T}_\rho g)(y) = \mathop{\mathbf{E}}_{\mathbf{y}'\sim \mathrm{T}_\rho y}\left[g(\mathbf{y}')\right].$$

We then tensorize this operator, i.e., consider $\mathrm{T}_\rho^{\otimes n}$ which acts on functions on n-variables, i.e. on $L^2(\Phi^n,\nu^{\otimes n})$. When clear from context, we drop the $\otimes n$ superscript from notation.

Definition 2.5. $\mathsf{Stab}_\rho^\nu(g) = \langle g, \mathrm{T}_\rho g\rangle_{\nu^{\otimes n}}$. *We drop the superscript ν from $\mathsf{Stab}_\rho^\nu(g)$, if it is clear from the context.*

Let $m = |\Phi|$ and write the multilinear expansion of g with respect to ν, i.e., $g(y) = \sum_{\boldsymbol{\sigma}\in\{0,1,\ldots,m-1\}^n} \widehat{g}(\boldsymbol{\sigma})\boldsymbol{\ell}_{\boldsymbol{\sigma}}(y)$, where $\boldsymbol{\ell}_0 \equiv 1$ is the trivial character.[9] Then $\mathsf{Stab}_\rho^\nu(g) = \sum_{\boldsymbol{\sigma}\in\{0,1,\ldots,m-1\}^n} \rho^{|\boldsymbol{\sigma}|}\widehat{g}(\boldsymbol{\sigma})^2$, where $|\boldsymbol{\sigma}|$ is the number of non-zero entries in $\boldsymbol{\sigma}$. Thus, if g has small weight on $\widehat{g}(\boldsymbol{\sigma})$ where $|\boldsymbol{\sigma}|$ is small then $\mathsf{Stab}_\rho^\nu(g)$ is small. Thus, we use the notion of small stability of a function as a proxy for high-degreeness of the function.

We now state the main analytical lemma that we use in the analysis of our dictatorship test. The lemma is proved in the next section (Section 3).

Lemma 2.6. *For all $m\in\mathbb{N}$, $\varepsilon,\alpha > 0$ there exist $\xi > 0$ and $\delta > 0$ such that the following holds. Suppose μ is a distribution over $\Sigma\times\Phi\times\Gamma$ whose support (a) is semi-rich, and (b) cannot be embedded in an Abelian group. Further suppose that $|\Sigma|, |\Phi|, |\Gamma| \leqslant m$, each atom in μ has probability at least α and marginals of μ on Σ,Φ and Γ have full support. If $f\colon \Sigma^n\to[-1,1]$, $g\colon\Phi^n\to[-1,1]$, $h\colon\Gamma^n\to[-1,1]$ are functions such that*

$$\text{either } \mathsf{Stab}_{1-\xi}(f)\leqslant\delta, \text{ or } \mathsf{Stab}_{1-\xi}(g)\leqslant\delta, \text{ or } \mathsf{Stab}_{1-\xi}(h)\leqslant\delta,$$

then $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \varepsilon$.

2.1.1 Constraint Satisfaction Problems

We will use the notation $[R]$ to denote the set $\{1,2,\ldots,R\}$. In our dictatorship test analysis, we are going to need a few lemmas from Raghavendra's thesis [Rag09] as black-box. Therefore, we try to use the same notations from his thesis. Our analytical lemma (Lemma 2.6) that we prove in the next section works only for the 3-ary CSPs. However, in this section, we work with general k-ary P-CSPs. If we have the analogous analytical lemma for any k-ary CSP, then the test designed in this section can be combined with it to get a result for k-ary CSPs.

Raghavendra considered CSPs with mixed predicates. In this work, we consider CSPs with one predicate $P:\Sigma^k\to\{0,1\}$ (or possibly mix of predicates with the same template P, as described below). We formally define the P-CSP instance in the following definition.

Definition 2.7. *For a given predicate $P:\Sigma^k\to\{0,1\}$, a P-CSP instance is given by $\mathfrak{I} = (\mathcal{V},\mathcal{P})$ where*

- *$\mathcal{V}$ is the set of variables.*
- *$\mathcal{P}$ is a probability distribution on the payoff functions $P':\Sigma^{\mathcal{V}}\to\{0,1\}$ of type,*

$$P'(\boldsymbol{y}) = P(y_{i_1}, y_{i_2},\ldots,y_{i_k}),$$

for some $i_1,i_2,\ldots,i_k\in\mathcal{V}$.

Remark 2.8. *A P-CSP instance actually consists of a mix of payoffs on the same template P. In the Boolean CSP, these mix of payoffs are formed by using literals (or negations). Here are few examples to illustrate this for Boolean CSPs as well as for general CSPs.*

[9] This is formally defined in the Section 2.1.2 below.

1. *In 3SAT, the template predicate P is $P : \{0,1\}^3 \to \{0,1\}$ where $P(x,y,z) = 0$ iff $x = y = z = 0$. However, a 3SAT instance contains 8 different payoffs, one for each literal pattern.*
2. *In 3LIN, the template predicate $P : \{0,1\}^3 \to \{0,1\}$ is such that $P(x,y,z) = 1$ iff $x \oplus y \oplus z = 1$. In this case, the instance also contains constraints of type $x \oplus y \oplus z = 0$.*
3. *In 3LIN equations over a non Abelian group $(G, \cdot)$, the predicate is $P : G^3 \to \{0,1\}$ such that $P(x,y,z) = 1$ iff $x \cdot y \cdot z = 1_G$, where 1_G is the identity element of G. The instance contains constraints of type $x \cdot y \cdot z = g$ for some $g \in G$.*

Without the mix of payoffs, certain P-CSPs are trivial; for instance, the all 1 assignment would satisfy every 3SAT and 3LIN instance. Therefore we allow the use of such mix of payoffs in our instances. Note that for certain predicates, like $3NAE : \{0,1\}^3 \to \{0,1\}$, defined as $3NAE(x,y,z) = 1$ iff x, y, z are not all the same, instances without any mix of payoffs are non-trivial to solve.

For a payoff P', the set of indices $i_1, i_2, \ldots, i_k \in \mathcal{V}$ on which it depends is denoted by $\mathcal{V}(P')$. Let $\mathtt{supp}(\mathcal{P})$ be the set of payoffs in $\mathfrak{I}$. Given a P-CSP instance $\mathfrak{I}$, the objective is to find an assignment $\boldsymbol{y} \in \Sigma^{\mathcal{V}}$ that maximizes the value of the instance which is defined as follows:

$$\mathbf{val}(\boldsymbol{y}) = \mathop{\mathbf{E}}_{P' \sim \mathcal{P}} [P'(\boldsymbol{y})].$$

The optimum value of the instance $\mathfrak{I} = (\mathcal{V}, \mathcal{P})$ is defined as:

$$\textsc{Opt}(\mathfrak{I}) = \max_{\boldsymbol{y} \in \Sigma^{\mathcal{V}}} \mathbf{val}(\boldsymbol{y}).$$

Let $\blacktriangle(\Sigma)$ be the set of probability distributions on Σ.

2.1.2 Functions on Product Spaces

Let (Ω, μ) be a probability space with $|\Omega| = q$ and μ has full support on Ω. Define the inner product between two functions $f, g : \Omega \to \mathbb{R}$ on this space as follows: $\langle f, g \rangle = \mathbf{E}_{x \sim \mu}[f(x)g(x)]$.

Definition 2.9. *An orthonormal ensemble consists of a basis of real orthonormal random variables $\mathcal{L} = \{\ell_0 \equiv \mathbf{1}, \ell_1, \ldots, \ell_{q-1}\}$, where $\mathbf{1}$ is the constant 1 function.*

Henceforth, we will sometimes refer to orthonormal ensembles as just ensembles. For an ensemble $\mathcal{L} = \{\ell_0 \equiv \mathbf{1}, \ell_1, \ldots, \ell_{q-1}\}$ of random variables, we will use $\mathcal{L}^R$ to denote the ensemble obtained by taking R independent copies of $\mathcal{L}$. Further $\mathcal{L}^{(i)} = \{\ell_0^{(i)}, \ell_1^{(i)}, \ldots, \ell_{q-1}^{(i)}\}$ will denote the i^{th} independent copy of $\mathcal{L}$.

Fix an ensemble $\mathcal{L} = \{\ell_0 \equiv \mathbf{1}, \ell_1, \ldots, \ell_{q-1}\}$ that forms a basis for $L^2(\Omega)$. Given such a basis for $L^2(\Omega)$, it induces a basis for the space $L^2(\Omega^R)$, given by the random variables

$$\left\{ \ell_{\boldsymbol{\sigma}} := \prod_{i=1}^{R} \ell_{\sigma_i}^{(i)} \;\middle|\; \boldsymbol{\sigma} \in \{0, 1, \ldots, q-1\}^R \right\}.$$

Therefore, any function $\mathcal{F} : \Omega^R \to \mathbb{R}$ has a multilinear expansion

$$\mathcal{F}(\boldsymbol{z}) = \sum_{\boldsymbol{\sigma} \in \{0,1,\ldots,q-1\}^R} \hat{\mathcal{F}}(\boldsymbol{\sigma}) \ell_{\boldsymbol{\sigma}}(\boldsymbol{z}),$$

where $\ell_{\boldsymbol{\sigma}}(\boldsymbol{z}) = \prod_{i=1}^R \ell_{\sigma_i}(z_i)$.

Definition 2.10. *A multi-index $\boldsymbol{\sigma}$ is a vector $(\sigma_1, \sigma_2, \ldots . \sigma_R) \in \{0, 1, \ldots, q-1\}^R$ and the degree of $\boldsymbol{\sigma}$ is denoted by $|\boldsymbol{\sigma}|$ which is equal to $|\boldsymbol{\sigma}| = |\{i \in [R] \mid \sigma_i \neq 0\}|$. Given a set of indeterminates $\boldsymbol{\mathcal{X}} = \{x_j^{(i)} | j \in \{0, 1, \ldots, q-1\}, i \in [R]\}$ and a multi-index $\boldsymbol{\sigma}$, define the monomial $x_{\boldsymbol{\sigma}}$ as*

$$x_{\boldsymbol{\sigma}} = \prod_{i=1}^{R} x_{\sigma_i}^{(i)}.$$

The degree of the monomial is given by $|\boldsymbol{\sigma}|$. A multilinear polynomial over such indeterminates is given by

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

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

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

We will always use the symbol $\mathcal{F}$ to denote real-valued function on a product probability space Ω^R. Further $F(\boldsymbol{x})$ will denote the formal multilinear polynomial corresponding to $\mathcal{F}$. Hence $F(\mathcal{L}^R)$ is a random variable obtained by substituting the random variables $\mathcal{L}^R$ in place of $\boldsymbol{x}$. For instance, the following equation holds in this notation:

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

We now define the notion of the influence of a variable.

Definition 2.11. *For a function $\mathcal{F} : \Omega^R \to \mathbb{R}$ over the space $(\Omega^R, \mu^{\otimes R})$, the influence of the j^{th} coordinate is given by:*

$$\mathbf{Inf}_j[\mathcal{F}; \mu^{\otimes R}] = \mathop{\mathbf{E}}_{\boldsymbol{z}^{(-j)} \in \Omega^{R-1}}[\mathbf{Var}_{\boldsymbol{z}^{(j)} \in \Omega}[\mathcal{F}(\boldsymbol{z})]],$$

where $\boldsymbol{z}^{(-j)}$ is a string missing the j^{th} coordinate.

We have the following proposition that relates the average value and the variance of a function to its Fourier coefficients.

Proposition 2.12. *For a function $\mathcal{F} : \Omega^R \to \mathbb{R}$ over the space $(\Omega^R, \mu^{\otimes R})$, if $\mathcal{F}(\boldsymbol{z}) = \sum_{\boldsymbol{\sigma}} \hat{\mathcal{F}}(\boldsymbol{\sigma})\ell_{\boldsymbol{\sigma}}(\boldsymbol{z})$ with respect to an orthonormal ensemble $\mathcal{L}$ of (Ω, μ), then $\mathbf{E}_{\boldsymbol{z} \in \Omega^R}[\mathcal{F}(\boldsymbol{z})] = \hat{\mathcal{F}}_{\mathbf{0}}$ and $\mathbf{Var}[\mathcal{F}] = \sum_{\boldsymbol{\sigma} \neq \mathbf{0}} \hat{\mathcal{F}}_{\boldsymbol{\sigma}}^2$.*

We also define the degree $\geqslant D$ weight of a function $\mathcal{F}$ as follows:

$$W^{\geqslant D}[\mathcal{F}; \mu^{\otimes R}] = \sum_{\boldsymbol{\sigma} : |\boldsymbol{\sigma}| \geqslant D} \hat{\mathcal{F}}_{\boldsymbol{\sigma}}^2.$$

Another way of writing a function on a probability space as sum of orthogonal functions is called the Efron-Stein decomposition.

Definition 2.13. *Let (Ω, μ) be a probability space and $(\Omega^R, \mu^{\otimes R})$ be the corresponding product space. For a function $f : \Omega^R \to \mathbb{R}$, the Efron-Stein decomposition of f with respect to the product space is given by*

$$f(z_1, \cdots, z_R) = \sum_{\beta \subseteq [R]} f_\beta(\boldsymbol{z}),$$

where f_β depends only on z_i for $i \in \beta$ and for all $\beta' \not\supseteq \beta, \boldsymbol{a} \in \Omega^{\beta'}$, $\mathbf{E}_{\boldsymbol{z} \in \mu^{\otimes R}}[f_\beta(\boldsymbol{z}) \mid \boldsymbol{z}|_{\beta'} = \boldsymbol{a}] = 0$.

We have the following facts about the Efron-Stein decomposition of functions.

Fact 2.14. *If $f(\boldsymbol{z}) = \sum_{\beta \subseteq [R]} f_\beta(\boldsymbol{z})$ and $g(\boldsymbol{z}) = \sum_{\beta \subseteq [R]} g_\beta(\boldsymbol{z})$ are the Efron-Stein decompositions of f and g respectively w.r.t. the product space $(\Omega^R, \mu^{\otimes R})$, then*

1. $T_\rho f(\boldsymbol{z}) = \sum_{\beta \subseteq [R]} \rho^{|\beta|} f_\beta(\boldsymbol{z})$ *and*
2. $\langle f, g \rangle_{\mu^{\otimes R}} = \sum_{\beta \subseteq [R]} \langle f_\beta, g_\beta \rangle_{\mu^{\otimes R}}$.

2.1.3 Vector valued functions

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

The notions of influence and degree $\geqslant D$ weight can be extended to the vector valued functions using the following definitions.

$$\mathbf{Inf}_i[\boldsymbol{\mathcal{F}}; \mu^{\otimes R}] = \sum_{j=1}^{q} \mathbf{Inf}_i[\mathcal{F}_j; \mu^{\otimes R}] \quad \text{and} \quad W^{\geqslant D}[\boldsymbol{\mathcal{F}}; \mu^{\otimes R}] = \sum_{j=1}^{q} W^{\geqslant D}[\mathcal{F}_j; \mu^{\otimes R}].$$

2.1.4 Invariance Principle

Define functions $f_{[0,1]} : \mathbb{R} \to \mathbb{R}$ and $\xi : \mathbb{R}^q \to \mathbb{R}$ as follows:

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

$$\xi(\boldsymbol{a}) = \sum_{j=1}^{q} (f_{[0,1]}(a_j) - a_j)^2.$$

A crucial step in the analysis of the dictatorship test is to replace the discrete inputs with correlated Gaussians. The following theorem from Mossel [Mos10] states that one can do this provided the functions do not have influential coordinates and the functions are low-degree.

Theorem 2.15 ([Mos10])**.** *Fix $0 < \alpha \leqslant 1/2$ and $d \in \mathbb{N}$. Let (Ω, μ), $|\Omega| = m$, be a finite probability space such that every atom has probability at least α. Let $\mathcal{L}^{(r)} = \{\ell_0^{(r)} \equiv \mathbf{1}, \ell_1^{(r)}, \ldots, \ell_{m-1}^{(r)}\}$ be an orthonormal ensemble of random variables over Ω and $\mathcal{G}^{(r)} = \{g_0^{(r)} \equiv \mathbf{1}, g_1^{(r)}, \ldots, g_{m-1}^{(r)}\}$ be an orthonormal ensemble of Gaussian random variables.*

Let $\boldsymbol{F} = (F_1, F_2, \ldots, F_d)$ denote a vector valued multilinear polynomial on Ω^R. If $\mathbf{Inf}_i[\boldsymbol{F}; \mu^{\otimes R}] \leqslant \tau$ for all $i \in [R]$, $W^{\geqslant D}[\boldsymbol{F}; \mu^{\otimes R}] \leqslant \delta$ and $\mathbf{Var}[F_j] \leqslant 1$ for all $j \in \{1, \ldots, d\}$, then the following holds.

1. *For every function $\psi : \mathbb{R}^d \to \mathbb{R}$ that is thrice differentiable with all its partial derivatives up to order 3 bounded uniformly by C_0,*

$$\left|\mathbf{E}[\psi(\boldsymbol{F}(\mathcal{L}^R))] - \mathbf{E}[\psi(\boldsymbol{F}(\mathcal{G}^R))]\right| \leqslant O\left(D\sqrt{\tau}\left(8\alpha^{-1/2}\right)^D\right) + O(\sqrt{\delta}).$$

2. *For the function ξ defined above,*

$$\left|\mathbf{E}[\xi(\boldsymbol{F}(\mathcal{L}^R))] - \mathbf{E}[\xi(\boldsymbol{F}(\mathcal{G}^R))]\right| \leqslant O\left(\sqrt{\tau}\left(10\alpha^{-1/2}\right)^D\right)^{2/3} + O(\sqrt{\delta}).$$

In both the cases, the $O(.)$ hides the constant C_0.

Proof. The theorem follows from Theorem 4.1 from [Mos10]. Truncate the polynomial $\boldsymbol{F}$ to degree D to get a polynomial $\boldsymbol{L}$. Using Theorem 4.1 of Mossel [Mos10], we have

$$\left|\mathbf{E}[\psi(\boldsymbol{L}(\mathcal{L}^R))] - \mathbf{E}[\psi(\boldsymbol{L}(\mathcal{G}^R))]\right| \leqslant 2DC_0 d^3 (8\alpha^{-1/2})^D \tau^{1/2} = O\left(D\sqrt{\tau}\left(8\alpha^{-1/2}\right)^D\right).$$

Since ψ is a smooth functional,

$$\left|\mathbf{E}[\psi(\boldsymbol{L}(\mathcal{L}^R))] - \mathbf{E}[\psi(\boldsymbol{F}(\mathcal{L}^R))]\right| \leqslant C_0 \|\boldsymbol{L}(\mathcal{L}^R) - \boldsymbol{F}(\mathcal{L}^R)\| = C_0\left(W^{\geqslant D}[\boldsymbol{F}; \mu^{\otimes R}]\right)^{1/2} \leqslant C_0\sqrt{\delta}.$$

Similarly, we get

$$\left|\mathbf{E}[\psi(\boldsymbol{L}(\mathcal{G}^R))] - \mathbf{E}[\psi(\boldsymbol{F}(\mathcal{G}^R))]\right| \leqslant C_0\sqrt{\delta}.$$

Combining the three inequalities, we get the required bound for (1).

The second item follows from Theorem 3.19 from [MOO05]. Here again, let $\boldsymbol{L}$ be the low-degree part of $\boldsymbol{F}$ truncated at degree D and let $\boldsymbol{H} = \boldsymbol{F} - \boldsymbol{L}$. Using Theorem 3.19 of [MOO05],

$$\left|\mathbf{E}[\xi(\boldsymbol{L}(\mathcal{L}^R))] - \mathbf{E}[\xi(\boldsymbol{L}(\mathcal{G}^R))]\right| \leqslant O\left(\sqrt{\tau}\left(10\alpha^{-1/2}\right)^D\right)^{2/3}.$$

Using Lemma 3.24 from [MOO05],

$$\begin{aligned}
\left|\mathbf{E}[\xi(\boldsymbol{F}(\mathcal{L}^R))] - \mathbf{E}[\xi(\boldsymbol{L}(\mathcal{L}^R))]\right| &\leqslant 2\,\mathbf{E}[\boldsymbol{L}(\mathcal{L}^R)\boldsymbol{H}(\mathcal{L}^R)] + \mathbf{E}[\boldsymbol{H}(\mathcal{L}^R)^2] \\
&\leqslant 2\sqrt{\mathbf{E}[\boldsymbol{L}(\mathcal{L}^R)^2]}\sqrt{\mathbf{E}[\boldsymbol{H}(\mathcal{L}^R)^2]} + \mathbf{E}[\boldsymbol{H}(\mathcal{L}^R)^2] \\
&\leqslant 2\sqrt{\mathbf{E}[\boldsymbol{H}(\mathcal{L}^R)^2]} + \mathbf{E}[\boldsymbol{H}(\mathcal{L}^R)^2] \leqslant 2\sqrt{\delta} + \delta \leqslant 3\sqrt{\delta},
\end{aligned}$$

where the second step follows from the Cauchy-Schwarz inequality. Similarly, we get,

$$\left|\mathbf{E}[\xi(\boldsymbol{F}(\mathcal{G}^R))] - \mathbf{E}[\xi(\boldsymbol{L}(\mathcal{G}^R))]\right| \leqslant 3\sqrt{\delta},$$

and the claim follows.

□

2.2 SDP Relaxation

Given an instance $\mathfrak{I} = (\mathcal{V}, \mathcal{P})$, the basic semi-definite programming relaxation of the instance is given in Figure 1. It consists of vectors $\{\boldsymbol{b}_{i,a}\}_{i\in\mathcal{V},a\in\Sigma}$, distributions $\{\mu_{P'}\}_{P'\in\mathsf{supp}(\mathcal{P})}$ over the local assignments (i.e., on $\Sigma^{\mathcal{V}(P')}$) and a unit vector $\boldsymbol{b}_0$. Let $\mathbf{val}(\boldsymbol{V}, \boldsymbol{\mu})$ be the objective value of the solution $(\boldsymbol{V}, \boldsymbol{\mu})$.

$$\begin{array}{lll r}
\text{maximize} & \mathop{\mathbf{E}}_{P'\sim\mathcal{P}} \mathop{\mathbf{E}}_{x\in\mu_{P'}}[P'(x)] & & \\
\text{subject to} & \langle \boldsymbol{b}_{i,a}, \boldsymbol{b}_{j,b}\rangle = \mathop{\mathbf{Pr}}_{x\sim\mu_{P'}}[x_i = a, x_j = b] & P' \in \mathsf{supp}(\mathcal{P}), \quad i,j \in \mathcal{V}(P'), \quad a,b\in\Sigma & (1) \\
& \langle \boldsymbol{b}_{i,a}, \boldsymbol{b}_0\rangle = \|\boldsymbol{b}_{i,a}\|_2^2 & \forall i \in \mathcal{V}, a\in\Sigma & (2) \\
& \|\boldsymbol{b}_0\|_2^2 = 1 & & (3) \\
& \mu_{P'} \in \blacktriangle(\Sigma^{\mathcal{V}(P')}) & P' \in \mathsf{supp}(\mathcal{P}) & (4)
\end{array}$$

Figure 1: Basic SDP relaxation of a P-CSP instance $\mathfrak{I} = (\mathcal{V}, \mathcal{P})$.

Following is a definition of $(1, s)$ integrality gap instance.

Definition 2.16. *An instance $\mathfrak{I} = (\mathcal{V}, \mathcal{P})$ is a $(1, s)$ SDP integrality gap instance if the optimal value of the instance is at most s and the optimal value of the basic SDP relaxation for $\mathfrak{I}$ is 1.*

For our dictatorship test to work, we require that the support of every local distribution $\mu_{P'}$ is semi-rich and it is not linearly embeddable in any Abelian group. Henceforth, we will assume that the SDP solution satisfies this property.

2.3 Dictatorship Test

In this section, we study the dictatorship test for P-CSP instances over a k-ary predicate P. Throughout this section, when $k = 3$, we restrict ourselves to the predicates P that are semi-rich and that do not satisfy any linear equation.

Let $\mathfrak{I} = (\mathcal{V}, \mathcal{P})$ be an instance of P-CSP, where $P : \Sigma^k \to \{0, 1\}$ and $|\Sigma| = q$. We will fix an arbitrary mapping from Σ to $\{1, 2, \dots, q\}$, denoted by $\varsigma : \Sigma \to \{1, 2, \dots, q\}$.

Let $(\boldsymbol{V}, \boldsymbol{\mu})$ be a solution for the basic SDP relaxation of $\mathfrak{I}$ which is semi-rich and which does not satisfy any linear equation. 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 assume that Ω_s has full support for every $s \in \mathcal{V}$. However, our proof works even when the support is a subset of Σ.

A function $F : \Sigma^R \to \Sigma$ is called a dictator function if $F(\boldsymbol{z}) = \boldsymbol{z}^{(i)}$ for some $i \in [R]$. In Figure 2, we give the dictatorship test $\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}$ for functions $F : \Sigma^R \to \Sigma$.

1. Sample a payoff $P' \sim \mathcal{P}$. Let $\mathcal{V}(P') = \{s_1, s_2, \dots, s_k\}$.
2. Sample $\boldsymbol{z}_{P'} = \{\boldsymbol{z}_{s_1}, \boldsymbol{z}_{s_2}, \dots, \boldsymbol{z}_{s_k}\}$ from the product distribution $\mu_{P'}^{\otimes R}$, i.e., independently for each $i \in [R]$, $(\boldsymbol{z}_{s_1}^{(i)}, \boldsymbol{z}_{s_2}^{(i)}, \dots, \boldsymbol{z}_{s_k}^{(i)}) \sim \mu_{P'}$.
3. Query the function values $F(\boldsymbol{z}_{s_1}), F(\boldsymbol{z}_{s_2}), \dots, F(\boldsymbol{z}_{s_k})$.
4. Accept iff $P'(F(\boldsymbol{z}_{s_1}), F(\boldsymbol{z}_{s_2}), \dots, F(\boldsymbol{z}_{s_k})) = 1$.

Figure 2: SDP integrality gap to a dictatorship test $\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}$.

Remark 2.17. *There is one main difference between our test and the dictatorship test given in [Rag09]. In [Rag09], in Step 2 (Figure 2), uniformly random noise is added from Σ^k. This step loses the perfect completeness of the dictatorship test.*

2.3.1 Completeness Analysis

The completeness of the test is defined as follows,

$$\text{Completeness}(\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}) = \min_{\substack{i \in [R], \\ F \text{ is the } i^{th} \text{ dictator}}} \mathbf{Pr}[F \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}].$$

If the function is a dictator function, then the test accepts with probability 1. The simple claim is proven below.

Lemma 2.18. *If* $\mathbf{val}(\boldsymbol{V}, \boldsymbol{\mu}) = 1$ *then*

$$\text{Completeness}(\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}) = 1.$$

Proof. Consider a dictator function $F(\boldsymbol{z}) = \boldsymbol{z}^{(j)}$ for some $j \in [R]$. In this case, $(F(\boldsymbol{z}_{s_1}), F(\boldsymbol{z}_{s_2}), \dots, F(\boldsymbol{z}_{s_k})) = (\boldsymbol{z}_{s_1}^{(j)}, \boldsymbol{z}_{s_2}^{(j)}, \dots, \boldsymbol{z}_{s_k}^{(j)})$. When the payoff $P' \sim \mathcal{P}$ is selected, then $(\boldsymbol{z}_{s_1}^{(j)}, \boldsymbol{z}_{s_2}^{(j)}, \dots, \boldsymbol{z}_{s_k}^{(j)})$ is distributed according to $\mu_{P'}$. As the SDP value is 1, the distribution $\mu_{P'}$ is fully supported on $P'^{-1}(1)$ and hence the test passes with probability 1. ☐

2.3.2 Soundness Analysis

We now move to prove the soundness analysis of the test. Here we formally define the functions which are far from dictator functions (also known as quasirandom functions). Let $\Delta_q := \{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_q\}$ where $\boldsymbol{e}_j$ is the j^{th} basis vector of $\mathbb{R}^q$.

Definition 2.19. *For a function $F : \Sigma^R \to \Sigma$, the corresponding Δ_q-representation is a function $\boldsymbol{\mathcal{F}} : \Sigma^R \to \Delta_q$ given by*

$$\boldsymbol{\mathcal{F}}(\boldsymbol{z}) = \boldsymbol{e}_{\varsigma(F(\boldsymbol{z}))}.$$

Therefore, in this setting F is a dictator function if $\boldsymbol{\mathcal{F}}(\boldsymbol{z}) = \boldsymbol{e}_{\varsigma(\boldsymbol{z}^{(i)})}$ for some $i \in [R]$. Any function $\boldsymbol{\mathcal{F}} : \Sigma^R \to \blacktriangle_q$ can be interpreted as a distribution on functions $\boldsymbol{\mathcal{F}}' : \Sigma^R \to \Delta_q$ as follows: For each $\boldsymbol{z} \in \Sigma^R$, set the value of $\boldsymbol{\mathcal{F}}'(\boldsymbol{z})$ independently as

$$\boldsymbol{\mathcal{F}}'(\boldsymbol{z}) = \boldsymbol{e}_j \quad \text{with probability } \boldsymbol{\mathcal{F}}(\boldsymbol{z})_j \text{ for all } j \in \{1, 2, \ldots, q\}.$$

Thus, for each $\boldsymbol{z} \in \Sigma^R$, we have $\boldsymbol{\mathcal{F}}(\boldsymbol{z}) = \mathbf{E}[\boldsymbol{\mathcal{F}}'(\boldsymbol{z})]$.

Fix a function $\boldsymbol{\mathcal{F}} : \Sigma^R \to \blacktriangle_q$. For each $s \in \mathcal{V}$, let $\boldsymbol{\mathcal{F}}_s$ denote the function $\boldsymbol{\mathcal{F}}$ interpreted as a function on the product probability space $(\Sigma^R, \mu_s^{\otimes R})$.

Definition 2.20. *A function $\boldsymbol{\mathcal{F}} : \Sigma^R \to \blacktriangle_q$ is said to be (τ, δ)-quasirandom if for each $s \in \mathcal{V}$, it holds that*

$$\max_{1 \leqslant i \leqslant R} \mathbf{Inf}_i[T_{1-\delta}\boldsymbol{\mathcal{F}}_s; \mu_s^{\otimes R}] \leqslant \tau,$$

where $\mathbf{Inf}_i[\boldsymbol{\mathcal{F}}_s; \mu_s^{\otimes R}] = \sum_{j=1}^{q} \mathbf{Inf}_i[\mathcal{F}_{s,j}; \mu_s^{\otimes R}]$ and $\mathcal{F}_{s,j}$ is $\boldsymbol{\mathcal{F}}_s$ restricted to the j^{th}-coordinate of $\blacktriangle_q$.

The domain of payoff P' can be extended from Σ^k to $\blacktriangle_q^k$. To see this, by the abuse of notation, first define a Δ_q-representation of a payoff $P' : \Sigma^k \to \{0, 1\}$ as $P' : \Delta_q^k \to \{0, 1\}$ where

$$P'(\boldsymbol{e}_{a_1}, \boldsymbol{e}_{a_2}, \ldots, \boldsymbol{e}_{a_k}) = P'(\varsigma^{-1}(a_1), \varsigma^{-1}(a_2), \ldots, \varsigma^{-1}(a_k)), \text{ for all } (a_1, a_2, \ldots, a_k) \in \{1, 2, \ldots, q\}^k.$$

The function P' can be extended to the domain $\blacktriangle_q^k$ by its multi-linear extension. Again, by abusing the notation, define the extension P' as:

$$P'(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k) = \sum_{\sigma \in \Sigma^k} P'(\sigma) \prod_{i=1}^{k} x_{i,\varsigma(\sigma_i)}, \text{ for all } \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k \in \blacktriangle_q. \tag{5}$$

Define the soundness of the test as:

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

Extending P' to $\mathbb{R}^{qk}$: We will extend the payoff function P' further to a real valued function on $(\mathbb{R}^q)^k$, by plugging the real values in the expansion of P' given in the Equation (5). This extension of P' is smooth in the following sense:

1. All the partial derivatives of P' up to order 3 are uniformly bounded by $C_0(q, k)$.
2. P' is a Lipschitz function with Lipschitz constant $C_0(q, k)$, i.e.,$\forall \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}, \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\} \in (\mathbb{R}^q)^k$,

$$|P'(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) - P'(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k)| \leqslant C_0(q, k) \sum_{i=1}^{k} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|_2.$$

Setting of parameters. Let $\xi > 0$ be the parameter from Lemma 2.6. Let $\delta > 0$ be a sufficiently small constant. Set $\eta \in (0, 1)$ to be the smallest constant such that for all $\ell \geqslant 0$,

$$(1 - \xi)^\ell (1 - (1 - \delta)^\ell)^2 \leqslant \eta.$$

Note that as $\delta \to 0$, $\eta(\delta) \to 0$. We will denote the smallest non-zero probability of an atom in the SDP local distribution by α. As the SDP instance is finite, we can assume that $\alpha > 0$ independent of R.

Input: An SDP solution $(\boldsymbol{V}, \boldsymbol{\mu})$.

Setup: For each $s \in \mathcal{V}$, the probability space $\Omega_s = (\Sigma, \mu_s)$ consists of atoms in Σ with the distribution $\mu_s(a) = \|b_{s,a}\|^2$. Let $\boldsymbol{\mathcal{F}}_s$ denote the function obtained by interpreting the function $\boldsymbol{\mathcal{F}} : \Sigma^R \to \blacktriangle_q$ as a function over Ω_s^R. Let $\boldsymbol{\mathcal{H}}_s = T_{1-\delta}\boldsymbol{\mathcal{F}}_s$ for all $s \in \mathcal{V}$. Let $\boldsymbol{F}_s, \boldsymbol{H}_s$ denote the multilinear polynomials corresponding to functions $\boldsymbol{\mathcal{F}}_s, \boldsymbol{\mathcal{H}}_s$ respectively.

Rounding Scheme:

Step I: Sample R Gaussian vectors $\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}, \ldots, \boldsymbol{\zeta}^{(R)}$ 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 $g_{s,0}^{(j)} \equiv \mathbf{1}$ and for $c \in \{1, \ldots, q-1\}$, set

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

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

2. Evaluate the multilinear polynomial $\boldsymbol{H}_s$ with $\boldsymbol{g}_s$ as inputs to obtain $\boldsymbol{p}_s \in \mathbb{R}^q$, i.e., $\boldsymbol{p}_s = \boldsymbol{H}_s(\boldsymbol{g}_s)$.

3. Round $\boldsymbol{p}_s$ to $\boldsymbol{p}_s^*$.

$$\boldsymbol{p}_s^* = \mathbf{Scale}(f_{[0,1]}((\boldsymbol{p}_s)_1), f_{[0,1]}((\boldsymbol{p}_s)_2), \ldots, f_{[0,1]}((\boldsymbol{p}_s)_q)),$$

where

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

and

$$\mathbf{Scale}(x_1, x_2, \ldots, x_q) = \begin{cases} \frac{1}{\sum_i x_i}(x_1, x_2, \ldots, x_q) & \text{if } \sum_i x_i \neq 0, \\ (1, 0, 0, \ldots, 0) & \text{if } \sum_i x_i = 0. \end{cases}$$

4. Assign the variable $s \in \mathcal{V}$ a value $a \in \Sigma$ with probability $(\boldsymbol{p}_s^*)_{\varsigma^{-1}(a)}$.

Step III: Output the assignment from Step II.

Figure 3: Rounding Scheme $\mathbf{Round}_{\boldsymbol{\mathcal{F}}}$.

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.

We also define the following global ensembles of random variables:

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

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

and $\boldsymbol{\zeta}$ is a normal Gaussian random vector of appropriate dimension.

The following lemma states that the local integral ensemble and the global Gaussian ensemble have matching first and second moments. We need this to apply the invariance principle in our analysis below.

Lemma 2.21. *For every $s \in \mathcal{V}$, $\boldsymbol{g}_s$ is an orthonormal ensemble w.r.t. the space Ω_s. Also, for any payoff $P' \in \mathcal{P}$, the global ensembles $\mathcal{G}$ match the following moments of the local integral ensembles $\mathcal{L}$:*

$$\mathop{\mathbf{E}}_{\boldsymbol{\zeta}}[g_{s,c}.g_{s',c'}] = \mathop{\mathbf{E}}_{(\omega,\omega')\sim\mu_{P'}|(s,s')}[\ell_{s,c}(\omega).\ell_{s',c'}(\omega')] \quad \forall c,c' \in \{1,\dots,q-1\}, s,s' \in \mathcal{V}(P'),$$

where $\mu_{P'}|(s,s')$ is the marginal distribution of $\mu_{P'}$ on the coordinates of s,s'.

Proof. For any $s,s' \in \mathcal{V}$ and $c,c' \in \{1,\dots,q-1\}$, we have

$$\begin{aligned}
\mathbf{E}[g_{s,c}.g_{s',c'}] &= \mathbf{E}\left[\sum_{\omega\in\Sigma} \ell_{s,c}(\omega)\langle \boldsymbol{b}_{s,\omega}, \boldsymbol{\zeta}\rangle \sum_{\omega'\in\Sigma} \ell_{s',c'}(\omega')\langle \boldsymbol{b}_{s',\omega'}, \boldsymbol{\zeta}\rangle\right] \\
&= \sum_{\omega,\omega'\in\Sigma} \ell_{s,c}(\omega)\ell_{s',c'}(\omega')\,\mathbf{E}\left[\langle \boldsymbol{b}_{s,\omega}, \boldsymbol{\zeta}\rangle\langle \boldsymbol{b}_{s',\omega'}, \boldsymbol{\zeta}\rangle\right] \\
&= \sum_{\omega,\omega'\in\Sigma} \ell_{s,c}(\omega)\ell_{s',c'}(\omega')\langle \boldsymbol{b}_{s,\omega}, \boldsymbol{b}_{s',\omega'}\rangle. \qquad (6)
\end{aligned}$$

Now, when $s = s'$, for $\omega \neq \omega'$, $\langle \boldsymbol{b}_{s,\omega}, \boldsymbol{b}_{s',\omega'}\rangle = 0$ because of the SDP constraints (1). Therefore, in this case

$$\mathbf{E}[g_{s,c}.g_{s,c'}] = \sum_{\omega\in\Sigma} \ell_{s,c}(\omega)\ell_{s,c'}(\omega)\|\boldsymbol{b}_{s,\omega}\|_2^2 = \mathop{\mathbf{E}}_{\omega\sim\mu_s}[\ell_{s,c}(\omega)\ell_{s,c'}(\omega)] = \langle \ell_{s,c}, \ell_{s,c'}\rangle_{\mu_s},$$

which is 1 when $c = c'$ and 0 otherwise. This shows the orthonormality of $\boldsymbol{g}_s$. Coming back to the Equation (6), again by the SDP constraints (1), the inner-product $\langle \boldsymbol{b}_{s,\omega}, \boldsymbol{b}_{s',\omega'}\rangle$ is precisely the probability of (ω,ω') according to the distribution $\mu_{P'}|(s,s')$ for any payoff P' containing s and s'. This proves the lemma. ☐

Let $\mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V},\boldsymbol{\mu})$ be the expected value of the assignment returned by the rounding algorithm in Figure 3. In this section, we prove the following soundness lemma.

Lemma 2.22. *Let $k = 3$ and assume that the SDP solution is semi-rich and does not satisfy any linear equation. Then, for any (τ,δ)-quasirandom function $\boldsymbol{\mathcal{F}}$,*

$$\mathrm{Soundness}(\mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}) \leqslant \mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V},\boldsymbol{\mu}) + o_{\delta,\tau}(1).$$

The notation $o_{\delta,\tau}(1)$ means that it goes to 0 as $\delta \to 0$ and $\tau \to 0$. Therefore, in this case the acceptance probability of the test is upper bounded by the integral value of the given instance. This shows that if there exists an $(1,s)$ integrality gap instance of Max-P-CSP, then there exists a dictatorship test with completeness 1 and soundness $s + \varepsilon$ for any constant $\varepsilon > 0$.

Remark 2.23. *If we can extend our main analytical lemma to other predicates, then we can remove the condition on the predicate from Lemma 2.22.*

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

$$\mathbf{Pr}[\boldsymbol{\mathcal{F}} \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}] = \mathop{\mathbf{E}}_{P'\sim\mathcal{P}}\mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}}[P'(\boldsymbol{\mathcal{F}}_{s_1}(\boldsymbol{z}_{s_1}), \boldsymbol{\mathcal{F}}_{s_2}(\boldsymbol{z}_{s_2}), \dots, \boldsymbol{\mathcal{F}}_{s_k}(\boldsymbol{z}_{s_k}))].$$

We will prove a series of claims which will help us relate the probability to $\mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V},\boldsymbol{\mu})$. We begin with the following claim which shows that we can replace $\boldsymbol{\mathcal{F}}$ with its *noisy* version $T_{1-\delta}\boldsymbol{\mathcal{F}}$. Here, we use the main analytical lemma (Lemma 2.6).

Claim 2.24 (Changing $\boldsymbol{\mathcal{F}}$ to $\boldsymbol{\mathcal{H}}$)**.** *Let $k = 3$ and assume that the SDP solution is semi-rich and does not satisfy any linear equation. Then for every $P' \in \mathcal{P}$,*

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} [P'(\boldsymbol{\mathcal{F}}_{s_1}(\boldsymbol{z}_{s_1}), \boldsymbol{\mathcal{F}}_{s_2}(\boldsymbol{z}_{s_2}), \ldots, \boldsymbol{\mathcal{F}}_{s_k}(\boldsymbol{z}_{s_k}))] - \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} [P'(\boldsymbol{\mathcal{H}}_{s_1}(\boldsymbol{z}_{s_1}), \boldsymbol{\mathcal{H}}_{s_2}(\boldsymbol{z}_{s_2}), \ldots, \boldsymbol{\mathcal{H}}_{s_k}(\boldsymbol{z}_{s_k}))] \right| \leqslant \eta(\delta).$$

Proof. Consider the following expression.

$$P'(\boldsymbol{\mathcal{F}}_{s_1}(\boldsymbol{z}_{s_1}), \boldsymbol{\mathcal{F}}_{s_2}(\boldsymbol{z}_{s_2}), \ldots, \boldsymbol{\mathcal{F}}_{s_k}(\boldsymbol{z}_{s_k})) = \sum_{\sigma \in \Sigma^k} P'(\sigma) \prod_{j=1}^{k} \mathcal{F}_{s_j, \sigma_j}(\boldsymbol{z}_{s_j}).$$

We will show that for all $P' \in \mathcal{P}$ and $\sigma \in \Sigma^k$,

$$\Gamma := \left| \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} \left[\prod_{j=1}^{k} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \right] - \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} \left[\prod_{j=1}^{k} \mathcal{H}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \right] \right| \leqslant \eta.$$

Let us define $\Gamma_{j'}$ for $j' = 1, \ldots, k$ as follows:

$$\Gamma_{j'} := \left| \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} \left[\prod_{j=1}^{j'-1} \mathcal{H}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \prod_{j=j'}^{k} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \right] - \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} \left[\prod_{j=1}^{j'} \mathcal{H}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \prod_{j=j'+1}^{k} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \right] \right|.$$

By triangle inequality, $\Gamma \leqslant \sum_{j'} \Gamma_{j'}$.

$$\begin{aligned}
\Gamma_{j'} &= \left| \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} \left[\prod_{j=1}^{j'-1} \mathcal{H}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \prod_{j=j'}^{k} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \right] - \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} \left[\prod_{j=1}^{j'} \mathcal{H}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \prod_{j=j'+1}^{k} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \right] \right| \\
&= \left| \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} \left[(\mathcal{F}_{s_{j'},\sigma_{j'}}(\boldsymbol{z}_{s_{j'}}) - \mathcal{H}_{s_{j'},\sigma_{j'}}(\boldsymbol{z}_{s_{j'}})) \cdot \prod_{j=1}^{j'-1} \mathcal{H}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \prod_{j=j'+1}^{k} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \right] \right| \\
&= \left| \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}} \left[(\mathbf{Id} - T_{1-\delta}) \mathcal{F}_{s_{j'},\sigma_{j'}}(\boldsymbol{z}_{s_{j'}}) \cdot \prod_{j=1}^{j'-1} \mathcal{H}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \prod_{j=j'+1}^{k} \mathcal{F}_{s_j,\sigma_j}(\boldsymbol{z}_{s_j}) \right] \right|.
\end{aligned}$$

Here, $\mathbf{Id}$ is the identity operator. Now, the function $Q := (\mathbf{Id} - T_{1-\delta})\mathcal{F}_{s_{j'},\sigma_{j'}}(\boldsymbol{z}_{s_{j'}})$ is a function $Q : \Sigma^R \to [0,1]$ that satisfies the property of being a 'high-degree' function: Using the Efron-Stein decomposition of Q and using Fact 2.14, we have

$$\langle Q, T_{1-\xi} Q \rangle_{\mu_{s_{j'}}^{\otimes R}} = \sum_{S \subseteq [R]} (1-\xi)^{|S|} (1 - (1-\delta)^{|S|})^2 \|(\mathcal{F}_{s_{j'},\sigma_{j'}})_S\|_2^2.$$

Now, $(1-\xi)^\ell (1 - (1-\delta)^\ell)^2 \leqslant \eta$ for every $\ell \geqslant 0$. Therefore,

$$\langle Q, T_{1-\xi} Q \rangle_{\mu_{s_{j'}}^{\otimes R}} \leqslant \eta \sum_{S \subseteq [R]} \|(\mathcal{F}_{s_{j'},\sigma_{j'}})_S\|_2^2 \leqslant \eta(\delta).$$

Hence, the product inside the expectation satisfies the hypothesis of Lemma 2.6, with $\mathsf{Stab}_{1-\xi}(Q) \leqslant \eta(\delta)$. Applying the lemma, we conclude that $\Gamma_{j'} \leqslant \eta(\delta)/k$, where $\eta(\delta) \to 0$ as $\delta \to 0$. Therefore, $\Gamma \leqslant \sum_{j'} \Gamma_{j'} \leqslant \eta(\delta)$. $\square$

We now switch to the multilinear polynomials. By definition, we have

$$\mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}}[P'(\boldsymbol{\mathcal{H}}_{s_1}(\boldsymbol{z}_{s_1}), \boldsymbol{\mathcal{H}}_{s_2}(\boldsymbol{z}_{s_2}), \ldots, \boldsymbol{\mathcal{H}}_{s_k}(\boldsymbol{z}_{s_k}))] = \mathop{\mathbf{E}}_{\mathcal{L}^R_{P'}}[P'(\boldsymbol{H}_{s_1}(\boldsymbol{\ell}_{s_1}), \boldsymbol{H}_{s_2}(\boldsymbol{\ell}_{s_2}), \ldots, \boldsymbol{H}_{s_k}(\boldsymbol{\ell}_{s_k}))].$$

Here, $\mathcal{L}_{P'}$ is the joint distribution of the local ensembles based on the distribution $\mu_{P'}$. We now apply the Invariance Principle to replace the Integral Ensembles with the Gaussian Ensembles.

Claim 2.25. *(Moving to the global Gaussian ensembles) Using the invariance principle, for every $P' \in \mathcal{P}$, we have*

$$\left| \mathop{\mathbf{E}}_{\mathcal{L}^R_{P'}}[P'(\boldsymbol{H}_{s_1}(\boldsymbol{\ell}_{s_1}), \boldsymbol{H}_{s_2}(\boldsymbol{\ell}_{s_2}), \ldots, \boldsymbol{H}_{s_k}(\boldsymbol{\ell}_{s_k}))] - \mathop{\mathbf{E}}_{\mathcal{G}^R_{P'}}[P'(\boldsymbol{H}_{s_1}(\boldsymbol{g}_{s_1}), \boldsymbol{H}_{s_2}(\boldsymbol{g}_{s_2}), \ldots, \boldsymbol{H}_{s_k}(\boldsymbol{g}_{s_k}))] \right| \leqslant \tau^{O_{\delta,\alpha}(1)}.$$

Proof. This claim follows directly from the Invariance Principle, i.e., from Theorem 2.15, and using Lemma 2.21. Here, the maximum influence of the functions is at most τ and any non-zero probability of an atom is at least α. Also, for $D = O(\log_{1-\delta} \tau)$, the degree $\geqslant D$ weight of the functions $\boldsymbol{H}_s$ is at most $O(\tau)$. This is as follows.

$$W^{\geqslant D}[\mathcal{H}_{s,i}; \mu^{\otimes R}] = \sum_{\boldsymbol{\sigma}:|\boldsymbol{\sigma}|\geqslant D} (\hat{\mathcal{H}_{s,i}})^2_{\boldsymbol{\sigma}} = \sum_{\boldsymbol{\sigma}:|\boldsymbol{\sigma}|\geqslant D} (1-\delta)^{|\boldsymbol{\sigma}|} (\hat{\mathcal{F}_{s,i}})^2_{\boldsymbol{\sigma}} \leqslant (1-\delta)^D \leqslant \tau.$$

Therefore, $W^{\geqslant D}[\mathcal{H}_s; \mu^{\otimes R}] = \sum_{j=1}^q W^{\geqslant D}[\mathcal{H}_{s,j}; \mu^{\otimes R}] \leqslant q \cdot \tau = O(\tau)$. ☐

The final claim shows that, as far as the multilinear polynomial evaluations are concerned, the rounding step (Step II (3)) does not change the expectation by much if the function $\boldsymbol{\mathcal{F}}$ is a quasirandom function.

Claim 2.26. *(Analyzing the loss due to truncation and scaling) For every payoff $P' \in \mathcal{P}$,*

$$\left| \mathop{\mathbf{E}}_{\mathcal{G}^R}[P'(\boldsymbol{H}_{s_1}(\boldsymbol{g}_{s_1})^\star, \boldsymbol{H}_{s_2}(\boldsymbol{g}_{s_2})^\star, \ldots, \boldsymbol{H}_{s_k}(\boldsymbol{g}_{s_k})^\star)] - \mathop{\mathbf{E}}_{\mathcal{G}^R}[P'(\boldsymbol{H}_{s_1}(\boldsymbol{g}_{s_1}), \boldsymbol{H}_{s_2}(\boldsymbol{g}_{s_2}), \ldots, \boldsymbol{H}_{s_k}(\boldsymbol{g}_{s_k}))] \right| \leqslant \tau^{O_{\delta,\alpha}(1)}.$$

Proof. $\boldsymbol{\mathcal{H}}_{s_j} = T_{1-\delta}\boldsymbol{\mathcal{F}}_{s_j}$ is over the domain Σ^R and has the range $\blacktriangle_q$. The difference between the first and the second expression (rounding error because of scaling and truncation) is bounded by $O(C_0, q) \cdot \sum_{s \in \mathcal{V}(P')} \mathbf{E}[\boldsymbol{\xi}(\boldsymbol{H}_s(\boldsymbol{g}_s))]$ [Rag09, Claim 7.4.2], where $\boldsymbol{\xi}(\boldsymbol{a}) = \sum_j (f_{[0,1]}(a_j) - a_j)^2$ and C_0 is an absolute constant from the smoothness property of the payoff P'. We know that $\mathbf{E}[\boldsymbol{\xi}(\boldsymbol{H}_s(\boldsymbol{\ell}_s))] = 0$, as $\boldsymbol{H}_s(\boldsymbol{\ell}_s) \in \blacktriangle_q$. Now, we can apply the invariance principle to conclude

$$\left| \mathop{\mathbf{E}}_{\mathcal{G}^R}[\boldsymbol{\xi}(\boldsymbol{H}_s(\boldsymbol{g}_s))] - \mathop{\mathbf{E}}_{\mathcal{L}^R_{P'}}[\boldsymbol{\xi}(\boldsymbol{H}_s(\boldsymbol{\ell}_s))] \right| \leqslant \tau^{O_{\delta,\alpha}(1)}.$$

As $\mathbf{E}[\boldsymbol{\xi}(\boldsymbol{H}_s(\boldsymbol{\ell}_s))] = 0$, the claim follows. ☐

Proof of Lemma 2.22. We are now ready to prove the soundness of the test: The value returned by the rounding scheme is

$$\mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V}, \boldsymbol{\mu}) = \mathop{\mathbf{E}}_{P' \in \mathcal{P}} \mathop{\mathbf{E}}_{\mathcal{G}^R} \left[P'(\boldsymbol{H}_{s_1}(\boldsymbol{g}_{s_1})^\star, \boldsymbol{H}_{s_2}(\boldsymbol{g}_{s_2})^\star, \ldots, \boldsymbol{H}_{s_k}(\boldsymbol{g}_{s_k})^\star)\right]$$

and the soundness of the test is given by the following expression:

$$\mathbf{Pr}[\boldsymbol{\mathcal{F}} \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}] = \mathop{\mathbf{E}}_{P' \sim \mathcal{P}} \mathop{\mathbf{E}}_{\boldsymbol{z}_{P'}}[P'(\boldsymbol{\mathcal{F}}_{s_1}(\boldsymbol{z}_{s_1}), \boldsymbol{\mathcal{F}}_{s_1}(\boldsymbol{z}_{s_2}), \ldots, \boldsymbol{\mathcal{F}}_{s_k}(\boldsymbol{z}_{s_k}))].$$

For $k = 3$, using the Claims 2.24, 2.25, 2.26 that we proved earlier, we can relate the two quantities as follows:

$$\mathbf{Pr}[\boldsymbol{\mathcal{F}} \text{ passes } \mathbf{Dict}_{\boldsymbol{V},\boldsymbol{\mu}}] \leqslant \mathbf{Round}_{\boldsymbol{\mathcal{F}}}(\boldsymbol{V}, \boldsymbol{\mu}) + \eta(\delta) + \tau^{O_{\delta,\alpha}(1)}.$$

Now, $\eta(\delta) \to 0$ as $\delta \to 0$. Therefore, we get

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

as required. $\square$

3 The main analytical lemma

In this section, we prove our main analytical lemma (Lemma 2.6). We begin by addressing a more specialized case, in which the requirement of semi-rich support of the distribution is replaced with the stronger condition that the support of the distribution is a *union of matchings*:

Definition 3.1. *We say a set $S \subseteq \Sigma \times \Phi \times \Gamma$ is a union of matchings if there exists $\Sigma' \subseteq \Sigma$ and a collection of matchings $M_x \subseteq \Phi \times \Gamma$, one for each $x \in \Sigma'$, such that*

$$S = \bigcup_{x \in \Sigma'} \{x\} \times M_x.$$

The version of Lemma 2.6 for union of matchings is Lemma 3.2 stated below; another difference is that below we introduce some asymmetry in the roles of f, g and h, and we need the stability of either g or h to be small. In Section 3.7 we explain the slight adaptations that allow our argument to go through in the case of semi-rich support, and then explain how to generalize the statement to the case the stability of f is small (thereby establishing Lemma 2.6).

Lemma 3.2. *For all $m \in \mathbb{N}$, $\varepsilon, \alpha > 0$ there exist $\xi > 0$ and $\delta > 0$ such that the following holds. Suppose μ is a distribution over $\Sigma \times \Phi \times \Gamma$ whose support (a) is a union of matchings, and (b) cannot be embedded in an Abelian group. Further suppose that $|\Sigma|, |\Phi|, |\Gamma| \leqslant m$ and each atom in μ has probability at least α. Then, if $f\colon \Sigma^n \to [-1,1]$ $g\colon \Phi^n \to [-1,1]$, $h\colon \Gamma^n \to [-1,1]$ are functions such that*

- $\mathsf{Stab}_{1-\xi}(g) \leqslant \delta$ *or* $\mathsf{Stab}_{1-\xi}(h) \leqslant \delta$.

Then $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \varepsilon$.

As the roles of g and h will be interchangeable in our arguments, without loss of generality we shall focus on the case that $\mathsf{Stab}_{1-\xi}(g) \leqslant \delta$ throughout this section. Before proceeding to the formal argument, we begin with a quick overview of the proof that outlines the main components involved.

Proof overview. The proof of Lemma 3.2 consists of several steps. We think of $\mathsf{supp}(\mu)$ as a graph between Φ and Γ, wherein edges are labeled by elements of Σ in the natural way. Our initial premise is that for each $x \in \Sigma$, the collection of edges labeled by x forms a matching, and we perform several steps in order to improve the structure we have on that graph (by possibly increasing the size of the alphabet Σ).

1. Let $T_x \in \{0,1\}^{\Phi\times\Gamma}$ be the permutation matrix corresponding to the matching labeled by x. First, we show that by moving to a different distribution μ' satisfying similar properties to μ', we may assume that not only the edges of T_x lie in the graph of μ', but rather also the edges of $T_{x_1}T_{x_2}^{t}T_{x_3}$ for any $x_1, x_2, x_3 \in \Sigma$. In other words, we may compose various matchings and "insert" them into the support of our distribution. Performing this step $\ell = O_m(1)$ times, we get that as the graph of μ is connected, we would end up with the complete bipartite graph between Φ and Γ. We now move on to a similar looking expectation to the one in the main lemma but for μ', which is a distribution over $\Sigma^{\ell} \times \Phi \times \Gamma$.
2. We next reduce the size of the alphabet Σ^{ℓ} to be smaller. Note that for each $\vec{x} \in \Sigma^{\ell}$, the edges in the graph of μ' labeled by $\vec{x}$ form a matching. We show that if for $\vec{x}, \vec{x}'$ these matchings are not edge disjoint, then we may glue together the symbols $\vec{x}, \vec{x}'$ and modify the distribution μ' and the functions f, g, h (in a way that preserves their various properties) so that the expectation does not drop too

much. The edges of the new symbols will consist of the union of the edges of the old symbols, and the new alphabet for x is $\Sigma' \subseteq \Sigma^{\ell}$

We note that in such operation, if the matchings corresponding to $\vec{x}, \vec{x}'$ were not identical, then the edges corresponding to the new symbol will not form a matching. We show that in that case, one may further do identification of symbols in Φ and Γ that preserve the properties of the distributions and the functions, and keeps the expectation high. Performing such identification steps sufficiently many times, one returns to the case wherein for each $x \in \Sigma'$ the edges corresponding to x form a matching. We note that each time we perform such step, the alphabet of y or z drops by at least 1, so in total we will have at most $2m$ such steps.

3. We thus reach new alphabets Σ', Φ', Γ'. We consider further operations of composing three x-matchings, i.e. moving from Σ' to Σ'^3. We say that this move is worthwhile if doing it, and then the subsequent identifications, the alphabets Φ', Γ' will shrink further. As long as performing this move is worthwhile, we do so and otherwise we proceed to the next step.

4. After performing $O_m(1)$ steps as in the previous item, we reach to the state wherein the alphabets are Σ'', Φ'' and Γ'', and it is no longer worthwhile to execute the previous step. This means that for every $(x_1, x_2, x_3) \in \Sigma''^3$ and $(x_4, x_5, x_6) \in \Sigma''^3$, the permutations $T_{x_1} T_{x_2}^t T_{x_3}$ and $T_{x_4} T_{x_5}^t T_{x_6}$ are either identical, or are edge disjoint (otherwise we would be able to execute the previous step once more). We use this structure in order to identify a non-Abelian group structure.

   More specifically, we construct a group $(G, \cdot)$ that has no representations of dimension 1 (besides the trivial representation), such that our expectation is $\mathbf{E}_{(g_1,g_2,g_3): g_3 = g_1 g_2}\left[f'(g_1) g'(g_2) h'(g_3)\right]$. Here, f', g', h' are really the same as the functions f, g, h we have, except that they interpret their input as elements from G. We argue that the fact that g' is highly noise sensitive implies that almost all of the mass of g' (with respect to the representation theoretic Fourier decomposition over G) lies on the high degrees. We use this fact along with basic Fourier analysis in order to give an upper bound on the expectation above that vanishes as $\xi, \delta \to 0$ (uniformly in n), and hence finish the proof.

3.1 Step 1: turning the graph into a bipartite clique

Throughout the proof, we suppose the support of μ on each one of the components is full, otherwise we may shrink Σ, Φ or Γ.

3.1.1 Reduction 1: assuming a formula for h

Given f, g, h, let $\tilde{h}(z) = \mathbf{E}_{(\mathbf{x},\mathbf{y},\mathbf{z}) \sim \mu^{\otimes n}}\left[f(\mathbf{x}) g(\mathbf{y}) \mid \mathbf{z} = z\right]$.

Claim 3.3. $\left|\mathbf{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})}\left[f(\mathbf{x}) g(\mathbf{y}) h(\mathbf{z})\right]\right|^2 \leqslant \left|\mathbf{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})}\left[f(\mathbf{x}) g(\mathbf{y}) \tilde{h}(\mathbf{z})\right]\right|$.

Proof. Note that the left hand side is $\mathbf{E}_{\mathbf{z} \sim \mu_z}\left[h(\mathbf{z}) \tilde{h}(\mathbf{z})\right]^2$, hence by Cauchy-Schwarz it is upper bounded by $\|h\|_{2;\mu_z}^2 \|\tilde{h}\|_{2;\mu_z}^2$. Using the fact that $\|h\|_{2;\mu_z} \leqslant 1$ and that the right hand side is equal to $\|\tilde{h}\|_{2;\mu_z}^2$, the proof is concluded. ☐

Thus, it suffices to prove Lemma 2.6 under the additional assumption that $h = \tilde{h}$, and we assume that henceforth.

3.1.2 Reduction 2: composing matchings

Consider the support of the distribution μ. The purpose of the current section is to show that one may assume without loss of generality that μ has full support on $\Phi \times \Gamma$, at the expense of enlarging the alphabet Σ. Towards this end, we begin with the following observation.

Consider the bipartite graph $G_\mu = (\Phi \cup \Gamma, E)$, where $E = \{\, (y, z) \mid \exists x \in \Sigma \text{ such that } (x, y, z) \in \mathsf{supp}(\mu)\}$.

Lemma 3.4. *If $\mathsf{supp}(\mu)$ cannot be embedded in an Abelian group, then G_μ is connected.*

Proof. Suppose G_μ is disconnected. Then it contains at least two connected components, and without loss of generality we may partition $\Phi = \Phi_1 \cup \Phi_2$, $\Gamma = \Gamma_1 \cup \Gamma_2$ such that $\Phi_1, \Gamma_1, \Gamma_2$ are non-empty and G_μ does not contain an edge from Γ_1 to any vertex in Φ_1.

We define $\sigma\colon \Sigma \to \mathbb{F}_3$, $\phi\colon \Phi \to \mathbb{F}_3$ and $\gamma\colon \Gamma \to \mathbb{F}_3$ by $\sigma(x) = 0$, $\phi(y) = i$ if $y \in \Phi_i$, and $\gamma(z) = i$ if $z \in \Gamma_i$. As $(x, y, z) \in \mathsf{supp}(\mu)$ implies that either ($y \in \Phi_1$ and $z \in \Gamma_2$) or ($y \in \Phi_2$ and $z \in \Gamma_1$), we get that $\sigma(x) + \phi(y) + \gamma(z) = 0 \pmod 3$, hence this is an embedding of $\mathsf{supp}(\mu)$ into $(\mathbb{F}_3, +)$. $\square$

We use Van-der Corput type argument to argue we may compose matchings as described in the proof overview.

Lemma 3.5. *For all $\alpha > 0$, there exists $\alpha' > 0$ such that the following holds. Given a distribution μ over $\Sigma \times \Phi \times \Gamma$, consider the distribution ν over $\Sigma^2 \times \Phi^2 \times \Gamma$ defined as: sample $\mathbf{z} \sim \mu_z$, then sample $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{z}^1), (\mathbf{x}^2, \mathbf{y}^2, \mathbf{z}^2) \sim \mu$ conditioned on $\mathbf{z}^1 = \mathbf{z}^2 = \mathbf{z}$, and output $(\mathbf{x}^1, \mathbf{x}^2, \mathbf{y}^1, \mathbf{y}^2, \mathbf{z})$. The distribution ν satisfies*

1. *The marginal distributions of $(\mathbf{x}^1, \mathbf{y}^1, \mathbf{z})$ and $(\mathbf{x}^2, \mathbf{y}^2, \mathbf{z})$ are μ.*
2. *If the probability of each atom in μ is at least α, then the probability of each atom in ν is at least α'.*
3. *If $f\colon \Sigma^n \to \mathbb{R}$, $g\colon \Phi^n \to \mathbb{R}$, $h\colon \Gamma^n \to \mathbb{R}$ are such that $\|h\|_{2;\mu_z} \leqslant 1$, then*

$$\left|\mathop{\mathbf{E}}_{\mu}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right|^2 \leqslant \mathop{\mathbf{E}}_{\nu}\left[f(\mathbf{x}^1)f(\mathbf{x}^2)g(\mathbf{y}^1)g(\mathbf{y}^2)\right].$$

Proof. The expectation on the left hand side is

$$\begin{aligned}
\left|\mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right|^2 &= \mathop{\mathbf{E}}_{\mathbf{z}\sim\mu_z}\left[h(\mathbf{z})\mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y})\sim\mu_{x,y|\mathbf{z}}}\left[f(\mathbf{x})g(\mathbf{y})\right]\right]^2 \\
&\leqslant \mathop{\mathbf{E}}_{\mathbf{z}\sim\mu_z}\left[h(\mathbf{z})^2\right]\mathop{\mathbf{E}}_{\mathbf{z}\sim\mu_z}\left[\left|\mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y})\sim\mu_{x,y|\mathbf{z}}}\left[f(\mathbf{x})g(\mathbf{y})\right]\right|^2\right] \\
&= \|h\|_{2;\mu_z}^2 \mathop{\mathbf{E}}_{(\mathbf{x}^1,\mathbf{x}^2,\mathbf{y}^1,\mathbf{y}^2,\mathbf{z})\sim\nu}\left[f(\mathbf{x}^1)f(\mathbf{x}^2)g(\mathbf{y}^1)g(\mathbf{y}^2)\right], \qquad (7)
\end{aligned}$$

where we used Cauchy-Schwarz. The first term is bounded as $\|h\|_{2;\mu_z} \leqslant 1$, completing the proof. $\square$

3.1.3 Composing an even number of matchings

Intuitively, Lemma 3.5 allows us to replace $g(y)h(z)$ that have an edge between them in G_μ by $g(y)g(y')$ that have a path of length 2 between them in G_μ. The next lemma, which we will use iteratively, allows us to replace such paths with longer ones.

Lemma 3.6. *For all $\alpha > 0$, there is $\alpha' > 0$ such that the following holds. Suppose we have finite alphabets Σ', Φ and a distribution ν over $(\mathbf{x}, \mathbf{y}^1, \mathbf{y}^2) \sim \Sigma' \times \Phi \times \Phi$. Consider the distribution ν' over $(\Sigma')^2 \times \Phi \times \Phi$ defined as: sample $\mathbf{y} \sim \nu'_{y_2}$, then $(\mathbf{x}, \mathbf{y}^1, \mathbf{y}^2) \sim \nu$, and $(\mathbf{x}', \mathbf{y}^{1'}, \mathbf{y}^{2'}) \sim \nu$ conditioned on $\mathbf{y}^2 = \mathbf{y}^{2'} = \mathbf{y}$ then output $(\mathbf{x}, \mathbf{x}', \mathbf{y}^{1'}, \mathbf{y}^1)$.*

1. *If the probability of each atom in ν is at least α, then the probability of each atom in ν' is at least α'.*
2. *for all $F\colon \Sigma'^n \to \mathbb{R}$, $g\colon \Phi^n \to \mathbb{R}$ such that $\|g\|_{2;\nu_y} \leqslant 1$ it holds that*

$$\left|\mathop{\mathbf{E}}_{\nu}\left[F(\mathbf{x})g(\mathbf{y}^1)g(\mathbf{y}^2)\right]\right|^2 \leqslant \mathop{\mathbf{E}}_{\nu'}\left[F(\mathbf{x})F(\mathbf{x}')g(\mathbf{y}^1)g(\mathbf{y}^{1'})\right].$$

Proof. The left hand side is

$$\left|\mathop{\mathbf{E}}_{\mathbf{y}^2}\left[g(\mathbf{y}^2)\mathbf{E}_{\mathbf{x},\mathbf{y}^1\sim\nu}\left[F(\mathbf{x})g(\mathbf{y}^1)\;\middle|\;\mathbf{y}^2\right]\right]\right|^2 \leqslant \|g\|_{2;\nu_{y^2}}\mathop{\mathbf{E}}_{\mathbf{y}^2}\left[\left|\mathbf{E}_{\mathbf{x},\mathbf{y}^1\sim\nu}\left[F(\mathbf{x})g(\mathbf{y}^1)\;\middle|\;\mathbf{y}^2\right]\right|^2\right],$$

where the last inequality is by Cauchy-Schwarz. Using $\|g\|_{2;\nu_{y^2}} \leqslant 1$ and expanding the last quantity finishes the proof. □

3.1.4 Composing an odd number of steps

Summarizing Lemma 3.5 and Lemma 3.6, we get the following conclusion. Let $\ell \in \mathbb{N}$ be a parameter, and consider the sequence of distributions $\nu_0, \ldots, \nu_\ell$ defined as follows. We have $\nu_0 = \mu$; ν_1 is the distribution ν from Lemma 3.5 for ν_0; ν_2 is the distribution ν' from Lemma 3.6 for ν_1 (also including the z's that were generated). Iteratively, once ν_r has been defined for $r < \ell$, ν_{r+1} is the distribution ν' from Lemma 3.6 for ν_r.

We now state an alternative way to view the distributions ν_r. For each $0 \leqslant r \leqslant \ell$, consider the distribution $\mathcal{D}_r$ over $(x^1, x^{1'}, x^2, x^{2'}, \ldots, x^{2^{r-1}}, x^{2^{r-1}'}, y^1, \ldots, y^{2^{r-1}+1}, z^1, \ldots, z^{2^{r-1}})$ defined as:

1. sample $y^1 \sim \mu_y$;
2. sample (x^1, z^1) from μ conditioned on y^1;
3. sample $(x^{1'}, y^2)$ from μ conditioned on z^1.
4. Iteratively, for $j \leqslant 2^{r-1}$, after sampling y^j sample (x^j, z^j) from μ conditioned on y^j.
5. Iteratively, for $j \leqslant 2^{r-1}$, after sampling z^j sample $(x^{j'}, y^{j+1})$ from μ conditioned on z^j.

A bit less precisely, $\mathcal{D}_r$ describes a random walk of length 2^r in the graph G_μ, starting at a point distributed according to μ_y (which is the stationary distribution of a random walk according to G_μ when weighted appropriately), and at each point we take a random step from our current location. The distribution $\mathcal{D}_r$ also records the label x of the edges that we use: the vectors x^i records the label of the edge taken in the random step from y^i, and the label $x^{i'}$ records the label of the edge taken in the random step from z^i.

We observe that for $r \geqslant 1$, the distribution ν_r, and the joint distribution of the x-part and $y^1, y^{2^{r-1}+1}$ parts of $\mathcal{D}_r$, are identical. For that, we first note that the distribution ν_r also described a random path of length 2^r (which a-priori may have a different distribution), and for convenience we add to the distribution ν_r the entire path as well as the labels it encountered during the walk. The essence of the reason the distributions ν_r and $\mathcal{D}_r$ are identical is the observation that given a bi-regular bipartite graph $G = (L \cup R, E)$, the following two ways of sampling paths are equivalent:

1. Sample $u \in L$ uniformly, sample a neighbour v of u uniformly, and sample a neighbour w of v uniformly; this is similar to the way the distributions $\mathcal{D}_r$ are defined.
2. Sample $v \in R$ uniformly, and sample two neighbours of it u and w independently uniformly; this is similar to the way the distributions ν_r are defined.

Claim 3.7. *The distributions ν_r (after modifying it to include all of the information of the walk), and $\mathcal{D}_r$ are identical.*

Proof. The proof is by induction on r, but first introduce a convenient notion. We say a distribution ν over paths is flip-able if sampling a path $\mathbf{P}$ according to ν and reversing it, say to a path $\mathsf{reverse}(\mathbf{P})$, the distribution of the reversed path $\mathsf{reverse}(\mathbf{P})$ is ν. We note that clearly, the each one of the distributions $\mathcal{D}_r$ is flip-able.

Base case. For $r = 1$, we have

$$\nu_1(x^1, x^{1'}, x^2, x^{2'}, y^1, y^2, z) = \mu_z(z)\mu|_z(x^1, y^1)\mu|_z(x^2, y^2) = \frac{\mu(x^1, y^1, z)\mu(x^2, y^2, z)}{\mu_z(z)},$$

and

$$\mathcal{D}_1(x^1, x^{1\prime}, x^2, x^{2\prime}, y^1, y^2, z) = \mu_y(y^1)\mu|_{y_1}(x^1, z)\mu|_z(x^2, y^2) = \frac{\mu(x^1, y^1, z)\mu(x^2, y^2, z)}{\mu_z(z)},$$

so they are the same.

The inductive step. Writing any path $P = (P_1, P_2)$ where P_1, P_2 have equal length and the endpoint of P_1 is the starting point of P_2, and y is the middle point of P, we note that

$$\mathcal{D}_{r+1}(P) = \mathcal{D}_r(P_1)\frac{\mathcal{D}_r(P_2)}{\mu_y(y)}.$$

Here, we used the fact that the marginal distribution of each y^i (and in particular its middle point) in $\mathcal{D}_{r+1}$ is μ_y. Thus by the flip-ability of $\mathcal{D}_r$ and inductive hypothesis we have

$$\mathcal{D}_{r+1}(P) = \mathcal{D}_r(P_1)\frac{\mathcal{D}_r(\mathsf{reverse}(P_2))}{\mu_y(y)} = \nu_r(P_1)\frac{\nu_r(\mathsf{reverse}(P_2))}{\mu_y(y)}.$$

We now argue that the last expression is exactly $\nu_{r+1}(P)$. To make a sample according to ν_{r+1}, by definition we pick $\mathbf{y} \sim \mu_y$, and then pick $\mathbf{P}_1 \sim \nu_r$, $\mathbf{P}_2 \sim \nu_r$ independently conditioned on their endpoints being $\mathbf{y}$, and then output the path concatenated path, denote it by $\mathbf{P}_1 \circ \mathsf{reverse}(\mathbf{P}_2)$. Thus, the probability of a path $P = (P_1, P_2)$ with middle point y is $\mu_y(y)\frac{\nu_r(P_1)}{\mu_y(y)}\frac{\nu_r(\mathsf{reverse}(P_2))}{\mu_y(y)}$, which is the same as the expression we have above. Here, we used the fact that the distribution of each y^i in the path described by ν_r is μ_y, which is true by the induction hypothesis and the fact that this holds for $\mathcal{D}_r$. □

Lemma 3.8.

$$\left|\mathop{\mathbf{E}}_{\mu}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right|^{2^\ell} \leqslant \mathop{\mathbf{E}}_{\mathcal{D}_\ell}\left[\left(\prod_{i=1}^{2^{\ell-1}-1} f(\mathbf{x}^i)f(\mathbf{x}^{i\prime})\right) f(\mathbf{x}^{2^{\ell-1}})g(\mathbf{y}^1)h(\mathbf{z}^{2^{\ell-1}})\right]$$

Proof. By Lemma 3.5 and iterating Lemma 3.6, we may bound:

$$\left|\mathop{\mathbf{E}}_{\mu}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right|^{2^\ell} \leqslant \mathop{\mathbf{E}}_{\nu_\ell}\left[g(\mathbf{y}^1)g(\mathbf{y}^{2^{\ell-1}+1})\prod_{i=1}^{2^{\ell-1}} f(\mathbf{x}^i)f(\mathbf{x}^{i\prime})\right] = \mathop{\mathbf{E}}_{\mathcal{D}_\ell}\left[g(\mathbf{y}^1)g(\mathbf{y}^{2^{\ell-1}+1})\prod_{i=1}^{2^{\ell-1}} f(\mathbf{x}^i)f(\mathbf{x}^{i\prime})\right],$$

where the last transition is by Claim 3.7. We pull the expectation over everything, except for the random variables $\mathbf{x}^{2^{\ell-1}\prime}$, $y^{2^{\ell-1}+1}$, outside. We get that the last expectation is equal to

$$\mathop{\mathbf{E}}_{\substack{\mathbf{x}^1,\ldots,\mathbf{x}^{2^{\ell-1}} \\ \mathbf{x}^{1\prime},\ldots,\mathbf{x}^{2^{\ell-1}-1\prime} \\ \mathbf{y}^1,\ldots,\mathbf{y}^{2^{\ell-1}} \\ \mathbf{z}^1,\ldots,\mathbf{z}^{2^{\ell-1}}}}\left[f(\mathbf{x}^{2^{\ell-1}})g(\mathbf{y}^1)\prod_{i=1}^{2^{\ell-1}-1} f(\mathbf{x}^i)f(\mathbf{x}^{i\prime})\mathbf{E}_{\mathbf{x}^{2^{\ell-1}\prime}, y^{2^{\ell-1}+1}}\left[f(\mathbf{x}^{2^{\ell-1}\prime})g(\mathbf{y}^{2^{\ell-1}+1})\,\middle|\,\mathbf{z}^{2^{\ell-1}}\right]\right].$$

Here, we used the fact that conditioned on $\mathbf{z}^{2^{\ell-1}}$, the distribution of $\mathbf{x}^{2^{\ell-1}\prime}, \mathbf{y}^{2^{\ell-1}+1}$ is independent of the rest of the random variables (this is clear from the description of the distribution $\mathcal{D}_\ell$). We now observe that by Section 3.1.1, the value of the inner most expectation is precisely $h(\mathbf{z}^{2^{\ell-1}})$, and the proof is concluded. □

Taking large enough ℓ so that the graph between y^1 and $z^{2^{\ell-1}}$ becomes full, we get from Lemma 3.8 that our original expectation is bounded by a similar-looking expectation, wherein the function f becomes $F = \left(\prod_{i=1}^{2^{\ell-1}-1} f(\mathbf{x}^i)f(\mathbf{x}^{i\prime})\right) f(\mathbf{x}^{2^{\ell-1}})$ and so the alphabet of F becomes $\Sigma' = \Sigma^{2^\ell-1}$. We note that this move from f to F preserves the range of the function being $[-1, 1]$, and as $\ell = O_m(1)$ the shift from Σ to Σ' will be essentially irrelevant for us. We may thus assume that in our initial expectation, the graph between y and z was already full.

3.2 Step 2: identifying symbols

In this section, we show that we may assume that in the distribution μ, for any (x, y, z) in the support, the value of any two coordinates implies the value of the third coordinate. Namely, for example, we show that one may assume that if (y, z) is in the support of $\mu_{y,z}$, then there is a unique x such that $(x, y, z) \in \mathsf{supp}(\mu)$.

Definition 3.9. *Suppose Σ, Φ, Γ are finite alphabets, and $P \subseteq \{\, (x, y, z) \mid x \in \Sigma, y \in \Phi, z \in \Gamma\}$. We say that (x, y) determine z in P if for all $x \in \Sigma, y \in \Phi$ there is at most one $z \in \Gamma$ such that $(x, y, z) \in P$. Similarly, we define the notions of (x, z) determining y and (y, z) determining x.*

Lemma 3.10. *For all $\alpha, \varepsilon, \delta, \xi > 0$, there exists $\gamma, \alpha', \xi', \varepsilon', \delta' > 0$ and $N \in \mathbb{N}$ such that the following holds. Let μ be a distribution over $\Sigma \times \Phi \times \Gamma$ in which each atom has probability at least α, $n \geqslant N$, and $f\colon \Sigma^n \to [-1, 1]$, $g\colon \Phi^n \to [-1, 1]$, $h\colon \Gamma^n \to [-1, 1]$ be such that*

1. $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}} \left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \geqslant \delta'$;
2. $\mathsf{Stab}_{1-\varepsilon'}(g) \leqslant \xi'$.

Then, there is $n' \geqslant \gamma n$, $\Sigma' \subseteq \Sigma$, a probability distribution μ' over $\Sigma' \times \Phi \times \Gamma$ and $f'\colon \Sigma'^{n'} \to [-1, 1]$, $g'\colon \Phi^{n'} \to [-1, 1]$, $h'\colon \Gamma^{n'} \to [-1, 1]$ such that

1. $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu'^{\otimes n'}} \left[f'(\mathbf{x})g'(\mathbf{y})h'(\mathbf{z})\right]\right| \geqslant \delta$;
2. $\mathsf{Stab}_{1-\varepsilon}(g') \leqslant \xi$;
3. *each atom in μ' has probability at least α';*
4. *(y, z) determine x;*
5. *if in μ, (y, z) does not determine x, then $\Sigma' \subsetneq \Sigma$.*

Moreover, if the support of μ is not linearly embedded, then the support of μ' is not linearly embedded.

Remark 3.11. *A few remarks are in order.*

1. *The probability distribution μ' is not arbitrary, and is very much related to μ. Roughly speaking, it is the result of partitioning Σ into disjoint groups $\Sigma_1, \ldots, \Sigma_\ell$, then treating all of the symbols in each one as being the same (thinking of the new symbol as the "or" of the old symbols, and in this way forming the support of μ'). In the case that (y, z) dos not determine x, at least one of these groups will be non-trivial, i.e. have size at least 2, in which case the size of Σ' is strictly smaller.*
2. *We intend to iterate this lemma, and variants of which will be discussed later on, several times. Our progress measure which makes sure this process terminates is the size of the alphabets Σ, Φ, Γ that we shift to, hence it is important for us to keep track of it.*
3. *The roles of x and z (and therefore of f and h) are completely symmetric in the above statement, hence we have a variant of the lemma which allows us to make sure that (x, y) determines z, and we will state it formally only later. The role of y (respectively of g) is somewhat distinctive as we wish to preserve the stability of g' to be small, however as we argue later on a slight adaptation of the argument will show that this is also possible, hence we will also be able to assume that (x, z) determines y.*

Before we prove the lemma, we need the following fact.

Fact 3.12. *Let $d \in \mathbb{N}$, $\varepsilon > 0$ be such that $d\varepsilon < 1$. Suppose we have $g\colon (\Phi^n, \nu^{\otimes n}) \to \mathbb{R}$, and choose $J \subseteq [n]$ randomly by including each coordinate with probability $1/d$. Then*

$$\mathop{\mathbf{E}}_{J, z\sim\nu^{\bar{J}}} \left[\mathsf{Stab}_{1-d\varepsilon}(g_{\bar{J}\to z})\right] = \mathsf{Stab}_{1-\varepsilon}(g).$$

Proof. Let $m = |\Phi|$ and write the Fourier expansion of g with respect to ν, i.e. $g(y) = \sum_{u \in [m]^n} \widehat{g}(u)\chi_u(y)$, where $\chi_0 \equiv 1$ is the trivial character. Then the right hand side is equal to $\sum_{u\in[m]^n} (1-\varepsilon)^{|u|}\widehat{g}(u)^2$, where $|u|$ is the number of non-zero entries in u. Similarly, the left hand side is

$$\mathop{\mathbf{E}}_{J,z\sim\nu^{\bar{J}}}\left[\sum_{v\in[m]^J}(1-d\varepsilon)^{|v|}\widehat{g_{\bar{J}\to z}}(v)^2\right] = \mathop{\mathbf{E}}_{J}\left[\sum_{v\in[m]^J}(1-d\varepsilon)^{|v|}\sum_{w\in[m]^n:\mathsf{supp}(w)\cap J=\mathsf{supp}(v)}\widehat{g}(w)^2\right]$$
$$= \sum_{w\in[m]^n}\widehat{g}(w)^2\mathop{\mathbf{E}}_{J}\left[(1-d\varepsilon)^{|\mathsf{supp}(w)\cap J|}\right].$$

Note that for each $i \in [n]$, $\mathbf{E}_J\left[(1-d\varepsilon)^{1_{i\in J}}\right] = \frac{d-1}{d} + \frac{1}{d}(1-d\varepsilon) = 1-\varepsilon$, so the above expectation is exactly $(1-\varepsilon)^{|w|}$, and the proof is concluded. $\square$

Corollary 3.13. *Let $d\in\mathbb{N}$, $\varepsilon>0$ be such that $d\varepsilon<1$, and let $\xi>0$. Suppose we have $g\colon(\Phi^n,\nu^{\otimes n})\to\mathbb{R}$ such that $\mathsf{Stab}_{1-\varepsilon}(g)\leqslant\xi$, and choose $J\subseteq[n]$ randomly by including each coordinate with probability $1/d$. Then*

$$\mathop{\mathbf{Pr}}_{J,z\sim\nu^{\bar{J}}}\left[\mathsf{Stab}_{1-d\varepsilon}(g_{\bar{J}\to z})\geqslant\sqrt{\xi}\right]\leqslant\sqrt{\xi}.$$

Proof. Note that $\mathsf{Stab}_{1-d\varepsilon}(g_{\bar{J}\to z})$ is always positive, and by Fact 3.12 its expectation is at most ξ. Thus, the result follows by Markov's inequality. $\square$

We are now ready to prove Lemma 3.10.

Proof of Lemma 3.10. Consider the averaging operator $T\colon L^2(\Sigma;\mu_x)\to L^2(\Phi\times\Gamma;\mu_{y,z})$ defined by

$$(Tf)(y,z) = \mathbf{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu}\left[f(\mathbf{x})\mid\mathbf{y}=y,\mathbf{z}=z\right],$$

and consider $S=T^*T\colon L^2(\Sigma;\mu_x)\to L^2(\Sigma;\mu_x)$. We note that S is also an averaging operator, and for each $a\in\Sigma$ denote by Sa the distribution over Σ that defines $Sf(a)$. We associate with S a graph $G=(\Sigma,E)$, wherein $a,a'\in\Sigma$ are adjacent if $a'\in\mathsf{supp}(Sa)$, noting that G contains a self loop on each $a\in\Sigma$. We also note that the number of connected components in G is $|\Sigma|$ if and only if (y,z) determine x in μ. Thus, if the number of connected components in G is $|\Sigma|$ there is nothing to prove, so assume that it is $\ell<|\Sigma|$, and denote them by $\Sigma=C_1\cup\ldots\cup C_\ell$.

Note that S is a symmetric operator; also, note that if a function $w\colon\Sigma\to\mathbb{R}$ is only supported on C_i, then Sw is also only supported on C_i. Thus, we may diagonalize S on each connected component separately. Namely, letting $s_i=|C_i|$, we can come up with a collection of $|\Sigma|$ functions, $\chi_{i,j}\colon\Sigma\to\mathbb{R}$, for $i=1,\ldots,\ell$, $j=0,\ldots,s_i-1$ such that

1. $\chi_{i,j}$ is only supported on C_i, and $\chi_{i,0}$ is constant on C_i;
2. the set $\{\chi_{i,j}\}$ is orthonormal with respect to the inner product defined by S;
3. each $\chi_{i,j}$ is an eigenvector of S.

Let $\lambda_{i,j}$ be the eigenvalue of S corresponding to $\chi_{i,j}$. We clearly have $\lambda_{i,0}=1$ for all $i=1,\ldots,\ell$, and as the induced graph on each C_i is connected and has self loops we have that $|\lambda_{i,j}|<1$ for all $i=1,\ldots,\ell$, $j\geqslant1$. Thus, we may pick an absolute constant $0\leqslant\lambda<1$ depending only on $|\Sigma|,\alpha$ such that $|\lambda_{i,j}|\leqslant\lambda$.

We shall use the functions $\chi_{i,j}$ in order to expand our function f. Namely, we can write

$$f(x) = \sum_{\substack{\vec{i}\in[\ell]^n\\ \vec{j}\in[s_{i_1}]\times\ldots\times[s_{i_n}]}}\widehat{f}(\vec{i},\vec{j})\chi_{\vec{i},\vec{j}}(x),$$

where $\chi_{\vec{i},\vec{j}}(x) = \prod_{k=1}^{n} \chi_{i_k,j_k}(x_k)$. We define the effective degree of $\chi_{\vec{i},\vec{j}}$, denoted by $\mathsf{effdeg}(\chi_{\vec{i},\vec{j}})$, to be $|\{k \mid j_k \geqslant 1\}|$, i.e. the number of terms in it that are not constant on their respective connected component, and then partition f into $f = f_L + f_H$ by

$$f_L(x) = \sum_{\substack{\vec{i}\in[\ell]^n \\ \vec{j}\in[s_{i_1}]\times\ldots\times[s_{i_n}] \\ \mathsf{effdeg}(\chi_{\vec{i},\vec{j}})<D}} \widehat{f}(\vec{i},\vec{j})\chi_{\vec{i},\vec{j}}(x), \qquad f_H(x) = \sum_{\substack{\vec{i}\in[\ell]^n \\ \vec{j}\in[s_{i_1}]\times\ldots\times[s_{i_n}] \\ \mathsf{effdeg}(\chi_{\vec{i},\vec{j}})\geqslant D}} \widehat{f}(\vec{i},\vec{j})\chi_{\vec{i},\vec{j}}(x),$$

where $D = 2\frac{\log(2/\delta)}{\log(1/\lambda)}$. Thus, we write

$$\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})] = \underbrace{\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}[f_L(\mathbf{x})g(\mathbf{y})h(\mathbf{z})]}_{(I)} + \underbrace{\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}[f_H(\mathbf{x})g(\mathbf{y})h(\mathbf{z})]}_{(II)}.$$

We begin by bounding the absolute value of (II). Letting $R(y,z) = g(y)h(z)$, we write

$$|(II)| = \left|\mathop{\mathbf{E}}_{\mathbf{y},\mathbf{z}\sim\mu_{y,z}^{\otimes n}}[Tf_H(\mathbf{y},\mathbf{z})R(\mathbf{y},\mathbf{z})]\right| = \left|\langle Tf_H, R\rangle_{\mu_{y,z}^{\otimes n}}\right| \leqslant \|Tf_H\|_{2;\mu_{y,z}^{\otimes n}}\|R\|_{2;\mu_{y,z}^{\otimes n}},$$

where we used Cauchy-Schwarz. We have $\|R\|_{2;\mu_{y,z}^{\otimes n}} \leqslant 1$. As for the first norm, we note

$$\|Tf_H\|_{2;\mu_{y,z}^{\otimes n}}^2 = \langle Tf_H, Tf_H\rangle_{\mu_{y,z}^{\otimes n}} = \langle T^*Tf_H, f_H\rangle_{\mu_x^{\otimes n}} = \langle Sf_H, f\rangle_{\mu_x^{\otimes n}} \leqslant \|Sf_H\|_{2;\mu_x^{\otimes n}}\|f\|_{2;\mu_x^{\otimes n}}.$$

Clearly $\|f\|_{2;\mu_x^{\otimes n}} \leqslant 1$, and by Parseval

$$\|Sf_H\|_{2;\mu_x^{\otimes n}}^2 = \sum_{\substack{\vec{i}\in[\ell]^n \\ \vec{j}\in[s_{i_1}]\times\ldots\times[s_{i_n}] \\ \mathsf{effdeg}(\chi_{\vec{i},\vec{j}})\geqslant D}} \widehat{f}(\vec{i},\vec{j})^2 \prod_{k=1}^{n} \lambda_{i_k,j_k}^2 \leqslant \lambda^{2D},$$

so $\|Sf_H\|_{2;\mu_x^{\otimes n}} \leqslant \lambda^D$. Plugging that back in, we get that $|(II)| \leqslant \frac{\delta}{2}$, and it follows by the premise of the lemma that $|(I)| \geqslant \frac{\delta}{2}$. For the rest of the proof, we assume without loss of generality that I is positive.

Next, we use random restrictions in order to reduce most of the Fourier mass of f to be on characters with effective degree 0. Let $d = \frac{64^2}{\delta^4}D$, choose $J \subseteq [n]$ randomly by including each element with probability $1/d$. Sample a restriction outside J by $(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\tilde{\mathbf{z}}) \sim \mu^{\bar{J}}$, and let $\tilde{f} = (f_L)_{\bar{J}\to\tilde{\mathbf{x}}}$, $\tilde{g} = g_{\bar{J}\to\tilde{\mathbf{y}}}$, $\tilde{h} = h_{\bar{J}\to\tilde{\mathbf{z}}}$. Denote

$$\phi(\tilde{f},\tilde{g},\tilde{h}) = \mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^J}\left[\tilde{f}(\mathbf{x})\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\right].$$

We wish to show that with positive probability over the choice of the restriction, the following five events hold:

1. E_1: The Fourier mass of $\tilde{f}$ on monomials of effective degree more than 0 is at most $\frac{\delta^2}{64}$;
2. E_2: $\tilde{g}$ remains a high-degree function, i.e. $\mathsf{Stab}_{1-\varepsilon}(\tilde{g}) \leqslant \xi$;
3. E_3: $|J| \geqslant \gamma n$ for $\gamma = \frac{1}{2d}$;
4. E_4: $\phi(\tilde{f},\tilde{g},\tilde{h}) \geqslant \frac{\delta}{4}$;
5. E_5: $\|\tilde{f}\|_2 \leqslant \frac{10}{\delta}$.

We will show that the first three events, as well as the fifth one, hold with probability close to 1, whereas the fourth one holds with probability bounded away from 0.

Bounding $\mathbf{Pr}[E_1]$. Computing the expectation of the Fourier mass on characters of effective degree at least 1, we see that

$$\mathbf{E}\left[\sum_{\vec{i}\in[\ell]^J,\vec{j}} \widehat{\tilde{f}}(\vec{i},\vec{j})^2 1_{\mathsf{effdeg}(\chi_{\vec{i},\vec{j}})\geqslant 1}\right] = \sum_{\vec{i}\in[\ell]^n,\vec{j}} \widehat{f_L}(\vec{i},\vec{j})^2 \mathbf{E}\left[1_{|\{k \mid i_k\in J, j_k\geqslant 1\}|\geqslant 1}\right] \leqslant \sum_{\vec{i}\in[\ell]^n,\vec{j}} \widehat{f_L}(\vec{i},\vec{j})^2 \frac{\mathsf{effdeg}(\chi_{\vec{i},\vec{j}})}{d},$$

which is at most $\frac{D}{d}$. Thus, by Markov's inequality the probability of E_1 is at least $1-\frac{64D}{\delta^2 d}\geqslant 1-\frac{\delta^2}{64}$.

Bounding $\mathbf{Pr}[E_2]$. We choose $\varepsilon'=\frac{\varepsilon}{d}$, $\xi'=\min(\xi^2,\delta^8/64^2)$. From Corollary 3.13 we have that

$$\mathop{\mathbf{Pr}}_{J,(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\tilde{\mathbf{z}})\sim\mu^{\bar{J}}}\left[\mathsf{Stab}_{1-\varepsilon}(\tilde{g})\geqslant\sqrt{\xi'}\right]\leqslant\sqrt{\xi'}\leqslant\frac{\delta^2}{64},$$

so $\mathbf{Pr}[E_2]\geqslant 1-\frac{\delta^2}{64}$.

Bounding $\mathbf{Pr}[E_3]$. Choosing $\gamma=\frac{1}{2d}$, it follows by Chernoff's inequality that $\mathbf{Pr}[E_3]\geqslant 1-e^{-\frac{\gamma}{100}n}$.

Bounding $\mathbf{Pr}[E_4]$. Note that $\mathbf{E}_{J,(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\tilde{\mathbf{z}})\sim\mu^{\bar{J}}}\left[\phi(\tilde{f},\tilde{g},\tilde{h})\right]\geqslant\frac{\delta}{2}$. Also, since $\tilde{g},\tilde{h}$ are bounded,

$$\mathop{\mathbf{E}}_{J,(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\tilde{\mathbf{z}})\sim\mu^{\bar{J}}}\left[\phi(\tilde{f},\tilde{g},\tilde{h})^2\right]\leqslant\mathop{\mathbf{E}}_{J,(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\tilde{\mathbf{z}})\sim\mu^{\bar{J}}}\left[\mathop{\mathbf{E}}_{\mathbf{x}\sim\mu_x^J}\left[\left|\tilde{f}(\mathbf{x})\right|\right]\right]=\|f_L\|_{1;\mu_x}\leqslant\|f_L\|_{2;\mu_x}\leqslant\|f\|_{2;\mu_x}\leqslant 1.$$

Hence, it follows that

$$\mathbf{Pr}[E_4]=\mathop{\mathbf{Pr}}_{J,(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\tilde{\mathbf{z}})\sim\mu^{\bar{J}}}\left[\phi(\tilde{f},\tilde{g},\tilde{h})\geqslant\frac{\delta}{4}\right]\geqslant\frac{\delta^2}{16}.$$

Bounding $\mathbf{Pr}[E_5]$. By Markov's inequality,

$$\mathbf{Pr}\left[\|\tilde{f}\|_2>\frac{10}{\delta}\right]=\mathbf{Pr}\left[\|\tilde{f}\|_2^2>\frac{100}{\delta^2}\right]\leqslant\frac{\mathbf{E}\left[\|\tilde{f}\|_2^2\right]}{100/\delta^2}=\frac{\|f_L\|_2^2}{100/\delta^2}\leqslant\frac{\|f\|_2^2}{100/\delta^2}\leqslant\frac{\delta^2}{100}.$$

Summarizing, we get that $\mathbf{Pr}[E_1\cap E_2\cap E_3\cap E_4\cap E_5]\geqslant\frac{\delta^2}{16}-2\frac{\delta^2}{64}-e^{-\frac{\gamma}{100}n}-\frac{\delta^2}{100}>0$, so we may find a restriction as desired. We fix such restriction henceforth.

Write $\tilde{f}=\tilde{f}_0+\tilde{f}_{\neq 0}$ where

$$\tilde{f}_0(x)=\sum_{\substack{\vec{i}\in[\ell]^J\\ \vec{j}\in[s_{i_1}]\times\ldots\times[s_{i_n}]\\ \mathsf{effdeg}(\chi_{\vec{i},\vec{j}})=0}}\widehat{\tilde{f}}(\vec{i},\vec{j})\chi_{\vec{i},\vec{j}}(x),\qquad \tilde{f}_{\neq 0}(x)=\sum_{\substack{\vec{i}\in[\ell]^J\\ \vec{j}\in[s_{i_1}]\times\ldots\times[s_{i_n}]\\ \mathsf{effdeg}(\chi_{\vec{i},\vec{j}})>0}}\widehat{\tilde{f}}(\vec{i},\vec{j})\chi_{\vec{i},\vec{j}}(x).$$

Then we have that $\frac{\delta}{4}\leqslant\phi(\tilde{f},\tilde{g},\tilde{h})=\phi(\tilde{f}_0,\tilde{g},\tilde{h})+\phi(\tilde{f}_{\neq 0},\tilde{g},\tilde{h})$. Note that by Hölder's inequality

$$\left|\phi(\tilde{f}_{\neq 0},\tilde{g},\tilde{h})\right|\leqslant\|\tilde{f}_{\neq 0}\|_2\|\tilde{g}\|_4\|\tilde{h}\|_4\leqslant\sqrt{\frac{\delta^2}{64}}\leqslant\delta/8.$$

Here, we used the fact that E_1 holds so that $\|\tilde{f}_{\neq 0}\|_2\leqslant\sqrt{\frac{\delta^2}{64}}$.

We conclude that $\phi(\tilde{f}_0,\tilde{g},\tilde{h})\geqslant\delta/8$. We have therefore found functions $\tilde{f}_0,\tilde{g},\tilde{h}$ that satisfy almost all of the conditions of the lemma, and it remains to explicitly state the distribution μ' and make $\tilde{f}_0$ bounded.

To define the distribution μ', for each connected component C_i in G pick a representative element, call it a_i. We define $\Sigma'=\{a_i \mid i\in[\ell]\}$, and define the distribution μ' as:

1. sample $(a,b,c)\sim\mu$;

2. let $C_{\mathbf{i}}$ be the connected component of a. Output (a_i,b,c).

It is clear that
$$\mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu'^J}\left[\tilde{f}_0(\mathbf{x})\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\right]=\phi(\tilde{f}_0,\tilde{g},\tilde{h}).$$
We define the function $\tilde{f}_0'\colon\Sigma'^J\to[-1,1]$ by
$$\tilde{f}_0'(x)=\mathbf{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu'^J}\left[\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\;\middle|\;\mathbf{x}=x\right].$$
As in Claim 3.3 we have that
$$\frac{\delta^2}{64}\leqslant\phi(\tilde{f}_0,\tilde{g},\tilde{h})^2=\mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu'^J}\left[\tilde{f}_0(\mathbf{x})\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\right]^2\leqslant\mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu'^J}\left[\tilde{f}_0(\mathbf{x})^2\right]\mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu'^J}\left[\tilde{f}_0'(\mathbf{x})\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\right],$$
and as E_5 holds we have
$$\mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu'^J}\left[\tilde{f}_0(\mathbf{x})^2\right]=\|\tilde{f}_0\|_2^2\leqslant\|\tilde{f}\|_2^2\leqslant\frac{100}{\delta^2},$$
so we get that
$$\mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu'^J}\left[\tilde{f}_0'(\mathbf{x})\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\right]\geqslant\frac{\delta^4}{6400}.$$

Concluding, we have established the first item of the lemma for $\delta'=\frac{\delta^4}{6400}$. The second item of the lemma follows since the marginal on y in both μ' and μ is the same, and the fact that the event E_2 holds. The third item is clear from the definition of μ' for $\alpha'=\alpha$, and the fourth item is clear since we have contracted each connected component of G into a single element. Finally, for the moreover statement, we note that any non-trivial embedding of μ' into an Abelian group naturally induces a non-trivial embedding of μ into an Abelian group (by assigning all elements in C_i the value of the embedding of a_i), and the proof is concluded. $\square$

In Lemma 3.10 allows us to identify symbols in Σ so that in the new distribution μ', (y,z) determines x. As explained in Remark 3.11, as the roles of z and x are symmetric, the lemma also allows us to identify symbols in Γ. The situation is somewhat different if we tried to identify symbols in Φ; the reason is that in very last step of the last argument, we have moved from the function $\tilde{f}_0$ to $\tilde{f}_0'$ (in order to make the function bounded), and if we were to do that for g (in place of f), it may lead to a violation of the stability of g being small. In the next lemma, we show that nevertheless, it is possible to do a small tweak in the end of the argument in order to overcome this issue.

Lemma 3.14. *For all $\alpha,\varepsilon,\delta,\xi>0$, there exists $\gamma,\alpha',\xi',\varepsilon',\delta'>0$ and $N\in\mathbb{N}$ such that the following holds. Let μ be a distribution over $\Sigma\times\Phi\times\Gamma$ in which each atom has probability at least α, $n\geqslant N$, and $f\colon\Sigma^n\to[-1,1]$, $g\colon\Phi^n\to[-1,1]$, $h\colon\Gamma^n\to[-1,1]$ be such that*

1. $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right|\geqslant\delta'$*;*

2. $\mathsf{Stab}_{1-\varepsilon'}(g)\leqslant\xi'$*.*

Then, there is $n'\geqslant\gamma n$, $\Phi'\subseteq\Phi$, a probability distribution μ' and $f'\colon\Sigma^{n'}\to[-1,1]$, $g'\colon\Phi'^{n'}\to[-1,1]$, $h'\colon\Gamma^{n'}\to[-1,1]$ such that

1. $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu'^{\otimes n'}}\left[f'(\mathbf{x})g'(\mathbf{y})h'(\mathbf{z})\right]\right|\geqslant\delta$*;*

2. $\mathsf{Stab}_{1-\varepsilon}(g')\leqslant\xi$*;*

3. *each atom in μ' has probability at least α';*

4. *(x,z) determine y;*

5. *if in μ, (x,z) does not determine y, then $\Phi' \subsetneq \Phi$.*

Moreover, if the support of μ is not linearly embedded, then the support of μ' is not linearly embedded.

Proof. We go through the proof of Lemma 3.10, replacing the roles of f and g. In the end, we have

$$\frac{\delta^2}{64} \leqslant \phi(\tilde{f}, \tilde{g}_0, \tilde{h})^2 = \mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu'^J}\left[\tilde{f}(\mathbf{x})\tilde{g}_0(\mathbf{y})\tilde{h}(\mathbf{z})\right]^2 = \langle \tilde{g}_0, \tilde{g}_0' \rangle^2,$$

and we inspect this quantity. Note that

$$\langle \tilde{g}_0, \tilde{g}_0' \rangle = \langle \tilde{g}_0, (I - T_{1-\varepsilon})\tilde{g}_0' \rangle + \langle \tilde{g}_0, T_{1-\varepsilon}\tilde{g}_0' \rangle .$$

We argue that the second inner product is negligible. Indeed,

$$\langle \tilde{g}_0, T_{1-\varepsilon}\tilde{g}_0' \rangle = \langle T_{1-\varepsilon}\tilde{g}_0, \tilde{g}_0' \rangle \leqslant \|T_{1-\varepsilon}\tilde{g}_0\|_2 \|\tilde{g}_0'\|_2 \leqslant \|T_{1-\varepsilon}\tilde{g}_0\|_2 \leqslant \sqrt{\mathsf{Stab}_{1-\varepsilon}(\tilde{g}_0)} \leqslant \sqrt{\xi} \leqslant \frac{\delta}{16}.$$

We thus conclude that

$$\frac{\delta}{16} \leqslant |\langle \tilde{g}_0, (I - T_{1-\varepsilon})\tilde{g}_0' \rangle| ,$$

and so

$$\frac{\delta^2}{256} \leqslant \langle \tilde{g}_0, (I - T_{1-\varepsilon})\tilde{g}_0' \rangle^2 \leqslant \|\tilde{g}_0\|_2^2 \|(I - T_{1-\varepsilon})\tilde{g}_0'\|_2^2,$$

and as $\|\tilde{g}_0\|_2^2 \leqslant \frac{100}{\delta^2}$ we get that

$$\|(I - T_{1-\varepsilon})\tilde{g}_0'\|_2^2 \geqslant \frac{\delta^4}{25,600}.$$

On the other hand,

$$\|(I - T_{1-\varepsilon})\tilde{g}_0'\|_2^2 = \langle (I - T_{1-\varepsilon})\tilde{g}_0', (I - T_{1-\varepsilon})\tilde{g}_0' \rangle = \left\langle (I - T_{1-\varepsilon})^2\tilde{g}_0', \tilde{g}_0' \right\rangle = \phi(\tilde{f}, (I - T_{1-\varepsilon})^2\tilde{g}_0', \tilde{h}).$$

We have thus proved the statement for $\delta' = \frac{\delta^4}{4\cdot 25,600}$ and the functions $f' = \tilde{f}$, $g' = \frac{1}{4}(I - T_{1-\varepsilon})^2\tilde{g}_0'$ and $h' = \tilde{h}$ (we have divided by 4 to make sure the range of all functions is $[-1,1]$), except that we need to verify that the stability of g' is small. Indeed, we observe that

$$\mathsf{Stab}_{1-\sqrt{\varepsilon}}(g') \leqslant \left\langle g', T_{1-\sqrt{\varepsilon}}g' \right\rangle \leqslant \|T_{1-\sqrt{\varepsilon}}g'\|_2 \leqslant \frac{1}{4}\|T_{1-\sqrt{\varepsilon}}(I - T_{1-\varepsilon})^2\tilde{g}_0'\|_2 \leqslant \frac{1}{4}\|T_{1-\sqrt{\varepsilon}}(I - T_{1-\varepsilon})\tilde{g}_0'\|_2.$$

The eigenvalues of the operator $T_{1-\sqrt{\varepsilon}}(I - T_{1-\varepsilon})$ are $(1-\sqrt{\varepsilon})^j(1-(1-\varepsilon)^j)$ for $j = 0, 1 \ldots, n$, and it is easy to see they are all $O(\sqrt{\varepsilon})$, hence we get that $\mathsf{Stab}_{1-\sqrt{\varepsilon}}(g') \leqslant O(\sqrt{\varepsilon}) \leqslant \xi$. In the last move we have used the fact that we may take ε to be smaller and only prove a stronger statement. ☐

3.3 Step 3: Moving to the uniform distribution

In this section, we show that one can assume without loss of generality that the underlying distribution on $\Sigma \times \Phi \times \Gamma$ is a uniform distribution, as far as bounding the expectation is concerned.

Lemma 3.15. *For all $\alpha, \varepsilon, \delta, \xi > 0$, there exist $\gamma, \xi', \varepsilon' > 0$ and $N \in \mathbb{N}$ such that the following holds. Let μ be a distribution over $\Sigma \times \Phi \times \Gamma$ fully supported on each one of its coordinates, in which each atom has probability at least α, $n \geqslant N$, and $f\colon \Sigma^n \to [-1,1]$, $g\colon \Phi^n \to [-1,1]$, $h\colon \Gamma^n \to [-1,1]$ be such that*

1. $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \geqslant \delta$, *and*

2. $\mathsf{Stab}^{\nu}_{1-\varepsilon'}(g) \leqslant \xi'$, *where ν be the marginal distribution of μ on Φ.*

Then, there is $n' \geqslant \gamma n$, and $f'\colon \Sigma^{n'} \to [-1,1]$, $g'\colon \Phi^{n'} \to [-1,1]$ and $h'\colon \Gamma^{n'} \to [-1,1]$ such that

1. $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mathbf{U}^{\otimes n'}}\left[f'(\mathbf{x})g'(\mathbf{y})h'(\mathbf{z})\right]\right| \geqslant \delta/2$*, where* $\mathbf{U}$ *is a uniform distribution on the support of* μ*, and*

2. $\mathsf{Stab}^{\mathbf{u}}_{1-\varepsilon}(g') \leqslant \xi$*, where* $\mathbf{u}$ *is the marginal distribution of* $\mathbf{U}$ *on* Φ*.*

Proof. Write $\mu = \frac{1}{2}\alpha\mathbf{U} + (1-\frac{1}{2}\alpha)\tau$, where $\mathbf{U}$ is the uniform distribution on the support of μ. The idea is to choose a set of coordinates J that includes each $i \in [n]$ with probability $(1-\frac{1}{2}\alpha)$, fix them according to τ, and take the rest of the coordinates to be according to $\mathbf{U}$.

More formally, select $J \subseteq [n]$ by including $i \in J$ with probability $(1-\frac{1}{2}\alpha)$ for each $i \in [n]$. Write $\mathbf{x} = (\mathbf{x}', \mathbf{x}'')$ where $\mathbf{x}'$ is the J-part of $\mathbf{x}$ and $\mathbf{x}''$ is the $\bar{J}$-part of $\mathbf{x}$, and similarly $\mathbf{y} = (\mathbf{y}', \mathbf{y}'')$ and $\mathbf{z} = (\mathbf{z}, \mathbf{z}'')$. We sample $(\mathbf{x}', \mathbf{y}', \mathbf{z}')$ according to τ^J, and think of $(\mathbf{x}'', \mathbf{y}'', \mathbf{z}'') \sim \mathbf{U}$. We note that this way, the distribution of $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is precisely $\mu^{\otimes n}$.

Define

$$f' = f_{J\to\mathbf{x}'} \qquad g' = g_{J\to\mathbf{y}'} \qquad h' = h_{J\to\mathbf{z}'},$$

and thus we have that

$$\begin{aligned}\delta \leqslant \left|\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| &= \left|\mathop{\mathbf{E}}_{\substack{J\\(\mathbf{x}',\mathbf{y}',\mathbf{z}')\sim\tau^J}}\left[\mathop{\mathbf{E}}_{(\mathbf{x}'',\mathbf{y}'',\mathbf{z}'')\sim\mathbf{U}^{\bar{J}}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right]\right| \\ &\leqslant \mathop{\mathbf{E}}_{\substack{J\\(\mathbf{x}',\mathbf{y}',\mathbf{z}')\sim\tau^J}}\left[\left|\mathop{\mathbf{E}}_{(\mathbf{x}'',\mathbf{y}'',\mathbf{z}'')\sim\mathbf{U}^{\bar{J}}}\left[f'(\mathbf{x}'')g'(\mathbf{y}'')h'(\mathbf{z}'')\right]\right|\right].\end{aligned}$$

As f, g and h are bounded, $\left|\mathbf{E}_{(\mathbf{x}'',\mathbf{y}'',\mathbf{z}'')\sim\mathbf{U}^{\bar{J}}}\left[f'(\mathbf{x}'')g'(\mathbf{y}'')h'(\mathbf{z}'')\right]\right|$ for any choice of J and $\mathbf{x}', \mathbf{y}'$ and $\mathbf{z}'$, hence we get that with probability at least $\delta/2$ we have that

$$\left|\mathop{\mathbf{E}}_{(\mathbf{x}'',\mathbf{y}'',\mathbf{z}'')\sim\mathbf{U}^{\bar{J}}}\left[f'(\mathbf{x}'')g'(\mathbf{y}'')h'(\mathbf{z}'')\right]\right| \geqslant \frac{\delta}{2}.$$

In the rest of the argument, we show that with probability at least $1-\delta/4$ we have that $n' \stackrel{def}{=} \left|\bar{J}\right| \geqslant \gamma n$, and $\mathsf{Stab}^{\mathbf{u}}_{1-\varepsilon}(g') \leqslant \xi$, which by the union bound finish the proof.

The probability that $n < \gamma n$ may be bounded by $e^{-\Omega_\alpha(n)}$ using Chernoff's bound, hence is smaller than $\delta/8$ provided that N is large enough as $n \geqslant N$.

The rest of the proof is devoted for showing that $\mathsf{Stab}^{\mathbf{u}}_{1-\varepsilon}(g') > \xi$ with probability at most $\delta/8$. The proof is similar in spirit to the proof of Corollary 3.13, but is a bit more delicate as we are changing the distribution.

$$\mathop{\mathbf{E}}_{J,\mathbf{y}''}\left[\mathsf{Stab}^{\mathbf{u}}_{1-\varepsilon}(g')\right] = \mathop{\mathbf{E}}_{J,\mathbf{y}''}\left[\mathop{\mathbf{E}}_{\mathbf{y}',\tilde{\mathbf{y}}'\ (1-\varepsilon)\text{ correlated}}\left[g'(\mathbf{y}')g'(\tilde{\mathbf{y}}')\right]\right] = \mathop{\mathbf{E}}_{J,\mathbf{y}''}\left[\mathop{\mathbf{E}}_{\mathbf{y}',\tilde{\mathbf{y}}'\ (1-\varepsilon)\text{ correlated}}\left[g(\mathbf{y}',\mathbf{y}'')g(\tilde{\mathbf{y}}',\mathbf{y}'')\right]\right].$$

Consider the Markov chain on Φ^n, that transitions from $\mathbf{y}', \mathbf{y}''$ to $\tilde{\mathbf{y}}', \mathbf{y}''$ above. Note that it is a tensor $T\otimes n$ of a basic Markov chain T defined as the transition on a single coordinate. Explicitly, sampling $\mathbf{a} \sim \mu_y$ and $\mathbf{b} \sim T_a$ can be done as: with probability $(1-\alpha/2)$ sample $\mathbf{a} \sim \tau_y$ and set $\mathbf{b} = \mathbf{a}$; otherwise with probability $(1-\varepsilon)$ take $\mathbf{a} = \mathbf{b}$ according to $\mathbf{U}_y$, and with probability ε take $\mathbf{a}, \mathbf{b}$ independently from $\mathbf{U}_y$. Thus we have

$$\begin{aligned}\mathop{\mathbf{E}}_{J,\mathbf{y}''}\left[\mathsf{Stab}^{\mathbf{u}}_{1-\varepsilon}(g')\right] = \left\langle g, T^{\otimes n}g\right\rangle = \sum_{S\subseteq[n]}\left\langle g^{=S}, (T^{\otimes n}g)^{=S}\right\rangle &\leqslant \sum_{S\subseteq[n]}\|g^{=S}\|_2\|(T^{\otimes n}g)^{=S}\|_2 \\ &\leqslant \sum_{S\subseteq[n]}\lambda_2(T)^{|S|}\|g^{=S}\|_2^2.\end{aligned}$$

Next, note that $T(a,b) \geqslant \frac{\alpha\varepsilon}{2}\alpha\frac{1}{m^6}$ for any $a, b \in \Phi$. Indeed, with probability $\alpha\varepsilon/2$ we are making independent samples from $\mathtt{supp}(\mu)$ and taking their y-part, and as $|\mathtt{supp}(\mu)| \leqslant m^3$ the result follows. It follows that looking at the transitions of T as a graph on Φ, the edge expansion of each set is at least $\frac{\alpha\varepsilon}{2m^6}$, hence by Cheeger's inequality

$$\lambda_2(T) \leqslant 1 - \frac{\alpha^2\varepsilon^2}{8m^6},$$

so we get that

$$\mathop{\mathbf{E}}_{J,\mathbf{y}''}\left[\mathsf{Stab}^{\mathbf{u}}_{1-\varepsilon}(g')\right] \leqslant \sum_{S\subseteq[n]}\left(1-\frac{\alpha^2\varepsilon^2}{8m^6}\right)^{|S|}\|g^{=S}\|_2^2 \leqslant \sum_{S\subseteq[n]}(1-\varepsilon')^{|S|}\|g^{=S}\|_2^2 = \mathsf{Stab}_{1-\varepsilon'}(g) \leqslant \xi',$$

where we chose $\varepsilon' = 1-\frac{\alpha^2\varepsilon^2}{8m^6}$ and used the hypothesis of the lemma. As $\mathsf{Stab}^{\mathbf{u}}_{1-\varepsilon}(g') \geqslant 0$ always, it follows from Markov's inequality that $\mathsf{Stab}^{\mathbf{u}}_{1-\varepsilon}(g') > \sqrt{\xi'}$ with probability at most $\sqrt{\xi'}$, and taking $\xi' = \min(\xi^2, \delta^2/64)$ finishes the proob. □

3.4 Step 4: Embedding into a group predicate

We now view the set of accepting assignments from $\Sigma\times\Phi\times\Gamma$ as a set of triples from a non Abelian group that satisfy a group equation. This is captured in the following lemma.

Lemma 3.16. *Let μ be any distribution on $\Sigma\times\Phi\times\Gamma$ whose support on each coordinate is full, such that $|\Sigma| = |\Phi| = |\Gamma| = n$. For every $x\in\Sigma$, let T_x denote the adjacency matrix of the edges labeled by x. Suppose μ satisfies the following properties.*

1. *For every $x\in\Sigma$, the collection of edges labeled by x forms a complete matching, i.e., T_x is a permutation matrix for every $x\in\Sigma$.*
2. *For every $(x_1,x_2,x_3)\in\Sigma^3$ the permutation $T_{x_1}T^t_{x_2}T_{x_3}$ belongs to $\{T_x\}_{x\in\Sigma}$.*

Then there exists a group $(G,\bullet)$ and permutation maps $\sigma:\Sigma\to G$, $\phi:\Phi\to G$ and $\gamma:\Gamma\to G$ such that the following holds:

$$\{(\sigma(x),\phi(y),\gamma(z)) \mid (x,y,z)\in\mathsf{supp}(\mu)\} = \{(a,b,c) \mid c = a\bullet b\}.$$

Moreover, if the support of μ is not linearly embedded, the G does not have any non-trivial representation of dimension 1.[10]

Proof. Suppose without loss of generality that $\Sigma=\Phi=\Gamma=[n]$. We choose a distinctive $x^\star\in\Sigma$ arbitrarily, and re-label Φ and Γ so that for any y,z such that $(x^\star,y,z)\in\mathsf{supp}(\mu)$ it holds that $y=z$, i.e. that $T_{x^\star}$ is the identity map.
Defining σ. We define $\sigma(x) = T_x$.
Defining γ. Choose some distinctive $y^\star\in\Phi$ arbitrarily, and for each z pick the unique x such that $(x,y^\star,z)\in\mathsf{supp}(\mu)$ and set $\gamma(z) = T_x$.
Defining ϕ. We set $\phi(y) = \gamma(y)$.

Group structure. Firstly, the set of matrices $\{T_x\}_{x\in\Sigma}$ is closed under multiplication and this follows from the second property, by taking $x_2 = x^\star$ as $T_{x^\star} = I$. Secondly, for every $x'\in\Sigma$, there exists an $\tilde{x}'\in\Sigma$ such that $T_{\tilde{x}'} = T^t_{x'}$. This is because for $(x^\star,x',x^\star)$ the corresponding $T_{x^\star}T^t_{x'}T_{x^\star}\in\{T_x\}_{x\in\Sigma}$ by the second property. But $T_{x^\star}T^t_{x'}T_{x^\star} = T^t_{x'}$ as $T_{x^\star} = I$. Thus for every x, there exists x' such that $T_xT_{x'} = I$. The associativity of $\{T_x\}_{x\in\Sigma}$ is obvious under matrix multiplication. Therefore, the collection of permutation matrices $\{T_x\}_{x\in\Sigma}$ form a group.

Triples in the support of μ satisfy the group equation. Let $(x,y,z)\in\mathsf{supp}(\mu)$. Then we have that the matching T_x has an edge between y and z. Let $x'\in\Sigma$ be the unique element such that $T_{x'}$ has an edge between $y^\star$ and z, so that $\gamma(z) = T_{x'}$, and $x''\in\Sigma$ be the unique element such that $T_{x''}$ has an edge between $y^\star$ and $y\in\Gamma$, so that $\gamma(y) = T_{x''}$. We claim that $T_{x''}T^t_{x'}T_x = I$. To see this, T_x maps y to z, $T^t_{x'}$ maps z

[10] See Section 3.5 for the definition of a representation of a group.

to $y^\star$ and $T_{x''}$ maps $y^\star$ back to y. But there is only one map that maps y to y and that map must be the identity map $T_{x^\star}$. Thus, for every $(x, y, z) \in \mathsf{supp}(\mu)$, we have

$$T_{x''} T^t_{x'} T_x = \phi(y)\gamma(z)^{-1}\sigma(x) = I,$$

which implies

$$(x, y, z) \in \mathsf{supp}(\mu) \implies \gamma(z) = \sigma(x)\phi(y),$$

as required.

We now prove the moreover part. Let $G = \{T_x\}_{x\in\Sigma}$ be the group and towards the contradiction, suppose G has a non-trivial representation $\rho : G \to \mathbb{C}$ of dimension 1. Note that for every $g \in G$, $|\rho(g)| = 1$. This follows as $1 = \rho(g) \cdot \rho(g^{-1}) = \rho(g)\overline{\rho(g)}$. Let $p > 1$ be the smallest number such that for every $g \in G$, $\rho(g) = \omega_p^{i(g)}$, where ω_p is the primitive p^{th} root of unity and $i(g) \in \{0, 1, \ldots, p-1\}$. Consider the following maps $\sigma' : \Sigma \to \mathbb{Z}_p$, $\phi' : \Phi \to \mathbb{Z}_p$ and $\gamma' : \Gamma \to \mathbb{Z}_p$ where

$$\sigma'(x) = i(\sigma(x)), \phi'(y) = i(\phi(y)), \text{ and } \gamma'(z) = -i(\gamma(x)).$$

As ρ is non-trivial, each map is non-constant. Now,

$$(x, y, z) \in \mathsf{supp}(\mu) \implies \gamma(z) = \sigma(x)\phi(y) \implies \sigma'(x) + \phi'(y) + \gamma'(z) = 0 \pmod{p}.$$

This contradicts the fact the the support of μ has no linear embedding. □

3.5 Step 5: Bounding the expectation using non-Abelian Fourier analysis

In this section, we finally bound the expectation using Fourier analysis over non-Abelian groups.

We begin by recalling some basic representation theory and non-Abelian Fourier analysis. See the monograph by Diaconis [Dia98, Chapter 2] for a more detailed treatment (with proofs).

We will be working with a finite group G and complex-valued functions $f : G \to \mathbb{C}$ on G. The *convolution* between two function $f, h : G \to \mathbb{C}$, denoted by $f * h$, is defined as follows:

$$(f * h)(x) := \mathop{\mathbf{E}}_{y}[f(xy^{-1})h(y)].$$

For any $p \geqslant 1$, the p-norm of any function $f : G \to \mathbb{C}$ is defined as

$$\|f\|_p^p := \mathop{\mathbf{E}}_{x}[|f(x)|^p].$$

Given a complex vector space V, we denote the vector space of linear operators on V by $\mathrm{End}(V)$. This space is endowed with the following inner product and norm (usually referred to as the *Hilbert-Schmidt* norm):

$$\text{For } A, B \in \mathrm{End}(V), \quad \langle A, B\rangle_{\mathrm{End}(V)} := \mathsf{tr}(A^*B) \quad \text{ and } \quad \|A\|_{\mathsf{HS}}^2 := \langle A, A\rangle_{\mathrm{End}(V)} = \mathsf{tr}(A^*A).$$

This norm is known to be submultiplicative (i.e., $\|AB\|_{\mathsf{HS}} \leqslant \|A\|_{\mathsf{HS}} \cdot \|B\|_{\mathsf{HS}}$).

Representations and Characters: A *representation* $\rho : G \to \mathrm{End}(V)$ is a homomorphism from G to the set of linear operators on ρ_V for some finite-dimensional vector space ρ_V over $\mathbb{C}$, i.e., for all $x, y \in G$, we have $\rho(xy) = \rho(x)\rho(y)$. The dimension of the representation ρ, denoted by d_ρ, is the dimension of the underlying $\mathbb{C}$-vector space ρ_V. The *character* of a representation ρ, denoted by $\chi_\rho : G \to \mathbb{C}$, is defined as $\chi_\rho(x) := \mathsf{tr}(\rho(x))$, where $\mathsf{tr}(.)$ is the trace of a matrix.

The representation $1 : G \to \mathbb{C}$ satisfying $1(x) = 1$ for all $x \in G$ is the *trivial* representation. A representation $\rho : G \to \mathrm{End}(V)$ is said to *reducible* if there exists a non-trivial subpsace $W \subset V$ such that for all $x \in G$, we have $\rho(x)W \subset W$. A representation is said to be *irreducible* otherwise. The set of all irreducible representations of G (up to equivalences) is denoted by $\mathsf{irrep}(G)$. For G^n, we define

$$\mathsf{irrep}(G^n) = \left\{ \bigotimes_{i=1}^{n} \rho_i \;\middle|\; \rho_i \in \mathsf{irrep}(G) \;\forall i = 1, \ldots, n \right\}.$$

For every representation $\rho\colon G \to \mathrm{End}(V)$, there exists an inner product $\langle , \rangle_V$ over V such that every $\rho(x)$ is unitary (i.e, $\langle \rho(x)u, \rho(x)v\rangle_V = \langle u, v\rangle_V$ for all $u, v \in V$ and $x \in G$). Hence, we may assume, without loss of generality, that all the representations we are considering are unitary.

The following are some well-known facts about representations and characters.

Proposition 3.17. *1. The group G is Abelian iff $d_\rho = 1$ for every irreducible representation ρ in $\mathsf{irrep}(G)$.*

2. *For any finite group G, $\sum_{\rho\in\mathsf{irrep}(G)} d_\rho^2 = |G|$.*
3. *[Orthogonality of characters] For any $\rho, \rho' \in \mathsf{irrep}(G)$ we have: $\mathbf{E}_x\left[\chi_\rho(x)\overline{\chi_{\rho'}(x)}\right] = \mathbf{1}[\rho = \rho']$.*

Non-Abelian Fourier analysis: Given a function $f\colon G \to \mathbb{C}$ and an irreducible representation $\rho \in \mathsf{irrep}(G)$, the Fourier transform is defined as follows:

$$\widehat{f}(\rho) := \mathop{\mathbf{E}}_x[f(x)\rho(x)].$$

The following proposition summarizes the basic properties of Fourier transform that we will need.

Proposition 3.18. *For any $f, h\colon G \to \mathbb{C}$, we have the following*

1. *[Fourier transform of trivial representation] $\widehat{f}(1) = \mathbf{E}_x[f(x)]$.*
2. *[Convolution] $\widehat{f * h}(\rho) = \widehat{f}(\rho)\cdot\widehat{h}(\rho)$.*
3. *[Fourier inversion formula] $f(x) = \sum_{\rho\in\mathsf{irrep}(G)} d_\rho \cdot \langle \widehat{f}(\rho), \rho(x)\rangle_{\mathrm{End}(\rho_V)}$.*
4. *[Parseval's identity] $\|f\|_2^2 = \sum_{\rho\in\mathsf{irrep}(G)} d_\rho \cdot \|\widehat{f}(\rho)\|_{\mathsf{HS}}^2$.*

We now prove the following lemma.

Lemma 3.19. *Let $(G, \bullet)$ be any non-Abelian group with no non-trivial representation of dimension 1. Let μ be the uniform distribution on the triples $\{(a, b, c) \mid a, b \in G, c^{-1} = a \bullet b\}$. For any $\varepsilon, \delta > 0$, if $f\colon G^n \to [-1, 1]$, $g\colon G^n \to [-1, 1]$ and $h\colon G^n \to [-1, 1]$ are functions such that $\mathsf{Stab}^\mu_{1-\varepsilon}(g) \leqslant \delta$ then*

$$\left|\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})]\right| \leqslant \gamma,$$

where $\gamma \to 0$ as $\delta \to 0$.

Proof. For a given $\mathbf{y} \in G^n$, let $\tilde{\mathbf{y}}$ be the $(1-\varepsilon)$-correlated copy of $\mathbf{y}$.

$$\begin{aligned}
\mathsf{Stab}^\mu_{1-\varepsilon}(g) &= \mathop{\mathbf{E}}_{\mathbf{y},\tilde{\mathbf{y}}}\left[g(\mathbf{y})\overline{g(\tilde{\mathbf{y}})}\right] \\
&= \mathop{\mathbf{E}}_{\mathbf{y},\tilde{\mathbf{y}}}\left[\sum_{\rho,\tau\in\mathsf{Irrep}(G^n)} d_\rho d_\tau \langle\widehat{g}(\rho), \rho(\mathbf{y})\rangle \cdot \overline{\langle\widehat{g}(\tau), \tau(\tilde{\mathbf{y}})\rangle}\right] \\
&= \mathop{\mathbf{E}}_{\mathbf{y},\tilde{\mathbf{y}}}\left[\sum_{\rho,\tau\in\mathsf{Irrep}(G^n)} d_\rho d_\tau \sum_{i,j\in[d_\rho]} \widehat{g}(\rho)_{ij}\overline{\rho(\mathbf{y})_{ij}} \sum_{k,\ell\in[d_\tau]} \overline{\widehat{g}(\tau)_{k\ell}}\tau(\tilde{\mathbf{y}})_{k\ell}\right].
\end{aligned}$$

For a fixed $\rho = \otimes\rho_t, \tau = \otimes\tau_t, i = (i_1, i_2, \ldots, i_n), j = (j_1, j_2, \ldots, j_n), k = (k_1, k_2, \ldots, k_n)$ and $\ell = (\ell_1, \ell_2, \ldots, \ell_n)$, we have

$$\mathop{\mathbf{E}}_{\mathbf{y},\tilde{\mathbf{y}}}\left[\overline{\rho(\mathbf{y})_{i,j}}\tau(\tilde{\mathbf{y}})_{k\ell}\right] = \prod_{t=1}^n \mathop{\mathbf{E}}_{y_t,\tilde{y}_t}\left[\overline{\rho_t(y_t)_{i_t,j_t}}\tau(\tilde{y}_t)_{k_t,\ell_t}\right].$$

For any fixed $t \in [n]$,

$$\mathop{\mathbf{E}}_{y_t,\tilde{y}_t}\left[\overline{\rho_t(y_t)_{i_t,j_t}}\tau(\tilde{y}_t)_{k_t,\ell_t}\right] = (1-\varepsilon)\mathop{\mathbf{E}}_{y_t}\left[\overline{\rho_t(y_t)_{i_t,j_t}}\tau(y_t)_{k_t,\ell_t}\right] + \varepsilon \mathop{\mathbf{E}}_{y_t,z_t\sim G}\left[\overline{\rho_t(y_t)_{i_t,j_t}}\tau(z_t)_{k_t,\ell_t}\right]$$
$$= (1-\varepsilon)1_{\rho_t=\tau_t \& i_t=k_t \& j_t=\ell_t} + \varepsilon 1_{\rho_t\equiv\tau_t\equiv\mathsf{triv}}.$$

Therefore,

$$\begin{aligned}
\mathsf{Stab}^{\mu}_{1-\varepsilon}(g) &= \mathop{\mathbf{E}}_{\mathbf{y},\tilde{\mathbf{y}}}\left[\sum_{\rho,\tau\in\mathsf{Irrep}(G^n)} d_\rho d_\tau \sum_{i,j\in[d_\rho]}\widehat{g}(\rho)_{ij}\overline{\rho(\mathbf{y})_{ij}}\sum_{k,\ell\in[d_\tau]}\overline{\widehat{g}(\tau)_{k\ell}}\tau(\tilde{\mathbf{y}})_{k\ell}\right]\\
&= \sum_{\rho\in\mathsf{Irrep}(G^n)} d_\rho^2 \sum_{i,j\in[d_\rho]}\widehat{g}(\rho)_{ij}\overline{\widehat{g}(\rho)_{ij}}(1-\varepsilon)^{|\rho|}\\
&= \sum_{\rho\in\mathsf{Irrep}(G^n)} d_\rho^2\cdot\|\widehat{g}(\rho)\|_{\mathsf{HS}}^2(1-\varepsilon)^{|\rho|},
\end{aligned}$$

where $|\rho|$ is the number of non-trivial representation on the representation $\rho = \otimes\rho_i$. Thus, if $\mathsf{Stab}^{\mu}_{1-\varepsilon}(g) \leqslant \delta$, then for every ρ with $|\rho| \leqslant D$, we have $\|\widehat{g}(\rho)\|_{\mathsf{HS}}^2(1-\varepsilon)^D \leqslant \delta$. For $D = \frac{1}{2}\frac{\log\delta}{\log(1-\varepsilon)}$, this implies that,

$$\max_{\rho,|\rho|\leqslant D}\{\|\widehat{g}(\rho)\|_{\mathsf{HS}}\} \leqslant \sqrt{\frac{\delta}{(1-\varepsilon)^D}} \leqslant \delta^{1/4}.$$

We now bound the expectation of the product of f, g, h. We can upper bound the expectation as follows:

$$\begin{aligned}
\left|\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})]\right| &= |(f*g*h)(1^n)|\\
&= \left|\sum_{\rho\in\mathsf{Irrep}(G^n)} d_\rho\cdot\mathsf{tr}(\widehat{f*g*h}(\rho))\right|\\
&\leqslant \sum_{\rho\in\mathsf{Irrep}(G^n)} d_\rho\cdot|\mathsf{tr}(\widehat{f*g*h}(\rho))|\\
&= \sum_{\rho\in\mathsf{Irrep}(G^n)} d_\rho\cdot|\mathsf{tr}(\widehat{f}(\rho)\widehat{g}(\rho)\widehat{h}(\rho))|\\
&= \sum_{\rho\in\mathsf{Irrep}(G^n)} d_\rho\cdot|\langle\widehat{f}(\rho)\widehat{g}(\rho),\widehat{h}(\rho)^*\rangle_{\mathrm{End}(\rho_V)}|\\
&= \sum_{\rho\in\mathsf{Irrep}(G^n)} d_\rho\cdot\|\widehat{f}(\rho)\widehat{g}(\rho)\|_{\mathsf{HS}}\|\widehat{h}(\rho)\|_{\mathsf{HS}}\\
&\leqslant \sum_{\rho\in\mathsf{Irrep}(G^n)} d_\rho\|\widehat{f}(\rho)\|_{\mathsf{HS}}\|\widehat{g}(\rho)\|_{\mathsf{HS}}\|\widehat{h}(\rho)\|_{\mathsf{HS}}.
\end{aligned}$$

We now split the summation based on the dimension of the representation ρ. Let $\gamma = 2\cdot\max\{\delta^{1/4}, \frac{1}{\sqrt{2^D}}\}$.

$$\Theta_L = \sum_{\rho,|\rho|\leqslant D} d_\rho\|\widehat{f}(\rho)\|_{\mathsf{HS}}\|\widehat{g}(\rho)\|_{\mathsf{HS}}\|\widehat{h}(\rho)\|_{\mathsf{HS}}, \qquad \Theta_H = \sum_{\rho,|\rho|>D} d_\rho\|\widehat{f}(\rho)\|_{\mathsf{HS}}\|\widehat{g}(\rho)\|_{\mathsf{HS}}\|\widehat{h}(\rho)\|_{\mathsf{HS}}.$$

We now bound Θ_L and Θ_H separately as follows.

$$\begin{aligned}
\Theta_L &\leqslant \max_{\rho,|\rho|\leqslant D}\{\|\widehat{g}(\rho)\|_{\mathsf{HS}}\}\sum_{\rho,|\rho|\leqslant D} d_\rho\|\widehat{f}(\rho)\|_{\mathsf{HS}}\|\widehat{h}(\rho)\|_{\mathsf{HS}}\\
&\leqslant \frac{\gamma}{2}\cdot\sum_{\rho,|\rho|\leqslant D} d_\rho\|\widehat{f}(\rho)\|_{\mathsf{HS}}\|\widehat{h}(\rho)\|_{\mathsf{HS}}
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \frac{\gamma}{2}\cdot\left(\sum_{\rho,|\rho|\leqslant D} d_\rho\|\widehat{f}(\rho)\|_{\mathsf{HS}}^2\right)^{1/2}\left(\sum_{\rho,|\rho|\leqslant D} d_\rho\|\widehat{h}(\rho)\|_{\mathsf{HS}}^2\right)^{1/2}\\
&\leqslant \frac{\gamma}{2}\cdot\|f\|_2\cdot\|h\|_2\leqslant\frac{\gamma}{2},
\end{aligned}
$$

where, the third inequality is by Cauchy-Schwarz. If $|\rho| = t$, then $d_\rho \geqslant 2^t$ as every non-trivial representation of G has dimension at least 2. Using this fact,

$$
\begin{aligned}
\Theta_H &= \sum_{\rho,|\rho|>D} d_\rho\|\widehat{f}(\rho)\|_{\mathsf{HS}}\|\widehat{g}(\rho)\|_{\mathsf{HS}}\|\widehat{h}(\rho)\|_{\mathsf{HS}}\\
&\leqslant \frac{1}{\sqrt{2^D}}\sum_{\rho,|\rho|>D} d_\rho^{3/2}\|\widehat{f}(\rho)\|_{\mathsf{HS}}\|\widehat{g}(\rho)\|_{\mathsf{HS}}\|\widehat{h}(\rho)\|_{\mathsf{HS}}\\
&\leqslant \frac{1}{\sqrt{2^D}}\left(\sum_{\rho,|\rho|>D} d_\rho^2\|\widehat{f}(\rho)\|_{\mathsf{HS}}^2\right)^{1/2}\cdot\left(\sum_{\rho,|\rho|>D} d_\rho^2\|\widehat{g}(\rho)\|_{\mathsf{HS}}^2\|\widehat{h}(\rho)\|_{\mathsf{HS}}^2\right)^{1/2} && \text{(Cauchy-Schwarz)}\\
&\leqslant \frac{1}{\sqrt{2^D}}\|f\|_2\cdot\left(\left(\sum_{\rho,|\rho|>D} d_\rho^2\|\widehat{g}(\rho)\|_{\mathsf{HS}}^2\right)\cdot\left(\sum_{\rho,|\rho|>D} d_\rho^2\|\widehat{h}(\rho)\|_{\mathsf{HS}}^2\right)\right)^{1/2}\\
&\leqslant \frac{1}{\sqrt{2^D}}\|f\|_2\cdot\|g\|_2\cdot\|h\|_2\\
&\leqslant \frac{1}{\sqrt{2^D}}=\frac{\gamma}{2}.
\end{aligned}
$$

Thus, we have

$$
\left|\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})]\right|\leqslant\gamma,
$$

as required, where $\gamma\to 0$, as $\delta\to 0$.

□

The following corollary follows from the above lemma by considering the function h' instead of h where $h'(\boldsymbol{z}) = h(\boldsymbol{z}^{-1})$.

Corollary 3.20. *Let $(G,\bullet)$ be any non-Abelian group with no non-trivial representation of dimension 1. Let μ be the uniform distribution of the triples $\{(a,b,c)\mid a,b\in G, c=a\bullet b\}$. For any $\varepsilon,\delta>0$, if $f\colon G^n\to[-1,1]$, $g\colon G^n\to[-1,1]$ and $h\colon G^n\to[-1,1]$ are functions such that $\mathsf{Stab}_{1-\varepsilon}^{\mu}(g)\leqslant\delta$ then*

$$
\left|\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})]\right|\leqslant\gamma,
$$

where $\gamma\to 0$ as $\delta\to 0$.

3.6 Proof of Lemma 3.2

In this section, we explain how the tools we have established so far can be used to prove Lemma 3.2. We start with ξ and δ that are small to be determined. We assume that $\mathsf{Stab}_{1-\xi}(g)\leqslant\delta$, otherwise the argument is analogous. We invoke Lemma 3.8 in order to get that the graph between y and z is full, followed by applications of Lemmas 3.10 and 3.14 to identify symbols. If the alphabet of y or z has shrunk we repeat this process, and note that our parameters ξ and δ degrade each time, but we can ensure they are small enough as long as we make at most $O_m(1)$ iterations by taking ξ and δ small enough to begin with.

We stop this process whenever doing it again does not shrink the alphabet of y or z further. This means that composing the matchings T_x, any two matchings we get would either be edge disjoint, or identical. In particular, the matching $T_{x^1}T_{x^2}^tT_{x^3}$ would already exist as a matching $T_{x'}$ for some x'.

At this point, using Lemma 3.16 we are able to view our expectation as an expectation over the predicate $P' = \left\{ (a,b,c) \in G^3 \,\middle|\, c = a \bullet b \right\}$ for some non-Abelian group $(G, \bullet)$ with no non-trivial representations of dimension 1. Finally, using Lemma 3.15 we are able to move to the case where the distribution over support of the predicate P' is uniform (as opposed to just full). We remark again that the invocations of these lemmas lead to degradation of our parameters ξ and δ, but we can make sure they are still small in the end by taking them sufficiently small in the beginning.

Eventually, the expectation we get is bounded by Lemma 3.19, completing the proof. □

3.7 Going from union of matchings to semi-rich supports

We now explain how to modify the proof of Lemma 3.2 to prove:

Lemma 3.21. *For all $m \in \mathbb{N}$, $\varepsilon, \alpha > 0$ there exist $\xi > 0$ and $\delta > 0$ such that the following holds. Suppose μ is a distribution over $\Sigma \times \Phi \times \Gamma$ whose support (a) is semi rich, and (b) cannot be embedded in an Abelian group. Further suppose that $|\Sigma|, |\Phi|, |\Gamma| \leqslant m$ and each atom in μ has probability at least α. Then, if $f\colon \Sigma^n \to [-1,1]$ $g\colon \Phi^n \to [-1,1]$, $h\colon \Gamma^n \to [-1,1]$ are functions such that*

- $\mathsf{Stab}_{1-\xi}(g) \leqslant \delta$ *(or $\mathsf{Stab}_{1-\xi}(h) \leqslant \delta$).*

Then $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \varepsilon$.

Proof. Let $S = \mathsf{supp}(\mu)$, and assume that μ is semi-rich. If S is not a union of matchings, then it means that there is an x such that

$$S_x = \{\,(y,z) \mid (x,y,z) \in S\}$$

is not a matching. However, by the semi-rich property, we know that the support of S_x on y, as well as on z is full, so this means that at least one of the following must occur:

1. There exist $y \in \Phi$ and distinct $z, z' \in \Gamma$ such that (y,z) and (y,z') are both in S_x;
2. there exist distinct $y, y' \in \Phi$ and $z \in \Gamma$ such that (y,z) and (y',z) are both in S_x.

Let us assume without loss of generality the first case occurs. We may then apply Lemmas 3.10 to merge the symbols z and z'; we note that merging preserves the semi rich property. Thus, we apply mergers as in Lemma 3.10 or Lemma 3.14 so long as we can (this process changes the functions f, g and h, but for convenience of notation we ignore this), eventually moving to a distribution μ' wherein no further merges are possible.

Furthermore, we note that the various conditions of Lemma 3.2 are preserved (no Abelian embedding, probability of atoms, stability of g or h etc.), so we have reduced the problem to the case where $\mathsf{supp}(\mu)$ is semi-rich and any two variables determine the last one. In this case, it is easily seen that $\mathsf{supp}(\mu)$ is a union of matchings and hence we may apply Lemma 3.2 on f, g, h we get that

$$\left|\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \varepsilon. \qquad \square$$

3.8 Deducing Lemma 2.6

Finally, we explain how to deduce Lemma 2.6, restated below.

Lemma 3.22 (Restatement of Lemma 2.6)**.** *For all $m \in \mathbb{N}$, $\varepsilon, \alpha > 0$ there exist $\xi > 0$, and $\delta > 0$ such that the following holds. Suppose μ is a distribution over $\Sigma \times \Phi \times \Gamma$ whose support (a) semi-rich, and (b) cannot be embedded in an Abelian group. Further suppose that $|\Sigma|, |\Phi|, |\Gamma| \leqslant m$ and each atom in μ has probability at least α. Then, if $f\colon \Sigma^n \to [-1,1]$ $g\colon \Phi^n \to [-1,1]$, $h\colon \Gamma^n \to [-1,1]$ are functions such that*

- $\mathsf{Stab}_{1-\xi}(f) \leqslant \delta$, $\mathsf{Stab}_{1-\xi}(g) \leqslant \delta$ *or* $\mathsf{Stab}_{1-\xi}(h) \leqslant \delta$.

Then $\left|\mathbf{E}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \varepsilon$.

Proof. The case that the stability of either g or h is small was already covered in Lemma 3.21, so assume that $\mathsf{Stab}_{1-\xi}(f) \leqslant \delta$. Consider the Efron-Stein decomposition according to each one of μ_x, μ_y and μ_z and define the corresponding noise operator on each one of these spaces, which we will denote (abusing notation) by $T_{1-\xi}$ and $T_{1-\xi'}$, where ξ' is a parameter to be determined (which should be thought of as much larger than ξ, say $\sqrt{\xi}$). Thus

$$\begin{aligned}\mathop{\mathbf{E}}_{x,y,z\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right] &= \mathop{\mathbf{E}}_{x,y,z\sim\mu^{\otimes n}}\left[f(\mathbf{x})(T_{1-\xi'}g)(\mathbf{y})(T_{1-\xi'}h)(\mathbf{z})\right]\\ &+ \mathop{\mathbf{E}}_{x,y,z\sim\mu^{\otimes n}}\left[f(\mathbf{x})((I-T_{1-\xi'})g)(\mathbf{y})(T_{1-\xi'}h)(\mathbf{z})\right]\\ &+ \mathop{\mathbf{E}}_{x,y,z\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})((I-T_{1-\xi'})h)(\mathbf{z})\right].\end{aligned}$$

Consider the second expectation; we wish to invoke Lemma 3.21. Let $g' = \frac{1}{2}((I-T_{1-\xi'})g)$ and $h' = \frac{1}{2}T_{1-\xi'}h$. We take $\tilde{\xi}$, p and δ from Lemma 3.21 for $\varepsilon/16$. We shall assume that $\xi', \tilde{\xi} \leqslant \delta^4$, since lowering $\tilde{\xi}$ only increases the stability of g'. We will also assume that $\xi' \leqslant \tilde{\xi}^4$, again by the same reasoning.

Note that

$$\mathsf{Stab}_{1-\tilde{\xi}}(g') = \sum_{S\subseteq[n]}(1-\tilde{\xi})^{|S|}\|g'^{=S}\|_2^2 = \sum_{S\subseteq[n]}(1-\tilde{\xi})^{|S|}(1-(1-\xi')^{|S|})\|g'^{=S}\|_2^2 \leqslant \|g'\|_2^2\max_j(1-\tilde{\xi})^j(1-(1-\xi')^j).$$

We have $\|g'\|_2^2 \leqslant 1$; for $j > \frac{1}{\sqrt{\xi'}}$ we have

$$\max_j(1-\tilde{\xi})^j(1-(1-\xi')^j) \leqslant e^{-j\tilde{\xi}} \leqslant e^{-\frac{\tilde{\xi}}{\sqrt{\xi'}}} \leqslant e^{-\xi'^{-1/4}} \leqslant \delta.$$

For $j < \frac{1}{\sqrt{\xi'}}$ we have

$$\max_j(1-\tilde{\xi})^j(1-(1-\xi')^j) \leqslant j\xi' \leqslant \delta.$$

Thus, $\mathsf{Stab}_{1-\tilde{\xi}}(g') \leqslant \delta$, and from Lemma 3.21

$$\left|\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})((I-T_{1-\xi'})g)(\mathbf{y})(T_{1-\xi'}h)(\mathbf{z})\right]\right| = 4\left|\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})g'(\mathbf{y})h'(\mathbf{z})\right]\right| \leqslant \frac{\varepsilon}{4}.$$

The same argument works for the third expectation.

Finally, we argue that the first expectation is small, and for that we exploit the relationship between ξ and ξ' (which for now is arbitrary). Write

$$\mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}}\left[f(\mathbf{x})(T_{1-\xi'}g)(\mathbf{y})(T_{1-\xi'}h)(\mathbf{z})\right] = \mathop{\mathbf{E}}_{\substack{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}\\ \mathbf{y}'\sim_{1-\xi'}\mathbf{y},\mathbf{z}'\sim_{1-\xi'}\mathbf{z}}}\left[f(\mathbf{x})g(\mathbf{y}')h(\mathbf{z}')\right] = \mathop{\mathbf{E}}_{(\mathbf{x},\mathbf{y}',\mathbf{z}')\sim\mu'^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y}')h(\mathbf{z}')\right].$$

Consider the operator $T^{\otimes n}\colon L_2(\Sigma^n;{\mu'}_x^{\otimes n}) \to L_2((\Phi\times\Gamma)^n;{\mu'}_{y,z}^{\otimes n})$ defined as

$$T^{\otimes n}f(y,z) = \mathbf{E}_{(\mathbf{x}',\mathbf{y}',\mathbf{z}')\sim\mu'^{\otimes n}}\left[f(\mathbf{x}) \mid \mathbf{y}'=y,\mathbf{z}'=z\right].$$

Then the above expectation may be written as

$$\left|\mathop{\mathbf{E}}_{(\mathbf{y},\mathbf{z})\sim{\mu'}_x^{\otimes n}}\left[g(\mathbf{y})h(\mathbf{z})T^{\otimes n}f(\mathbf{y},\mathbf{z})\right]\right| \leqslant \sqrt{\mathop{\mathbf{E}}_{(\mathbf{y},\mathbf{z})\sim{\mu'}_x^{\otimes n}}[g(\mathbf{y})^2h(\mathbf{z})^2]}\sqrt{\mathop{\mathbf{E}}_{(\mathbf{y},\mathbf{z})\sim{\mu'}_x^{\otimes n}}[T^{\otimes n}f(\mathbf{y},\mathbf{z})^2]}.$$

The first expectation is bounded by Cauchy-Schwarz by

$$\|g\|_{4;{\mu'}_y}\|h\|_{4;{\mu'}_z} = \|g\|_{4;\mu_y}\|h\|_{4;\mu_z} \leqslant 1.$$

The second expectation is

$$\sqrt{\langle T^{\otimes n} f, T^{\otimes n} f\rangle} = \sqrt{\langle f, T^{\otimes n *} T^{\otimes n} f\rangle} \leqslant \sqrt{\|f\|_{2;{\mu'}_x} \|S^{\otimes n} f\|_2},$$

where $S = T^* T \colon L_2(\Sigma; {\mu'}_x) \to L_2(\Sigma; {\mu'}_x)$. Thus, overall we get that

$$\left| \mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}} [f(\mathbf{x})(T_{1-\xi'} g)(\mathbf{y})(T_{1-\xi'} h)(\mathbf{z})] \right| \leqslant \sqrt{\|S^{\otimes n} f\|_2}.$$

Consider the operator S; we think about it also as a natural Markov process that samples (a, b) with probability $S(a, b)$ by sampling $(a, y, z) \sim \mu'$, and then sampling $(b, y', z') \sim \mu'$ conditioned on $y' = y$, $z' = z$. Note that ${\mu'}_x$ is the stationary distribution of S, and also note that for any a, b we have that

$$S(a, b) \geqslant \alpha^2 {\xi'}^4.$$

Indeed, this follows as for each y, z, the support of μ'_x conditioned on y, z is full, and each atom there has probability at least $\alpha {\xi'}^2$. Thus, thinking of S as the adjacency operator of a graph over Σ, we get that the edge expansion of each set is at least $\alpha^2 {\xi'}^4$, so by Cheeger's inequality

$$\lambda_2(S) \leqslant 1 - \frac{1}{2}\alpha^4 {\xi'}^8.$$

We thus have that

$$\|S^{\otimes n} f\|_2^2 = \sum_{T\subseteq[n]} \|S^{\otimes n} f^{=T}\|_2^2 \leqslant \sum_{T\subseteq[n]} \left(1 - \frac{1}{2}\alpha^4 {\xi'}^8\right)^{2|T|} \|f^{=T}\|_2^2 \leqslant \sum_{T\subseteq[n]} (1-\xi)^{2|T|} \|f^{=T}\|_2^2 = \mathsf{Stab}_{1-\xi}(f),$$

where the third transition determines the value of ξ', i.e. $\frac{1}{2}\alpha^4 {\xi'}^8 = \xi$. Thus by the premise we get that $\|S^{\otimes n} f\|_2^2 \leqslant \delta$.

Overall we got that

$$\left| \mathop{\mathbf{E}}_{\mathbf{x},\mathbf{y},\mathbf{z}\sim\mu^{\otimes n}} [f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})] \right| \leqslant \delta^{1/4} + 2\frac{\varepsilon}{4} \leqslant \varepsilon,$$

where we used the fact that we may take δ sufficiently small compared to ε. □

References

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. (Preliminary version in *33rd FOCS*, 1992).

[AM09] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998. (Preliminary version in *33rd FOCS*, 1992).

[BG21] Joshua Brakensiek and Venkatesan Guruswami. The quest for strong inapproximability results with perfect completeness. *ACM Trans. Algorithms*, 17(3), 2021.

[BK21] Amey Bhangale and Subhash Khot. Optimal Inapproximability of Satisfiable k-LIN over Non-Abelian Groups. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, page 1615–1628, 2021.

[BKLM22] Mark Braverman, Subhash Khot, Noam Lifshitz, and Dor Minzer. An invariance principle for the multi-slice, with applications. In *IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 228–236, 2022.

[BKM21] Mark Braverman, Subhash Khot, and Dor Minzer. On Rich 2-to-1 Games. In 12*th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 185, pages 27:1–27:20, 2021.

[Bul17] Andrei A. Bulatov. A Dichotomy Theorem for Nonuniform CSPs. In *IEEE* 58*th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 319–330, 2017.

[Dia98] Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *IMS Lecture Notes Monogr. Ser.* Institute of Mathematical Statistics, 1998.

[EH05] Lars Engebretsen and Jonas Holmerin. Three-query PCPs with perfect completeness over non-Boolean domains. *Random Structures & Algorithms*, 27(1):46–75, 2005.

[FGL+96] Uriel Feige, Shafi Goldwasser, Laszlo Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.

[FV98] 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.

[GW95] Michel X. Goemans and David P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the ACM*, 42(6):1115–1145, 1995.

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, July 2001. (Preliminary version in *29th STOC*, 1997).

[HK05] Johan Håstad and Subhash Khot. Query Efficient PCPs with Perfect Completeness. *Theory of Computing*, 1(7):119–148, 2005.

[Hua14] Sangxia Huang. Approximation Resistance on Satisfiable Instances for Sparse Predicates. *Theory of Computing*, 10(14):359–388, 2014.

[Hå14] Johan Håstad. On the NP-Hardness of Max-Not-2. *SIAM Journal on Computing*, 43(1):179–193, 2014.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the* 34*th Annual ACM symposium on Theory of computing (STOC)*, pages 767–775. ACM, 2002.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal Inapproximability Results for MAX-CUT and Other 2-Variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Proceedings* 46*th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 21–30, 2005.

[Mos10] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010.

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the* 14*th Annual ACM symposium on Theory of computing (STOC)*, pages 245–254, 2008.

[Rag09] Prasad Raghavendra. *Approximating NP-hard problems efficient algorithms and their limits.* University of Washington, 2009.

[Sch78] Thomas J. Schaefer. The complexity of satisfiability problems. In *Proceedings of the* 10*th Annual ACM Symposium on Theory of Computing (STOC)*, page 216–226, 1978.

[Tan09] Linqing Tang. Conditional hardness of approximating satisfiable Max 3CSP-q. In *International Symposium on Algorithms and Computation*, pages 923–932. Springer, 2009.

[Zhu20] Dmitriy Zhuk. A Proof of the CSP Dichotomy Conjecture. *Journal of the ACM*, 67(5), August 2020.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il