On Approximability of Satisfiable k-CSPs: II

Amey Bhangale
University of California, Riverside
Riverside, USA
ameyb@ucr.edu

Subhash Khot*
New York University
New York, USA
khot@cs.nyu.edu

Dor Minzer†
Massachusetts Institute of Technology
Cambridge, USA
khot@cs.nyu.edu

ABSTRACT

Let Σ be an alphabet and μ be a distribution on Σ^k for some $k \geqslant 2$. Let $\alpha > 0$ be the minimum probability of a tuple in the support of μ (denoted $\mathrm{supp}(\mu)$). Here, the support of μ is the set of all tuples in Σ^k that have a positive probability mass under μ. We treat the parameters Σ, k, μ, α as fixed and constant.

We say that the distribution μ has a linear embedding if there exist an Abelian group G (with the identity element 0_G) and mappings $\sigma_i : \Sigma \to G$, $1 \leqslant i \leqslant k$, such that at least one of the mappings is non-constant and for every $(a_1, a_2, \ldots, a_k) \in \mathrm{supp}(\mu)$, $\sum_{i=1}^{k} \sigma_i(a_i) = 0_G$.

Let $f_i : \Sigma^n \to [-1, 1]$ be bounded functions, such that at least one of the functions f_i essentially has degree at least d, meaning that the Fourier mass of f_i on terms of degree less than d is negligible, say at most δ. In particular, $|\mathbb{E}[f_i]| \leqslant \delta$. The Fourier representation is w.r.t. the marginal of μ on the i^{th} co-ordinate, denoted (Σ, μ_i). If μ has no linear embedding (over any Abelian group), then is it necessarily the case that

$$|\mathbb{E}_{(x_1,x_2,\ldots,x_k)\sim\mu^{\otimes n}}[f_1(x_1)f_2(x_2)\cdots f_k(x_k)] = o_{d,\delta}(1),$$

where the right hand side $\to 0$ as the degree $d \to \infty$ and $\delta \to 0$?

In this paper, we answer this analytical question fully and in the affirmative for $k = 3$. We also show the following two applications of the result. The first application is related to hardness of approximation. We show that for every 3-ary predicate $P : \Sigma^3 \to \{0, 1\}$ such that P has no linear embedding, an *SDP integrality gap instance* of a P-CSP instance with gap $(1, s)$ can be translated into a dictatorship test with completeness 1 and soundness $s + o(1)$, under certain additional conditions on the instance. The second application is related to additive combinatorics. We show that if the distribution μ on Σ^3 has no linear embedding, marginals of μ are uniform on Σ, and $(a, a, a) \in \mathrm{supp}(\mu)$ for every $a \in \Sigma$, then every large enough subset of Σ^n contains a triple (x_1, x_2, x_3) from $\mu^{\otimes n}$ (and in fact a significant density of such triples).

CCS CONCEPTS

• **Theory of computation → Problems, reductions and completeness.**

KEYWORDS

constraint satisfaction problems, hardness of approximation

ACM Reference Format:
Amey Bhangale, Subhash Khot, and Dor Minzer. 2023. On Approximability of Satisfiable k-CSPs: II. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC '23), June 20–23, 2023, Orlando, FL, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3564246.3585120

1 INTRODUCTION

The motivation for this paper is to study the following quantity associated with the product of functions $f_1, f_2, \ldots, f_k : \Sigma^n \to \mathbb{R}$,

$$\mathop{\mathbb{E}}_{(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k)\sim\mu^{\otimes n}}[f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)\cdots f_k(\mathbf{x}_k)], \quad (1)$$

where each coordinate of $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ is distributed independently, according to the same distribution μ on Σ^k. We assume that all the functions are bounded, i.e., $\|f_i\|_\infty \leqslant 1$. This expression appears naturally in many areas including additive combinatorics, social choice, pseudorandomenss and hardness of approximation. Here are a few examples.

(1) Example 1: For $1 \leqslant i \leqslant 3$, let $f_i : \mathbb{Z}_p^n \to \{0, 1\}$ be the indicator functions of the sets $A_i \subseteq \mathbb{Z}_p^n$. Let μ be the uniform distribution on the three-term arithmetic progressions $(x, x+y, x+2y)$ in $\mathbb{Z}_p$. Then the quantity

$$\mathop{\mathbb{E}}_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\sim\mu^{\otimes n}}[f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)f_3(\mathbf{x}_3)],$$

up to a normalization factor, precisely counts the number of arithmetic progressions $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ from $\mathbb{Z}_p^n$ such that $\mathbf{x}_i \in A_i$ for every $i \in [3]$.

(2) Example 2: Consider a Boolean function $f : \{-1, +1\}^n \to \{-1, +1\}$. For a given $\rho \in [-1, 1]$, the stability of f, $\mathrm{Stab}_\rho(f)$, is defined as $\mathbb{E}[f(\mathbf{x})f(\mathbf{y})]$ where for each $i \in [n]$, $\mathbf{x}_i$ and $\mathbf{y}_i$ are uniformly distributed, and $\mathbb{E}[\mathbf{x}_i\mathbf{y}_i] = \rho$. The Majority is Stablest Theorem [18], which is instrumental in the area of hardness of approximation and the theory of social choice, is about estimating $\mathrm{Stab}_\rho(f)$ for the class of so-called low-influence functions.

(3) Example 3: Fix a predicate $P : \Sigma^k \to \{0, 1\}$ and a distribution μ on Σ^k. Dictatorship tests corresponding to a predicate P and a distribution μ are extensively studied in hardness of approximation. Here, one is given a function $f : \Sigma^n \to \Sigma$ and the acceptance probability of the test is precisely

$$\Pr_{(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k)\sim\mu^{\otimes n}}\left[(f(\mathbf{x}_1), f(\mathbf{x}_2), \cdots f(\mathbf{x}_k)) \in P^{-1}(1)\right].$$

*Supported by the NSF Award CCF-1422159, NSF Award CCF-2130816, and the Simons Investigator Award.
†Supported by a Sloan Research Fellowship and NSF Award CCF-2227876.

STOC '23, June 20–23, 2023, Orlando, FL, USA

ACM ISBN 978-1-4503-9913-5/23/06.
https://doi.org/10.1145/3564246.3585120

One is interested in estimating this probability for the class of low influence functions. Using the multilinear expansions of P and f, the above expectation can be expressed as a linear combination of expectations of the form (1).

Let $c = \Pr_{(a_1,a_2,\ldots,a_k)\sim\mu}[(a_1, a_2, \ldots, a_k) \in P^{-1}(1)]$. It is seen that the test accepts any Dictatorship function, namely functions of the form $f(\mathbf{x}) = \mathbf{x}_{i_0}$ for a fixed co-ordinate $i_0 \in [n]$, with probability c. While tests with imperfect completeness, namely with $c < 1$, are interesting and well-studied in hardness of approximation,[1] in the current paper, we exclusively focus on tests with perfect completeness, namely with $c = 1$. That is, we assume that $\text{supp}(\mu) \subseteq P^{-1}(1)$. In fact, we will generally assume that μ has full support, i.e. $\text{supp}(\mu) = P^{-1}(1)$ and then talk interchangeably in terms of either the predicate P or the distribution μ. In terms of hardness of approximation, this amounts to studying approximability of Constraint Satisfaction Problems (CSPs) on (fully) satisfiable instances, and this indeed has been the main motivation for authors' work in [5], continuing in the current paper.

One way to analyze the expectation from (1) is to write each function f_i as the sum of two functions $g_i + h_i$, where g_i is the structured part of f_i and h_i is the remaining unstructured part (resembling noise). The idea is that whenever the term h_i appears in the product of functions, then the expectation is negligible. Therefore, the expectation can be estimated by replacing each f_i by its structured part g_i. For instance, in Example 1, Roth's Theorem [22] estimates the desired density of arithmetic progressions; therein, the structured part is taken as all the heavy-weight Fourier terms of f_i. It is shown that the contribution of the unstructured part is negligible; formally, if we let $\hat{f}_i$ be the Fourier terms of f_i, then we have

$$\left| \mathop{\mathbb{E}}_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\sim\mu^{\otimes n}} [f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)f_3(\mathbf{x}_3)] \right| \leqslant \min_{1\leqslant i\leqslant 3} \|\hat{f}_i\|_\infty.$$

On the other hand, it is often useful (especially in hardness of approximation) to take the structured part as the low-degree part of f_i. In this case, after replacing the functions f_i by their low degree parts g_i, provided that g_i are low influence functions, it is possible to estimate the expectation well using invariance principles. Here, one replaces the discrete inputs from Σ^n by Gaussian inputs and then the expectation is estimated using bounds in the Gaussian space. Still, the question remains as to when one can argue that the expectation is negligible for the unstructured, i.e. the high-degree, part of the functions.

Specifically, one is naturally led to the following analytic question.

Question 1. (Informal) *Find the necessary and sufficient condition on the distribution μ on Σ^k, such that*

$$\left| \mathop{\mathbb{E}}_{(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k)\sim\mu^{\otimes n}} [f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)\cdots f_k(\mathbf{x}_k)] \right| \to 0 \quad \text{as} \quad d \to \infty, \tag{2}$$

where the functions are bounded in $[-1, 1]$ and at least one function (essentially) has degree at least d.

Mossel [17] showed a sufficient condition: if the distribution μ is *connected*, then Conclusion (2) as above holds. The connectedness condition is defined as follows: for every pair of tuples $(a_1, a_2, \ldots, a_k) \in \text{supp}(\mu)$ and $(a'_1, a'_2, \ldots, a'_k) \in \text{supp}(\mu)$, there is a way to convert the first tuple to the second by replacing only one coordinate at a time such that every intermediate tuple remains in $\text{supp}(\mu)$.

The connectedness condition however is not necessary. An example is noted implicitly in [4]. Let G be a non-Abelian group with no dimension one representation. Consider the group-equation predicate $P : G^3 \to \{0, 1\}$, $P^{-1}(1) = \{(x, y, z) | x \cdot y \cdot z = 1_G\}$, along with the distribution μ that is uniform on $P^{-1}(1)$. The distribution μ is (clearly) not connected and Conclusion (2) still holds as can be shown using basic representation theory.

A certain necessary condition was observed in [5] (for Conclusion (2) to hold), namely that the distribution μ has no linear embedding as defined below. To illustrate that this condition is necessary, one considers the contra-positive: if the distribution μ does have a linear embedding (in particular, it is not connected), then there do exist high-degree, bounded functions that make the expectation in (2) non-negligible.

Definition 1. *We say that a distribution μ on Σ^k has a linear embedding (or that μ satisfies a linear equation or simply that μ is linear) if there exists an Abelian group G and mappings $\sigma_i : \Sigma \to G$, $1 \leqslant i \leqslant k$, such that (i) at least one of the maps σ_i is non-constant and (ii) for every $(a_1, a_2, \ldots, a_k) \in \text{supp}(\mu)$, $\sum_{i=1}^k \sigma_i(a_i) = 0_G$.*

The illustration is as follows. Suppose μ does have a linear embedding as in the definition. We show that it is possible to achieve non-negligible expectation in (2). To see this, let χ be any non-trivial character of the Abelian group G, namely a non-trivial group homomorphism $\chi : G \to \mathbb{C}$, and define $f_i(\mathbf{x}_i) = \prod_{j=1}^n \chi(\sigma_i((\mathbf{x}_i)_j))$. Now,

$$\begin{aligned} f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)\cdots f_k(\mathbf{x}_k) &= \prod_{i=1}^k \prod_{j=1}^n \chi(\sigma_i((\mathbf{x}_i)_j)) \\ &= \prod_{j=1}^n \prod_{i=1}^k \chi(\sigma_i((\mathbf{x}_i)_j)) \\ &= \prod_{j=1}^n \chi\left(\sum_{i=1}^k \sigma_i((\mathbf{x}_i)_j)\right) \\ &= \prod_{j=1}^n \chi(0_G) \\ &= 1. \end{aligned}$$

Here one uses the multiplicativity of the character χ and that $\chi(0_G) = 1$. For every $1 \leqslant j \leqslant n$, we have $\sum_{i=1}^k \sigma_i((\mathbf{x}_i)_j) = 0_G$ noting that the tuple $((\mathbf{x}_1)_j, \ldots, (\mathbf{x}_k)_j) \in \text{supp}(\mu)$ and using the definition of the linear embedding. Moreover, for large n, whenever σ_i is non-constant, the corresponding f_i is a (essentially) high-degree function.[2]

[1] Indeed, Example 2 corresponds to the hardness of approximation result for the Max-Cut problem. Here the predicate is $x \neq y$ over a binary alphabet, μ is the ρ-correlated distribution on $\{-1, 1\}^2$ as mentioned, completeness $c = \frac{1-\rho}{2}$, and $-1 < \rho < 0$.

[2] The functions here are complex valued with absolute value 1; one can take their real part if one insists on having real valued functions.

Motivated by these examples and certain long-term applications to approximability of constraint satisfaction problems (CSPs) on satisfiable instances, authors of [5] hypothesized that the non-linearity is indeed the necessary and sufficient condition. We state the hypothesis below.

Hypothesis 1. (Informal): *The necessary and sufficient condition on a distribution μ on Σ^k so that the Conclusion (2) holds is that μ has no linear embedding over any Abelian group.*

In [5], the authors were able to prove the hypothesis for a subclass of 3-ary predicates referred to therein as *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 $(x, y, z) \in P^{-1}(1)$ and also, for every $(x, z) \in \Sigma \times \Sigma$, there exists a $y \in \Sigma$ such that $(x, y, z) \in P^{-1}(1)$. We recall that while considering predicates, we always have an underlying distribution μ (in this case on Σ^3) such that $\text{supp}(\mu) = P^{-1}(1)$ and we may interchangeably talk in terms of either the predicate P or the distribution μ.

In this paper, we prove the hypothesis for **all** 3-ary predicates. The result, referred to as the Main Lemma in the rest of the paper, is stated below. It is more convenient (and general) to work with distributions μ on $\Sigma \times \Gamma \times \Phi$, allowing a different alphabet for each co-ordinate. In this case, a linear embedding consists of maps into an Abelian group G, $\sigma : \Sigma \to G$, $\gamma : \Gamma \to G$, $\phi : \Phi \to G$, not all constant, such that $\sigma(x)+\gamma(y)+\phi(z) = 0_G$ for all $(x, y, z) \in \text{supp}(\mu)$. We assume, unless stated otherwise, that the marginals of μ have full support on Σ, Γ, Φ respectively. In the following, m denotes the maximum size of Σ, Γ, Φ and $\alpha > 0$ denotes the minimum probability of a tuple in $\text{supp}(\mu)$. We always treat μ as fixed and m, α as fixed constants.

Lemma 1 (Main Analytical Lemma). *Suppose $|\Sigma|, |\Gamma|, |\Phi| \leqslant m$ and μ is a distribution over $\Sigma \times \Gamma \times \Phi$ such that*

- *The support of μ cannot be linearly embedded.*
- *$\mu(x, y, z) \geqslant \alpha$ for some $\alpha > 0$ and all $(x, y, z) \in \text{supp}(\mu)$.*
- *Marginals of μ (denoted as μ_x, μ_y, μ_z resp.) have full support on Σ, Γ, Φ respectively.*

Considering m and α as fixed, for all $\varepsilon > 0$, there are $\xi, \delta > 0$ such that the following holds. If $f\colon \Sigma^n \to [-1, 1]$, $g\colon \Gamma^n \to [-1, 1]$, $h\colon \Phi^n \to [-1, 1]$ and $\text{Stab}_{1-\xi}(h; \mu_z) \leqslant \delta$, then we have that

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

We clarify the condition that $\text{Stab}_{1-\xi}(h) \leqslant \delta$. Note that we have dropped μ_z from the notation for convenience. The parameter $\text{Stab}_{1-\xi}(h)$ denotes the stability of h under the noise parameter ξ. It is defined as $\langle h, T_{1-\xi} h\rangle$ where $T_{1-\xi}$ is the standard Beckner (noise) operator. We refer to the full-version of the paper for all analytic definitions and basic tools.

The condition that $\text{Stab}_{1-\xi}(h) \leqslant \delta$ serves as a proxy for the condition that the function h is essentially of high degree. Indeed, if $\text{Stab}_{1-\xi}(h) \leqslant \delta$, it implies that the Fourier mass of h on terms of degree less than $\frac{1}{\xi}$ is at most $O(\delta)$. Conversely, if the Fourier mass on terms of degree less than $O(\frac{1}{\xi}\log(\frac{1}{\delta}))$ is at most $\frac{\delta}{2}$, then $\text{Stab}_{1-\xi}(h) \leqslant \delta$. Hence the low-stability condition is a proxy for the high-degree condition and turns out to be more convenient to work with.

One may wonder when a function h is bounded in $[-1, 1]$ as well as essentially of high degree. A natural example is when $h' : \Phi^n \to [-1, 1]$ is an arbitrary function and $h = h' - T_{1-\xi}h'$. In this case, since h' is bounded and $T_{1-\xi}$ is an averaging operator, h is also bounded. In addition, the operator $T_{1-\xi}$, roughly speaking, retains only the low-degree part of h', and hence $h = h' - T_{1-\xi}h'$, roughly speaking, corresponds to the high-degree part of h'. More precisely, the Fourier mass of h on terms of degree less than $\frac{\delta}{\xi}$ is at most δ.[3] In applications, it is almost always the case that the lemma is applied with $h = h' - T_{1-\xi}h'$ for some bounded function h'. One refers to h as a soft-truncation of h', as opposed to a hard-truncation that would simply drop terms of degree less than a certain degree threshold. The advantage of using soft-truncation is that it preserves boundedness of functions whereas the hard-truncation in general does not.

Applications. In this section, we state a couple of applications of our main analytical lemma.

Hardness of approximation: Our first application is new results on dictatorship tests from integrality gap instances of constraint satisfaction problems (CSPs). 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. 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 $\alpha \cdot$ Opt fraction of the constraints, where Opt is the value of the optimum assignment.

Assuming the Unique Games Conjecture [15], Raghavendra [21] gave optimal hardness of approximation result for every Max-P-CSP. His work can be succinctly described as a two-step scheme:

$$\text{SDP integrality gap} \implies \text{A dictatorship test} \implies \text{A hardness of approximation result.}$$

However in his work, one necessarily loses perfect completeness and the hardness result does not hold on CSP instances that are (fully) satisfiable.

In order to prove hardness results on satisfiable instances, one would need a similar scheme that preserves perfect completeness in both the steps. Towards this goal, the Rich 2-to-1 Games Conjecture was introduced in [7] and further explored in [6]. Under this conjecture, [6, 7] showed how to convert, in certain specific cases, dictatorship test with completeness 1 and soundness s to a hardness result on satisfiable CSP instances with hardness threshold $s + \varepsilon$, for every constant $\varepsilon > 0$. This result can be interpreted as fulfilling the second step in the scheme above (albeit only morally speaking, since the implication is not entirely seamless and general yet).

[3] Given the connection between stability and degree before, h also has low stability, albeit with somewhat different parameters.

It thus remains to fulfill the first step in the scheme while preserving perfect completeness. The authors [5] made progress on this question, showing that a $(1, s)$ integrality gap instance for certain CSPs can be converted into a dictatorship test with completeness 1 and soundness $s+\varepsilon$. Their result however was limited to (non-linear) 3-ary predicates satisfying the aforementioned semi-richness condition, and this was because in [5], the authors were able to prove analytic Lemma 1 only under the additional semi-richness condition. Since we are now able to prove the lemma for all (non-linear) 3-ary predicates, we now get the intergality gap to dictatorship test implication for all such predicates. The formal statement of our result appears below (one wishes that the condition ($2b$) therein could be dropped; if so, we would have a full-proof implication).

For definitions and a more detailed discussion, we refer to Section 3 and the introductory section of [5].

Theorem 1. *Let $P\colon \Sigma^3 \to \{0,1\}$ be any predicate that satisfies the following conditions: (1) P has no linear embedding, (2a) there exists an instance of Max-P-CSP that has a $(1,s)$-integrality gap for the basic SDP relaxation, (2b) on every constraint, the local distribution in the SDP solution is not linearly embeddable. Then for every $\varepsilon > 0$, there is a dictatorship test for P-CSP that has perfect completeness and soundness $s+\varepsilon$.*

Counting Progressions: In additive combinatorics, finding a certain fixed progression (i.e. a pattern) in a subset of a given group is a cornerstone question. Such questions have had huge implications in understanding the pseudo-random properties of subsets of a group. Below we list a few of these results answering this question in different settings.

Fix a finite Abelian group $(G, +)$. A subset $A \subseteq G$ is said to be three term arithmetic progression (3-AP) free if there is no arithmetic progression of size 3 in A. In other words, there are no elements $x, y, z \in A$ such that $x + z = 2y$. The famous Roth's Theorem [22] shows that any 3-AP free subset of $\mathbb{Z}_N$ must be of size $o(N)$. In the contrapositive, any constant density subset of $\mathbb{Z}_N$ contains a 3-term AP. Szemerédi [23] generalized Roth's Theorem to any k-term AP. In these and similar results quoted next, one actually shows that a density δ subset of the group contains an ε fraction of all the progressions; the precise dependence of ε as a function of δ is also interesting, but for the sake of conciseness, we skip quantitative statements to that effect.

Now let $(G, \cdot)$ be a finite group that is not necessarily Abelian. A subset of G is called product free if it does not contain three elements x, y, z with $x \cdot y = z$. If G is any Abelian group, then it is easy to come up with product-free sets of constant density. Gowers [12] showed that this is not true for a class of non-Abelian groups called *quasirandom groups.*[4] That is, every constant density subset of a quasirandom group contains the progression (x, y, xy). Tao [24] extended Gowers' result to other progressions of the form (x, xg, xg^2) and (x, xg, xg^2, xg^3) for some very specific quasirandom groups. Bergelson and Tao [2] established it for progressions (x, xg, gx) and (g, x, xg, gx) for every quasirandom group. Recently, following the work by Peluse [19], Bhangale, Harsha and Roy [3] established it for the progression (x, xg, xg^2) for every quasirandom group. In a high-dimensional setting, finding the largest size of the 3-AP free set in $\mathbb{F}_3^n$ has received considerable attention [1, 8, 16]. Ellenberg and Gijswijt [11], building on a beautiful work by Croot, Lev, Pach [10], obtained a substantial quantitative improvement over Roth's Theorem (applied to $\mathbb{F}_3^n$).

[4] A group (or rather a family of groups) is quasirandom if the minimum dimension of any non-trivial group representation grows with the size of the group.

We now state our general theorem that establishes a similar result in high-dimensional setting for arbitrary 3-ary progression provided that the progression has no linear embedding (along with a couple of other conditions).

Theorem 1. *Suppose μ is a distribution over Σ^3 such that (1) the marginal distributions μ_x, μ_y, μ_z are uniform on Σ, (2) $\{(x,x,x) \mid x \in \Sigma\} \subseteq \mathrm{supp}(\mu)$, and (3) $\mathrm{supp}(\mu)$ cannot be linearly embedded. Then for all $\delta > 0$, there exists $\varepsilon > 0$ such that for $S \subseteq \Sigma^n$ with $|S| \geqslant \delta|\Sigma|^n$,*

$$\Pr_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}} [\mathbf{x} \in S, \mathbf{y} \in S, \mathbf{z} \in S] \geqslant \varepsilon.$$

Note that the condition (2) is necessary for such a conclusion to hold. This can be seen by the following example. Consider $\Sigma = \{0,1,2\}$ and μ be uniform on $\Sigma^3 \setminus \{(0,0,0\}$. It is easy to check that μ is not linearly embeddable. Now, if we take $S \subseteq \Sigma^n$ to be $S = \{\mathbf{x} \in \Sigma^n \mid x_1 = 0\}$, then clearly the conclusion does not hold. Our theorem is comparable to the result by Hązła, Holenstein and Mossel [13] with the same conclusion under the additional condition that the distribution μ is *connected.* As there are distributions that are not linearly embeddable as well as not connected, Theorem 1 extends their result.

2 TECHNIQUES

In this section, we elaborate on the ideas involved in the proof of Lemma 1. We focus only on a few high-level ideas here. Since we will skip many technical (and even conceptual) details, there might be some discrepancies between the high-level exposition here and formal proofs appearing later.

Let μ be a distribution on $\Sigma \times \Gamma \times \Phi$ such that $\mathrm{supp}(\mu)$ is not linearly embeddable. We wish to show that

$$\left| \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}} [f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})] \right| \approx 0, \tag{3}$$

where $f : \Sigma^n \to [-1,1]$, $g : \Gamma^n \to [-1,1]$, $h : \Phi^n \to [-1,1]$, are ℓ_∞-bounded and at least one of the functions essentially has high degree. We begin by sketching Mossel's proof [17] that works in the 2-ary case, i.e. for a (non-linear) distribution μ on $\Sigma \times \Gamma$. This will help us understand various hurdles and new ideas needed to overcome these hurdles in our proof of the 3-ary case as above.

2.1 The 2-ary Case: Sketch of Mossel's Proof

Let μ be a distribution on $\Sigma \times \Gamma$ such that $\mathrm{supp}(\mu)$ is not linearly embeddable. It is easily seen that the non-linearity condition, in this special 2-ary case, is same as saying that $\mathrm{supp}(\mu)$, viewed as a bipartite graph G_μ on the vertex set $\Sigma \cup \Gamma$, is connected. Indeed, if this graph were disconnected, with components $C_0 \cup D_0, \ldots, C_{r-1} \cup D_{r-1}$, then an embedding $\sigma : C_j \to j, \gamma : D_j \to -j$ is an embedding of Σ and Γ respectively into $\mathbb{Z}_r$ and for all $(x,y) \in \mathrm{supp}(\mu)$ (i.e. the edges of the graph G_μ), we have $\sigma(x) + \gamma(y) = 0$ in $\mathbb{Z}_r$.

We intend to show that if $f : \Sigma^n \to \mathbb{R}, g : \Gamma^n \to \mathbb{R}$ are n-dimensional ℓ_∞-bounded functions where g has high degree, then $\left|\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}} [f(\mathbf{x})g(\mathbf{y})]\right|$ is small. For simplicity of exposition, we assume that g in fact has full degree n.[5] In this case, we are able to show that $\left|\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})\right]\right| \leqslant (1-\tau)^n\|f\|_2\|g\|_2$ for some constant $\tau = \tau(\mu) > 0$. We emphasize here that one gets an upper bound in terms of the ℓ_2-norm of the functions. This of course implies an upper bound in terms of the ℓ_∞-norms. Thus we really do not need the n-dimensional functions to be ℓ_∞-bounded in the 2-ary case. This is one aspect (among many) in which the 3-ary case is fundamentally different, where one does need the n-dimensional functions to be ℓ_∞-bounded (as we will soon demonstrate via an example).

Continuing the consideration of the 2-ary case, the proof proceeds in two steps: first establishing a base case inequality (for $n = 1$) and then observing that the inequality tensorizes, leading to an inductive proof and the desired bound for the general case of n-dimensional functions. The base case inequality is necessarily an ℓ_2-inequality and this fact is essential for the inductive proof (and the same holds in the 3-ary case).

Towards stating the base case inequality, let $f : \Sigma \to \mathbb{R}, g : \Gamma \to \mathbb{R}$ be functions. By Cauchy-Schwarz,

$$\left|\mathop{\mathbb{E}}_{(x,y)\sim\mu}\left[f(x)g(y)\right]\right| \leqslant \|f\|_2\|g\|_2.$$

We refer to this essentially trivial inequality as the (base case) sanity check inequality. The inequality that is actually needed is that when $\mathbb{E}[f] = \mathbb{E}[g] = 0$, we in fact have the improvement

$$\left|\mathop{\mathbb{E}}_{(x,y)\sim\mu}\left[f(x)g(y)\right]\right| \leqslant (1-\tau)\|f\|_2\|g\|_2, \qquad \mathbb{E}[f] = \mathbb{E}[g] = 0, \quad (4)$$

for some constant $\tau = \tau(\mu) > 0$. It is not difficult to see that this follows from the connectedness of the distribution μ (or equivalently the graph G_μ), but we skip the proof. An equivalent way to express the inequality is that the operator $T : \tilde{L}_2(\Gamma;\mu_y) \to \tilde{L}_2(\Sigma;\mu_x)$ defined as $Tg(x) = \mathbb{E}_{(x',y)\sim\mu}\left[g(y)|x' = x\right]$ has operator norm at most $1-\tau$. Here $\tilde{L}_2(\Gamma;\mu_y)$ denotes the subspace of $L_2(\Gamma;\mu_y)$ consisting of those functions g for which $\mathbb{E}[g] = 0$ (and similarly for $\tilde{L}_2(\Sigma;\mu_x)$). The operator norm of T, denoted $\|T\| = \max_{g:\mathbb{E}[g]=0}\|Tg\|_2/\|g\|_2$, is at most $1-\tau$ according to the equivalent interpretation of the inequality (4), which can then be derived as:

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{(x,y)\sim\mu}\left[f(x)g(y)\right]\right| &= |\langle f, Tg\rangle| \\ &\leqslant \|f\|_2\|Tg\|_2 \\ &\leqslant \|f\|_2\|T\|\|g\|_2 \leqslant (1-\tau)\|f\|_2\|g\|_2.\end{aligned}$$

Now we consider the n-dimensional case. Let $f : \Sigma^n \to \mathbb{R}, g : \Gamma^n \to \mathbb{R}$ be n-dimensional functions. As mentioned before, we assume that g has full degree, which amounts to saying that $g \in \tilde{L}_2(\Gamma;\mu_y)^{\otimes n}$. In this case, it follows directly that

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})\right]\right| \leqslant (1-\tau)^n\|f\|_2\|g\|_2,$$

using the well-known fact that the operator norm is multiplicative (i.e. it tensorizes), namely that $\|T^{\otimes n}\| = \|T\|^n \leqslant (1-\tau)^n$. Using this fact, one immediately concludes that

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})\right]\right| &= \left|\left\langle f, T^{\otimes n}g\right\rangle\right| \\ &\leqslant \|f\|_2\|T^{\otimes n}g\|_2 \\ &\leqslant \|f\|_2\|T^{\otimes n}\|\|g\|_2 \leqslant (1-\tau)^n\|f\|_2\|g\|_2,\end{aligned}$$

as desired. If one wishes, one can prove the multiplicativity of operator norm by induction and view the overall proof as an inductive proof, using the base case inequality (4) and "gaining" a factor $1-\tau$ in each step of the induction. While we don't demonstrate it here, we mention it because the proof for the 3-ary case proceeds along similar lines, albeit with many conceptual and technical hurdles. Therein, it is rather challenging even to formulate the "correct" base case inequality.

2.2 Towards 3-ary Base Case: Restoring Sanity First

Moving onto the 3-ary case, let μ be a distribution on $\Sigma\times\Gamma\times\Phi$ such that supp(μ) is not linearly embeddable. One hopes to write down a suitable base case inequality and use it towards an inductive proof. However, it turns out that even the sanity check inequality fails in general! That is, for $f : \Sigma \to \mathbb{R}, g : \Gamma \to \mathbb{R}, h : \Phi \to \mathbb{R}$, while we desire a base case inequality (say when $\mathbb{E}[f] = 0$) of the form

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}\left[f(x)g(y)h(z)\right]\right| \leqslant (1-\tau)\|f\|_2\|g\|_2\|h\|_2, \qquad (5)$$

it may actually happen that

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}\left[f(x)g(y)h(z)\right]\right| > \|f\|_2\|g\|_2\|h\|_2.$$

In other words, we may not even have the upper bound of the expression $\|f\|_2\|g\|_2\|h\|_2$ in the 3-ary case whereas the corresponding upper bound in the 2-ary case is the essentially trivial application of Cauchy-Schwarz! Here is an example.

Suppose that $\Sigma = \Gamma = \Phi$, $|\Sigma| = m \geqslant 54$, and μ has a probability mass of $1-\varepsilon$ uniformly spread on the triples $\{(x,x,x)|x \in \Sigma\}$ and the remaining probability mass of ε uniformly spread on all the remaining triples in Σ^3. Clearly, supp$(\mu) = \Sigma^3$ and hence μ is not linearly embeddable. The marginals of μ are uniform on Σ. We can certainly construct a function $f : \Sigma \to \mathbb{R}$ such that $\mathbb{E}[f(x)] = 0$ and $\mathbb{E}\left[f(x)^3\right] > \|f\|_2^3$. For instance, f could take the values $2m, -m, -m$ at three distinct points in Σ and zero at the remaining points in Σ. In this case, $\mathbb{E}[f(x)] = 0, \mathbb{E}\left[f(x)^2\right] = 6m$, and $\mathbb{E}\left[f(x)^3\right] = 6m^2$, and thus $\mathbb{E}\left[f(x)^3\right] \geqslant \sqrt{m/6}\cdot\|f\|_2^3 \geqslant 3\,\|f\|_2^3$. Letting $f = g = h$ and recalling that the triples (x,x,x) receive $1-\varepsilon$ of the probability mass, it follows that

$$\begin{aligned}\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}\left[f(x)g(y)h(z)\right] &\geqslant (1-\varepsilon)\mathbb{E}\left[f(x)^3\right] - \varepsilon\cdot O_m(1) \\ &\geqslant 2\cdot\|f\|_2^3 = 2\cdot\|f\|_2\|g\|_2\|h\|_2,\end{aligned}$$

by making ε sufficiently small. This example also shows that in order to claim the desired bound for n-dimensional functions as in Equation (3), we must use the fact that the functions are ℓ_∞-bounded (i.e. in $[-1,1]$)! Indeed, consider the same example here

[5] This amounts to saying that after restricting any $n-1$ co-ordinates, the expectation of g over the remaining co-ordinate is zero.

and let n-dimensional functions $\tilde{f} = \tilde{g} = \tilde{h} : \Sigma^n \to \mathbb{R}$ be all equal to $f^{\otimes n}/\|f^{\otimes n}\|_2$. Then these have all ℓ_2-norm 1, whereas

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[\tilde{f}(\mathbf{x})\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\right]\\ &= \mathop{\mathbb{E}}_{(x,y,z)\sim\mu}[f(x)g(y)h(z)]^n \cdot \frac{1}{\|f\|_2^{3n}}\\ &\geqslant 2^n.\end{aligned}$$

We thus face a seemingly intractable hurdle and a contradictory set of constraints: (i) we do need the ℓ_∞-boundedness of the n-dimensional functions, (ii) an inductive proof is some form of tensorization argument and hence inherently an ℓ_2-proof; consequently, the intermediate functions arising during the induction can only be assumed to have ℓ_2 norm at most 1, (iii) the inductive argument requires a base case ℓ_2-inequality such as (5) which actually happens to fail miserably!

We now show how to overcome this hurdle step-by-step. This is achieved in a round-about manner, by carefully transforming the distribution and the alphabet $(\Sigma \times \Gamma \times \Phi, \mu)$ to another distribution and alphabet $(\tilde{\Sigma} \times \tilde{\Gamma} \times \tilde{\Phi}, \tilde{\mu})$. Formally, we show that

- If μ was not linearly embeddable to begin with, then $\tilde{\mu}$ isn't either.
- If Lemma 1 (i.e. our Main Lemma/Result) holds for $\tilde{\mu}$, then it also holds for μ.

In this sense, we are able to reduce our task of proving the lemma for the original distribution μ to proving the same lemma for the new distribution $\tilde{\mu}$. In fact, there will be a series of such transformations. The (first) transformation will ensure that the marginal of $\tilde{\mu}$ on $\tilde{\Gamma}\times\tilde{\Phi}$ is a uniform, product distribution. Once we have this additional property, we at least have the (base case) sanity check inequality as demonstrated next. For the sake of notational convenience, we rename the new distribution and the alphabet again as $(\Sigma\times\Gamma\times\Phi, \mu)$ and assume that the marginal of μ on $\Gamma \times \Phi$ is a uniform, product distribution. If so, it is easily seen that we get the (base case) sanity check inequality, namely that for $f : \Sigma \to \mathbb{R}, g : \Gamma \to \mathbb{R}, h : \Phi \to \mathbb{R}$, we have

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}[f(x)g(y)h(z)]\right| \leqslant \|f\|_2\|g\|_2\|h\|_2.$$

Indeed, by Cauchy-Schwarz,

$$\begin{aligned}&\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}[f(x)g(y)h(z)]^2 \qquad (6)\\ &\leqslant \mathop{\mathbb{E}}_{x\sim\mu_x}\left[f(x)^2\right]\mathop{\mathbb{E}}_{(y,z)\sim\mu_{y,z}}\left[g(y)^2h(z)^2\right]\\ &= \mathop{\mathbb{E}}_{x\sim\mu_x}\left[f(x)^2\right]\mathop{\mathbb{E}}_{y\sim\mu_y}\left[g(y)^2\right]\mathop{\mathbb{E}}_{z\sim\mu_z}\left[h(z)^2\right]\\ &= \|f\|_2^2\|g\|_2^2\|h\|_2^2, \qquad (7)\end{aligned}$$

where in the second step, we used the property that (y, z) are uniform and independent! It is also possible to ensure (after the transformation) another property of μ that is quite convenient: for all pairs $(y, z) \in \Gamma \times \Phi$, there is a unique $x \in \Sigma$ such that $(x, y, z) \in \text{supp}(\mu)$ (we then say that (y, z) determine x). The details of this transformation and related proofs appear in the full-version of the paper are borrowed from authors' earlier work [5].

2.3 The 3-ary Relaxed Base Case: Overcoming the Horn-SAT Obstruction

We will henceforth assume that the distribution μ on $\Sigma \times \Gamma \times \Phi$ has no linear embedding and has uniform marginal on $\Gamma \times \Phi$. Now that we at least have the sanity check inequality, we ask ourselves whether we can claim the desired base case inequality as below:

Question 2. (Desired, Hypothetical Base Case Inequality:) *If μ has no linear embedding and has uniform marginal on $\Gamma \times \Phi$, is it necessarily the case that for $f : \Sigma \to \mathbb{R}, g : \Gamma \to \mathbb{R}, h : \Phi \to \mathbb{R}$,*

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}[f(x)g(y)h(z)]\right| \leqslant (1-\tau(\theta))\|f\|_2\|g\|_2\|h\|_2, \qquad (8)$$

where $|\mathbb{E}[f]| \leqslant (1-\theta)\|f\|_2$. To avoid the trivial case when f, g, h are all constant functions, we added here the condition that f is non-constant and has some variance, the condition captured by the requirement $|\mathbb{E}[f]| \leqslant (1-\theta)\|f\|_2$.

We note that such a base case inequality seems necessary towards an inductive proof since one hopes to "gain" a factor of $1-\tau$ in each step of the induction. However it turns out that such an inequality need not necessarily hold and there could be an obstruction that we refer to as the Horn-SAT obstruction (and this is the only possible obstruction).

Definition 2. *Assume that a distribution μ on $\Sigma \times \Gamma \times \Phi$ has no linear embedding and its marginal on $\Gamma \times \Phi$ is uniform. We say that μ has a Horn-SAT embedding if there are Boolean functions $f : \Sigma \to \{0, 1\}, g : \Gamma \to \{0, 1\}, h : \Phi \to \{0, 1\}$, such that*

- *For all $(x, y, z) \in \text{supp}(\mu)$, we have $f(x) = g(y)h(z)$.*
- *f is non-constant (and in that case so must be g and h).*

The condition $f(x) = g(y)h(z)$ for Boolean functions is equivalent to the conjunction of clauses $\overline{f(x)} \vee g(y)$, $\overline{f(x)} \vee h(z)$, $f(x) \vee \overline{g(y)} \vee \overline{h(z)}$. These are all Horn-SAT clauses (i.e. having at most one positive literal), explaining the term *Horn-SAT embedding*. We now make several remarks towards understanding how a Horn-SAT embedding is an obstruction towards the desired inequality (8) and how it is the only possible obstruction.

- Firstly, we note that having a Horn-SAT embedding violates inequality (8). Indeed, since $f(x) = g(y)h(z)$ in $\text{supp}(\mu)$ and (y, z) are uniform and independent, we have $\|f\|_2 = \|g\|_2\|h\|_2$ and then
$$\begin{aligned}\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}[f(x)g(y)h(z)] &= \mathop{\mathbb{E}}_{(y,z)\sim\mu_{y,z}}\left[g(y)^2h(z)^2\right]\\ &= \|g\|_2^2\|h\|_2^2 = \|f\|_2\|g\|_2\|h\|_2.\end{aligned}$$
One also notes that since f is Boolean and non-constant, it does have constant variance.
- Secondly, we note that if the inequality (8) is not possible, then there is necessarily a Horn-SAT embedding. A sketch of the proof is as follows. For a fixed θ, suppose that there are functions that violate the inequality for all $\tau \to 0$. Then by standard compactness argument, there are three functions $f : \Sigma \to \mathbb{R}, g : \Gamma \to \mathbb{R}, h : \Phi \to \mathbb{R}$, such that
$$\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}[f(x)g(y)h(z)] = \|f\|_2\|g\|_2\|h\|_2,$$

i.e. achieving an exact equality. This means that the application of Cauchy-Schwarz in Equation (7) must be tight and therefore $f(x) = g(y)h(z)$ in $\mathrm{supp}(\mu)$ (as equality of real numbers). If $f(x)$ is always non-zero, then so are $g(y)$ and $h(z)$. In this case, if at least one of f, g, h has a non-constant sign (i.e. positive and negative), then turning $f(x), g(y), h(z)$ into their $\{+1, -1\}$- signs, we have

$$\mathrm{sign}(f(x)) = \mathrm{sign}(g(y))\mathrm{sign}(h(z)),$$

which yields a linear embedding of μ, a contradiction. On the other hand, if f, g, h all have constant sign, then w.l.og. this sign is positive, and then $\log f(x) = \log g(y) + \log h(z)$ gives a linear emebdding, again a contradiction. Here f has some variance, so $\log f(x)$ is non-constant and the embedding is non-trivial. The embedding is not into a finite Abelian group, but this is not difficult to fix. One concludes therefore that $f(x)$ takes the zero value for some $x \in \Sigma$ and of course also takes a non-zero value for some $x' \in \Sigma$. We can now define the Horn-SAT embedding by turning $f(x), g(y), h(z)$ into Boolean 1 if the value is non-zero and Boolean 0 if the value is zero!

- In the definition, if f is non-constant, then so must be g and h. Let's suppose on the contrary that g is constant (the same proof applies for h). If $g \equiv 0$, then the condition $f(x) = g(y)h(z)$ implies that $f \equiv 0$, reaching a contradiction. If $g \equiv 1$, then one concludes that $f(x) = h(z)$ for all $(x, z) \in \mathrm{supp}(\mu_{x,z})$. Since μ is not linearly embeddable, its marginals are not linearly embeddable either.[6] In particular, $\mu_{x,z}$ has no linear embedding and hence is connected, implying that both f and h are constant, again a contradiction.

Considering these remarks, if μ does not have a Horn-SAT embedding, then we do have the base case inequality (8) and we can hope to carry out the induction. However, if μ does have a Horn-SAT embedding as in Definition 2, then the embedding serves as a violation of the inequality and we are stuck with a similar hurdle as before. The Horn-SAT embedding leads to n-dimensional functions $\tilde{f} = f^{\otimes n}/\|f^{\otimes n}\|$, $\tilde{g} = g^{\otimes n}/\|g^{\otimes n}\|$, $\tilde{h} = h^{\otimes n}/\|h^{\otimes n}\|$, with ℓ_2-norm 1, and

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[\tilde{f}(\mathbf{x})\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\right] = 1.$$

As before, this hinders the possibility of proving the n-dimensional inequality (3) by induction: there is no base case inequality and there is a counter-example if one allows functions to have ℓ_2 norm 1 instead of ℓ_∞ norm 1.

We overcome this hurdle in a similar manner as before, albeit with even more subtleness. We carefully transform the distribution and the alphabet $(\Sigma\times\Gamma\times\Phi, \mu)$ to another distribution and alphabet $(\tilde{\Sigma}\times\tilde{\Gamma}\times\tilde{\Phi}, \tilde{\mu})$. Formally, we show that

- If Lemma 1 (i.e. our Main Lemma/Result) holds for $\tilde{\mu}$, then it also holds for μ.
- All the key properties of μ are retained by $\tilde{\mu}$ which has further additional properties.

[6] This is seen easily from the definition of a linear embedding, Definition 1. If marginal of μ on a subset of co-ordinates has a linear embedding, then so does μ by letting the embedding on other co-ordinates to be 0_G.

In this sense, we are able to reduce our task of proving the lemma for the original distribution μ to proving the same lemma for the new distribution $\tilde{\mu}$. Now we state what additional properties $\tilde{\mu}$ has. For the sake of notational convenience, we rename the new distribution and the alphabet as $(\Sigma\times\Gamma\times\Phi, \mu)$ again. The key additional property is stated below, referred to as the relaxed base case inequality.

Definition 3. *(Relaxed Base Case Inequality) Suppose a distribution μ on $\Sigma\times\Gamma\times\Phi$ has no linear embedding and has uniform support on $\Gamma\times\Phi$. We say that μ satisfies the relaxed base case inequality if:*

- *There is some $\Sigma' \subseteq \Sigma$, $|\Sigma'| \geqslant 2$, and constants $C > 0$ and $0 < c < 1$ such that the following holds. For all $\tau > 0$, let functions $f: \Sigma \to \mathbb{R}$, $g: \Gamma \to \mathbb{R}$ and $h: \Phi \to \mathbb{R}$ be such that f has variance at least $\tau\|f\|_2^2$ on Σ', that is*

$$\mathop{\mathbb{E}}_{x,x'\in\Sigma'}\left[(f(x) - f(x'))^2\right] \geqslant \tau\|f\|_2^2.$$

Then

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}[f(x)g(y)h(z)]\right| \leqslant \max(1-\tau^C, c)\|f\|_2\|g\|_2\|h\|_2.$$

- *Furthermore, the distribution on $\Sigma'\times\Gamma\times\Phi$, derived as $(x, y, z) \sim \mu$ conditioned on $x \in \Sigma'$, cannot be linearly embedded.*

We remark that if μ did not have a Horn-SAT embedding, no transformation is needed, and one can simply take $\Sigma' = \Sigma$ in the above definition. However in general there might be a Horn-SAT embedding and the transformation would be needed. The transformation is rather subtle and while we do consider it to be one of the key ideas, we skip the discussion here and refer to the full-version of the paper for details. To summarize, we reduce the task of proving our Main Lemma 1 to the same task with the additional property that μ satisfies the relaxed base case inequality, i.e. to the task of proving the lemma stated below. In the following lemma, properties numbered 1 and 2 are as before, 3 and 4 can be assumed from the authors' earlier work as discussed in Section 2.2, and that numbered 5 is the key relaxed base case inequality.

Lemma 2. (Main Analytical Lemma under Relaxed Base Case Inequality) *Suppose $|\Sigma|, |\Gamma|, |\Phi| \leqslant m$ and μ is a distribution over $\Sigma\times\Gamma\times\Phi$ such that:*

1. *$\mu(x, y, z) \geqslant \alpha$ for some $\alpha > 0$ and all $(x, y, z) \in \mathrm{supp}(\mu)$.*
2. *$\mathrm{supp}(\mu)$ cannot be linearly embedded.*
3. *The marginal $\mu_{y,z}$ is uniform and independent over $\Gamma\times\Phi$.*
4. *For all $(y, z) \in \Gamma\times\Phi$, there is a unique $x \in \Sigma$ such that $(x, y, z) \in \mathrm{supp}(\mu)$ (i.e. y, z determine x).*
5. *μ satisfies the relaxed base case inequality as in Definition 3.*

Then for all $\varepsilon > 0$, there are $\xi, \delta > 0$ such that the following holds. If $f: \Sigma^n \to [-1, 1]$, $g: \Gamma^n \to [-1, 1]$ and $h: \Phi^n \to [-1, 1]$ satisfy that either $\mathrm{Stab}_{1-\xi}(g) \leqslant \delta$ or $\mathrm{Stab}_{1-\xi}(h) \leqslant \delta$, then we have that

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

2.4 The Inductive Argument (Without the Horn-SAT Obstruction)

Armed with the "correct" relaxed base case inequality, we now give an overview of the inductive proof (of Lemma 2). It is instructive and less cumbersome to first consider the special case when there is no Horn-SAT embedding and we already have the base case inequality as in (8). We will indicate how to incorporate the relaxed base case inequality later. Formal proofs appear in the full-version of the paper.

So let us focus on this special case and assume the base case inequality (8) holds. The inductive proof proceeds in several steps. We emphasize again that an inductive proof must necessarily work with ℓ_2 norms of functions that arise as intermediate functions during the induction and we have no control over their ℓ_∞ norms.[7] We are given that either g or h has essentially high degree, so let's say this holds for g, formalized in terms of its low-stability. The first step towards the inductive proof is to note that it is sufficient (and necessary as far as our proof goes) to focus on the case when f, g, h are homogenous functions. We will skip details regarding how this is sufficient towards the general case. Therefore let's assume that f, g, h are homogenous and define the parameter

$$\beta_{n,d_1,d_2,d_3} = \sup_{f,g,h} \frac{\left|\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right|}{\|f\|_2\|g\|_2\|h\|_2},$$

where the maximum is taken over all $f\colon \Sigma^n \to \mathbb{R}$, $g\colon \Gamma^n \to \mathbb{R}$, $h\colon \Phi^n \to \mathbb{R}$ homogenous of degrees d_1, d_2, d_3 respectively. Since we assumed that g had high degree, we think of d_2 as (roughly) the largest among the degrees. Indeed, it is sufficient to consider the case when$d_1, d_3 \leqslant 10d_2$, and we make this assumption skipping the details. We will be able to show an exponential decay, namely

$$\beta_{n,d_1,d_2,d_3} \leqslant (1-\Omega_{\alpha,m}(1))^{d_2},$$

completing the proof. We now describe how this exponential decay is proved. First, we reduce the dimension n so that $n \leqslant O(d_2)$. Then comes the core inductive argument, where we "gain" a factor $1-\Omega_{\alpha,m}(1)$ in each step of the induction, reducing the degree d_2 by one, until we have reduced it to say $\frac{d_2}{2}$.

Reducing Dimension: We show here that it is sufficient to consider the case when $n \leqslant O(d_2)$ (and we already assume that $d_1, d_3 \leqslant 10d_2$). The idea is as follows. As long as $n \gg d_2$, we can find a coordinate $i \in [n]$ which has very small "influence" on f, g and h; assume without loss of generality that this co-ordinate is $i = n$.

If the influence was zero, then f, g and h would only be functions of the first $n-1$ co-ordinates, and hence we would conclude that $\beta_{n,d_1,d_2,d_3} \leqslant \beta_{n-1,d_1,d_2,d_3}$, making "progress" in reducing n. However in general, that influence may be very small but still non-zero. In that case one may write the decompositions

$$f = f_1 + f', \quad g = g_1 + g', \quad h = h_1 + h',$$

where f_1, g_1, h_1 depend only on the first $n-1$ co-ordinates, and f', g', h' do depend on the n^{th} co-ordinate but have very small ℓ_2-norm (which is precisely what influence is). Since f', g', h' have very small norm, one doesn't expect them to contribute much, and one still hopes to deduce that $\beta_{n,d_1,d_2,d_3} \leqslant \beta_{n-1,d_1,d_2,d_3}$. Alas, this doesn't quite work. While their contribution is very small, it is still non-zero, and a naive application of this idea would only give $\beta_{n,d_1,d_2,d_3} \leqslant \beta_{n-1,d_1,d_2,d_3} + o(1)$, and the $o(1)$ error terms will keep accumulating in successive inductive steps. To overcome this difficulty, we perform a more detailed analysis, and need more refined decompositions of f, g and h. For the sake of simplicity, we consider only a specialized scenario that allows us to write

$$f = f_1 + f_2 f_2', \qquad g = g_1 + g_2 g_2', \qquad h = h_1 + h_2 h_2',$$

where f_1, g_1, h_1 depend only on the first $n-1$ coordinates and have the same degrees as f, g, h, the functions f_2, g_2, h_2 also depend only on the first $n-1$ coordinates but have degrees one less than f, g, h respectively, and f_2', g_2', h_2' are functions that only depend on the last coordinate and have very small ℓ_2-norm. Using this decomposition, we can write

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\\
&= \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n-1}}\left[f_1(\mathbf{x})g_1(\mathbf{y})h_1(\mathbf{z})\right]\\
&+ \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n-1}}\left[f_2(\mathbf{x})g_2(\mathbf{y})h_1(\mathbf{z})\right] \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu}\left[f_2'(\mathbf{x})g_2'(\mathbf{y})\right]\\
&+ \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n-1}}\left[f_2(\mathbf{x})g_1(\mathbf{y})h_2(\mathbf{z})\right] \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu}\left[f_2'(\mathbf{x})h_2'(\mathbf{z})\right]\\
&+ \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n-1}}\left[f_2(\mathbf{x})g_2(\mathbf{y})h_2(\mathbf{z})\right] \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu}\left[f_2'(\mathbf{x})g_2'(\mathbf{y})h_2'(\mathbf{z})\right]\\
&+ \text{Other terms.}
\end{aligned}$$

The other terms are zero thanks to the fact that $\mu_{y,z}$ is uniform and independent. The first term is the dominant term, the second and the third terms constitute as error terms, and the fourth term can be ignored when compared to the second and third terms. Roughly speaking, the reason is that if ε denotes the small norm of f_2', g_2', h_2', then the corresponding expectations are of the order ε^2 in the second and third terms, and of the order ε^3 in the fourth term.

The second and third terms are error terms, which however cannot be ignored altogether (as said before) and require care. Skipping many details, it turns out that the key is to bound the expectation

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu}\left[f_2'(\mathbf{x})(g_2'(\mathbf{y})+h_2'(\mathbf{z}))\right].$$

This can be upper bounded by $(1-\Omega(1))\|f_2'\|_2\sqrt{\|g_2'\|_2^2+\|h_2'\|_2^2}$. We emphasize here that this is an inequality on functions of a single co-ordinate. It is referred to as the additive base case inequality. Using this bound, one can obtain an effective enough bound on the second and third terms above, somehow recover the loss from these error terms and get that $\beta_{n,d_1,d_2,d_3} \leqslant \beta_{n-1,d_1,d_2,d_3}$ as desired.

The Core Induction: We now show the core inductive step giving the exponential decay, namely that $\beta_{n,d_1,d_2,d_3} \leqslant (1-\Omega_{\alpha,m}(1))^{d_2}$. We assume that $n \leqslant O(d_2)$ as discussed and that $d_1, d_3 \leqslant 10d_2$. Skipping details, it is sufficient to assume further that $d_1 \geqslant \Omega(d_2)$ as well. It follows from these assumptions that average influence of a coordinate on f is $\frac{d_1}{n} \geqslant \Omega(1)$. Let us assume that the coordinate n has influence $\Omega(1)$ on f. For the sake of simplicity, consider furthermore only a specialized scenario that allows us to write f, g and h as

$$f = f_1 f_1', \qquad g = g_1 g_1' \qquad h = h_1 h_1',$$

where f_1, g_1, h_1 depend only on the first $n-1$ co-ordinates and have degrees one less than f, g, h, and the functions f_1', g_1', h_1' depend

[7] When there is no Horn-SAT embedding, we do not need an ℓ_∞ bound on the original functions either. This is indeed the special case we are considering here. When there is a Horn-SAT embedding, as noted before, we must somehow use the fact that the original functions f, g, h do have ℓ_∞ norm at most 1. We still have no control however over the ℓ_∞ norm of the intermediate functions. This issue is addressed later.

only on the single coordinate n, and f_1' has constant norm (which amounts to the said influence). In this case, we would have that

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})] = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n-1}}[f_1(\mathbf{x})g_1(\mathbf{y})h_1(\mathbf{z})]\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu}\left[f_1'(\mathbf{x})g_1'(\mathbf{y})h_1'(\mathbf{z})\right].$$

By the inductive hypothesis, the first term is at most the quantity $\beta_{n-1,d_1-1,d_2-1,d_3-1}$ and by the base case inequality,we have $|\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu}\left[f_1'(\mathbf{x})g_1'(\mathbf{y})h_1'(\mathbf{z})\right]| \leqslant \lambda = 1-\Omega(1)$. Hence we get that

$$\beta_{n,d_1,d_2,d_3} \leqslant \lambda\beta_{n-1,d_1-1,d_2-1,d_3-1},$$

as desired, and iterating this gives an exponential decay.

In general, the main complication is that f, g and h need not take the specialized form as above, and instead one has to decompose them in a more complicated manner (amounting to decomposing a tensor into a sum of mutually orthogonal rank one tensors). Using a more complicated argument (but vaguely similar in spirit) one can still recover that $\beta_{n,d_1,d_2,d_3} \leqslant \lambda\beta_{n-1,d_1-1,d_2-1,d_3-1}$.

2.5 The Inductive Argument (Incorporating the Relaxed Base Case Inequality)

As discussed before, in general the base case inequality (8) does not hold and we are able to use only the relaxed base case inequality in Definition 3. We now indicate the main modification necessary in the inductive proof, skipping most other details from this overview.

Let $\Sigma' \subseteq \Sigma$ be the subset that exhibits the relaxed base case inequality in Definition 3. We consider the *effective influence* and *effective degree* of the function $f : \Sigma^n \to \mathbb{R}$. We recall that the standard influence of the i^{th} co-ordinate is

$$\mathop{\mathbb{E}}_{\substack{\mathbf{x}_{-i},\\ x_i,x_i'\in\Sigma}}\left[(f(\mathbf{x}_{-i},x_i)-f(\mathbf{x}_{-i},x_i'))^2\right].$$

That is, the influence is the variance of the function on the i^{th} co-ordinate after randomly restricting the rest of the co-ordinates. We define the effective influence as

$$\mathop{\mathbb{E}}_{\substack{\mathbf{x}_{-i},\\ x_i,x_i'\in\Sigma'}}\left[(f(\mathbf{x}_{-i},x_i)-f(\mathbf{x}_{-i},x_i'))^2\right],$$

which is similar, except that the variance is considered only over the subset Σ'.

We also indicate the related notion of the effective degree of f. We set up a suitable orthonormal basis $\mathbf{B}$ of *characters* for (single co-ordinate) functions in $L_2(\Sigma;\mu_x)$. We ensure that $\mathbf{B} = \mathbf{B}_1 \cup \mathbf{B}_2$ so that characters in $\mathbf{B}_1$ span all functions that are constant on Σ' (including the All-1 function), and characters in $\mathbf{B}_2$ are zero outside Σ'. The effective degree of a monomial is then the degree when only the characters in $\mathbf{B}_2$ are counted towards the degree. The inductive proof is now carried out assuming that f not only has high degree, but also has high effective degree.

We do mention a crucial detail here. We do need to argue that starting with the original function $f : \Sigma^n \to [-1,1]$ that has essentially high degree, we can "reduce" to the case where it has high effective degree as well. This argument does need that the original functions f, g, h are ℓ_∞-bounded.[8] As noted before, Lemmas 1, 2 could simply be false (for certain distributions μ) if only ℓ_2-norm of the functions is assumed to be 1.

[8]One needs ℓ_∞-boundedness also while transforming the original distribution μ to achieve additional properties.

3 APPLICATIONS

In this section, we give a few applications of our main analytical lemma.

3.1 Hardness of Approximation of CSPs

In this section we use our main analytical lemma to get optimal dictatorship tests with completeness 1 for a large class of 3-ary predicates.

Definition 4. *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 $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \Sigma^n$ be the rows of the matrix. The test accepts if $(f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_k))$ is also a satisfying assignment to P.*

We now describe the dictatorship test that was studied in [5]. The test is given in Figure 1. 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.[9] 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.

Let $P : \Sigma^k \to \{0,1\}$ be the predicate. 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 $P(f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_k)) = 1$.

Figure 1: Dictatorship test for the predicate P.

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 exactly the same as the one given in [5]. If we use our main analytical lemma, Lemma 1, to analyze the above dictatorship test, then we have the following theorem on the soundness of the above test.

Theorem 2 (Restatement of Theorem 1). *Let $P: \Sigma^3 \to \{0,1\}$ be any predicate that satisfies the following conditions. (1) P does not satisfy any linear embedding, and (2) there exists an instance of*

[9]We refer the readers to [5, 20] for detailed information on the semidefinite program, its value and the local distributions.

P-CSP that has a $(1, s)$-integrality gap for the basic SDP relaxation and every local distribution is not linearly embeddable. Then for every $\varepsilon > 0$, there is a dictatorship test for P that has perfect completeness and soundness $s + \varepsilon$.

The proof of this theorem is identical to the proof of [5, Theorem1.1]. The only difference is that in the proof of [5, Theorem1.1], Lemma 1 with the added condition that the distribution μ is semi-rich was used. As the proof is identical to the proof of [5, Theorem1.1], we skip the proof of Theorem 2 in this version.

3.2 Counting Lemmas

Theorem 2. *Suppose μ is a distribution over $\Sigma \times \Gamma \times \Phi$ such that $\mathrm{supp}(\mu)$ cannot be linearly embedded. Then for all $\delta > 0$, there exist $d \in \mathbb{N}$, $\tau > 0$, $\varepsilon > 0$ and $N \in \mathbb{N}$ such that for $n \geqslant N$, if $f\colon \Sigma^n \to [0,1]$, $g\colon \Gamma^n \to [0,1]$, $h\colon \Phi^n \to [0,1]$ are functions with average at least δ and $\max_i(I_i[f^{\leqslant d}], I_i[g^{\leqslant d}], I_i[h^{\leqslant d}]) \leqslant \tau$, then*

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})] \geqslant \varepsilon.$$

Proof. Let $0 \ll \tau \ll d^{-1} \ll \xi \ll \nu \ll \kappa \ll \eta \ll \varepsilon \ll \delta$; first, we argue that

$$\left|\begin{array}{c}\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})]- \\ \mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[\mathrm{T}_{1-\xi}f(\mathbf{x})\mathrm{T}_{1-\xi}g(\mathbf{y})\mathrm{T}_{1-\xi}h(\mathbf{z})\right]\end{array}\right| \leqslant \eta.$$

Here, it is understood that the operator $\mathrm{T}_{1-\xi}$ applied on each one of the functions refers to the standard noise operator with respect to the marginal distribution of μ on that coordinate. This is done by a hybrid argument, wherein we switch at each time a single function to a noisy version of it and bound the difference. For example, we argue that

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[(I-\mathrm{T}_{1-\xi})f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \frac{\eta}{3}.$$

Indeed, note that

$$\begin{aligned}\mathsf{Stab}_{1-\nu}((I-\mathrm{T}_{1-\xi})f) &= \|\mathrm{T}_{1-\nu}(I-\mathrm{T}_{1-\xi})f\|_2^2 \\ &\leqslant \max_j (1-\nu)^j(1-(1-\xi)^j),\end{aligned}$$

as these are the eigenvalues of $\mathrm{T}_{1-\nu}(I-\mathrm{T}_{1-\xi})$. As $\xi \ll \nu$, these eigenvalues smaller than κ, and the bound follows from Lemma 1.

Consider the distribution μ' defined as follows:

(1) Sample $(x, y, z) \sim \mu$;
(2) sample x' by taking $x' = x$ with probability $\sqrt{1-\nu}$ and otherwise resample it according to μ_x;
(3) sample y' by taking $y' = y$ with probability $\sqrt{1-\nu}$ and otherwise resample it according to μ_y;
(4) sample z' by taking $z' = z$ with probability $\sqrt{1-\nu}$ and otherwise resample it according to μ_z;
(5) output (x', y', z').

Note that

$$\begin{aligned}&\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[\mathrm{T}_{1-\xi}f(\mathbf{x})\mathrm{T}_{1-\xi}g(\mathbf{y})\mathrm{T}_{1-\xi}h(\mathbf{z})\right] \\ &= \mathop{\mathbb{E}}_{(\mathbf{x}',\mathbf{y}',\mathbf{z}')\sim\mu'^{\otimes n}}\left[\mathrm{T}_{\sqrt{1-\xi}}f(\mathbf{x}')\mathrm{T}_{\sqrt{1-\xi}}g(\mathbf{y}')\mathrm{T}_{\sqrt{1-\xi}}h(\mathbf{z}')\right].\end{aligned}$$

Also note that the distribution μ' is connected and each atom has probability $\Omega_{\nu,\alpha}(1)$, and also that the individual influences are at most $\tau + (1-\xi)^d$. Hence by [17, Theorem 1.14] it follows that this expectation is at least ε, provided τ is small enough. □

Using regularity lemma for low-degree influences, one may remove the assumption on influences in some cases.

Lemma 3. *For all $\alpha > 0$, $m \in \mathbb{N}$, if μ is a distribution over Σ in which each atom has probability at least α, $|\Sigma| \leqslant m$, then the following holds. For all $\varepsilon > 0$, $d \in \mathbb{N}$ and $\tau > 0$ there exists $D \in \mathbb{N}$ such for every $f\colon \Sigma^n \to [0,1]$, there exists a decision tree $\mathcal{T}$ of depth at most D such that sampling a root to path leaf in it $(\mathbf{I}, \mathbf{x}')$ yields*

$$\Pr_{(\mathbf{I},\mathbf{x}')}\left[I_i^{\leqslant d}[f_{\mathbf{I}\to\mathbf{x}'};\mu] \leqslant \tau \ \forall i \in [n]\setminus\mathbf{I}\right] \geqslant 1-\varepsilon.$$

Proof. We omit the full details of the proof, as it is virtually identical to the proof of Jones' regularity lemma [14] (see also [9] for details). □

Theorem 3 (Restatement of Theorem 1). *Suppose μ is a distribution over Σ^3 such that (1) the three marginal distributions μ_x, μ_y, μ_z are identical, (2) $\{(x,x,x) \mid x \in \Sigma\} \subseteq \mathrm{supp}(\mu)$, and (3) $\mathrm{supp}(\mu)$ cannot be linearly embedded. Then for all $\delta > 0$, there exists $\varepsilon > 0$ and $N \in \mathbb{N}$ such that for $n \geqslant N$ and $S \subseteq \Sigma^n$ with $|S| \geqslant \delta|\Sigma|^n$,*

$$\Pr_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}[\mathbf{x} \in S, \mathbf{y} \in S, \mathbf{z} \in S] \geqslant \varepsilon.$$

Proof. Let $f = 1_S$ and $0 \ll \varepsilon \ll D^{-1} \ll \tau \ll d^{-1} \ll \xi \ll \nu \ll \kappa \ll \eta \ll \delta$. By Lemma 3 we may find a decision tree $\mathcal{T}$ of depth at most $D(d,\tau,\delta)$ such that sampling a path on it according μ_x, i.e. a subset I of at most D variables and $\mathbf{x}' \sim \mu_x^I$, we get that $I_i^{\leqslant d}[f_{I\to\mathbf{x}'}] \leqslant \tau$ for all except with probability $\delta/100$. We denote the process that samples a path on it by $(\mathbf{I}, \mathbf{x}')$.

Note that by an averaging argument, $\mu(f_{I\to x'}) \geqslant \delta/2$ with probability at least $\delta/2$, hence we get that with probability at least $\delta/4$ we have that all influences are small and the average is at least $\delta/2$; we refer to this event by E. If we denote,

$$Z_{(\mathbf{I},\mathbf{x}')} := \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{[n]\setminus\mathbf{I}}}[f_{\mathbf{I}\to\mathbf{x}'}(\mathbf{x})f_{\mathbf{I}\to\mathbf{x}'}(\mathbf{y})f_{\mathbf{I}\to\mathbf{x}'}(\mathbf{z})],$$

then, we get that

$$\begin{aligned}&\Pr_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}[\mathbf{x} \in S, \mathbf{y} \in S, \mathbf{z} \in S] \\ &\geqslant \mathop{\mathbb{E}}_{(\mathbf{I},\mathbf{x}')}\left[1_E \mathbb{E}_{(\mathbf{y},\mathbf{z})\in\mu_{y,z}^{\mathbf{I}}}\left[1_{\mathbf{y}=\mathbf{z}=\mathbf{x}'}\cdot Z_{(\mathbf{I},\mathbf{x}')}\,\middle|\,\mathbf{x}'\right]\right] \\ &\geqslant \frac{\delta}{4}\alpha^D \mathop{\mathbb{E}}_{\substack{(\mathbf{x}',\mathbf{I}) \\ (\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{[n]\setminus\mathbf{I}}}}[f_{\mathbf{I}\to\mathbf{x}'}(\mathbf{x})f_{\mathbf{I}\to\mathbf{x}'}(\mathbf{y})f_{\mathbf{I}\to\mathbf{x}'}(\mathbf{z}) \mid E] \\ &\geqslant \varepsilon,\end{aligned}$$

where the last inequality is by Theorem 2. □

For example, Theorem 1 may be applied to find progressions of the form $(\mathbf{x}, \mathbf{x}+\mathbf{a}, \mathbf{x}+\mathbf{a}^2)$ in dense subsets of $\mathbb{F}_p^n$; we omit the details.

REFERENCES

[1] Michael Bateman and Nets Katz. 2012. New bounds on cap sets. *Journal of the American Mathematical Society* 25, 2 (2012), 585–613. https://doi.org/10.1090/S0894-0347-2011-00725-X

[2] Vitaly Bergelson and Terence Tao. 2014. Multiple recurrence in quasirandom groups. *Geometric and Functional Analysis* 24, 1 (2014), 1–48. https://doi.org/10.1007/s00039-014-0252-0

[3] Amey Bhangale, Prahladh Harsha, and Sourya Roy. 2022. Mixing of 3-Term Progressions in Quasirandom Groups. In *13th Innovations in Theoretical Computer Science Conference (ITCS)*, Vol. 215. Dagstuhl, Germany, 20:1–20:9. https://doi.org/10.4230/LIPIcs.ITCS.2022.20

[4] Amey Bhangale and Subhash Khot. 2021. Optimal Inapproximability of Satisfiable K-LIN over Non-Abelian Groups. In *Proceedings of the 53rd annual Symposium on Theory of Computing (STOC)* (Virtual, Italy). ACM, New York, NY, USA, 1615–1628. https://doi.org/10.1145/3406325.3451003

[5] Amey Bhangale, Subhash Khot, and Dor Minzer. 2022. On Approximability of Satisfiable K-CSPs: I. In *Proceedings of the 54th annual Symposium on Theory of Computing (STOC)* (Rome, Italy). ACM, New York, NY, USA, 976–988. https://doi.org/10.1145/3519935.3520028

[6] Mark Braverman, Subhash Khot, Noam Lifshitz, and Dor Minzer. 2021. An Invariance Principle for the Multi-slice, with Applications. In *Proceedings of the 62nd annual symposium on Foundations of Computer Science (FOCS)*. IEEE, 228–236. https://doi.org/10.1109/FOCS52979.2021.00030

[7] Mark Braverman, Subhash Khot, and Dor Minzer. 2021. On Rich 2-to-1 Games. In *12th Innovations in Theoretical Computer Science Conference (ITCS)*, Vol. 185. Dagstuhl, Germany, 27:1–27:20. https://doi.org/10.4230/LIPIcs.ITCS.2021.27

[8] T.C Brown and J.P Buhler. 1982. A density version of a geometric ramsey theorem. *Journal of Combinatorial Theory, Series A* 32, 1 (1982), 20–34. https://doi.org/10.1016/0097-3165(82)90062-0

[9] Gilad Chase, Yuval Filmus, Dor Minzer, Elchanan Mossel, and Nitin Saurabh. 2022. Approximate Polymorphisms. In *Proceedings of the 54th annual Symposium on Theory of Computing (STOC)* (Rome, Italy). ACM, New York, NY, USA, 195–202. https://doi.org/10.1145/3519935.3519966

[10] Ernie Croot, Vsevolod F. Lev, and Péter Pál Pach. 2017. Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small. *Annals of Mathematics* 185, 1 (2017), 331–337. https://doi.org/10.4007/annals.2017.185.1.7

[11] Jordan S Ellenberg and Dion Gijswijt. 2017. On large subsets of with no three-term arithmetic progression. *Annals of Mathematics* (2017), 339–343. https://doi.org/10.4007/annals.2017.185.1.8

[12] W.T. Gowers. 2008. Quasirandom Groups. *Combinatorics, Probability and Computing* 17, 3 (2008), 363–387. https://doi.org/10.1017/S0963548307008826

[13] Jan Hązła, Thomas Holenstein, and Elchanan Mossel. 2018. Product Space Models of Correlation: Between Noise Stability and Additive Combinatorics. *Discrete Analysis* (2018), 6513. https://doi.org/10.19086/da.6513

[14] Chris Jones. 2016. A Noisy-Influence Regularity Lemma for Boolean Functions. *CoRR* abs/1610.06950 (2016). arXiv:1610.06950 http://arxiv.org/abs/1610.06950

[15] Subhash Khot. 2002. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th annual Symposium on Theory of Computing (STOC)*. ACM, 767–775. https://doi.org/10.1145/509907.510017

[16] Roy Meshulam. 1995. On subsets of finite abelian groups with no 3-term arithmetic progressions. *Journal of Combinatorial Theory, Series A* 71, 1 (1995), 168–172. https://doi.org/10.1016/0097-3165(95)90024-1

[17] Elchanan Mossel. 2010. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis* 19, 6 (2010), 1713–1756. https://doi.org/10.1007/s00039-010-0047-x

[18] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. 2005. Noise stability of functions with low influences: invariance and optimality. In *Proceedings of the 46th annual symposium on Foundations of Computer Science (FOCS)*. IEEE, 21–30. https://doi.org/10.1109/SFCS.2005.53

[19] Sarah Peluse. 2018. Mixing for three-term progressions in finite simple groups. *Mathematical Proceedings of the Cambridge Philosophical Society* 165, 2 (2018), 279–286. https://doi.org/10.1017/S0305004117000482

[20] Prasad Raghavendra. 2008. Optimal algorithms and inapproximability results for every CSP?. In *Proceedings of the 14th annual Symposium on Theory of Computing (STOC)*. 245–254. https://doi.org/10.1145/1374376.1374414

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

[22] K. F. Roth. 1953. On Certain Sets of Integers. *Journal of the London Mathematical Society* s1-28, 1 (1953), 104–109. https://doi.org/10.1112/jlms/s1-28.1.104

[23] Endre Szemerédi. 1975. On sets of integers containing k elements in arithmetic progression. *Acta Arithmetica* 27, 1 (1975), 199–245.

[24] Terence Tao. 2013. Mixing for progressions in nonabelian groups. In *Forum of Mathematics, Sigma*, Vol. 1. Cambridge University Press, e2. https://doi.org/10.1017/fms.2013.2

Received 2022-11-07; accepted 2023-02-06