Combinatorics, Probability and Computing (2025), 1–70
doi:10.1017/S0963548325100114

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

On approximability of satisfiable *k*-CSPs: II

Amey Bhangale[1], Subhash Khot[2] and Dor Minzer[3]

[1]Department of Computer Science and Engineering, University of California, Irvine, CA, USA, [2]Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA, and [3]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA
Corresponding author: Dor Minzer; Email: minzer.dor@gmail.com

(Received 25 August 2023; revised 27 May 2024; accepted 27 May 2024)

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 $\mathrm{supp}(\mu)$). 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 $(a_1, a_2, \ldots, a_k) \in \mathrm{supp}(\mu)$, $\sum_{i=1}^{k} \sigma_i(a_i) = 0_G$. In [Bhangale-Khot-Minzer, STOC 2022], the authors asked the following analytical question. 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 at most δ. If μ has no linear embedding (over any Abelian group), then is it necessarily the case that

$$\left| \mathop{\mathbb{E}}_{(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) \sim \mu^{\otimes n}} [f_1(\mathbf{x}_1) f_2(\mathbf{x}_2) \cdots f_k(\mathbf{x}_k)] \right| = o_{d,\delta}(1),$$

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.

1. The first application is related to hardness of approximation. Using the reduction from [5], we show that for every 3-ary predicate $P:\Sigma^3 \to \{0, 1\}$ such that P has no linear embedding, an *SDP (semi-definite programming) integrality gap instance* of a P-Constraint Satisfaction Problem (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.
2. 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 $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ from $\mu^{\otimes n}$ (and in fact a significant density of such triples).

Keywords: Abelian embeddings; Analysis of Boolean functions; Hardness of approximation

2020 MSC Codes: 06E30

1. Introduction

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

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

where each one of the n coordinates 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, that is, $\|f_i\|_\infty \leqslant 1$. 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 (3-AP) $(x, x+y, x+2y)$ in $\mathbb{Z}_p$. Then the quantity $\mathbb{E}_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\sim\mu^{\otimes n}}\left[f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)f_3(\mathbf{x}_3)\right]$, up to a normalisation factor, precisely counts the number of AP$(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ 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, $\mathsf{Stab}_\rho(f)$, is defined as $\mathbb{E}\left[f(\mathbf{x})f(\mathbf{y})\right]$ where for each $i \in [n]$, $\mathbf{x}_i$, and $\mathbf{y}_i$ are uniformly distributed, and $\mathbb{E}\left[\mathbf{x}_i\mathbf{y}_i\right] = \rho$. 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 $\mathsf{Stab}_\rho(f)$ 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
$$\mathbb{P}_{(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_k)\sim\mu^{\otimes n}}\left[(f(\mathbf{x}_1), f(\mathbf{x}_2), \cdots f(\mathbf{x}_k)) \in P^{-1}(1)\right].$$
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 $c = \mathbb{P}_{(a_1,a_2,\ldots,a_k)\sim\mu}\left[(a_1, a_2, \ldots, a_k) \in P^{-1}(1)\right]$. Observe that the test accepts any Dictatorship function, namely $f(\mathbf{x}) = \mathbf{x}_{i_0}$ 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 $\mathsf{supp}(\mu) \subseteq P^{-1}(1)$. In fact, we will generally assume that μ has full support, that is, $\mathsf{supp}(\mu) = P^{-1}(1)$ 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 analyse 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 [23] estimates the desired density of AP; 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
$$\left|\mathop{\mathbb{E}}_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\sim\mu^{\otimes n}}\left[f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)f_3(\mathbf{x}_3)\right]\right| \leqslant \min_{1\leqslant i\leqslant 3} \|\hat{f}_i\|_\infty.$$

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

[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 $-1 < \rho < 0$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

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, that is, the high-degree, part of the functions. Specifically, one is naturally led to the following analytic question.

Question 1.1. (Informal) *Find the necessary and sufficient condition on the distribution μ on Σ^k, such that*

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

where the functions are complex valued, 1-bounded 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 $(a_1, a_2, \ldots, a_k) \in \mathsf{supp}(\mu)$ and $(a'_1, a'_2, \ldots, a'_k) \in \mathsf{supp}(\mu)$, 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 $\mathsf{supp}(\mu)$.

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 1.2. *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 $(a_1, a_2, \ldots, a_k) \in \mathsf{supp}(\mu)$, $\sum_{i=1}^k \sigma_i(a_i) = 0_G$.*

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 $f_i(\mathbf{x}_i) = \prod_{j=1}^n \chi(\sigma_i((\mathbf{x}_i)_j))$.[2] Now, for all $(\mathbf{x}_1, \ldots, \mathbf{x}_k) \in \mathsf{supp}(\mu^{\otimes n})$ we have

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

[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.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

$$= \prod_{j=1}^{n} \chi\left(\sum_{i=1}^{k} \sigma_i((\mathbf{x}_i)_j)\right)$$
$$= \prod_{j=1}^{n} \chi(0_G)$$
$$= 1.$$

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 $((\mathbf{x}_1)_j, \ldots, (\mathbf{x}_k)_j) \in \mathsf{supp}(\mu)$ 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.[3]

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

Hypothesis 1.3. (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 sub-class 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 $\mathsf{supp}(\mu) = P^{-1}(1)$ 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 $G, \sigma\colon \Sigma \to G, \gamma\colon \Gamma \to G, \phi\colon \Phi \to G$, not all constant, such that $\sigma(x) + \gamma(y) + \phi(z) = 0_G$ for all $(x, y, z) \in \mathsf{supp}(\mu)$. 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 $\mathsf{supp}(\mu)$. We always treat μ as fixed and m, α as fixed constants.

Lemma 1.4 (Main Analytical Lemma, Informal Version). *For $k = 3$, the necessary and sufficient condition for (2) to hold for all 1-bounded f_1, f_2, f_3 such that at least one of them essentially has degree at least d, is that μ has no linear embedding over any Abelian group.*

One may wonder when a function h is 1-bounded as well as essentially of high degree. A natural example is when $h'\colon \Phi^n \to \mathbb{C}$ is an arbitrary 1-bounded function and $h = h' - T_{1-\xi}h'$, where $T_{1-\xi}$ is the standard Beckner (noise) operator. 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.[4] 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.

[3]This follows from the fact that weight of f_i up to level d is at most the noise stability of f_i with noise rate $1/d$. Since f_i is a product of functions each depending on a different coordinate, its noise stability is the n-th power of the noise stability of $\chi \circ \sigma_i$, which is equal to $(1 - \Omega(1/d))^n = o(1)$ as the noise stability of $\chi \circ \sigma_i$ with noise rate $1/d$ is $1 - \Omega(1/d)$.

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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

The advantage of using soft-truncation is that it preserves boundedness of functions whereas the hard-truncation in general does not.

1.1 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 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 [22] 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).

It thus remains to fulfil 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 $s + \varepsilon$. Here and throughout, an integrality gap for a CSP is an instance of it whose integral value (i.e. the maximum fraction of constraints that can be satisfied) is at most s, wherein the value of the SDP relaxation of the instance is 1. The result of [5] 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 2.1 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 integrality gap to dictatorship test implication for all such predicates. The formal statement of our result appears below (one wishes that the condition (2*b*) therein could be dropped; if so, we would have a full-proof implication). For definitions and a more detailed discussion, we refer to Section 9 and the introductory section of [5].

Theorem 1.5. *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 $s + \varepsilon$.*

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Counting patterns: In additive combinatorics, finding a certain fixed 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 $(G, +)$, one often studied the pattern of AP. A subset $A \subseteq G$ is said to be 3-AP free if there is no AP of size 3 in A. In other words, there are no elements $x, y, z \in A$ such that $x + z = 2y$. Roth's Theorem [23] 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 [24] generalised 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. In the finite field 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] observed that one may apply the methods from a beautiful work by Croot, Lev, and Pach [10] to obtain a substantial quantitative improvement over Roth's Theorem (applied to $\mathbb{F}_3^n$).

Suppose now that $(G, \cdot)$ is a finite group that is not necessarily Abelian. In this setting there are many more patterns that have been studied. 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 (these usually go by the name sum-free sets in the Abelian setting). Gowers [12] showed that this is not true for a class of non-Abelian groups called *quasirandom groups*.[5] That is, every constant density subset of a quasirandom group contains the pattern (x, y, xy). Tao [25] extended Gowers' result to the patterns (x, xg, xg^2) and (x, xg, xg^2, xg^3) for some very specific quasirandom groups. Bergelson and Tao [2] established it for the patterns (x, xg, gx) and (g, x, xg, gx) for every quasirandom group. Recently, following the work by Peluse [20], Bhangale, Harsha, and Roy [3] established it for the pattern (x, xg, xg^2) for every quasirandom group.

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

Theorem 1.6. *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 $|S| \geqslant \delta|\Sigma|^n$,*

$$\mathbb{P}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})\sim\mu^{\otimes n}}\left[\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in S\right] \geqslant \varepsilon.$$

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 Hazła, 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 1.6 extends their result.

2. Our techniques

In this section, we elaborate on the ideas involved in proving Lemma 1.4, and we state a more formal version below. We focus only on a few high-level ideas here, and thus skip many technical (and even conceptual) details. This leads to some discrepancies between the high-level exposition here and formal proofs appearing later.

[5] 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.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Lemma 2.1 (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* $(x, y, z) \in \mathsf{supp}(\mu)$.
- *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* $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$, $h\colon \Phi^n \to \mathbb{C}$ *are* 1*-bounded functions and* $\mathsf{Stab}_{1-\xi}(h; \mu_z) \leqslant \delta$, *then we have that*

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

We clarify the condition that $\mathsf{Stab}_{1-\xi}(h) \leqslant \delta$. Note that we have dropped μ_z from the notation for convenience. The parameter $\mathsf{Stab}_{1-\xi}(h)$ 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 Section 3 for all analytic definitions and basic tools. The condition that $\mathsf{Stab}_{1-\xi}(h) \leqslant \delta$ serves as a proxy for the condition that the function h is essentially of high degree. Indeed, if $\mathsf{Stab}_{1-\xi}(h) \leqslant \delta$, 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 $\mathsf{Stab}_{1-\xi}(h) \leqslant \delta$. Hence the low-stability condition is a proxy for the high-degree condition and turns out to be more convenient to work with.

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

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

where $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$, $h\colon \Phi^n \to \mathbb{C}$, are 1-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, that is, 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 $\mathsf{supp}(\mu)$ is not linearly embeddable. It is easily seen that the non-linearity condition, in this special 2-ary case, is same as saying that $\mathsf{supp}(\mu)$, 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 $\sigma\colon C_j \to j$, $\gamma\colon D_j \to -j$ is an embedding of Σ and Γ, respectively, into $\mathbb{Z}_r$ and for all $(x, y) \in \mathsf{supp}(\mu)$ (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 $f\colon \Sigma^n \to \mathbb{C}, g\colon \Gamma^n \to \mathbb{C}$ are n-dimensional ℓ_∞-bounded functions where g has high degree, then $\left| \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mu^{\otimes n}} \left[f(\mathbf{x}) g(\mathbf{y}) \right] \right|$ is small. For simplicity of exposition, we assume that g in fact has full degree n.[6] 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 emphasise 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

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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

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 $f:\Sigma \to \mathbb{C}, g:\Gamma \to \mathbb{C}$ 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 $\mathbb{E}\left[f\right] = \mathbb{E}\left[g\right] = 0$, we in fact have the improvement

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

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 $\mathbb{E}\left[g\right] = 0$ (and similarly for $\tilde{L}_2(\Sigma;\mu_x)$). The operator norm of T, denoted $\|T\| = \max_{g:\mathbb{E}\left[g\right]=0} \|Tg\|_2/\|g\|_2$, is at most $1-\tau$ according to the equivalent interpretation of the inequality (4), which can then be derived as:

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

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

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

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.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

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 $\mathsf{supp}(\mu)$ 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 $f:\Sigma \to \mathbb{C}, g:\Gamma \to \mathbb{C}, h:\Phi \to \mathbb{C}$, while we desire a base case inequality (say when $\mathbb{E}\left[f\right] = 0$) of the form

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

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 $\|f\|_2\|g\|_2\|h\|_2$ 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, $\mathsf{supp}(\mu) = \Sigma^3$ 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 $\mathbb{E}\left[f(x)\right] = 0$ 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, $\mathbb{E}\left[f(x)\right] = 0, \mathbb{E}\left[f(x)^2\right] = 6m$, 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

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

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! Indeed, consider the same example here 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

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

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 2.1 (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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

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 $f:\Sigma \to \mathbb{C}, g:\Gamma \to \mathbb{C}, h:\Phi \to \mathbb{C}$, we have

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

Indeed, by Cauchy-Schwarz,

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

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 $(x, y, z) \in \mathsf{supp}(\mu)$ (we then say that (y, z) determine x). The details of this transformation and related proofs appear in Section 8; some of its ingredients 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 2.2. (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 $f:\Sigma \to \mathbb{C}, g:\Gamma \to \mathbb{C}, h:\Phi \to \mathbb{C}$,*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu} \left[f(x)g(y)h(z) \right] \right| \leqslant (1 - \tau(\theta))\|f\|_2 \|g\|_2 \|h\|_2, \qquad |\mathbb{E}[f]| \leqslant (1-\theta)\|f\|_2. \tag{7}$$

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 2.3. *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 $(x, y, z) \in \mathsf{supp}(\mu)$, 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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

understanding how a Horn-SAT embedding is an obstruction towards the desired inequality (7) and how it is the only possible obstruction.

- First, we note that having a Horn-SAT embedding violates inequality (7). Indeed, since $f(x) = g(y)h(z)$ in $\mathsf{supp}(\mu)$ and (y, z) are uniform and independent, we have $\|f\|_2 = \|g\|_2\|h\|_2$ and then
$$\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}\left[f(x)g(y)h(z)\right] = \mathop{\mathbb{E}}_{(y,z)\sim\mu_{y,z}}\left[g(y)^2h(z)^2\right] = \|g\|_2^2\|h\|_2^2 = \|f\|_2\|g\|_2\|h\|_2.$$
One also notes that since f is Boolean and non-constant, it does have constant variance.
- Secondly, we note that if the inequality (7) 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 functions $f:\Sigma \to \mathbb{C}$, $g:\Gamma \to \mathbb{C}$, $h:\Phi \to \mathbb{C}$, such that
$$\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}\left[f(x)g(y)h(z)\right] = \|f\|_2\|g\|_2\|h\|_2,$$
that is, achieving an exact equality. This means that the application of Cauchy-Schwarz in Equation (6) must be tight and therefore $f(x) = s\overline{g(y)h(z)}$ in $\mathsf{supp}(\mu)$ (as equality of complex numbers), where $s \in \mathbb{C}$ is a complex number of absolute value 1. If $f(x)$ is always non-zero, then so are $g(y)$ and $h(z)$. In this case, one can choose a branch of the logarithm function and get an embedding into addition modulo $2\pi i$, which is an Abelian group.[7] 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 $(x, z) \in \mathsf{supp}(\mu_{x,z})$. Since μ is not linearly embeddable, its marginals are not linearly embeddable either.[8] 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 (7) and we can hope to carry out the induction. However, if μ does have a Horn-SAT embedding as in Definition 2.3, 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

[7] Strictly speaking, the embedding is not into a finite Abelian group, but this is not difficult to fix.

[8] This is seen easily from the definition of a linear embedding, Definition 1.2. 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.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

- If Lemma 2.1 (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.

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 2.4. *(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 $f\colon \Sigma \to \mathbb{C}$, $g\colon \Gamma \to \mathbb{C}$ and $h\colon \Phi \to \mathbb{C}$ be such that f has variance at least $\tau \|f\|_2^2$ on Σ', that is*

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

  Then

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

- *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 Section 8.3 for details. To summarise, we reduce the task of proving our Main Lemma 2.1 to the same task with the additional property that μ satisfies the relaxed base case inequality, that is, 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 2.5. *(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 $(x, y, z) \in \mathsf{supp}(\mu)$.*
2. *$\mathsf{supp}(\mu)$ 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 $(x, y, z) \in \mathsf{supp}(\mu)$ (i.e. y, z determine x).*
5. *μ satisfies the relaxed base case inequality as in Definition 2.4.*

Then for all $\varepsilon > 0$, there are $\xi, \delta > 0$ such that the following holds. If $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$ and $h\colon \Phi^n \to \mathbb{C}$ are 1-bounded functions satisfying that either $\mathsf{Stab}_{1-\xi}(g) \leqslant \delta$ or $\mathsf{Stab}_{1-\xi}(h) \leqslant \delta$, then we have that

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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

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 2.5). 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 (7). We will indicate how to incorporate the relaxed base case inequality later. Formal proofs appear in Sections 4, 5, 6, and 7.

So let us focus on this special case and assume the base case inequality (7) holds. The inductive proof proceeds in several steps. We emphasise 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.[9] We are given that either g or h has essentially high degree, so let's say this holds for g, formalised 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 homogeneous functions. We will skip details regarding how this is sufficient towards the general case. Therefore let's assume that f, g, h are homogeneous 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 $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$, $h\colon \Phi^n \to \mathbb{C}$ homogeneous 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

[9] 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.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

decompositions of f, g, and h. For the sake of simplicity, we consider only a specialised scenario that allows us to write

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

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 emphasise here that this is an inequality on functions of a single co-ordinate. It is referred to as the additive base case inequality (see Lemma 3.18). 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 specialised scenario that allows us to write f, g, and h as

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

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 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

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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

By the inductive hypothesis, the first term is at most $\beta_{n-1,d_1-1,d_2-1,d_3-1}$ and by the base case inequality, $|\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu}\left[f_1'(\mathbf{x})g_1'(\mathbf{y})h_1'(\mathbf{z})\right]| \leqslant \lambda = 1 - \Omega(1)$. 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 specialised 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 (7) does not hold and we are able to use only the relaxed base case inequality in Definition 2.4. 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 2.4. 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

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

That is, the influence is the variance of the function on the i^{th} coordinate after randomly restricting the rest of the coordinates. We define the effective influence as

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

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 1-bounded function $f:\Sigma^n \to \mathbb{C}$ 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.[10] As noted before, Lemmas 2.1, 2.5 could simply be false (for certain distributions μ) if only ℓ_2-norm of the functions is assumed to be 1.

2.6 Organisation

The rest of the paper is organised as follows. We start with preliminaries in Section 3. We set up the necessary machinery in Sections 4 and 5 that are needed to formulate the inductive statement towards proving the main analytical lemma under relaxed base case inequality. The proof of this lemma, which is divided into two parts, spans Sections 6 and 7. Finally, in Section 8 we derive our main analytical lemma, Lemma 2.1, from Lemma 2.5. In this section, we show how to get around the issue of the Horn-SAT embedding.

Section 9 is devoted to proving applications of our main analytical lemma.

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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

3. Preliminaries

In this section, we record some basic definitions and tools from analysis of Boolean functions that will be used throughout the paper (see O'Donnell's book [19] for reference). We begin with some notations.

We denote $A \lesssim B$ to refer to the fact that $A \leqslant C \cdot B$ for some absolute constant $C > 0$; we denote $A \gtrsim B$ to refer to the fact that $A \geqslant c \cdot B$ for some absolute constant $c > 0$. If this constant depends on some parameter, say m, we denote this fact by $A \lesssim_m B$. We use the normal letters x, y, z to denote elements from the domain Σ, Γ, Φ, respectively, and the bold face letters $\mathbf{x}, \mathbf{y}, \mathbf{z}$ to denote strings from $\Sigma^n, \Gamma^n, \Phi^n$, respectively. We say a function $f\colon \Sigma^n \to \mathbb{C}$ is C-bounded if $\left|f(x)\right| \leqslant C$ for all $x \in \Sigma^n$.

3.1 Degrees and homogeneous functions

We start with the definitions of a monomial and a degree-d monimial.

Definition 3.1. *Let Γ be a finite set, and let ν be some probability measure over Γ. A monomial over Γ is a function $\chi : (\Gamma, \nu) \to \mathbb{C}$ whose expectation according to ν is* 0.

Definition 3.2. *Let Γ be a finite set, $n \geqslant 1$, and let ν be some probability measure over Γ. A function $\chi : (\Gamma^n, \nu^{\otimes n}) \to \mathbb{C}$ is a degree d monomial if there are distinct indices $i_1, \ldots, i_d$ and monomials $\chi_{i_1}, \ldots, \chi_{i_d} : \Gamma \to \mathbb{C}$ with respect to ν, such that*

$$\chi(\mathbf{y}) = \prod_{j=1}^{d} \chi_{i_j}(y_{i_j}).$$

Based on these definitions, we now define homogeneous functions of degree d.

Definition 3.3. *Let Γ be a finite set, $n \geqslant 1$, and let ν be some probability measure over Γ. A function $g\colon (\Gamma^n, \nu^{\otimes n}) \to \mathbb{C}$ is a homogeneous degree d function if it can be written as a linear combination of monomials of degree d with respect to ν.*

3.2 Efron–Stein decomposition

For a product space $(\Gamma^n, \nu^{\otimes})$, we will use the standard Efron-Stein decomposition. Given a function $g\colon (\Gamma^n, \nu^{\otimes n}) \to \mathbb{C}$, one may write (in a unique manner)

$$g(\mathbf{y}) = \sum_{i=0}^{n} g^{=i}(\mathbf{y}),$$

where $g^{=i}$ is a homogeneous function of degree i. We denote by $V^{=i}(\Gamma^n, \nu^{\otimes n})$ the space of homogeneous functions of degree i, and often omit the domain and the measure if these are clear from the context. Hence $g^{=i} \in V^{=i}$. The Efron-Stein decomposition is a refinement of the above. For each $i = 0, \ldots, n$, one may write

$$g^{=i}(\mathbf{y}) = \sum_{S \subseteq [n], |S| = i} g^{=S}(\mathbf{y}),$$

where $g^{=S}(\mathbf{y})$ is a homogeneous function of degree i whose value on $\mathbf{y}$ only depends on $\mathbf{y}_S$, namely the co-ordinates in the set S. We denote by $V^{=S}(\Gamma^n, \nu^{\otimes n})$ the space of degree $|S|$ homogeneous functions depending only on coordinates in S. So, $g^{=S} \in V^{=S}$. Furthermore, this decomposition is

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

unique, and satisfies that $g^{=S}$ and $g^{=T}$ are orthogonal for any $S \neq T$, that is,

$$\left\langle g^{=S}, g^{=T} \right\rangle = \mathop{\mathbb{E}}_{\mathbf{y}\sim\nu^{\otimes n}} \left[g^{=S}(\mathbf{y}) \overline{g^{=T}(\mathbf{y})} \right] = 0.$$

We define

$$g^{\supseteq T}(\mathbf{y}) = \sum_{S \supseteq T} g^{=S}(\mathbf{y}), \quad g^{\leqslant d}(\mathbf{y}) = \sum_{i=0}^{d} g^{=i}(\mathbf{y}), \quad g^{>d}(\mathbf{y}) = \sum_{i=d+1}^{n} g^{=i}(\mathbf{y}).$$

Next, we define the standard notion of the influence of a variable and the total influence of a function as follows:

Definition 3.4. *The influence of a variable i on $g\colon (\Gamma^n, \mu^{\otimes n}) \to \mathbb{C}$ is defined as*

$$I_i[g] = \mathop{\mathbb{E}}_{\mathbf{y}_{-i}\sim\mu^{n-1},\, a,b\sim\mu} \left[\left| g(\mathbf{y}_{-i}, y_i = a) - g(\mathbf{y}_{-i}, y_i = b) \right|^2 \right].$$

Definition 3.5. *The total influence of g is defined as $I[g] = \sum_{i=1}^{n} I_i[g]$.*

The following fact relates the total influence with the Efron-Stein decomposition of a function.

Fact 3.6. $I[g] = 2 \sum_{S\subseteq[n]} |S| \, \|g^{=S}\|_2^2 = 2 \sum_{i=1}^{n} i \, \|g^{=i}\|_2^2$.

Definition 3.7. *The variance of $g\colon (\Gamma^n, \mu^{\otimes n}) \to \mathbb{C}$ is defined as*

$$\mathsf{var}(g) = \frac{1}{2} \mathop{\mathbb{E}}_{\mathbf{y},\mathbf{y}'\sim\mu^{\otimes n}} \left[|g(\mathbf{y}) - g(\mathbf{y}')|^2 \right] = \mathop{\mathbb{E}}_{\mathbf{y}} \left[\left| g(\mathbf{y}) \right|^2 \right] - \left| \mathop{\mathbb{E}}_{\mathbf{y}} \left[g(\mathbf{y}) \right] \right|^2.$$

3.3 The noise operator and stability

For a parameter $\rho \in [0, 1]$, a measure μ over Σ and a point $x \in \Sigma$ we define the distribution over points that are ρ-correlated with x as: take $y = x$ with probability ρ, and otherwise sample $y \sim \mu$. We denote the distribution over ρ-correlated points with x as $y \sim \mathrm{T}_\rho x$. Tensorizing, for $n \in \mathbb{N}$ and $x \in \Sigma^n$, the distribution over ρ-correlated points with x is denoted by $\mathrm{T}_\rho^{\otimes n} x$, and is sampled by taking $y_i \sim \mathrm{T}_\rho x_i$ for each i independently.

We may think of the operator $\mathrm{T}_\rho^{\otimes n}$ as acting on $L_2(\Sigma^n; \mu^{\otimes n})$ by mapping a function f to the function $\mathrm{T}_\rho^{\otimes n} f$ defined as

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

Definition 3.8. *The noise stability of $f\colon (\Sigma^n, \mu^{\otimes n}) \to \mathbb{C}$ with respect to correlation parameter ρ is defined as* $\mathsf{Stab}_\rho(f) = \langle f, \mathrm{T}_\rho f \rangle$.

We remark that a straightforward computation shows that the spaces $V^{=i}$ are eigenspaces of the operator $\mathrm{T}_\rho^{\otimes n}$ with eigenvalue ρ^i, and in particular one gets the Fourier analytic formula $\mathsf{Stab}_\rho(f) = \sum_S \rho^{|S|} \|f^{=S}\|_2^2$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

3.4 A markov chain lemma

We mention below a bound on the second largest eigenvalue of a Markov chain. We include the proof since we could not find this specific bound in literature (we need dependence on ξ to be linear and not quadratic).

Lemma 3.9. *Let Σ be a finite alphabet of size at most m and $\mu(x,y)$ be a symmetric, connected distribution over $\Sigma \times \Sigma$. Suppose that $\mu(x) := \sum_{y\in\Sigma} \mu(x,y) \geqslant \alpha$ for all $x \in \Sigma$ and whenever $\mu(x,y) > 0$, we have $\mu(x,y) \geqslant \xi$. Let T be the associated Markov chain on Σ.*[11] *Then the second largest eigenvalue $\lambda_2(T) \leqslant 1 - \Omega_{\alpha,m}(\xi)$.*

Proof. Note that μ is the stationary distribution of T. Let $f\colon \Sigma \to \mathbb{R}$ be the eigenvector of T corresponding to $\lambda_2(\mathrm{T})$. We have,

$$\mathop{\mathbb{E}}_{\mathbf{x}\sim\mu}\left[f(\mathbf{x})\right] = 0, \quad \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu}\left[f(\mathbf{x})^2\right] = 1, \quad \lambda_2(\mathrm{T}) = \langle f, \mathrm{T}f\rangle.$$

It follows that there is $x \in \Sigma$ such that $\left|f(x)\right| \geqslant 1$, and without loss of generality $f(x) \geqslant 1$. Since $\mathbb{E}\left[f\right] = 0$, there is y such that $f(y) \leqslant 0$. Since T is connected, there is a path $x = x^0 \to x^1 \to \ldots \to x^\ell = y$, where $\ell \leqslant |\Sigma| \leqslant m$, so that $\mu(x^i, x^{i+1}) > 0$ for all $i \leqslant \ell - 1$. We note that

$$1 \leqslant \left|f(x) - f(y)\right| \leqslant \sum_{i=0}^{\ell-1} \left|f(x^{i+1}) - f(x^i)\right|,$$

so it follows that there is i such that $\left|f(x^{i+1}) - f(x^i)\right| \geqslant \frac{1}{\ell} \geqslant \frac{1}{m}$. Therefore,

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})\sim\mu}\left[(f(\mathbf{x}) - f(\mathbf{y}))^2\right] \geqslant \mu(x,y)\,\frac{1}{m^2} \geqslant \frac{\xi}{m^2}.$$

On the other hand, the left hand side is $2\|f\|_2^2 - 2\langle f, \mathrm{T}f\rangle$, so we get that $\lambda_2(\mathrm{T}) = \langle f, \mathrm{T}f\rangle \leqslant 1 - \frac{\xi}{2m^2}$, and the proof is concluded. □

Lemma 3.10. *In the setting of Lemma 3.9, if $f\colon (\Sigma^n, \mu^{\otimes n}) \to \mathbb{C}$ is any function, then*

$$\|T^{\otimes n} f^{=S}\|_2 \leqslant \left(1 - \Omega_{\alpha,m}(\xi)\right)^{|S|} \|f^{=S}\|_2.$$

Proof. Assume without loss of generality that $S = \{1, \ldots, t\}$.

$$\|\mathrm{T}^{\otimes n} f^{=S}\|_2 = \|\mathrm{T}_1((\mathrm{T}_2 \circ \ldots \circ \mathrm{T}_t) f^{=S})\|_2 \leqslant (1 - \Omega_{\alpha,m}(\xi))\|(\mathrm{T}_2 \circ \ldots \circ \mathrm{T}_t) f^{=S}\|_2,$$

where we used Lemma 3.9 and the fact that $(\mathrm{T}_2 \circ \ldots \circ \mathrm{T}_t) f^{=S}$ has expectation 0 over x_1 for any setting of the remaining co-ordinates. The proof is concluded by iterating the above inequality over $x_2, \ldots, x_t$. □

3.5 Effective degrees and effective influences

Fix a distribution μ on $\Sigma \times \Gamma \times \Phi$ as in Lemma 2.5, and fix $\Sigma' \subseteq \Sigma$ to be the subset evidencing the fact that it satisfies the relaxed base case, Definition 2.4. We may choose an orthonormal basis of $L_2(\Sigma;\mu_x)$ as $B = B_1 \cup B_2$ where B_1 consists of functions that are constant on Σ', and B_2 consists of functions only supported on Σ' and orthogonal to B_1. Thus, a basis for $L_2(\Sigma^n; \mu_x^n)$ is given by

$$B^{\otimes n} = \left\{ \chi(\mathbf{x}) = \prod_i \chi_i(x_i) \;\middle|\; \chi_i \in B_1 \cup B_2 \;\forall i \in [n] \right\},$$

[11] Alternately, T is the adjacency matrix of a random walk on Σ where one leaves a vertex x with the edge probabilities $\mu(x,y)/\mu(x)$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

and a given function $f\colon \Sigma^n \to \mathbb{C}$ can be uniquely written as

$$f(\mathbf{x}) = \sum_{\chi \in B^{\otimes n}} \widehat{f}(\chi)\, \chi(\mathbf{x}), \quad \text{where } \widehat{f}(\chi) = \langle f, \chi \rangle.$$

We now define the effective degree of a character χ, which is the number of components from $\{\chi_i\}$ that are not constant on Σ'.

Definition 3.11. *Given $\chi \in B^{\otimes n}$, we define the effective degree of χ as*

$$\mathsf{effdeg}(\chi) = |\{i \in [n] \mid \chi_i \in B_2\}|.$$

To be compatible with the standard notion of degree, we will introduce a special notation for the trivial character in B_1 which is the constant 1 function on Σ, and we denote it by χ_{const}. Thus, we note that degree of a character χ is $|\{i \mid \chi_i \neq \chi_{\mathsf{const}}\}|$, and as χ_{const} is in B_1, one has that $\mathsf{effdeg}(\chi) \leqslant \mathsf{deg}(\chi)$ for all $\chi \in B^{\otimes n}$.

We also define the effective influence of a variable analogous to Definition 3.4, except we only resample the assignment to the variable if it belongs to Σ'.

Definition 3.12. *For a function $f\colon \Sigma^n \to \mathbb{C}$ and $i \in [n]$, we define*

$$I_{i,\,\mathsf{effective}}[f] = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y})}\left[\left|f(\mathbf{x}) - f(\mathbf{y})\right|^2\right],$$

where we sample $\mathbf{x} \sim \mu_x^{\otimes n}$, take $y_j = x_j$ for all $j \neq i$, and for the i^{th} co-ordinate, if $x_i \in \Sigma \setminus \Sigma'$ we take $y_i = x_i$, and if $x_i \in \Sigma'$, we sample $y_i \in \Sigma'$ independently.

Definition 3.13. *For a function $f\colon \Sigma^n \to \mathbb{C}$, the total effective influence is defined as*

$$I_{\mathsf{effective}}[f] = \sum_{i=1}^{n} I_{i,\,\mathsf{effective}}[f].$$

Based on how we defined the effective degree and the effective influences, we have the following fact analogous to Fact 3.6.

Fact 3.14. *For a function $f\colon \Sigma^n \to \mathbb{C}$ and $i \in [n]$, we have*

$$I_{i,\,\mathit{effective}}[f] = 2 \sum_{\chi : \chi_i \in B_2} \left|\widehat{f}(\chi)\right|^2, \quad I_{\mathit{effective}}[f] = 2 \sum_{\chi} \mathit{effdeg}(\chi)\, \left|\widehat{f}(\chi)\right|^2.$$

3.6 High degree is preserved under random restrictions

We will often consider random restrictions of functions and would need to argue that if the original function has high degree, then so is the restricted function w.h.p. We recall that having high degree is formalised in terms of having low-stability. In the lemma below, a random restriction of a function $g\colon (\Sigma^n, \mu^n) \to \mathbb{C}$ includes every co-ordinate in the set I with probability s and then samples each co-ordinate in I according to ν_1. For the restricted function $g_{I \to \mathbf{y}}$ on $[n] \setminus I$, each co-ordinate has marginal ν_2. It is easily checked that to get the marginals 'correct', we need $\mu = s\nu_1 + (1-s)\nu_2$.

Lemma 3.15. *Let ν_1, ν_2 be distributions over Σ whose support is full and the probability of each atom is at least α and $|\Sigma| \leqslant m$. Let $\mu = s\nu_1 + (1-s)\nu_2$. Then for some $c = c(\alpha) > 0$ we have*

$$\mathop{\mathbb{E}}_{\substack{|I| \sim sn \\ \mathbf{y} \sim \nu_1^I}}\left[\mathit{Stab}_{1-\xi}(g_{I \to \mathbf{y}}; \nu_2)\right] \leqslant \mathit{Stab}_{1-c(1-s)\xi}(g; \mu).$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Proof. The left hand side is

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{x}')}\left[g(\mathbf{x})\overline{g(\mathbf{x}')}\right],$$

where $(\mathbf{x}, \mathbf{x}')$ are sampled by taking, for each $i \in [n]$ independently, $x_i = x_i' \sim \nu_1$ with probability s; otherwise with probability $1-\xi$ we take $x_i = x_i' \sim \nu_2$, and with probability ξ we take $x_i, x_i' \sim \nu_2$ independently. We denote this Markov chain by T, and note that the stationary distribution of it is μ, and also that as the support of ν_2 is full, T is connected. Additionally, for every $a, b \in \Sigma$ we have that $\mathrm{T}(a,b) \geqslant (1-s)\xi\alpha^2$. We may now write the above expectation as

$$\langle g, \mathrm{T}^{\otimes n} g\rangle = \sum_{S,T}\langle g^{=S}, \mathrm{T}^{\otimes n} g^{=T}\rangle = \sum_{S}\langle g^{=S}, \mathrm{T}^{\otimes n} g^{=S}\rangle \leqslant \sum_{S}\|g^{=S}\|_2\|\mathrm{T}^{\otimes n} g^{=S}\|_2.$$

Using Lemma 3.10, we get that

$$\langle g, \mathrm{T}^{\otimes n} g\rangle \leqslant \sum_{S}\left(1-\Omega_{\alpha,m}((1-s)\xi)\right)^{|S|}\|g^{=S}\|_2^2 = \mathsf{Stab}_{1-c(1-s)\xi}(g),$$

where $c = c(m,\alpha) > 0$. □

3.7 Embedding into the infinite cyclic group

In our proofs, towards arriving at a contradiction, we will often get an embedding of $\mathsf{supp}(\mu)$ into an infinite cyclic group. Here, we argue that if $\mathsf{supp}(\mu)$ cannot be embedded into a finite Abelian group, then it also cannot be embedded into an infinite cyclic group (say, $[0,1)$ with addition mod 1). Therefore, the finiteness of the Abelian group is not essential to the definition of linear embeddability.

Claim 3.16. *Suppose that a finite set $S \subseteq \Sigma \times \Gamma \times \Phi$ cannot be linearly embedded. Then S cannot be embedded non-trivially into the infinite cyclic group.*

Proof. Suppose towards contradiction it can be. Then there are $\sigma\colon \Sigma \to [0,1)$, $\phi\colon \Phi \to [0,1)$ and $\gamma\colon \Gamma \to [0,1)$ not all constant such that $\sigma(x)+\phi(y)+\gamma(z) = 0 \pmod 1$. Without loss of generality, σ is non-constant. Consider the set of numbers $S = \mathsf{Image}(\sigma) \cup \mathsf{Image}(\phi) \cup \mathsf{Image}(\gamma)$, let $r = |\Sigma| + |\Phi| + |\Gamma|$ and let $N = N(r) \in \mathbb{N}$ to be determined. Then $|S| \leqslant r$, so by Dirichlet's approximation theorem we may find integers p_i, q such that for each $s_i \in S$ we have that $\left|s_i - \frac{p_i}{q}\right| \leqslant \frac{1}{qN^{1/r}}$.

Let

$$\alpha = \min_{x,x'\sigma(x)\neq\sigma(x')}\min_{z\in\mathbb{Z}}\left|z+\sigma(x)-\sigma(x')\right|.$$

We choose $N = \left(\frac{3}{\alpha}\right)^r$, define σ' by $\sigma'(x) = \frac{p_i}{q} \pmod 1$ if $\sigma(x) = s_i$, and similarly define ϕ', γ'.

1. First, we show that σ', ϕ', γ' is an embedding. Fix $(x,y,z) \in \mathsf{supp}(\mu)$; then we have
$$\sigma'(x)+\phi'(y)+\gamma'(z) = \sigma(x)+\phi(y)+\gamma(z)+\Delta,$$
where $|\Delta| \leqslant \frac{3}{qN^{1/r}}$. Noting that $\sigma(x)+\phi(y)+\gamma(z)$ is an integer (as it is 0 mod 1), it follows that $\sigma'(x)+\phi'(y)+\gamma'(z)$ is very close to an integer, up to $\frac{3}{qN^{1/r}} < 1/q$. On the other hand, by definition of σ', ϕ', γ' it is a number of the form P/q for some integer P, hence it can either be an integer or at least $1/q$ far from one. It follows that it is an integer, so $\sigma'(x)+\phi'(y)+\gamma'(z) = 0 \pmod 1$.
2. Second, we show that at least one of them is not constant. Indeed, take $x, x' \in \Sigma$ on which σ differs, and suppose towards contradiction that $\sigma'(x) - \sigma'(x') = 0$. Let i, j be such that

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

$\sigma'(x) = p_i/q, \sigma'(x') = p_j/q$, and $\sigma(x) = s_i, \sigma(x') = s_j$. Then we get that $(p_i - p_j)/q$ is an integer, and as $s_i - s_j$ is Δ-close to it for $|\Delta| \leqslant \frac{2}{qN^{1/r}}$, we get that $s_i - s_j$ is $\frac{\alpha}{3}$ close to an integer. This contradicts the definition of α. □

3.8 The additive base case

In this section, we deduce a certain auxiliary (base case) inequality that follows solely under the assumption that the distribution μ has no linear embedding. We emphasise that it holds irrespective of whether or not μ has a Horn-SAT embedding. The inequality is used while reducing the dimension and making it comparable to the degree during our inductive proof (as in the overview Section 2.4).

Claim 3.17. *Let μ be a distribution on $\Sigma \times \Gamma \times \Phi$ that has no linear embedding. Then there exists $c_1 = c_1(\mu) > 0$, such that for $f\colon \Sigma \to \mathbb{C}$, $g\colon \Phi \to \mathbb{C}$ and $h\colon \Gamma \to \mathbb{C}$ that each have average equal to 0 we have that*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu} \left[f(x)(g(y) + h(z)) \right] \right| \leqslant (1 - c_1)\|f\|_2 \|g + h\|_2.$$

Proof. Assume this is not the case, so that we may find a sequence of functions (f_m, g_m, h_m) such that $\|f_m\|_2 = \|g_m + h_m\|_2 = 1$, $\mathbb{E}[f_m] = \mathbb{E}[g_m] = \mathbb{E}[h_m] = 0$, and $\left|\mathbb{E}_{(x,y,z)\sim\mu}\left[f_m(x)(g_m(y) + h_m(z))\right]\right| \geqslant 1 - \frac{1}{m}$. Passing to subsequences, we may assume that f_m converges to a function f, g_m converges to a function g and h_m converges to a function h, so that we get $\mathbb{E}[f] = \mathbb{E}[g] = \mathbb{E}[h]$, $\|f\|_2 = \|g + h\|_2 = 1$ and $\left|\mathbb{E}_{(x,y,z)\sim\mu}\left[f(x)(g(y) + h(z))\right]\right| \geqslant 1$. By Cauchy-Schwarz we have that

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu} \left[f(x)(g(y) + h(z)) \right] \right| \leqslant \sqrt{\mathop{\mathbb{E}}_{(x,y,z)\sim\mu} \left[f(x)^2 \right]} \sqrt{\mathop{\mathbb{E}}_{(x,y,z)\sim\mu} \left[\left| g(y) + h(z) \right|^2 \right]} = \|f\|_2 \|g + h\|_2 = 1,$$

hence we get that Cauchy-Schwarz is tight and so $\overline{f(x)} = \theta(g(y) + h(z))$ for some $\theta \in \mathbb{C}$ of absolute value 1, for all $(x, y, z) \in \mathsf{supp}(\mu)$. As the 2-norm of f is 1 and its average is 0, the function f is not constant, and so after diving them by a sufficiently large constant and adding a constant (so that their image is contained in $[0, 1]$), we get that either the real part and the imaginary part of $\overline{f}, -\theta g, -\theta h$ form a non-trivial embedding of μ. Together with Claim 3.16, this contradicts the assumption that μ has no linear embedding. □

Lemma 3.18. *Let μ be a distribution on $\Sigma \times \Gamma \times \Phi$ that has no linear embedding and $\mu_{y,z}$ is uniform. Then there exists $c_1 = c_1(\mu) > 0$, such that for $f\colon \Sigma \to \mathbb{C}$, $g\colon \Phi \to \mathbb{C}$ and $h\colon \Gamma \to \mathbb{C}$ we have that*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu} \left[f(x)(g(y) + h(z)) \right] \right| \leqslant \|f\|_2 \sqrt{\left| \mathbb{E}[g] + \mathbb{E}[h] \right|^2 + (1 - c_1)(W_{=1}[g] + W_{=1}[h])}.$$

Here $W_{=1}[g]$ denotes the variance $\mathbb{E}\left[|g|^2\right] - \left|\mathbb{E}[g]\right|^2$ (and similarly for $W_{=1}[h]$).

Proof. Write $f = \mathbb{E}[f] + f^{=1}$, $g = \mathbb{E}[g] + g^{=1}$ and $h = \mathbb{E}[h] + h^{=1}$ so that

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

By Claim 3.17 we get that

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu} \left[f^{=1}(x)(g^{=1}(y) + h^{=1}(z)) \right] \right| \leqslant (1 - c_1)\|f^{=1}\|_2 \|g^{=1} + h^{=1}\|_2,$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

and so

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu}\left[f(x)(g(y)+h(z))\right]\right| &\leqslant \left|\mathbb{E}\left[f\right]\right|\left|\mathbb{E}\left[g\right]+\mathbb{E}\left[h\right]\right|+(1-c_1)\|f^{=1}\|_2\|g^{=1}+h^{=1}\|_2\\ &\leqslant \sqrt{\left|\mathbb{E}\left[f\right]\right|^2+\|f^{=1}\|_2^2}\sqrt{\left|\mathbb{E}\left[g\right]+\mathbb{E}\left[h\right]\right|^2+(1-c_1)^2\|g^{=1}+h^{=1}\|_2^2}\\ &= \|f\|_2\sqrt{\left|\mathbb{E}\left[g\right]+\mathbb{E}\left[h\right]\right|^2+(1-c_1)^2\|g^{=1}+h^{=1}\|_2^2}\end{aligned}$$

where we used Cauchy-Schwarz. As $\mu_{y,z}$ is uniform we get that $\|g^{=1}+h^{=1}\|_2^2=\|g^{=1}\|_2^2+\|h^{=1}\|_2^2=W_{=1}[g]+W_{=1}[h]$ and the proof is concluded □

4. The main homogeneous statement

In this section, we reduce Lemma 2.5 to a similar statement about homogeneous functions. Namely, we state Lemma 4.1, and show that it implies Lemma 2.5. There are two key differences between the two lemmas. The first of which is that whereas Lemma 2.5 is only concerned with bounded functions, Lemma 4.1 applies to general complex-valued functions. It can be seen, however, that if the base case fails, that is, if there are non-trivial functions for which (7) fails with $\tau=1$, then Lemma 2.5 would not hold for unbounded functions. Therefore, in exchange for relaxing ℓ_∞-boundedness to ℓ_2-boundedness, we get to assume that the functions f,g,h are all homogeneous, and the effective degree of f is significant. A precise statement follows.

Lemma 4.1. *There are $C(m,\alpha)>0$ and $D(m,\alpha)\in\mathbb{N}$ such that the following holds for all $d\geqslant D$. Suppose μ is a distribution over $\Sigma\times\Gamma\times\Phi$ satisfying the conditions of Lemma 2.5, and $f\colon\Sigma^n\to\mathbb{C}$, $g\colon\Gamma^n\to\mathbb{C}$, $h\colon\Phi^n\to\mathbb{C}$ are functions such that*

1. *f is homogeneous of degree at most $d\log^{10}d$, g is homogeneous of degree at most $d\log^{10}d$ and at least $\frac{d}{\log^{10}d}$, and h is homogeneous of degree at most $2d\log^{10}d$.*
2. *The effective degree of f is at least $\frac{d}{\log^{200}d}$. Namely, for all $\chi\in B^n$ such that $\widehat{f}(\chi)\neq 0$, we have that $\mathsf{effdeg}(\chi)\geqslant\frac{d}{\log^{200}d}$.*

Then

$$\left|\mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})\sim\mu^{\otimes n}}\left[f(\boldsymbol{x})g(\boldsymbol{y})h(\boldsymbol{z})\right]\right|\leqslant(1-\delta)^{d/\log^C(d)}\|f\|_2\|g\|_2\|h\|_2.$$

The rest of this section is devoted to showing that Lemma 4.1 implies Lemma 2.5. We prove the implication using the following sequence.

$$\text{Lemma 4.1}\implies\text{Lemma 4.3}\implies\text{Lemma 4.2}\implies\text{Lemma 2.5}. \tag{8}$$

Then, in Sections 5, 6, and 7, we prove Lemma 4.1.

4.1 Soft truncations

In this section, we define a couple of noise operators that will be used in stating the intermediate lemmas from (8).

The noise operators. Let $\xi\in(0,1]$ be some parameter.

1. We define the operator $\mathrm{T}_{1-\xi}$ acting on $L_2(\Sigma)$ as follows. Consider the Markov chain on Σ that on $x\in\Sigma$ generates $x'\sim\mathrm{T}_{1-\xi}x$ by: with probability $1-\xi$ we take $x'=x$, and otherwise we re-samples $x'\sim\mu_x$. We let $\mathrm{T}_{1-\xi}$ be the corresponding averaging operator on $L_2(\Sigma)$, that is, $\mathrm{T}_{1-\xi}f(x)=\mathbb{E}_{\mathbf{x}'\sim\mathrm{T}_{1-\xi}x}\left[f(\mathbf{x}')\right]$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

2. We define analogues of the operator $T_{1-\xi}$ on $L_2(\Gamma)$, $L_2(\Phi)$ in the same way. For notational convenience, we will use the same notation for them, that is, $T_{1-\xi}$, and it will be clear from the context which operator is applied.
3. We define the operators $E_{1-\xi}$. Let $\Sigma' \subseteq \Sigma$ be evidencing the fact that μ satisfies the relaxed base case. Consider the Markov chain on Σ that on $x \in \Sigma$ generates $x' \sim E_{1-\xi}x$ by: if $x \in \Sigma \setminus \Sigma'$, we take $x' = x$. Otherwise, with probability $1-\xi$ we take $x' = x$, and with probability ξ we resample $x' \sim \mu_x \mid \Sigma'$. We let $E_{1-\xi}$ be the corresponding averaging operator on $L_2(\Sigma)$, that is, $E_{1-\xi}f(x) = \mathbb{E}_{x' \sim E_{1-\xi}x}\left[f(x')\right]$.

The noise operator $T_{1-\xi}$ when applied on a function f dampens (in the ℓ_2 measure) the high-degree terms from f. Analogously, as we will see, the operator $E_{1-\xi}$ when applied on a function dampens the high *effective degree* terms from the function.

4.2 Intermediate lemmas

With these operators, we can now state the relaxed analogues of Lemma 4.1 wherein the degree conditions are replaced with analogous soft truncations (but in return, we still retain the boundedness of the functions).

4.2.1 Softly truncating g from both sides and f from above

In the first relaxation, we softly truncate the degree of the function g from both sides and the degree of the function f from above using the noise operators $T_{1-\xi}$.

Lemma 4.2. *For all $\alpha > 0$, $m \in \mathbb{N}$, and $M > 0$, there is $\xi_0 > 0$ such that the following holds for all $0 < \xi \leqslant \xi_0$. Suppose μ is a distribution over $\Sigma \times \Gamma \times \Phi$ satisfying the conditions of Lemma 2.5, and $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$, $h\colon \Phi^n \to \mathbb{C}$ are 1-bounded functions. Then*

$$\left| \mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \sim \mu^{\otimes n}} \left[T_{1-M\xi/\log(1/\xi)^3} f(\boldsymbol{x}) (T_{1-\xi/2} - T_{1-\xi}) g(\boldsymbol{y}) h(\boldsymbol{z}) \right] \right| \lesssim \frac{1}{\log^6(1/\xi)}.$$

4.2.2 Softly truncating the effective degree of f from below

In the next relaxation, we further softly truncate the *effective degree* of f from below. We do this using the operator $E_{1-\xi}$.

Lemma 4.3. *For all $\alpha > 0$, $m \in \mathbb{N}$, and $M > 0$, there is $\xi_0 > 0$ such that the following holds for all $0 < \xi \leqslant \xi_0$. Suppose μ is a distribution over $\Sigma \times \Gamma \times \Phi$ satisfying the conditions of Lemma 2.5, and $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$, $h\colon \Phi^n \to \mathbb{C}$ are 1-bounded functions. Then*

$$\left| \mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \sim \mu^{\otimes n}} \left[(I - E_{1-M\xi\log(1/\xi)^{100}}) T_{1-M\xi/\log(1/\xi)^3} f(\boldsymbol{x}) (T_{1-\xi/2} - T_{1-\xi}) g(\boldsymbol{y}) h(\boldsymbol{z}) \right] \right| \lesssim \frac{1}{\log^6(1/\xi)}.$$

We defer the proofs of the implications Lemma 4.1 $\Longrightarrow$ Lemma 4.3 $\Longrightarrow$ Lemma 4.2 $\Longrightarrow$ Lemma 2.5 to the Appendix. All these proofs follow from the standard arguments that use the fact that the noise operator essentially gets rid of the high degree part of the functions.

5. The main inductive statement, and set up

In this section, we re-phrase and state a sharper version of Lemma 4.1, which will be more convenient for us to work with.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

5.1 The parameter β_{n,d_1,d_2,d_3}

We begin by defining the parameter β'_{n,d_1,d_2,d_3} for all $d_1, d_2, d_3, n \in \mathbb{N}$ (which is a close variant of β_{n,d_1,d_2,d_3} which will be shortly defined):

$$\beta'_{n,d_1,d_2,d_3} = \max_{\substack{f\colon \Sigma^n \to \mathbb{C} \text{ degree } d_1 \text{ homogenous,} \\ \mathsf{effdeg}(f) \geqslant d_1/\log^{20}(d_1) \\ g\colon \Gamma^n \to \mathbb{C} \text{ degree } d_2 \text{ homogenous,} \\ h\colon \Phi^n \to \mathbb{C} \text{ degree at most } d_3 \text{ homogenous}}} \frac{\left|\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right|}{\|f\|_2\|g\|_2\|h\|_2}.$$

Using Cauchy-Schwarz, it is clear that $\beta_{n,d_1,d_2,d_3} \leqslant 1$ always. The following lemma asserts that if d_1, d_2, d_3 are roughly the same up to poly-logs, then β_{n,d_1,d_2,d_3} is actually almost exponentially decaying in d.

Lemma 5.1. *For all $K, m \in \mathbb{N}$ and $\alpha > 0$, there are $C, d_0 \in \mathbb{N}$ such that the following holds. Let μ be a distribution over $\Sigma \times \Gamma \times \Phi$ as in Lemma 4.1, and let $d_1, d_2, d_3 \leqslant n$ be such that $d_i \leqslant d_j \log^K(d_j)$ for all $i, j = 1, 2, 3$, and $d_i \geqslant d_0$ for all i. Then*

$$\beta'_{n,\,d_1,\,d_2,\,d_3} \leqslant 2^{-\frac{d_1}{\log^C(d_1)}}.$$

We note that Lemma 5.1 immediately implies Lemma 4.1, hence we will focus henceforth on proving that 5.1 is true. We will actually need to adjust the parameter $\beta'_{n,\,d_1,\,d_2,\,d_3}$ a bit and state a statement similar to Lemma 5.1 but stronger. The main difference will be that instead of working with the function f over $\mathbf{x}$, we will work with the function $F(\mathbf{y}, \mathbf{z})$. Given a function f, as in μ we have that $\mathbf{y}, \mathbf{z}$ implies $\mathbf{x}$, we may define $F\colon (\Gamma^n \times \Phi^n, \mu_{y,z}^{\otimes n}) \to \mathbb{C}$ by

$$F(\mathbf{y}, \mathbf{z}) = f(\mathbf{x}),$$

where $\mathbf{x}$ is the unique point in Σ^n such that $(x_i, y_i, z_i) \in \mathsf{supp}(\mu)$ for all $i \in [n]$. We denote this as an operator

$$W\colon L_2(\Sigma^n, \mu_x^{\otimes n}) \to L_2(\Gamma^n \times \Phi^n, \mu_{y,z}^{\otimes n}), \text{ so that } F = Wf.$$

It will be more convenient for us to work with F since the distribution over y, z is uniform (whereas the distribution over x may not be), but to facilitate that we need to translate the information that f is homogeneous and has high effective degree.

5.1.1 Setting up the basis to define the effective degree of F

Consider any two basis elements $\chi, \chi' \in B_1 \cup B_2$, and note that

$$\langle W\chi, W\chi' \rangle = \langle \chi, \chi' \rangle = 1_{\chi=\chi'},$$

so $\{W\chi\}_{B_1 \cup B_2}$ form a partial orthonormal basis for $L_2(\Gamma^n \times \Phi^n, \mu_{y,z}^{\otimes n})$. We may complete it to an orthonormal basis using some set $\{\tilde{\chi}\}_{\tilde{\chi} \in C}$. Thus, any function $F\colon \Gamma^n \times \Phi^n \to \mathbb{C}$ may be written as

$$F(\mathbf{y}, \mathbf{z}) = \sum_{\tilde{\chi} \in (WB_1 \cup WB_2 \cup C)^{\otimes n}} \widehat{F}(\tilde{\chi})\tilde{\chi}(\mathbf{y}, \mathbf{z}),$$

where

$$\tilde{\chi}(\mathbf{y}, \mathbf{z}) = \prod_{i=1}^{n} \tilde{\chi}_i(y_i, z_i), \quad \text{and } \widehat{F}(\tilde{\chi}) = \langle F, \tilde{\chi} \rangle.$$

Noting that the function $W\chi_{\mathsf{const}}$ is the all 1 function, we define the following notion of degree for monomials over $\mathbf{y}, \mathbf{z}$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Definition 5.2. *The degree of* $\tilde{\chi} \in (WB_1 \cup WB_2 \cup C)^{\otimes n}$ *is defined to be the number of coordinates* i *such that* $\tilde{\chi}_i \neq W\chi_{\text{const}}$.

We also define the effective degree.

Definition 5.3. *The effective degree of* $\tilde{\chi} \in (WB_1 \cup WB_2 \cup C)^{\otimes n}$ *is defined to be the number of coordinates* i *such that* $\tilde{\chi}_i \in WB_2$.

Finally, we define a property of functions $F(\mathbf{y}, \mathbf{z})$ which is equivalent to it being in the image of the operator W.

Definition 5.4. *Consider the graph* $G = (V, E)$ *whose vertices are* $\Gamma \times \Phi$ *and* (y, z) *and* (y', z') *are adjacent if there is some* $x \in \Sigma$ *such that* $(x, y, z), (x, y', z') \in \mathsf{supp}(\mu)$. *Consider the graph* $G^{\otimes n} = (V^{\otimes n}, E')$ *where* $E' = \left\{ ((\mathbf{y}, \mathbf{z}), (\mathbf{y}', \mathbf{z}')) \;\middle|\; ((y_i, z_i), (y_i', z_i')) \in E \; \forall i \right\}$.

We say a function $F\colon \Gamma^n \times \Phi^n \to \mathbb{C}$ *is constant on connected components if* F *is constant on all of the connected components of* $G^{\otimes n}$.

We have the following claim.

Claim 5.5. *For a function* $F\colon \Gamma^n \times \Phi^n \to \mathbb{C}$, *the following are equivalent:*

1. F *is constant on connected components;*
2. *There is* $f\colon \Sigma^n \to \mathbb{C}$ *such that* $F = Wf$.
3. $\widehat{F}(\chi) = 0$ *for* $\chi \notin (WB_1 \cup WB_2)^{\otimes n}$.

Proof. It is clear that the third item implies the second item, and that the second item implies the first item. We next show that the first item implies the third item. Towards this end, assume that F is constant on connected component, and let $\chi \notin (WB_1 \cup WB_2)^{\otimes n}$. Then $\chi_i \in C$ for some $i = 1, \ldots, n$, without loss of generality assume that $i = 1$. Then

$$\widehat{F}(\chi) = \mathop{\mathbb{E}}_{(\mathbf{y}_{-1}, \mathbf{z}_{-1}) \sim \mu_{y,z}^{\otimes n-1}} \left[\prod_{i=2}^{n} \overline{\chi_i(y_i, z_i)} \mathop{\mathbb{E}}_{(y_1, z_1) \sim \mu_{y,z}} \left[F(\mathbf{y}, \mathbf{z}) \overline{\chi_1(y_1, z_1)} \right] \right].$$

Fix $\mathbf{y}_{-1} = y_{-1}$ and $\mathbf{z}_{-1} = z_{-1}$; we show that the inner expectation is 0. To see this, first note that since $F_{\{2,\ldots,n\} \to (y_{-1}, z_{-z})}(y_1, z_1)$, as a function of y_1, z_1, is constant on connected components, it suffices to show that each basis element in C is perpendicular to functions that are constant on connected components. To see that, it suffices to show that for each connected component C of G, the indicator of it $1_C(y_1, z_1)$ is in $\mathsf{span}(WB_1 \cup WB_2)$, and we next show this is true.

Indeed, let C be some connected component. Define

$$f(x) = 1_C(y, z)$$

where (y, z) are chosen so that $(x, y, z) \in \mathsf{supp}(\mu)$. We note that this is well defined, since if we have two (y, z) and (y', z') such that (x, y, z) and (x, y', z') are both in $\mathsf{supp}(\mu)$, then they lie in the same connected component of H and hence $1_C(y, z) = 1_C(y', z')$. It follows that $1_C = Wf$, hence $1_C \in \mathsf{span}(WB_1 \cup WB_2)$. □

The following claim encapsulates the properties we need about $F = Wf$.

Claim 5.6. *Suppose that* $f\colon \Sigma^n \to \mathbb{C}$ *is a degree* d *homogeneous function, and for all* χ *such that* $\widehat{f}(\chi) \neq 0$ *we have* $\mathsf{effdeg}(\chi) \geqslant d'$. *Then* $F = Wf$ *is a degree* d *homogeneous function, it is constant on connected components and for all* $\tilde{\chi}$ *such that* $\widehat{F}(\tilde{\chi}) \neq 0$ *we have that* $\mathsf{effdeg}(\chi) \geqslant d'$.

Proof. Note that writing $f(\mathbf{x}) = \sum_{\chi \in (B_1 \cup B_2)^{\otimes n}} \widehat{f}(\chi)\chi(\mathbf{x})$, we get that

$$F(\mathbf{y}, \mathbf{z}) = Wf(\mathbf{y}, \mathbf{z}) = \sum_{\chi \in (B_1 \cup B_2)^{\otimes n}} \widehat{f}(\chi)(W\chi)(\mathbf{y}, \mathbf{z}),$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

so we get that $\widehat{F}(\tilde{\chi})$ is non-zero only if $\tilde{\chi} \in (WB_1 \cup WB_2)^{\otimes n}$, and it is equal to $\widehat{f}(\chi)$ where $\tilde{\chi} = W\chi$. From this and Claim 5.5, the assertions of the claim immediately follow. □

5.1.2 The main inductive statement

We are now ready to define β_{n,d_1,d_2,d_3} and state a stronger form of Lemma 5.1. We define

$$\beta_{n,d_1,d_2,d_3} = \max_{\substack{F\colon \Gamma^n \times \Phi^n \to \mathbb{C} \text{ degree } d_1 \text{ homogenous,} \\ F \text{ is constant on connected components,} \\ \mathrm{effdeg}(F) \geqslant d_1/\log^{20}(d_1), \\ g\colon \Gamma^n \to \mathbb{C} \text{ degree } d_2 \text{ homogenous,} \\ h\colon \Phi^n \to \mathbb{C} \text{ degree at most } d_3 \text{ homogenous.}}} \frac{\left|\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[F(\mathbf{y},\mathbf{z})g(\mathbf{y})h(\mathbf{z})\right]\right|}{\|F\|_2\|g\|_2\|h\|_2}.$$

Lemma 5.7. *For all $K, m \in \mathbb{N}$, and $\alpha > 0$, there are $C, d_0 \in \mathbb{N}$ such that the following holds. Let μ be a distribution over $\Sigma \times \Gamma \times \Phi$ as in Lemma 4.1, and let $d_1, d_2, d_3 \leqslant n$ be such that $d_i \leqslant d_j \log^K(d_j)$ for all $i, j = 1, 2, 3$, and $d_i \geqslant d_0$ for all i. Then*

$$\beta_{n,d_1,d_2,d_3} \leqslant 2^{-\frac{d_1}{\log^C(d_1)}}.$$

Lemma 5.1 follows from Lemma 5.7 by the virtue of Claim 5.6.

The proof of Lemma 5.7 spans Sections 6, 7, and in the rest of the current section we set up some machinery and explain the high level overview of the argument. Our overall approach to the proof of Lemma 5.7 is inductive, and so we will need to be able to write down a function over n variables as linear combination of products of a function of $n-1$ variables and a function over a single variable. Towards this end, we employ the singular-value decomposition as presented in the next section.

Throughout, we will have a partition the set of coordinates $[n]$ into $I \cup J$ where $|I| = n-1$ and $|J| = 1$, which for now will be arbitrary (we will later explain how to choose it so that it satisfies several additional properties that we need).

5.2 SVD decompositions

In this section, we state claims asserting an SVD decomposition for homogeneous and non-homogeneous functions. The proofs of these claims are deferred to the Appendix. Throughout, I, J is a partition of $[n]$ wherein $|J| = 1$ and $|I| = n-1$.

5.2.1 The SVD decompositions for homogeneous functions

The following decomposition claim is phrased in terms of the function g. However, it applies to functions over z, as well as to functions over y, z. We will use it for both g and h, and may use it also for F. However, for F we need a few additional properties, which we establish in Claim 5.10.

Claim 5.8. *If $g\colon \Gamma^n \to \mathbb{C}$ is a homogeneous function of degree d and $\|g\|_2 = 1$, then we may write*

$$g(\boldsymbol{y}) = \sum_{r=1}^{m} \lambda_r g_r(\boldsymbol{y}_I) g'_r(\boldsymbol{y}_J),$$

and $R = \{i \mid \lambda_i \neq 0\}$ where

1. *For $r \in R$, $g_r\colon \Gamma^I \to \mathbb{C}$ is an orthonormal set of functions.*
2. *If $1 \in R$, then g_1 is homogeneous of degree d and for $r \geqslant 2$ in R the function g_r is homogeneous of degree $d-1$.*

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

3. *For $r \in R$, $g'_r \colon \Gamma^J \to \mathbb{C}$ is an orthonormal set of functions.*
4. *g'_1 is constant.*
5. *Each λ_i is a non-negative real number and $\sum_{r=1}^{m} \lambda_i^2 = 1$.*

Proof. We defer this proof to the Appendix. □

A natural question is what can be said about the coefficients λ_i in the above SVD decomposition, and indeed this will help us in choosing an appropriate partition. We have

Claim 5.9. *Let $g \colon \Gamma^n \to \mathbb{C}$ is a homogeneous function of degree d_1 and $\|g\|_2 = 1$, and write*

$$g(\boldsymbol{y}) = \sum_{r=1}^{m} \lambda_r g_r(\boldsymbol{y}_I) g'_r(\boldsymbol{y}_J),$$

as in Claim 5.8. If j is the unique variable in the set J in the partition $[n] = I \cup J$, then

$$\lambda_1^2 = 1 - \frac{1}{2} I_j[g].$$

Proof. Consider $I_j[g]$, and note that

$$I_j[g] = \mathop{\mathbb{E}}_{\substack{\mathbf{y} \sim \mu^{n-1} \\ a,b}} \left[\left| g(\mathbf{y}_I, a) - g(\mathbf{y}_I, b) \right|^2 \right] = \mathop{\mathbb{E}}_{\substack{\mathbf{y} \sim \mu^{n-1} \\ a,b}} \left[\left| \sum_r \lambda_r g_r(\mathbf{y}_I)(g'_r(a) - g'_r(b)) \right|^2 \right]$$
$$= \sum_{r_1, r_2} \lambda_{r_1} \overline{\lambda_{r_2}} \langle g_{r_1}, g_{r_2} \rangle \mathop{\mathbb{E}}_{a,b} \left[(g'_{r_1}(a) - g'_{r_1}(b)) \overline{(g'_{r_2}(a) - g'_{r_2}(b))} \right].$$

For $r_1 \neq r_2$ we have $\langle g_{r_1}, g_{r_2} \rangle = 0$, so the last sum is equal to

$$\sum_r \lambda_r^2 \mathop{\mathbb{E}}_{a,b} \left[\left| g'_r(a) - g'_r(b) \right|^2 \right] = 2 \sum_r \lambda_r^2 \operatorname{var}(g'_r) = 2 \sum_{r \neq 1} \lambda_r^2,$$

as the variance of g'_1 is 0, and the variance of any other g'_r is 1. Hence

$$1 - \frac{I_j[g_r]}{2} = 1 - \sum_{r \neq 1} \lambda_r^2 = \lambda_1^2.$$

□

We next state an SVD decomposition statement that addresses the function F.

Claim 5.10. *Suppose $F \colon \Gamma^n \times \Phi^n \to \mathbb{C}$ is a homogeneous function of degree d which is constant on connected components, $\|F\|_2 = 1$, and the effective degree of each monomial in F is at least d'. Then we may write*

$$F(\boldsymbol{y}, \boldsymbol{z}) = \sum_{t=1}^{m} \gamma_t F_t(\boldsymbol{y}_I, \boldsymbol{z}_I) F'_t(\boldsymbol{y}_J, \boldsymbol{z}_J),$$

and $T = \{ t \mid \gamma_t \neq 0 \}$ where

1. *For $t \in T$, $F_t \colon \Gamma^I \times \Phi^I \to \mathbb{C}$ is an orthonormal set of functions.*
2. *If $1 \in T$, then F_1 is homogeneous of degree d and for $t \geqslant 2$ in T the function F_t is homogeneous of degree $d - 1$.*
3. *If $1 \in T$, then in F_1 the effective degree of monomial is at least d' and for $T \geqslant 2$ in T the effective degree of each monomial in F_t is at least $d' - 1$.*
4. *The functions F_t and F'_t are constant on connected components for all t.*

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

5. *For $t \in T$, $F'_t \colon \Gamma^J \times \Phi^J \to \mathbb{C}$ is an orthonormal set of functions.*
6. *F'_1 is constant.*
7. *Each γ_t is a non-negative real number and $\sum_{t=1}^{m} \gamma_t^2 = 1$.*

Proof. The proof of this claim is also deferred to the Appendix. □

5.2.2 The SVD decompositions for non-homogeneous functions

In this section, we state the decomposition for non-homogeneous functions. For such functions, we do not get the guarantee that one of the functions in the decomposition is the constant 1 function (and hence we do not have one of the functions in the decomposition to have *full degree*).

Claim 5.11. *If $g \colon \Gamma^n \to \mathbb{C}$ satisfies that $g^{\leqslant d} \equiv 0$ and $\|g\|_2 = 1$, then we may write*

$$g(\boldsymbol{y}) = \sum_{r=1}^{m} \lambda_r g_r(\boldsymbol{y}_I) g'_r(\boldsymbol{y}_J),$$

and $R = \{ i \mid \lambda_i \neq 0 \}$ where

1. *For $r \in R$, $g_r \colon \Gamma^I \to \mathbb{C}$ is an orthonormal set of functions.*
2. *For $r \in R$ we have that $(g_r)^{\leqslant d-1} \equiv 0$.*
3. *For $r \in R$, $g'_r \colon \Gamma^J \to \mathbb{C}$ is an orthonormal set of functions.*
4. *Each λ_i is a non-negative real number and $\sum_{r=1}^{m} \lambda_i^2 = 1$.*

Proof. The proof is similar to the proof of Claim 5.8 and is deferred to the Appendix. □

Claim 5.12. *Suppose $F \colon \Gamma^n \times \Phi^n \to \mathbb{C}$ satisfies $F^{\leqslant d} \equiv 0$, is constant on connected components, $\|F\|_2 = 1$, and the effective degree of each monomial in F is at least d'. Then we may write*

$$F(\boldsymbol{y}, \boldsymbol{z}) = \sum_{t=1}^{m} \gamma_t F_t(\boldsymbol{y}_I, \boldsymbol{z}_I) F'_t(\boldsymbol{y}_J, \boldsymbol{z}_J),$$

and $T = \{ t \mid \gamma_t \neq 0 \}$ where

1. *For $t \in T$, $F_t \colon \Gamma^I \times \Phi^I \to \mathbb{C}$ is an orthonormal set of functions.*
2. *For $t \in T$, $(F_t)^{\leqslant d-1} \equiv 0$, and each monomial in F_t has effective degree at least $d' - 1$.*
3. *The functions F_t and F'_t are constant on connected components for all t.*
4. *For $t \in T$, $F'_t \colon \Gamma^J \times \Phi^J \to \mathbb{C}$ is an orthonormal set of functions.*
5. *Each γ_t is a non-negative real number and $\sum_{t=1}^{m} \gamma_t^2 = 1$.*

Proof. The proof is similar to the proof of Claim 5.10, as is omitted. □

5.3 Roadmap of the proof of Lemma 5.7

The proof of Lemma 5.7 comprises two steps. In this section, we give an overview of these two steps. In Section 6, we formally show how to perform the first step of reducing to the case of nearly-linear degree and in Section 7, we show how to prove Lemma 5.7 when the degrees of the functions are nearly-linear in the number of variables.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

The first step: reducing to the case of nearly-linear degree. In the first step, we show an inductive step that manages to show (roughly speaking) if n is much larger than $\max(d_1, d_2, d_3)$, then one has

$$\beta_{n,d_1,d_2,d_3} \leqslant \min\left(\beta_{n-1,d_1,d_2,d_3}, (1-c)\beta_{n-1,d_1-1,d_2-1,d_3-1}\right),$$

where $c = c(m,\alpha) > 0$. Iterating this bound, we either manage to prove that β_{n,d_1,d_2,d_3} is at most $(1-c)^{d_1/2}$, or else we reduce n to be small enough so that the inductive step can no longer be made. In the first case we are done, and in the second case we have shown that

$$\beta_{n,d_1,d_2,d_3} \leqslant \beta_{n',d_1',d_2',d_3'}$$

where $d_i' \geqslant d_i/2$ for $i = 1,2,3$ and $n' \lesssim \max(d_1', d_2', d_3')$. Since d_1, d_2, d_3 are originally of the same order up to poly-logs, we are reduced to proving a variant of Lemma 5.7 in the case that n', d_1', d_2', d_3' are of the same order up to poly-logs; we refer to this case as the 'nearly-linear degree case'.

The overarching idea of our argument is that since n is much larger than d_1, d_2, d_3, we may choose the partition $I \cup J$ so that the variable in J has small influence in all of F, g, h, and hence when we use the SVD decompositions from Claims 5.8 and 5.10 (and using Claim 5.9), we get that most of the mass of the functions lies on F_1, g_1 and h_1 which are functions on one variable less that are homogeneous of the same degree, and the lower bound on the effective degree of F_1 still holds. If all of the mass lied only on these functions, we would immediately get that $\beta_{n,d_1,d_2,d_3} \leqslant \beta_{n-1,d_1,d_2,d_3}$. However in general there may be some small mass outside these. Intuitively, since this is a very small mass, it shouldn't matter that much; indeed our arguments show that something along these lines is true.

The second step: the case of nearly-linear degree. Thus, it remains to bound β_{n,d_1,d_2,d_3} when n, d_1, d_2 and d_3 are all the same, up to poly-logs. In this case, when we take F, g, h that achieve β_{n,d_1,d_2,d_3}, it follows that each monomial in F has effective degree at least $\tilde{\Omega}(d_1)$, and from this it follows that there is a variable (actually, many of them) i such that coordinate i has significant effective influence on F (more precisely, of the order $\Omega\left(\frac{1}{\log^C d_1}\right)$). We show that this implies that when we write F according to its SVD decomposition as in Claim 5.10 according to the partition $[n] = I \cup J$ where J has size 1 and it contains a variable with significant effective influence, we have

$$F(\mathbf{y}, \mathbf{z}) = \sum_t \gamma_t F_t(\mathbf{y}_I, \mathbf{z}_I) F_t'(\mathbf{y}_J, \mathbf{z}_J),$$

and we show that there is some t such that $\gamma_t F_t'$ has significant variance in $\Sigma' \subseteq \Sigma$ (here Σ' is chosen so as to satisfy the conditions of the relaxed base case). This means that when bounding the expectation of $F_t' g_r' h_s'$ over $\mathbf{y}_J, \mathbf{z}_J$, we may appeal to the relaxed base case to obtain a stronger bound than we originally had (the trivial bound is 1). Indeed, our argument proceeds in a similar way (though not exactly in this way due to technical reasons), and we (morally) prove that

$$\beta_{n,d_1,d_2,d_3} \leqslant \left(1 - \Omega\left(\frac{1}{\log^C d_1}\right)\right) \beta_{n-1,d_1-1,d_2-1,d_3-1}. \tag{9}$$

Iterating this argument for $d_1/2$ times, we get that

$$\beta_{n,d_1,d_2,d_3} \leqslant \left(1 - \Omega\left(\frac{1}{\log^C d_1}\right)\right)^{d_1/2} \beta_{n-d_1/2,d_1/2,d_2-d_1/2,d_3-d_1/2} \leqslant 2^{-\frac{d_1}{\log^{C'}(d_1)}},$$

concluding the proof.

Our formal argument follows the same spirit as the moral argument above, with two distinctions. First, we do not know how to establish inequality (9). Our inductive argument

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

necessitates appealing to an inductive assumption on functions over $n-1$ variables that are non-homogeneous functions (even if the original functions F, g, and h were homogeneous). Hence we cannot really work with the parameter β_{n,d_1,d_2,d_3}, and we define a similar parameter to it, $\gamma_{n,d}$, where the conditions that the function F is homogeneous of degree d is replaced with the condition that $F^{\leqslant d} \equiv 0$ (i.e. that F only contains monomials of degree d and more), and the conditions about g and h are dropped altogether.[12] The upside of moving to the parameter $\gamma_{n,d}$ is that it facilitates an inductive argument as described above. The downside of moving to the parameter $\gamma_{n,d}$ is that we may no longer use SVD decompositions as given in Claims 5.8, 5.10 due to the lack of homogeneity. Nevertheless, one still has a similar SVD decomposition as given in Claims 5.11, 5.12, where the main difference is that we can only say that the degree of each F_t is at least $d-1$. Thus, we lose 1 in the degree parameter for each iteration. Such tradeoff would not be beneficial in the context of the previous step, since there we do not necessarily manage to show that the parameter β_{n,d_1,d_2,d_3} actually decreases when we increase d_1. In the context of this step however, we manage to gain a factor of $\left(1-\Omega\left(\frac{1}{\log^C d_1}\right)\right)$ from each iteration, and since (as we show) we can perform $\Omega(d)$ such iterations, the move to $\gamma_{n,d}$ is affordable and leads to an exponentially decaying bound.

6. The first step in the proof of Lemma 5.7: reducing to $n \lesssim \max(d_1, d_2, d_3)$

In this section, we prove the following lemma. It states that either we reduce the number of variables by one without reducing any of the degrees and without gaining any factor, *or* we reduce at least one of the degrees by one along with the reduction in the number of variables by one and gain an additional multiplicative factor of $(1-\varepsilon)$ for some constant $\varepsilon > 0$.

Lemma 6.1. *For all $\alpha > 0$ and $m \in \mathbb{N}$, there exist $\varepsilon > 0$ and $L \in \mathbb{N}$ such that the following holds. Suppose $n, d_1, d_2, d_3 \in \mathbb{N}$ are parameters, and $n \geqslant L \cdot \max(d_1, d_2, d_3)$. Then letting*

$$\beta' = \max\Big(\beta_{n-1,d_1-1,d_2,d_3}, \beta_{n-1,d_1,d_2-1,d_3}, \beta_{n-1,d_1,d_2,d_3-1}, \beta_{n-1,d_1-1,d_2-1,d_3},$$
$$\beta_{n-1,d_1,d_2-1,d_3-1}, \beta_{n-1,d_1-1,d_2,d_3-1}, \beta_{n-1,d_1-1,d_2-1,d_3-1}, \beta_{n-1,d_1,d_2,d_3}\Big)$$

We have that

$$\beta_{n,d_1,d_2,d_3} \leqslant \max(\beta_{n-1,d_1,d_2,d_3}, (1-\varepsilon)\beta').$$

As can be observed easily, if we iterate the above lemma, then either we gain the $(1-\varepsilon)$ factor enough number of times to get the conclusion in Lemma 5.7, or we arrive at a situation when the degrees are nearly-linear in the number of variables and thereby finishing the first step in the proof of Lemma 5.7. This is formally shown in Corollary 6.6 later.

The rest of this section is devoted to the proof of Lemma 6.1. The proof proceeds by an inductive argument over n. Below, we consider the functions g and h of 2-norm 1 that achieve that value β_{n,d_1,d_2,d_3}, and partition the set of coordinates $[n]$ into $I \cup J$ where $|I| = n-1$ and $|J| = 1$.

Fix F, g, h achieving β_{n,d_1,d_2,d_3}.

6.1 Warm-up

To motivate the argument, we begin with considering a simplistic case in which F can be written as $F'(\mathbf{y}_I, \mathbf{z}_I)F''(\mathbf{y}_J, \mathbf{z}_J)$, g can be written as $g'(\mathbf{y}_I)g''(\mathbf{y}_J)$ and h could be written as well as $h'(\mathbf{z}_I)h''(\mathbf{z}_J)$ (Note that using the SVD decompositions, F, g, and h may be written as sum of such terms). The

[12] It is also possible to replace the condition on g with a condition of the form $g^{\leqslant d_2} \equiv 0$, but this is not necessary for us.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

inductive step would be very easy. Indeed, we then have that

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

It is always clear, by Cauchy-Shcwarz, that

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

1. If either g'' or h'' are constant, we show using our additive base cases, for example,Lemma 3.18, that this bound may be improved to $(1-\varepsilon)\|F''\|_2\|g''\|_2\|h''\|_2$; combining this with
$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{I}}\left[F'(\mathbf{y},\mathbf{z})g'(\mathbf{y})h'(\mathbf{z})\right]\leqslant\beta'$$
that follows by the inductive step, gives that $\beta_{n,d_1,d_2,d_3}\leqslant(1-\varepsilon)\beta'$.
2. If neither g'' nor h'' are constant, it follows that if F'' is constant then
$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{J}}\left[F''(\mathbf{y},\mathbf{z})g''(\mathbf{y})h''(\mathbf{z})\right]=0,$$
using the fact that (y,z) are independent and there is nothing to prove.

Thus, the only case left to consider is the case that F'', g'', and h'' are all non-constant. We do not know how to effectively handle this case (since the degrees of F',g', and h' have reduced but we do not know how to gain a $(1-\varepsilon)$ factor), and hence we will try to avoid having to give effective bounds on such terms. Indeed, such terms would have small weight (this is the way we will choose the partition $I\cup J$ and will hence be negligible when compared to terms from the item (1) above.

6.2 The actual proof

We proceed with the general case in which F, g, and h need not take this special form as above. Using the singular-value decomposition we may write F, g, and h as sum of at most m such functions satisfying some orthogonality properties (see Claim 5.8 and Claim 5.10 for precise statements):

$$F(\mathbf{y},\mathbf{z})=\sum_{t\in T}\gamma_t F_t(\mathbf{y}_I,\mathbf{z}_I)F'_t(\mathbf{y}_J,\mathbf{z}_J),\quad g(\mathbf{y})=\sum_{r\in R}\lambda_r g_r(\mathbf{y}_I)g'_r(\mathbf{y}_J),\quad h(\mathbf{z})=\sum_{s\in S}\mu_s h_s(\mathbf{y}_I)h'_s(\mathbf{y}_J),\quad(10)$$

where each one of the sets $\{F_t\}_{t\in T},\{F'_t\}_{t\in T},\{g_r\}_r,\{g'_r\}_r,\{h_s\}_s,\{h'_s\}_s$ is orthonormal, and $F'_1,g'_1,h'_1\equiv 1$, and $\sum_t\gamma_t^2=\sum_r\lambda_r^2=\sum_s\mu_s^2=1$. Thus, we have that

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[F(\mathbf{y},\mathbf{z})g(\mathbf{y})h(\mathbf{z})\right]=\sum_{r,s,t}\gamma_t\lambda_r\mu_s\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[F_t(\mathbf{y},\mathbf{z})g_r(\mathbf{y})h_s(\mathbf{z})\right]\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[F'_t(\mathbf{y},\mathbf{z})g'_r(\mathbf{y})h'_s(\mathbf{z})\right],$$

which appears as a sort of weighted sum over the cases considered in the above simplistic case. Indeed, we identify certain parts of this sum which are 0, certain parts which are negligible, and use additive base cases to bound the rest; the main point of the argument is to have the gain from the additive base case overcome the error terms, and we manage to achieve that.

It will be convenient for us to denote $\widehat{F_t}(r,s)=\langle F_t,\overline{g_rh_s}\rangle$ as well as $\widehat{F'_t}(r,s)=\langle F'_t,\overline{g'_rh'_s}\rangle$. This is justified because $(\overline{g_rh_s})_{r\in R,s\in S}$ forms an orthonormal set in $L_2(y_I,z_I;\mu_{y,z}^{\otimes I})$, and it can be completed to an orthonormal basis, in which case the coefficient $\widehat{F_t}(r,s)$ appears in front of g_rh_s in the representation of F_t. In particular, as g_rh_s is an orthonormal set of functions in $L_2(y,z)$, it follows by Bessel's inequality that

$$1=\|F_t\|_2^2\geqslant\sum_{r\in R,s\in S}\left|\widehat{F_t}(r,s)\right|^2.$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Similar reasoning applies to the notation $\widehat{F'_t}(r,s)$. Thus, we get that

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[F(\mathbf{y},\mathbf{z})g(\mathbf{y})h(\mathbf{z})\right]=\sum_{r,s,t}\gamma_t\lambda_r\mu_s\widehat{F_t}(r,s)\widehat{F'_t}(r,s). \tag{11}$$

Throughout the proof, we will assume that

$$\beta_{n,d_1,d_2,d_3}\geqslant\max\left(\beta_{n-1,d_1,d_2,d_3},(1-\varepsilon)\beta'\right),$$

since otherwise we are done. We will also assume that $\gamma_1,\lambda_1,\mu_1\geqslant 1-\eta$, and next show that as long as n is much larger than $\max(d_1,d_2,d_3)$, we may choose a partition $[n]=I\cup J$ so that this occurs.

6.2.1 The parameters and choosing the partition

We will use several parameters throughout this section, obeying the following relations:

$$0\ll\eta\ll c\ll m^{-1},\alpha\leqslant 1. \tag{12}$$

We show that if $n\geqslant\frac{10}{\eta}\max(d_1,d_2,d_3)$, then we may find a partition $[n]=I\cup J$ with $|J|=1$ so that in the SVD decomposition in (10), we have $\gamma_1,\lambda_1,\mu_1\geqslant 1-\eta$. To see that, choose the partition randomly, and note that by Claim 5.9 we have that $\mathbb{E}_{I,J}\left[1-\gamma_1^2\right]=\frac{1}{2}\mathbb{E}_{J=\{j\}}\left[I_j[f]\right]\leqslant\frac{d_1}{n}\leqslant\frac{\eta}{10}$ and similarly

$$\mathop{\mathbb{E}}_{I,J}\left[1-\lambda_1^2\right]=\frac{1}{2}\mathop{\mathbb{E}}_{J=\{j\}}\left[I_j[g]\right]\leqslant\frac{d_2}{n}\leqslant\frac{\eta}{10},\quad\mathop{\mathbb{E}}_{I,J}\left[1-\mu_1^2\right]=\frac{1}{2}\mathop{\mathbb{E}}_{J=\{j\}}\left[I_j[h]\right]\leqslant\frac{d_3}{n}\leqslant\frac{\eta}{10}.$$

By Markov's inequality, we get that

$$\mathbb{P}_{I,J}\left[1-\gamma_1^2\geqslant\eta\vee 1-\lambda_1^2\geqslant\eta\vee 1-\mu_1^2\geqslant\eta\right]\leqslant\frac{3}{10}<1,$$

so we may find a partition $[n]=I\cup J$ with $|J|$ such that $\gamma_1^2,\lambda_1^2,\mu_1^2\geqslant 1-\eta$. We fix this partition henceforth.

6.2.2 The main inductive argument

We wish to use the fact that $\gamma_1^2,\lambda_1^2,\mu_1^2\geqslant 1-\eta$ in order to effectively give an upper bound on β_{n,d_1,d_2,d_3}. Towards this, we split the expression from (11) as follows.

$$\beta_{n,d_1,d_2,d_3}=\sum_t\gamma_t A_t, \tag{13}$$

where A_t is defined as

$$A_t=\lambda_1\mu_1\widehat{F_t}(1,1)\widehat{F'}_t(1,1)+\sum_{r\neq 1}\lambda_r\mu_1\widehat{F_t}(r,1)\widehat{F'}_t(r,1)+\sum_{s\neq 1}\lambda_1\mu_s\widehat{F_t}(1,s)\widehat{F'}_t(1,s)+B_t,$$

$$B_t=\sum_{r,s\neq 1}\lambda_r\mu_s\widehat{F_t}(r,s)\widehat{F'}_t(r,s).$$

To simplify notation, we will omit the $r\neq 1$ and $s\neq 1$ from the sums below, as whenever we sum over r or s in this subsection, the summation does not include 1. We make a couple of preliminary observations.

1. Note that for $t\neq 1$, as F'_t is orthogonal to $F'_1\equiv 1$, we get that its average is 0 and so $\widehat{F'}_t(1,1)=0$ for $t\neq 1$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

2. As F'_1 is constant, and y, z are independent, we get that $\widehat{F'}_1(r,s) = 0$ for all $(r,s) \neq (1,1)$, and $\widehat{F'}_1(1,1) = 1$.

The next two claims give bounds on A_is and B_is. This following claim will let us ignore the contribution from B_t in (13).

Claim 6.2. *(Bounding B_is) We have:*

1. $B_1 = 0$.
2. *For $t \neq 1$, $|B_t| \leqslant \sqrt{(1-\mu_1^2)(1-\lambda_1^2)}\beta'$.*

The next claim handles A_is. In this claim, we use the additive base case inequality, Lemma 3.18, in order to gain enough so that the contributions from B_t in (13) becomes negligible.

Claim 6.3. *(Bounding A_is) There exists a constant $c = c(\mu) > 0$ such that the following holds.*

1. $|A_1| \leqslant |\lambda_1 \mu_1|\, \beta_{n-1,d_1,d_2,d_3}$
2. *For all $t \neq 1$*

$$|A_t| \leqslant \sqrt{(1-c)\mu_1^2 \sum_r \lambda_r^2 \left|\widehat{F}_t(r,1)\right|^2 + (1-c)\lambda_1^2 \sum_s \mu_s^2 \left|\widehat{F}_t(1,s)\right|^2} + E_t,$$

where $E_t \lesssim_m \sqrt{\left|1-\mu_1^2\right| \left|1-\lambda_1^2\right|}\beta'$.

Before we see the proofs of the above two claims, let us see why these are enough to prove Lemma 6.1.

Proof of lemma 6.1. Let $T_1 = \{1\}$, $T_2 = T \setminus T_1$. Then

$$\sum_t \gamma_t A_t = \underbrace{\sum_{t \in T_1} \gamma_t A_t}_{(I)} + \underbrace{\sum_{t \in T_2} \gamma_t A_t}_{(II)}.$$

By Claim 6.3 we have that $|(I)| \leqslant \gamma_1 \lambda_1 \mu_1 \beta_{n-1,d_1,d_2,d_3}$. For (II), we use Claim 6.3 and upper bound its absolute value by:

$$\sum_{t \in T_2} \gamma_t \sqrt{(1-c)\left(\mu_1^2 \sum_r \lambda_r^2 \left|\widehat{F}_t(r,1)\right|^2 + \lambda_1^2 \sum_s \mu_s^2 \left|\widehat{F}_t(1,s)\right|^2\right)} + \gamma_t E_t$$

where $E_t \lesssim_m \sqrt{(1-|\lambda_1|^2)(1-|\mu_1|^2)}\beta'$. Thus, by Cauchy–Schwarz, the absolute value of (II) is at most

$$\sqrt{\sum_{t \in T_2} \gamma_t^2} \sqrt{(1-c)\left(\mu_1^2 \sum_r \lambda_r^2 \sum_{t \in T_2} \left|\widehat{F}_t(r,1)\right|^2 + \lambda_1^2 \sum_s \mu_s^2 \sum_{t \in T_2} \left|\widehat{F}_t(1,s)\right|^2\right)}$$
$$+ O_m\left(\sqrt{(1-\gamma_1^2)(1-\lambda_1^2)(1-\mu_1^2)}\beta'\right).$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Note that for all r,

$$\sum_{t\in T_2}\left|\widehat{F}_t(r,1)\right|^2=\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^I}\left[\tilde{F}g_r h_1\right]\right|^2,$$

where $\tilde{F}=\frac{\sum_{t\in T_2}\widehat{F}_t(r,1)F_t}{\sqrt{\sum_{t\in T_2}|\widehat{F}_t(r,1)|^2}}$, and so by definition $\sum_{t\in T_2}\left|\widehat{F}_t(r,1)\right|^2\leqslant\beta^2_{n-1,d_1-1,d_2-1,d_3}\leqslant\beta'^2$. Similarly, $\sum_{t\in T_2}\left|\widehat{F}_t(1,s)\right|^2\leqslant\beta'^2$, so we get that

$$|(II)|\leqslant\sqrt{1-\gamma_1^2}\sqrt{(1-c)\beta'^2\left(\mu_1^2\sum_r\lambda_r^2+\lambda_1^2\sum_s\mu_s^2\right)}+O_m\left(\sqrt{(1-\gamma_1^2)(1-\lambda_1^2)(1-\mu_1^2)}\beta'\right),$$

and so

$$|(II)|\leqslant\sqrt{1-\gamma_1^2}\sqrt{(1-c)\beta'^2\left(\mu_1^2(1-\lambda_1^2)+\lambda_1^2(1-\mu_1^2)\right)}+O_m\left(\sqrt{(1-\gamma_1^2)(1-\lambda_1^2)(1-\mu_1^2)}\beta'\right).$$

Set $\beta''=\max\left(\beta_{n-1,d_1,d_2,d_3},(1-c/8)\beta'\right)$. Combining the bounds on (I) and (II) and using Cauchy-Schwarz, we get

$$\begin{aligned}
&|(I)+(II)|\\
&\leqslant\sqrt{1-\gamma_1^2+\gamma_1^2}\sqrt{\lambda_1^2\mu_1^2\beta^2_{n-1,d_1,d_2,d_3}+(1-\frac{c}{2})\beta'^2\left(\mu_1^2(1-\lambda_1^2)+\lambda_1^2(1-\mu_1^2)\right)}\\
&+O_m\left(\sqrt{(1-\gamma_1^2)(1-\lambda_1^2)(1-\mu_1^2)}\beta'\right)\\
&\leqslant\beta''\sqrt{1-\frac{c}{4}\left(\mu_1^2(1-\lambda_1^2)+\lambda_1^2(1-\mu_1^2)\right)}+O_m\left(\sqrt{(1-\gamma_1^2)(1-\lambda_1^2)(1-\mu_1^2)}\beta''\right)\\
&\leqslant\beta''\left(1-\frac{c}{8}\left(\mu_1^2(1-\lambda_1^2)+\lambda_1^2(1-\mu_1^2)\right)\right)+O_m\left(\sqrt{(1-\gamma_1^2)(1-\lambda_1^2)(1-\mu_1^2)}\beta''\right)\\
&\leqslant\beta'',
\end{aligned}$$

where the last inequality is because

$$\sqrt{(1-\gamma_1^2)(1-\lambda_1^2)(1-\mu_1^2)}\leqslant\sqrt{\eta}\left((1-\lambda_1^2)+(1-\mu_1^2)\right)\lesssim\sqrt{\eta}\left((1-\lambda_1^2)\mu_1^2+(1-\mu_1^2)\lambda_1^2\right),$$

and $\eta\ll c$. Plugging this into (13) finishes the proof of Lemma 6.1.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Bounding A_is and B_is. We now prove the bounds on A_is and B_is.

Claim 6.4 (Restatement of Claim 6.2, Bounding B_is). *We have:*

1. $B_1 = 0$.
2. *For* $t \neq 1$, $|B_t| \leqslant \sqrt{(1-\mu_1^2)(1-\lambda_1^2)}\beta'$.

Proof. The first item follows since $\widehat{F'}_1(r,s) = 0$ for all $(r,s) \neq (1,1)$. For the second item, we have

$$\begin{aligned}|B_t| &\leqslant \sum_{r,s} |\lambda_r \mu_s| \left|\widehat{F}_t(r,s)\right| \left|\widehat{F'}_t(r,s)\right| \leqslant \sqrt{\sum_{r,s} \lambda_r^2 \mu_s^2 \left|\widehat{F}_t(r,s)\right|^2} \sqrt{\sum_{r,s} \left|\widehat{F'}_t(r,s)\right|^2} \\ &\leqslant \sqrt{\sum_{r,s} \lambda_r^2 \mu_s^2 \beta'^2} \sqrt{1} \\ &\leqslant \sqrt{(1-\mu_1^2)(1-\lambda_1^2)}\beta'.\end{aligned}$$

The first transition is by Cauchy-Schwarz. The second transition is because

$$\widehat{F}_t(r,s) = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes I}} \left[F_t(\mathbf{y},\mathbf{z})\overline{g_r(\mathbf{y})h_s(\mathbf{z})}\right],$$

and as F_t, g_r, h_s are all homogeneous of degrees $d_1 - 1, d_2 - 1, d_3 - 1$ we get by definition that

$$\left|\widehat{F}_t(r,s)\right| \leqslant \beta_{n-1,d_1-1,d_2-1,d_3-1},$$

which is at most β'. □

Claim 6.5. *(Restatement of Claim 6.3, Bounding A_is) There exists a constant $c = c(\mu) > 0$ such that the following holds.*

1. $|A_1| \leqslant \lambda_1\mu_1\beta_{n-1,d_1,d_2,d_3}$
2. *For all* $t \neq 1$

$$|A_t| \leqslant \sqrt{(1-c)\mu_1^2 \sum_r \lambda_r^2 \left|\widehat{F}_t(r,1)\right|^2 + (1-c)\lambda_1^2 \sum_s \mu_s^2 \left|\widehat{F}_t(1,s)\right|^2} + E_t,$$

where $E_t \lesssim_m \sqrt{\left|1-\mu_1^2\right| \left|1-\lambda_1^2\right|}\beta'$.

Proof. Note that $A_1 = \lambda_1\mu_1\widehat{F}_1(1,1)$. The result follows since by definition we have that

$$\left|\widehat{F}_1(1,1)\right| = \left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^I} \left[F_1(\mathbf{y},\mathbf{z})\overline{g_1(\mathbf{y})h_1(\mathbf{z})}\right]\right| \leqslant \beta_{n-1,d_1,d_2,d_3},$$

as F_1 is homogeneous, constant on connected component, has the same degree and effective degree as F (see Claim 5.10), and g_1, h_1 are homogeneous of degrees d_2, d_3 respectively.

For the second item, consider

$$\tilde{g}(y) = \sum_r \lambda_r \overline{\widehat{F}_t(r,1)} g'_r, \quad \tilde{h}(y) = \sum_s \mu_s \overline{\widehat{F}_t(1,s)} h'_s.$$

First, note that

$$\langle F'_t, \tilde{g} + \tilde{h}\rangle = \sum_r \lambda_r \widehat{F}_t(r,1)\widehat{F'}_t(r,1) + \sum_s \mu_s \widehat{F}_t(1,s)\widehat{F}'_t(1,s).$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Therefore,

$$\left|\langle \tilde{g}+\tilde{h}, F'_t\rangle - \sum_r \mu_1\lambda_r \widehat{F}_t(r,1)\widehat{F'}_t(r,1) - \sum_s \lambda_1\mu_s \widehat{F}_t(1,s)\widehat{F'}_t(1,s)\right| \leqslant O\left(((1-\lambda_1^2)\sqrt{1-\mu_1^2}+(1-\mu_1^2)\sqrt{1-\lambda_1^2})\beta'\right). \tag{14}$$

In the last transition, we used the fact that

$$\left|\sum_r (1-\mu_1)\lambda_r \widehat{F}_t(r,1)\widehat{F'}_t(r,1)\right| \leqslant (1-\mu_1)\sqrt{\sum_r \lambda_r^2}\sqrt{\sum_r \left|\widehat{F'}_t(r,1)\right|^2}\max_r \left|\widehat{F}_t(r,1)\right|,$$

and $\sum_r \lambda_r^2 = 1-\lambda_1^2 \geqslant 1-\lambda_1$, $\max_r \left|\widehat{F}_t(r,1)\right| \leqslant \beta'$, and $\sum_r \left|\widehat{F'}_t(r,1)\right|^2 \leqslant \|F'_t\|_2^2 = 1$. Similarly, we also have that

$$\left|\sum_s (1-\lambda_1)\mu_s \widehat{F}_t(1,s)\widehat{F'}_t(1,s)\right| \lesssim (1-\lambda_1)\sqrt{1-\mu_1}\beta'.$$

Combining (14) with Claim 6.2 gives $|A_t| = \left|\langle \tilde{g}+\tilde{h}, F'_t\rangle\right| + O\left(\sqrt{(1-\lambda_1)(1-\mu_1)}\beta'\right)$.

By Lemma 3.18,

$$\left|\langle \tilde{g}+\tilde{h}, F'_t\rangle\right| \leqslant \sqrt{(1-c)\sum_r \lambda_r^2 \left|\widehat{F}_t(r,1)\right|^2 + (1-c)\sum_s \mu_s^2 \left|\widehat{F}_t(1,s)\right|^2},$$

implying that

$$\left|\langle \tilde{g}+\tilde{h}, F'_t\rangle\right| \leqslant \sqrt{(1-c)\mu_1^2\sum_r \lambda_r^2 \left|\widehat{F}_t(r,1)\right|^2 + (1-c)\lambda_1^2\sum_s \mu_s^2 \left|\widehat{F}_t(1,s)\right|^2 + E},$$

where $E \lesssim \left|1-\mu_1^2\right|\left|1-\lambda_1^2\right|\beta'^2$. □

6.3 *Concluding this section*

To conclude this section, we use Lemma 6.1 in order to show that for given parameters n, d_1, d_2, d_3, one either has the conclusion of the lemma or else can assume that n is at most $O(\max(d_1, d_2, d_3))$. Formally:

Corollary 6.6. *For all $\alpha > 0$ and $m \in \mathbb{N}$, there are $C > 0$ and $\varepsilon' > 0$ such that the following holds. For parameters $n, d_1, d_2, d_3 \in \mathbb{N}$, we either have that*

1. $\beta_{n,d_1,d_2,d_3} \leqslant (1-\varepsilon')^{\min(d_1,d_2,d_3)}$, *or else*
2. $\beta_{n,d_1,d_2,d_3} \leqslant \beta_{n',d'_1,d'_2,d'_3}$ *where* $n' \leqslant C\min(d'_1, d'_2, d'_3)$ *and* $d'_1 \geqslant d_1/2$, $d'_2 \geqslant d_2/2$ *and* $d'_3 \geqslant d_3/2$.

Proof. We iterate Lemma 6.1, and divide iterations into two types: those in which we gain a factor of $(1-\varepsilon)$, and those that we do not (in which case we leave d_1, d_2, d_3 as is and decrease n). Eventually, we stop at n', d'_1, d'_2, d'_3 where $n' \geqslant C\max(d'_1, d'_2, d'_3)$. Hence, there must have been at least $\max(d_1 - d'_1, d_2 - d'_2, d_3 - d'_3)$ iterations in which we gained a factor of $1-\varepsilon$. Thus, if $d'_1 \leqslant d_1/2$, $d'_2 \leqslant d_2/2$ or $d'_3 \leqslant d_3/3$, then we gained $(1-\varepsilon)^{\min(d_1/2, d_2/2, d_3/2)} \leqslant (1-\varepsilon')^{\min(d_1,d_2,d_3)}$, and we are done. □

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Looking at Lemma 5.7, we see that in the first case of Corollary 6.6 we get that

$$\beta_{n,d_1,d_2,d_3} \leqslant (1-\varepsilon)^{\min(d_1,d_2,d_3)} \leqslant (1-\varepsilon)^{d_1/\log^K(d_1)} \leqslant 2^{-d_1/\log^{C'}(d_1)},$$

and the conclusion of the Lemma 5.7 holds. Thus, we turn our attention to handling the second case of Corollary 6.6, in which case we get that $\beta_{n,d_1,d_2,d_3} \leqslant \beta_{n',d_1',d_2',d_3'}$ where $d_i' \leqslant d_j' \log^K(d_j')$ for all i, j and $n' \leqslant C \max(d_1', d_2', d_3')$.

7. The second step in the proof of lemma 5.7: the near-linear degree case

In our argument so far, we have not used the relaxed base case inequality. In this section, we crucially use the inequality in the second step of the inductive proof.

Now that we are in a case where the number of variables is of the same magnitude as the degrees of the functions, we can work with a simpler parameter that is defined next.

7.1 The parameter $\gamma_{n,d}$

Define the parameter

$$\gamma_{n,d} = \max_{\substack{F\colon \Gamma^n \times \Phi^n \to \mathbb{C}, F^{<d} \equiv 0 \\ F \text{ is constant on connected components,} \\ \mathsf{effdeg}(F) \geqslant d/\log^{20}(d) \\ g\colon \Gamma^n \to \mathbb{C} \\ h\colon \Phi^n \to \mathbb{C}}} \frac{\left|\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[F(\mathbf{y},\mathbf{z})g(\mathbf{y})h(\mathbf{z})\right]\right|}{\|F\|_2\|g\|_2\|h\|_2}.$$

Clearly $\beta_{n',d_1',d_2',d_3'}$ from the previous section is at most $\gamma_{n',d_1'}$. Our goal in this section is to prove the following result.

Lemma 7.1. *For all $\alpha > 0$, $m \in \mathbb{N}$ and $C > 0$ there exists $R > 0$ and d_0 such that if $n, d \in \mathbb{N}$ are parameters such that $n \leqslant d \log^C d$ and $d \geqslant d_0$, then*

$$\gamma_{n,d} \leqslant \left(1 - \frac{1}{\log^R(d)}\right)\gamma_{n-1,d-1}.$$

Before proving Lemma 7.1, we show the quick derivation of Lemma 5.7 from it (using the previous section).

Proof of Lemma 5.7. From Corollary 6.6 (see the discussion after it), we are either done by Corollary 6.6 or else we find d_1', d_2', d_3' and n' as there such that $\beta_{n,d_1,d_2,d_3} \leqslant \beta_{n',d_1',d_2',d_3'} \leqslant \gamma_{n',d_1'}$. Note that $n' \lesssim \max(d_1', d_2', d_3')$, so $n' \leqslant \frac{d}{2}\log^{C'}(\frac{d}{2})$, for $d = d_1'$. We now apply Lemma 7.1 $d/2$ times and get that

$$\gamma_{n',d} \leqslant \left(1 - \frac{1}{\log^R(d)}\right)\gamma_{n'-1,d-1} \leqslant \ldots \leqslant \left(1 - \frac{1}{\log^R(d)}\right)^{d/2}\gamma_{n'-d/2,d/2} \leqslant \left(1 - \frac{1}{\log^R(d)}\right)^{d/2},$$

which is at most $2^{-d/\log^{R'}(d)}$. □

The rest of this section is devoted to the proof of Lemma 7.1. We assume henceforth that

$$\gamma_{n,d} \geqslant \frac{\gamma_{n,d-1}}{2},$$

since otherwise we are done.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

7.2 Set up for the proof of Lemma 7.1

In this section, we present the set up for the proof of Lemma 7.1. We fix n and d as there. We will use the following hierarchy of parameters:

$$0 \ll R_4^{-1} \ll R_3^{-1} \ll R_2^{-1} \ll R_1^{-1} \ll R_0^{-1} \ll R^{-1} \ll c \ll m^{-1}, \alpha, C^{-1} \leqslant 1,$$
$$0 < \delta = \frac{1}{\log^{R_4} d} \leqslant \tau = \frac{1}{\log^{R_3} d} \leqslant \zeta = \frac{1}{\log^{R_2} d} \leqslant \nu = \frac{1}{\log^{R_1} d} \leqslant u = \frac{1}{\log^{R_0} d} \leqslant w = \frac{1}{\log^{R} d} \leqslant 1. \tag{15}$$

By definition, we may find $F\colon \Sigma^n \to \mathbb{C}$ which is constant on connected components, has 2-norm 1, and satisfies that $F^{<d} \equiv 0$ and all monomials in F have effective degree at least $d' = d/\log^{20} d$, as well as $g\colon \Gamma^n \to \mathbb{C}$, $h\colon \Phi^n \to \mathbb{C}$ of 2-norm equal to 1 such that

$$\gamma_{n,d} = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^n}\left[F(\mathbf{y},\mathbf{z})g(\mathbf{y})h(\mathbf{z})\right].$$

We employ an SVD decomposition according to Claims 5.11 and 5.12; if we apply it using a partition $[n] = I \cup J$ where $|J| = 1$, we get

$$F(\mathbf{y},\mathbf{z}) = \sum_{t\in T} \gamma_t F_t(\mathbf{y}_I,\mathbf{z}_I)F'_t(\mathbf{y}_J,\mathbf{z}_J), \quad g(\mathbf{y}) = \sum_{r\in R} \lambda_r g_r(\mathbf{y}_I)g'_r(\mathbf{y}_J), \quad h(\mathbf{z}) = \sum_{s\in S} \mu_s h_s(\mathbf{y}_I)h'_s(\mathbf{y}_J), \tag{16}$$

where each one of the sets $\{F_t\}_{t\in T}$, $\{F'_t\}_{t\in T}$, $\{g_r\}_r$, $\{g'_r\}_r$, $\{h_s\}_s$, $\{h'_s\}_s$ is orthonormal, and for all $t \in T$ the function F_t has $(F_t)^{<d-1} \equiv 0$ and all monomials in it have effective degree at least $d' - 1$, and $\sum_t \gamma_t^2 = \sum_r \lambda_r^2 = \sum_s \mu_s^2 = 1$. Then we have (following the Fourier coefficients notation from the last section) that

$$\gamma_{n,d} = \sum_{r,s,t} \gamma_t\lambda_r\mu_s\widehat{F}_t(r,s)\widehat{F'}_t(r,s). \tag{17}$$

We will want our partition I, J to satisfy a certain property that will be helpful for us. Namely, if $J = \{j\}$, we will want the variable j to have significant *effective influence* on F. Formally, we have defined effective influences for functions over $\mathbf{x}$, but as we explain now the definition makes sense for functions that are constant over connected components, such as F. Indeed, as F is constant on connected components we can view it, using Claim 5.5, as $F = Wf$ for some $f\colon \Sigma^n \to \mathbb{C}$ of the same 2-norm. We can thus define the effective influences and effective degrees of F by the corresponding measures in f: we define $I_{j,\,\mathsf{effective}}[F] = I_{j,\,\mathsf{effective}}[f]$, and $\mathsf{effdeg}(F) = \mathsf{effdeg}(f)$. Note that

$$\frac{1}{n}\sum_{j=1}^{n} I_{j,\,\mathsf{effective}}[F] = \frac{1}{n} I_{\,\mathsf{effective}}[f] = \frac{2}{n}\sum_{\chi} \mathsf{effdeg}(\chi)\left|\widehat{f}(\chi)\right|^2 \geqslant \frac{2d'}{n}\|f\|_2^2 = \frac{2d'}{n}$$

where we used Fact 3.14. Since $n \leqslant d\log^C d$ and $d \leqslant d'\log^{20} d$ it follows that $\frac{1}{n}\sum_{j=1}^{n} I_{j,\,\mathsf{effective}}[F] \geqslant \frac{1}{\log^{20+C} d}$, and therefore we may find a variable j such that $I_{j,\,\mathsf{effective}}[F] \geqslant \frac{1}{\log^{20+C} d}$. We denote $\nu = \frac{1}{\log^{20+C} d}$, and choose the partition $J = \{j\}$ and $I = [n] \setminus J$.

Since F'_t are all constant on connected components, there are unique $f'_t\colon \Sigma \to \mathbb{C}$ of 2-norm 1 such that $F'_t = Wf'_t$. We fix such f'_t; similarly, we have $F_t = Wf_t$, hence

$$F = W\sum_t \gamma_t f_t f'_t.$$

Note that

$$\langle f_{t_1}, f_{t_2}\rangle_{\mu_x} = \langle Wf_{t_1}, Wf_{t_2}\rangle_{\mu_{y,z}} = \langle F_{t_1}, F_{t_2}\rangle_{\mu_{y,z}} = 1_{t_1=t_2},$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

and similarly for f'_{t_1} and f'_{t_2}.

Next, we note that since F has effective influence at least ν, the variance of f'_t over Σ' (this is $\Sigma' \subseteq \Sigma$ for which the relaxed base case holds) is significant. For notational convenience, we denote

$$\mathsf{var}_{\Sigma'}(f'_t) = \mathop{\mathbb{E}}_{x,x' \in \Sigma'} \left[\left| f'_t(x) - f'_t(x') \right|^2 \right].$$

Claim 7.2. $\sum_t \gamma_t^2 \, var_{\Sigma'}(f'_t) \geqslant w$.

Proof. Consider $I_{j,\,\mathsf{effective}}[F]$. On the one hand, it is at least ν by choice. On the other hand, consider the distribution over (a, b) where $a \sim \mu_x$, and $b = a$ if $a \in \Sigma \setminus \Sigma'$ and otherwise $b \sim \mu_x \mid b \in \Sigma'$. Then $I_{j,\,\mathsf{effective}}[F]$ is equal to

$$\begin{aligned} I_{j,\,\mathsf{effective}}[f] &= \mathop{\mathbb{E}}_{\mathbf{x} \sim \mu_x^{\otimes n-1}, a, b} \left[\left| f(\mathbf{x}_I, a) - f(\mathbf{x}_I, b) \right|^2 \right] = \mathop{\mathbb{E}}_{\mathbf{x} \sim \mu_x^{\otimes n-1}, a, b} \left[\left| \sum_t \gamma_t f_t(\mathbf{x}_I)(f'_t(a) - f'_t(b)) \right|^2 \right] \\ &= \sum_{t_1, t_2} \gamma_{t_1} \gamma_{t_2} \langle f_{t_1}, f_{t_2} \rangle \mathop{\mathbb{E}}_{a,b} \left[(f'_{t_1}(a) - f'_{t_1}(b)) \overline{(f'_{t_2}(a) - f'_{t_2}(b))} \right]. \end{aligned}$$

For $t_1 \neq t_2$ we have $\langle f_{t_1}, f_{t_2} \rangle = 0$, so the last sum is equal to

$$\sum_t \gamma_t^2 \mathop{\mathbb{E}}_{a,b} \left[\left| f'_t(a) - f'_t(b) \right|^2 \right] \lesssim_{\alpha,m} \sum_t \gamma_t^2 \, \mathsf{var}_{\Sigma'}(f'_t).$$

□

Getting a gap between values close to 0 and bounded away from 0. Consider the values of λ_r, μ_s, and γ_t. We claim that there are $0 \ll \delta \ll \zeta$ (as in (15)) such that none of these values fall in the interval $[\delta, \zeta)$. Indeed, start from sufficiently small $\zeta_1 \ll \tau$ and take $\zeta_2 \ll \zeta_1$, if one of these values falls in the interval $[\zeta_2, \zeta_1)$, then we proceed with taking $\zeta_3 \ll \zeta_2$ and considering the interval $[\zeta_3, \zeta_2)$, and continue with the choice iteratively. Since these intervals are disjoint, and there are at most $3m$ distinct values of λ_r, μ_s, γ_t, we will reach an interval not containing any of them after at most $2m + m^2 + 1$ steps as required.

In other words, for all r, we either have $\lambda_r \geqslant \zeta$ or $\lambda_r \leqslant \delta$ and similarly for all s. Define

$$R' = \{ r \in R \mid \lambda_r \geqslant \zeta \}, \quad S' = \{ s \in S \mid \mu_s \geqslant \zeta \}, \quad T' = \{ t \in T \mid \gamma_t \geqslant \zeta \}.$$

Intuitively, one should think of r outside R' as having its associated masses λ_r as being 0; we cannot quite say that, but for the argument to go through it suffices to have a sufficiently large gap between ζ and δ. The same goes for μ_s, γ_t.

Getting rid of the error term. We would like to replace the right hand side of (17) by restricting the sum to go only over R', S', T', and argue that his only incurs a small loss. Consider for example the sum where $r \in R \setminus R'$ and $s \in S$, $t \in T$. Then we can bound

$$\left| \sum_{r \in R \setminus R'} \lambda_r \mu_s \gamma_t \widehat{F}_t(r, s) \widehat{F}'_t(r, s) \right| \leqslant \delta \max_{r,s,t} \left| \widehat{F}_t(r, s) \right| \lesssim_m \delta \gamma_{n-1,d-1}.$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Similarly, the other sums are also bounded by $\lesssim_m \delta\gamma_{n-1,d-1}$, hence we get that (17) implies that

$$\gamma_{n,d} \leqslant \left| \sum_{r\in R', s\in S', t\in T'} \lambda_r \mu_s \gamma_t \widehat{F}_t(r,s)\widehat{F}'_t(r,s) \right| + O_m\left(\delta\gamma_{n,d}\right). \tag{18}$$

7.3 Proof of Lemma 7.1

We begin by observing that $\gamma_{n,d} \leqslant \gamma_{n-1,d-1} + O_m\left(\delta\gamma_{n,d}\right)$. Indeed, by Cauchy–Schwarz

$$\begin{aligned}
\left| \sum_{r\in R', s\in S', t\in T'} \lambda_r \mu_s \gamma_t \widehat{F}_t(r,s)\widehat{F}'_t(r,s) \right| &\leqslant \sqrt{\sum_{r\in R', s\in S', t\in T'} \lambda_r^2\mu_s^2 \left|\widehat{F}_t(r,s)\right|^2} \sqrt{\sum_{r\in R', s\in S', t\in T'} \gamma_t^2 \left|\widehat{F}'_t(r,s)\right|^2} \\
&\leqslant \sqrt{\sum_{r\in R', s\in S'} \lambda_r^2\mu_s^2 \sum_{t\in T'} \left|\widehat{F}_t(r,s)\right|^2} \sqrt{\sum_{t\in T'} \gamma_t^2 \sum_{r\in R', s\in S'} \left|\widehat{F}'_t(r,s)\right|^2} \\
&\leqslant \sqrt{\sum_{r\in R', s\in S'} \lambda_r^2\mu_s^2\gamma_{n-1,d-1}^2} \sqrt{\sum_{t\in T'} \gamma_t^2} \\
&\leqslant \gamma_{n-1,d-1}.
\end{aligned} \tag{19}$$

In a high level, our proof inspects the potential equality cases in the above chain of inequalities, and demonstrates it cannot happen. Indeed, we have three inequalities that potentially could be improved. The first one is the Cauchy–Schwarz inequality, and if we knew that an improved version of it held, that is, that

$$\left| \sum_{r,s,t} \lambda_r \mu_s \cdot \widehat{F}_t(r,s)\gamma_t \widehat{F}'_t(r,s) \right| \leqslant (1-\tau) \sqrt{\sum_{r,s,t} \lambda_r^2\mu_s^2 \left|\widehat{F}_t(r,s)\right|^2} \sqrt{\sum_{r,s,t} \gamma_t^2 \left|\widehat{F}'_t(r,s)\right|^2},$$

we would be done. Indeed, if so we would then get that

$$\left| \sum_{r\in R', s\in S', t\in T'} \lambda_r \mu_s \gamma_t \widehat{F}_t(r,s)\widehat{F'}_t(r,s) \right| \leqslant (1-\tau)\gamma_{n-1,d-1},$$

so by (18) we get that

$$\gamma_{n,d} \leqslant (1-\tau)\gamma_{n-1,d-1} + O_m\left(\delta\gamma_{n,d}\right) \leqslant (1-\varepsilon)\gamma_{n-1,d-1},$$

as $\delta \ll \tau$ and $\varepsilon \ll \tau$. We therefore assume henceforth that

$$\begin{aligned}
&\left| \sum_{r\in R', s\in S', t\in T'} \lambda_r \mu_s \cdot \widehat{F}_t(r,s)\gamma_t \widehat{F}'_t(r,s) \right| \\
&\geqslant (1-\tau) \sqrt{\sum_{r\in R', s\in S', t\in T'} \lambda_r^2\mu_s^2 \left|\widehat{F}_t(r,s)\right|^2} \sqrt{\sum_{r\in R', s\in S', t\in T'} \gamma_t^2 \left|\widehat{F}'_t(r,s)\right|^2}.
\end{aligned} \tag{20}$$

The second inequality we have is that $\sum_{r\in R', s\in S'} \left|\widehat{F}'_t(r,s)\right|^2 \leqslant 1$ for all $t \in T'$, and if we were able to improve on it for $t \in T'$ that have sufficient weight, that is, if we had that

$$\sum_{t\in T'} \gamma_t^2 \sum_{r\in R', s\in S'} \left|\widehat{F}'_t(r,s)\right|^2 \leqslant (1-\tau) \sum_t \gamma_t^2 = 1-\tau,$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

then we would also be done again in a similar manner. We henceforth assume that

$$\sum_{t\in T'}\gamma_t^2\sum_{r\in R',s\in S'}\left|\widehat{F'_t}(r,s)\right|^2\geqslant 1-\tau. \tag{21}$$

Finally, in a similar way we may assume that

$$\sum_{r\in R',s\in S'}\lambda_r^2\mu_s^2\sum_{t\in T'}\left|\widehat{F_t}(r,s)\right|^2\geqslant(1-\tau)\gamma_{n-1,d-1}^2. \tag{22}$$

High level description of the argument. Taking inequalities (20), (21), and (22) to the extreme, we conclude that the case the argument is tight, that is, that $\lambda_r\mu_s\widehat{F_t}(r,s)$ is proportional to $\gamma_t\overline{\widehat{F'_t}(r,s)}$, $\sum_{t\in T'}\left|\widehat{F_t}(r,s)\right|^2=\gamma_{n-1,d-1}^2$, and $\sum_{r\in R',s\in S'}\left|\widehat{F'_t}(r,s)\right|^2=1$, and use these to reach a contradiction. We show that considering the $n-1$ variate functions $\tilde{g}=\sum_{r\in R'}\lambda_r\psi_r g_r$, $\tilde{h}=\sum_{s\in S'}\mu_s\tilde{\psi}_s h_s$, and $\tilde{F}=\frac{\sum_t\overline{\langle F_t,\tilde{g}\tilde{h}\rangle}F_t}{\sqrt{\sum_t\left|\langle F_t,\tilde{g}\tilde{h}\rangle\right|^2}}$ where ψ_r and $\tilde{\psi}_s$ are random variables that preserve the 2-norm of the function (which can be thought of as random signs for now, though that doesn't quite work), one may make an appropriate choice of the 'signs' so that $\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^I}\left[\tilde{F}(\mathbf{y},\mathbf{z})\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\right]$ exceeds $\gamma_{n-1,d-1}$.

The formal argument. Denote $M_1=\sum_{r\in R'}\lambda_r^2$ and $M_2=\sum_{s\in S'}\mu_s^2$. We shall consider collections of complex numbers $(\psi_r)_{r\in R'}$ and $(\tilde{\psi}_s)_{s\in S'}$ such that

$$\sum_{r\in R'}\lambda_r^2\left|\psi_r\right|^2=M_1,\quad \sum_{s\in S'}\mu_r^2\left|\tilde{\psi}_s\right|^2=M_2. \tag{23}$$

Later on, we will consider a distribution over $(\psi_r)_{r\in R'}$ and $(\tilde{\psi}_s)_{s\in S'}$ satisfying these equations. Namely, we choose the distributions so that the vector $(\lambda_r\psi_r)_{r\in R'}$ is distributed uniformly over vectors in $\mathbb{C}^{R'}$ with 2-norm equal to $\sqrt{M_1}$, and similarly of $(\mu_s\tilde{\psi}_s)_{s\in S'}$ is distributed uniformly over vectors in $\mathbb{C}^{S'}$ of 2-norm equal to $\sqrt{M_2}$. The point of these equations is that defining the functions $\tilde{g}\colon\Gamma^I\to\mathbb{C}$ and $\tilde{h}\colon\Phi^I\to\mathbb{C}$ as above we have that $\|\tilde{g}\|_2^2=M_1\leqslant 1$, $\|\tilde{h}\|_2^2=M_2\leqslant 1$. Thus, $\left|\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^I}\left[\tilde{F}(\mathbf{y},\mathbf{z})\tilde{g}(\mathbf{y})\tilde{h}(\mathbf{z})\right]\right|\leqslant\gamma_{n-1,d-1}$ for all choices of ψ_r and $\tilde{\psi}_s$ satisfying (23). Consider $\tilde{F}=\frac{\sum_t\overline{\langle F_t,\tilde{g}\tilde{h}\rangle}F_t}{\sqrt{\sum_t\left|\langle F_t,\tilde{g}\tilde{h}\rangle\right|^2}}$, and note that $\|\tilde{F}\|_2^2=1$. Thus, we may define $p((\psi_r)_{r\in R'},(\tilde{\psi}_s)_{s\in S'})$ as

$$p((\psi_r)_{r\in R'},(\tilde{\psi}_s)_{s\in S'})=\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^I}\left[\tilde{F}(\mathbf{y},\mathbf{z})\overline{\tilde{g}(\mathbf{y})}\overline{\tilde{h}(\mathbf{z})}\right]^2,$$

and get that

$$p((\psi_r)_{r\in R'},(\tilde{\psi}_s)_{s\in S'})\leqslant\gamma_{n-1,d-1}^2 \tag{24}$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

for every input satisfying (23). On the other hand, we note that

$$\begin{aligned} p((\psi_r)_{r\in R'}, (\tilde{\psi}_s)_{s\in S'}) &= \sum_t \left|\langle F_t, \tilde{g}\tilde{h}\rangle\right|^2 \\ &= \sum_t \left|\sum_{r,s} \lambda_r \psi_r \mu_s \tilde{\psi}_s \widehat{F}_t(r,s)\right|^2 \\ &= \sum_{\substack{r,r'\in R'\\ s,s'\in S'}} \lambda_r\lambda_{r'}\mu_s\mu_{s'}\psi_r\overline{\psi_{r'}}\tilde{\psi}_s\overline{\tilde{\psi}_{s'}} \sum_{t\in T'} \widehat{F}_t(r,s)\overline{\widehat{F}_t(r',s')}, \end{aligned}$$

so $p((\psi_r)_{r\in R'}, (\tilde{\psi}_s)_{s\in S'})$ is a polynomial in its input variables. We summarise this discussion with the following claim.

Claim 7.3. *$p((\psi_r)_{r\in R'}, (\tilde{\psi}_s)_{s\in S'})$ is a real-valued function equal to*

$$\sum_{\substack{r,r'\in R'\\ s,s'\in S'}} \lambda_r\lambda_{r'}\mu_s\mu_{s'}\psi_r\overline{\psi_{r'}}\tilde{\psi}_s\overline{\tilde{\psi}_{s'}} \sum_{t\in T'} \widehat{F}_t(r,s)\overline{\widehat{F}_t(r',s')}$$

that satisfies that $\left|p((\psi_r)_{r\in R'}, (\tilde{\psi}_s)_{s\in S'})\right| \leqslant \gamma^2_{n-1,d-1}$ for every input satisfying (23).

Roughly speaking, in the rest of the proof we will show that the expected value of p over a uniform choice of input satisfying (23) is very close to $\gamma^2_{n-1,d-1}$, hence p is close to being a constant function. On the other hand, we reach a contradiction by directly arguing that the variance of p is large, hence concluding the proof.

For $r \in R'$ and $s \in S'$, define the vector $V_{r,s} \in \mathbb{C}^{T'}$ by $V_{r,s}(t) = \widehat{F}_t(r,s)$. Then we may write p as

$$p((\psi_r)_{r\in R'}, (\tilde{\psi}_s)_{s\in S'}) = \sum_{\substack{r,r'\in R'\\ s,s'\in S'}} \lambda_r\lambda_{r'}\mu_s\mu_{s'}\psi_r\overline{\psi_{r'}}\tilde{\psi}_s\overline{\tilde{\psi}_{s'}}\langle V_{r,s}, V_{r',s'}\rangle. \tag{25}$$

7.3.1 Analysing the expectation of p and upper bounding the variance of p

Claim 7.4. *For all $r \in R'$ and $s \in S'$ we have that $\left|\|V_{r,s}\|_2^2 - \gamma^2_{n-1,d-1}\right| \lesssim \frac{\tau}{\zeta^4}\gamma^2_{n-1,d-1}$.*

Proof. It is clear that $\|V_{r,s}\|_2^2 \leqslant \gamma^2_{n-1,d-1}$ by definition of $\gamma_{n-1,d-1}$, and in the rest of the argument we show the lower bound. From (22) we get that

$$\sum_{r,s} \lambda_r^2\mu_s^2(\gamma^2_{n-1,d-1} - \|V_{r,s}\|_2^2) \leqslant \tau\gamma^2_{n-1,d-1},$$

hence for $r \in R'$ and $s \in S'$ we have that $\left|\gamma^2_{n-1,d-1} - \|V_{r,s}\|_2^2\right| \leqslant \frac{\tau}{\zeta^4}\gamma^2_{n-1,d-1}$. □

As a corollary, we get:

Claim 7.5. $\mathbb{E}_{\psi,\tilde{\psi}}\left[p(\psi,\tilde{\psi})\right] \geqslant (1-\sqrt{\tau})\gamma^2_{n-1,d-1}$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Proof. Note that for $r \neq r'$, the expectation of $\psi_r \psi_{r'}$ is 0, since the distribution of (ψ_r) is invariant under changing a sign of any ψ_r. Thus,

$$\begin{aligned}\mathop{\mathbb{E}}_{\psi,\tilde{\psi}}\left[p(\psi,\tilde{\psi})\right] &= \mathbb{E}\left[\sum_{r\in R', s\in S'} \lambda_r^2 \left|\psi_r\right|^2 \mu_s^2 \left|\tilde{\psi}_s\right|^2 \left\|V_{r,s}\right\|_2^2\right] \\ &\geqslant (1-O_\zeta(\tau))\gamma_{n-1,d-1}^2 \mathbb{E}\left[\sum_{r\in R', s\in S'} \lambda_r^2 \left|\psi_r\right|^2 \mu_s^2 \left|\tilde{\psi}_s\right|^2\right] \\ &= (1-O_\zeta(\tau))\gamma_{n-1,d-1}^2 M_1 M_2.\end{aligned}$$

As $M_1, M_2 \geqslant 1 - O_m(\delta)$, we get that

$$\mathop{\mathbb{E}}_{\psi,\tilde{\psi}}\left[p(\chi,\tilde{\psi})\right] \geqslant (1-O_\zeta(\tau))(1-O_m(\delta))\gamma_{n-1,d-1}^2 \geqslant (1-\sqrt{\tau})\gamma_{n-1,d-1}^2.$$

□

We can now upper bound the variance of p as:

Claim 7.6. *var*$(p) \leqslant 2\sqrt{\tau}\gamma_{n-1,d-1}^4$

Proof. By definition,

$$\mathsf{var}(p) = \mathbb{E}\left[p^2\right] - \mathbb{E}\left[p\right]^2.$$

Note that by (24), $\mathbb{E}\left[p^2\right] \leqslant \gamma_{n-1,d-1}^4$, whereas by Claim 7.5 $\mathbb{E}\left[p\right] \geqslant (1-\sqrt{\tau})\gamma_{n-1,d-1}^2$, and the result follows. □

7.3.2 Lower bounding the variance of p

Inspecting (25) it becomes apparent that to lower bound the variance of p, we must show that the vectors $V_{r,s}$ cannot be mutually orthogonal. We prove the following lemma in the next section. The proof proceeds by showing that if (21) holds and if the vectors are mutually (almost) orthogonal, then this contradicts Claim 7.2. This step crucially uses **the relaxed base case** condition.

Lemma 7.7. *There are* $r, r' \in R'$, *and* $s, s' \in S'$ *such that* $(r,s) \neq (r',s')$ *and* $\left|\langle V_{r,s}, V_{r',s'}\rangle\right| \gtrsim c\gamma_{n-1,d-1}^2$.

We are now ready to prove a lower bound, assuming the above lemma, on the variance of p. Towards this end, we write $p = p_1 + p_2 + p_3 + p_4$ so that

$$p_1 = \sum_{r\in R', s\neq s'\in S'} \lambda_r^2 \mu_s \mu_{s'} \left|\psi_r\right|^2 \tilde{\psi}_s \overline{\tilde{\psi}_{s'}} \langle V_{r,s}, V_{r,s'}\rangle, \quad p_2 = \sum_{r\neq r'\in R', s\in S'} \lambda_r \lambda_{r'} \mu_s^2 \psi_r \overline{\psi_{r'}} \left|\tilde{\psi}_s\right|^2 \langle V_{r,s}, V_{r',s}\rangle,$$

$$p_3 = \sum_{r\neq r'\in R', s\neq s'\in S'} \lambda_r \lambda_{r'} \mu_s \mu_{s'} \psi_r \overline{\psi_{r'}} \tilde{\psi}_s \overline{\tilde{\psi}_{s'}} \langle V_{r,s}, V_{r',s'}\rangle, \quad p_4 = \sum_{r\in R', s\in S'} \lambda_r^2 \mu_s^2 \left|\psi_r\right|^2 \left|\tilde{\psi}_s\right|^2 \langle V_{r,s}, V_{r,s}\rangle.$$

Claim 7.8. *var*$(p) \geqslant \frac{1}{2}\left(\mathbb{E}\left[\left|p_1\right|^2\right] + \mathbb{E}\left[\left|p_2\right|^2\right] + \mathbb{E}\left[\left|p_3\right|^2\right]\right) - \tau^{1/8}\gamma_{n-1,d-1}^4$.

Proof. Note that by Claim 7.4

$$\begin{aligned}\left|p_4 - M_1 M_2 \gamma_{n-1,d-1}^2\right| &\leqslant \sum_{r\in R', s\in S'} \lambda_r^2 \mu_s^2 \left|\psi_r\right|^2 \left|\tilde{\psi}_s\right|^2 \left|\left\|V_{r,s}\right\|_2^2 - \gamma_{n-1,d-1}^2\right| \\ &\lesssim_\zeta \tau \gamma_{n-1,d-1}^2,\end{aligned}$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

and by Claim 7.5 and Claim 7.3

$$\left|\mathbb{E}\left[p\right] - M_1 M_2 \gamma_{n-1,d-1}^2\right| \lesssim \sqrt{\tau}\gamma_{n-1,d-1}^2,$$

so together we get that $\left|\mathbb{E}\left[p\right] - p_4\right| \lesssim \sqrt{\tau}\gamma_{n-1,d-1}^2$. Thus, we get that

$$\begin{aligned}\mathsf{var}(p) &= \mathbb{E}\left[\left|p_1 + p_2 + p_3 + p_4 - \mathbb{E}\left[p\right]\right|^2\right] \\ &= \mathbb{E}\left[\left|p_1 + p_2 + p_3\right|^2\right] + 2\,\mathsf{Re}\left(\mathbb{E}\left[(p_1 + p_2 + p_3)\overline{(p_4 - \mathbb{E}\left[p\right])}\right]\right) + \mathbb{E}\left[\left|p_4 - \mathbb{E}\left[p\right]\right|^2\right].\end{aligned}$$

We bound

$$\begin{aligned}\left|\mathbb{E}\left[(p_1 + p_2 + p_3)(p_4 - \mathbb{E}\left[p\right])\right]\right| &\leqslant \sqrt{\mathbb{E}\left[\left|p_1 + p_2 + p_3\right|^2\right]\mathbb{E}\left[\left|p_4 - \mathbb{E}\left[p\right]\right|^2\right]} \\ &\leqslant \tau^{1/8}\mathbb{E}\left[\left|p_1 + p_2 + p_3\right|^2\right] + \tau^{-1/8}\mathbb{E}\left[\left|p_4 - \mathbb{E}\left[p\right]\right|^2\right],\end{aligned}$$

so

$$\begin{aligned}\mathsf{var}(p) &\geqslant (1 - \tau^{1/8})\mathbb{E}\left[\left|p_1 + p_2 + p_3\right|^2\right] - \tau^{-1/8}\mathbb{E}\left[\left|p_4 - \mathbb{E}\left[p\right]\right|^2\right] \\ &\geqslant (1 - \tau^{1/8})\mathbb{E}\left[\left|p_1 + p_2 + p_3\right|^2\right] - \tau^{1/8}\gamma_{n-1,d-1}^4.\end{aligned}$$

Finally, we note that $\mathbb{E}\left[p_1\overline{p_2}\right] = \mathbb{E}\left[p_1\overline{p_3}\right] = \mathbb{E}\left[p_2\overline{p_3}\right] = 0$ by the invariance of $\psi_r, \tilde{\psi}_s$ to sign change. The result follows. □

Claim 7.9. *var*$(p) \gtrsim_m \zeta^8\gamma_{n-1,d-1}^4$.

Proof. Let $r_1, r_2 \in R'$, and $s_1, s_2 \in S'$ be from Lemma 7.7 so that $(r_1, s_1) \neq (r_2, s_2)$ and $\left|\langle V_{r_1,s_1}, V_{r_2,s_2}\rangle\right| \geqslant c\gamma_{n-1,d-1}^2$. There are several cases, depending on if $r_1 = r_2$, $s_1 = s_2$ or none of them occur.

The case that $r_1 \neq r_2$ and $s_1 \neq s_2$

We get that

$$\mathbb{E}\left[\left|p_3\right|^2\right] = \mathbb{E}\left[\sum_{r\neq r'\in R', s\neq s'\in S'} \lambda_r^2\lambda_{r'}^2\mu_s^2\mu_{s'}^2 \left|\psi_r\right|^2\left|\psi_{r'}\right|^2\left|\tilde{\psi}_s\right|^2\left|\tilde{\psi}_{s'}\right|^2\left|\langle V_{r,s}, V_{r',s'}\rangle\right|^2\right]$$

as the other terms vanish by invariance under changing signs or under multiplying by $\mathbf{i}$. This is at least

$$\gtrsim_m \gamma_{n-1,d-1}^4\mathbb{E}\left[\lambda_{r_1}^2\lambda_{r_2}^2\mu_{s_1}^2\mu_{s_2}^2\left|\psi_{r_1}\right|^2\left|\psi_{r_2}\right|^2\left|\tilde{\psi}_{s_1}\right|^2\left|\tilde{\psi}_{s_2}\right|^2\right],$$

which is at least

$$\gtrsim_m \zeta^8\gamma_{n-1,d-1}^4\mathbb{E}\left[\left|\psi_{r_1}\right|^2\left|\psi_{r_2}\right|^2\left|\tilde{\psi}_{s_1}\right|^2\left|\tilde{\psi}_{s_2}\right|^2\right] = \zeta^8\gamma_{n-1,d-1}^4\mathbb{E}\left[\left|\psi_{r_1}\right|^2\left|\psi_{r_2}\right|^2\right]\mathbb{E}\left[\left|\tilde{\psi}_{s_1}\right|^2\left|\tilde{\psi}_{s_2}\right|^2\right].$$

By Claim 7.13 we get that this is at least $\gtrsim_m \zeta^8\gamma_{n-1,d-1}^4$, so overall we get that $\mathbb{E}\left[\left|p_3\right|^2\right] \gtrsim_m \zeta^8\gamma_{n-1,d-1}^4$. Hence by Claim 7.8 and (15) we get that $\mathsf{var}(p) \gtrsim_m \zeta^8\gamma_{n-1,d-1}^4$.

The case that $r_1 = r_2$ and $s_1 \neq s_2$.

Consider the event E that the variables ψ_r's and $\tilde{\psi}_s$ satisfy that

$$\left|\lambda_{r_1}\psi_{r_1}\right| \geqslant 1 - \frac{c^{100}}{m^{100}}, \qquad \left|\mu_{s_1}\tilde{\psi}_{s_1}\right| \geqslant \frac{1}{\sqrt{2}} - \frac{c^{100}}{m^{100}}, \qquad \left|\mu_{s_2}\tilde{\psi}_{s_2}\right| \geqslant \frac{1}{\sqrt{2}} - \frac{c^{100}}{m^{100}}.$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Note that in that case, for any $r \neq r_1$, $s \neq s_1$ we have that

$$|\lambda_r \psi_r| \leqslant \sqrt{1 - |\lambda_{r_1}\psi_{r_1}|^2} \lesssim \frac{c^{50}}{m^{50}}, \qquad \left|\mu_s \tilde{\psi}_s\right| \leqslant \sqrt{1 - \left|\mu_{s_1}\tilde{\psi}_{s_1}\right|^2 - \left|\mu_{s_2}\tilde{\psi}_{s_2}\right|^2} \lesssim \frac{c^{50}}{m^{50}}.$$

It follows that whenever E holds,

$$\begin{aligned}|p_1| &\geqslant \left|\lambda_{r_1}\mu_{s_1}\mu_{s_2}\psi_{r_1}^2\tilde{\psi}_{s_1}\tilde{\psi}_{s_2}\right| \left|\langle V_{r_1,s_1}, V_{r_2,s_2}\rangle\right| - O\left(m^4 \frac{c^{50}}{m^{50}} \max_{r\in R', s\in S'} \|V_{r,s}\|_2^2\right)\\ &\gtrsim \Omega(c\gamma_{n-1,d-1}^2) - m^4 \frac{c^{50}}{m^{50}}\gamma_{n-1,d-1}^2\\ &\gtrsim c\gamma_{n-1,d-1}^2,\end{aligned}$$

where we used Claim 7.4. It follows that $\mathbb{E}\left[|p_1|^2\right] \gtrsim_c \mathbb{P}[E] c\gamma_{n-1,d-1}^4 \gtrsim_{m,c} \gamma_{n-1,d-1}^4$, where we used the fact that $\mathbb{P}[E] \geqslant \Omega_{m,c}(1)$. The proof is concluded by Claim 7.8.

The case that $r_1 \neq r_2$ and $s_1 = s_2$.

Analogous to the previous case. □

7.3.3 Finishing the proof

Claims 7.6 and 7.9 directly contradict each other by (15). That means that our initial hypothesis is false, that is, not all of (20), (21) and (22) can hold, and therefore as explained in the beginning of Section 7.3, the conclusion of Lemma 7.1 follows.

7.3.4 Showing $\{V_{r,s}\}_{r\in R', s\in S'}$ cannot be roughly orthogonal

This section is devoted to the proof of Lemma 7.7. We assume towards contradiction that this is false, that is, that $\left|\langle V_{r,s}, V_{r',s'}\rangle\right| < c\gamma_{n-1,d-1}^2$ for all $(r,s) \neq (r',s')$.

Claim 7.10. $\sum_{r\in R', s\in S'} \left|\widehat{F'}_t(r,s)\right|^2 \geqslant 1 - \sqrt{\tau}$ *for all* $t \in T'$.

Proof. Note that

$$\sum_{t\in T'} \gamma_t^2 \geqslant 1 - \sum_{t\in T\setminus T'} \gamma_t^2 \geqslant 1 - m^2\delta^2.$$

Thus, by (21) we get that

$$\sum_{t\in T'} \gamma_t^2 \left(1 - \sum_{r\in R', s\in S'} \left|\widehat{F'_t}(r,s)\right|^2\right) \lesssim_m \delta^2 + \tau \lesssim_m \tau.$$

In particular, for all $t \in T'$ we have

$$\gamma_t^2 \left(1 - \sum_{r\in R', s\in S'} \left|\widehat{F'_t}(r,s)\right|^2\right) \lesssim_m \tau,$$

so

$$1 - \sum_{r\in R', s\in S'} \left|\widehat{F'_t}(r,s)\right|^2 \lesssim_m \frac{\tau}{\zeta^2} \leqslant \sqrt{\tau},$$

establishing the claim. □

Claim 7.11. *For all* r, s *we have* $\sum_{t\in T'} \left|\widehat{F'}_t(r,s)\right|^2 > 1 - \nu$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Proof. Assume this is not the case, and that there are r, s such that $\sum_{t\in T'} \left|\widehat{F'}_t(r,s)\right|^2 \leqslant 1-\nu$.

Claim 7.12. $\left|T'\right| \leqslant \left|R'\right| \left|S'\right| - 1$.

Proof. Summing Claim 7.10 over $t \in T'$ yields

$$\left|T'\right| \leqslant \frac{1}{1-\sqrt{\tau}} \sum_{t\in T'} \sum_{r\in R', s\in S'} \left|\widehat{F'_t}(r,s)\right|^2 = \sum_{r\in R', s\in S'} \sum_{t\in T'} \left|\widehat{F'_t}(r,s)\right|^2.$$

For all $r \in R'$ and $s \in S'$, we have that $\sum_{t\in T'} \left|\widehat{F'_t}(r,s)\right|^2 \leqslant 1$, and by assumption for some r, s we have $\sum_{t\in T'} \left|\widehat{F'_t}(r,s)\right|^2 \leqslant 1-\nu$. Hence,

$$\left|T'\right| \leqslant \frac{1}{1-\sqrt{\tau}} \sum_{r\in R', s\in S'} \sum_{t\in T'} \left|\widehat{F'_t}(r,s)\right|^2 \leqslant \frac{\left(\left|R'\right| \left|S'\right| - 1\right) + 1 - \nu}{1-\sqrt{\tau}} = \left|R'\right| \left|S'\right| - \frac{\sqrt{\tau}\left|R'\right| \left|S'\right| - \nu}{1-\sqrt{\tau}}$$
$$< \left|R'\right| \left|S'\right|,$$

as $\sqrt{\tau}\left|R'\right| \left|S'\right| \leqslant m^2\sqrt{\tau} < \nu$. As $\left|T'\right|$ is an integer, the statement of Claim 7.12 follows. □

Thus, applying Lemma 7.14 we get that

$$\sum_{(r,s)\neq(r',s')} \left|\langle \frac{V_{r,s}}{\|V_{r,s}\|_2}, \frac{V_{r',s'}}{\|V_{r',s'}\|_2} \rangle\right|^2 \geqslant \frac{1}{|T'|} \geqslant \frac{1}{m^2},$$

and therefore there are $(r,s) \neq (r',s')$ such that

$$\left|\langle V_{r,s}, V_{r',s'}\rangle\right| \geqslant \frac{1}{m^3}\|V_{r,s}\|_2\|V_{r',s'}\|_2 > c\gamma^2_{n-1,d-1},$$

where we used Claim 7.4. This is a contradiction, hence proving the assertion of Claim 7.11. □

Proof of Lemma 7.7. Let $V'_{r,s} = (\widehat{F'_t}(r,s))_{t\in T'}$, and define

$$\tilde{F}_{r,s} = \frac{\sum_{t\in T'} \overline{\widehat{F'_t}(r,s)} F'_t}{\sqrt{\sum_{t\in T'} \left|\widehat{F'_t}(r,s)\right|^2}}.$$

Note that as $\tilde{F}_{r,s}$ is constant on connected components, we may write $\tilde{F}_{r,s} = W\tilde{f}_{r,s}$. From Claim 7.11, it follows that for all r, s

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^J} \left[\tilde{f}_{r,s}(\mathbf{x}) g'_r(\mathbf{y}) h'_s(\mathbf{z})\right]^2 = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^J} \left[\tilde{F}_{r,s}(\mathbf{y},\mathbf{z}) g'_r(\mathbf{y}) h'_s(\mathbf{z})\right]^2 \geqslant 1-\nu,$$

hence by **the relaxed base case** we get that $\mathsf{var}_{\Sigma'}(\tilde{f}_{r,s}) \leqslant u$. We now show that this implies that $\mathsf{var}_{\Sigma'}(f'_t)$ is small contradicting Claim 7.2.

We have that

$$\|\tilde{F}_{r,s} - g'_{r'}h'_{s'}\|_2^2 = \|\tilde{F}_{r,s}\|_2^2 + \|g'_{r'}h'_{s'}\|_2^2 - 2 \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^J} \left[\tilde{F}_{r,s}(\mathbf{y},\mathbf{z}) g'_r(\mathbf{y}) h'_s(\mathbf{z})\right] \leqslant 2\nu.$$

For all $t \in T'$ we have

$$\|F'_t - \sum_{r,s} \widehat{F'_t}(r,s)\tilde{F}_{r,s}\|_2 \leqslant \|F'_t - \sum_{r,s} \widehat{F'_t}(r,s) g'_r h'_s\|_2 + \|\sum_{r,s} \widehat{F'_t}(r,s)(g'_r h'_s - \tilde{F}_{r,s})\|_2. \tag{26}$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Note that

$$\|F'_t - \sum_{r,s} \widehat{F}'_t(r,s) g'_r h'_s\|_2^2 = \|F'_t\|_2^2 - \sum_{r,s} \left|\widehat{F}'_t(r,s)\right|^2 = 1 - \sum_{r,s} \widehat{F}'_t(r,s)^2 \leqslant \sqrt{\tau},$$

where in the last inequality we used Claim 7.10. Also,

$$\|\sum_{r,s} \widehat{F}'_t(r,s)(g'_r h'_s - \tilde{F}_{r,s})\|_2 \leqslant \sum_{r,s} \left|\widehat{F}'_t(r,s)\right| \|g'_r h'_s - \tilde{F}_{r,s}\|_2 \leqslant \sum_{r,s} \left|\widehat{F}'_t(r,s)\right| 2\nu \leqslant 2m^2\nu.$$

Plugging these bounds into (26) yields $\|F'_t - \sum_{r,s} \widehat{F}'_t(r,s)\tilde{F}_{r,s}\|_2 \leqslant \tau^{1/4} + 2m^2\nu \leqslant \sqrt{\nu}$. Thus, as

$$\mathsf{var}_{\Sigma'}(\sum_{r,s} \widehat{F}'_t(r,s)\tilde{f}_{r,s}) \lesssim_m u,$$

we get that

$$\mathsf{var}_{\Sigma'}(W^{-1}F'_t) \lesssim_m u + \sqrt{\nu}$$

for all $t \in T'$, which contradicts Claim 7.2 as $u, \nu \lesssim w$ due to (15). □

7.4 Auxiliary statements

Claim 7.13. $\mathbb{E}\left[\left|\psi_{r_1}\right|^2 \left|\psi_{r_2}\right|^2\right] \gtrsim_m 1$

Proof. Note that the random vector $(\lambda_r \psi_r)$ is distributed uniformly over the unit sphere in $\mathbb{C}^{R'}$, hence

$$\mathbb{E}\left[\left|\lambda_{r_1}\psi_{r_1}\right|^2 \left|\lambda_{r_2}\psi_{r_2}\right|^2\right] \gtrsim_m 1,$$

and $\mathbb{E}\left[\left|\psi_{r_1}\right|^2 \left|\psi_{r_2}\right|^2\right]$ is at least as large. □

Lemma 7.14. *Suppose we have a collection of unit vectors $v_1, \ldots, v_q \in \mathbb{C}^T$ such that $q > T$. Then*

$$\sum_{i \neq j} \left|\langle v_i, v_j\rangle\right|^2 \geqslant \frac{q(q-T)}{T}.$$

Proof. Consider the matrix M whose (i,j) entry is $\langle v_i, v_j\rangle$. Then M is Hermitian and the rank of M is at most T. Note that

$$\sum_{i,j} \left|\langle v_i, v_j\rangle\right|^2 = \mathsf{Tr}(M^2) = \sum_{\ell=1}^{T} \lambda_\ell(M)^2 \geqslant \frac{1}{T}\left(\sum_{\ell=1}^{T} \lambda_\ell(M)\right)^2 = \frac{1}{T}\mathsf{Tr}(M)^2.$$

Therefore,

$$\sum_{i \neq j} \left|\langle v_i, v_j\rangle\right|^2 \geqslant \frac{1}{T}\mathsf{Tr}(M)^2 - q = \frac{1}{T}q^2 - q = \frac{q(q-T)}{T}.$$

□

8. Proof of reductions

In this section, we prove Lemma 2.1 assuming the correctness of Lemma 2.5. As discussed in the introduction, there are two main differences between these lemma statements. The first difference is that we assume the marginal distribution on (y,z) is uniform and independent. The second difference is in the use of the relaxed base case. In Section 8.1, we show that the marginal distribution

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

on (y, z) can be assumed to be uniform and independent. In Section 8.2 we show that the relaxed base case inequality is sufficient towards proving the main analytical lemma.

We use the following lemma as an intermediate lemma in showing this. Compared to Lemma 2.1, in the following lemma, we get to assume that the distribution $\mu_{y,z}$ is uniform over $\Gamma \times \Phi$.

Lemma 8.1. *For all $\alpha > 0$, $m \in \mathbb{N}$, there exists $\xi > 0$ such that for all $\varepsilon > 0$ there is $\delta > 0$ such that the following holds.*

Suppose $|\Sigma|, |\Gamma|, |\Phi| \leqslant m$ and μ is a distribution over $\Sigma \times \Gamma \times \Phi$ such that (a) the support of μ cannot be linearly embedded, (b) $\mu(x, y, z) \geqslant \alpha$ for all $(x, y, z) \in \mathsf{supp}(\mu)$, (c) the distribution $\mu_{y,z}$ is uniform over $\Gamma \times \Phi$. Then for all 1-bounded functions $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$, and $h\colon \Phi^n \to \mathbb{C}$ satisfying that $\mathsf{Stab}_{1-\xi}(g;\mu_y) \leqslant \delta$ we have that

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

8.1 Lemma 8.1 implies Lemma 2.1

Let f, g, h be functions as in the statement of Lemma 2.1, and suppose without loss of generality that $\mathsf{Stab}_{1-\xi}(g) \leqslant \delta$.

8.1.1 The path trick

The argument herein is virtually identical to the arguments in [[5],Sections 3.1, 3.2, 3.3], but we give it for sake of completeness. We let

$$0 \ll \delta \ll \varepsilon, \xi \ll \alpha, m^{-1} \leqslant 1,$$

and fix f, g, and h as in the statement of the lemma. Without loss of generality we assume that $\mathsf{Stab}_{1-\xi}(g) \leqslant \delta$.

Our first goal will be to reduce to the case the distribution μ has full support over $\Gamma \times \Phi$. Define

$$\tilde{h}(z) = \overline{\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})}\left[f(\mathbf{x}) g(\mathbf{y}) \mid \mathbf{z} = z \right]}.$$

We note that $\tilde{h}\colon \Phi^n \to \mathbb{C}$ is 1-bounded, and also that by Cauchy–Schwarz

$$\left| \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu} \left[f(\mathbf{x}) g(\mathbf{y}) h(\mathbf{z}) \right] \right|^2 = \left| \mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_z} \left[h(\mathbf{z}) \overline{\tilde{h}(\mathbf{z})} \right] \right|^2 \leqslant \|h\|_2^2 \|\tilde{h}\|_2^2 \leqslant \|\tilde{h}\|_2^2 = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu} \left[f(\mathbf{x}) g(\mathbf{y}) \tilde{h}(\mathbf{z}) \right],$$

so it suffices to prove the statement for $f, g, \tilde{h}$. To simplify notations, we henceforth assume that $h = \tilde{h}$ to begin with.

For each $r \geqslant 1$, consider the distribution $\mathcal{D}_r$ over

$$(x^1, x^{1'}, x^2, x^{2'}, \ldots, x^{2^{r-1}}, x^{2^{r-1}{}'}, y^1, y^2, y^3, \ldots, y^{2^{r-1}+1}, z^1, z^2, \ldots, z^{2^{r-1}})$$

defined as follows:

1. Sample $y^1 \sim \mu_y$;
2. Sample $(x^1, z^1) \sim \mu | y^1$;
3. Sample $(x^{1'}, y^2) \sim \mu | z^1$;
4. Iteratively, for $j \leqslant 2^{r-1}$, after sampling y^j we sample $(x^j, z^j) \sim \mu | y^j$;
5. Iteratively, for $j \leqslant 2^{r-1}$, after sampling z^j we sample $(x^{j'}, y^{j+1}) \sim \mu | z^j$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

This distribution can be viewed as a labelled random walk of length 2^r between the y and z side on the bipartite graph on $H = (\Gamma \cup \Phi, E)$ whose directed edges are $E = \mathrm{supp}(\mu_{y,z})$; each directed edge is labelled by an x showing that the 3-tuple is in the support of μ. We note that reversing the order of the random walk, that is, viewing this as a random walk from $y^{2^{r-1}}$ to y^1, yields the same distribution.

Claim 8.2. *For all $r \geqslant 1$ it holds that*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu} \left[\tilde{f}(x) g(y) h(z) \right] \right|^{2^r} \leqslant \mathop{\mathbb{E}}_{(\vec{x},\vec{y},\vec{z})\sim\mathcal{D}_r} \left[F(\vec{x}) g(y^1) \overline{g(y^{2^{r-1}+1})} \right],$$

where $F(\vec{x}) = \prod_{i=1}^{2^{r-1}} f(x^i)\overline{f(x^{i'})}$.

Proof. This is repeated application of Cauchy–Schwarz, and is done by induction on r. For $r = 1$, this is true as

$$\left| \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu} \left[f(\mathbf{x}) g(\mathbf{y}) h(\mathbf{z}) \right] \right|^2 = \left| \mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_z} \left[h(\mathbf{z}) \mathbb{E}_{(\mathbf{x},\mathbf{y},\bar{\mathbf{z}})\sim\mu} \left[f(\mathbf{x}) g(\mathbf{y}) \mid \bar{\mathbf{z}} = \mathbf{z} \right] \right] \right|^2,$$

and using Cauchy–Schwarz over z yields this is at most

$$\mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_z} \left[\left| \mathbb{E}_{(\mathbf{x},\mathbf{y},\bar{\mathbf{z}})\sim\mu} \left[f(\mathbf{x}) g(\mathbf{y}) \mid \bar{\mathbf{z}} = \mathbf{z} \right] \right|^2 \right] = \mathop{\mathbb{E}}_{(\vec{\mathbf{x}},\vec{\mathbf{y}},\vec{\mathbf{z}})\sim\mathcal{D}_1} \left[f(\mathbf{x}^1)\overline{f(\mathbf{x}^{1'})} g(\mathbf{y}^1)\overline{g(\mathbf{y}^2)} \right].$$

Suppose the statement is true for r, and prove for $r+1$. Then by induction hypothesis

$$\begin{aligned} \left| \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu} \left[f(\mathbf{x}) g(\mathbf{y}) h(\mathbf{z}) \right] \right|^{2^{r+1}} &\leqslant \mathop{\mathbb{E}}_{(\vec{\mathbf{x}},\vec{\mathbf{y}},\vec{\mathbf{z}})\sim\mathcal{D}_r} \left[F_r(\vec{\mathbf{x}}) g(\mathbf{y}^1) \overline{g(\mathbf{y}^{2^{r-1}+1})} \right]^2 \\ &= \mathop{\mathbb{E}}_{y^{2^{r-1}+1}} \left[\overline{g(\mathbf{y}^{2^{r-1}+1})} \mathbb{E}_{(\vec{\mathbf{x}},\vec{y},\vec{z})\sim\mathcal{D}_r} \left[F_r(\vec{\mathbf{x}}) g(\mathbf{y}^1) \mid y^{2^{r-1}+1} \right] \right]^2, \end{aligned}$$

where $F_r(\vec{x}) = \prod_{i=1}^{2^{r-1}} f(x^i)\overline{f(x^{i'})}$. Hence by Cauchy–Schwarz this is bounded by

$$\begin{aligned} &\|g\|_2^2 \mathop{\mathbb{E}}_{y^{2^{r-1}+1}} \left[\left| \mathbb{E}_{(\vec{\mathbf{x}},\vec{y},\vec{z})\sim\mathcal{D}_r} \left[F_r(\vec{\mathbf{x}}) g(\mathbf{y}^1) \mid y^{2^{r-1}+1} \right] \right|^2 \right] \\ &\leqslant \mathop{\mathbb{E}}_{y^{2^{r-1}+1}} \left[\mathop{\mathbb{E}}_{\substack{(\vec{\mathbf{x}},\vec{y},\vec{z})\sim\mathcal{D}_r \\ (\vec{\mathbf{x}}'',\vec{y}'',\vec{z}'')\sim\mathcal{D}_r}} \left[F_r(\vec{\mathbf{x}})\overline{F_r(\vec{\mathbf{x}}'')} g(\mathbf{y}^1)\overline{g(\mathbf{y}''^1)} \;\middle|\; \mathbf{y}^{2^{r-1}+1} = \mathbf{y}''^{2^{r-1}+1} = y^{2^{r-1}+1} \right] \right]. \end{aligned}$$

We may view $(\vec{x}, \vec{y}, \vec{z})$ as a random walk starting at $y^{2^{r-1}+1}$, and $(\vec{x}'', \vec{y}'', \vec{z}'')$ as an independently chosen random walk starting at $y^{2^{r-1}+1}$, therefore $(\vec{x},\ \mathrm{reverse}(\vec{x}''), \vec{y},\ \mathrm{reverse}(\vec{y}''), \vec{z},\ \mathrm{reverse}(\vec{z}''))$ describes a random walk of length $2 \cdot 2^{r-1} = 2^r$. We note that $\vec{y}$ and $\mathrm{rev}(\vec{y}'')$ overlap in their starting point, and after removing this overlap we get that this random walk matches the distribution $\mathcal{D}_{r+1}$, so the inductive proof is complete. □

Claim 8.3. *For all $r \geqslant 1$ it holds that*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu} \left[f(x) g(y) h(z) \right] \right|^{2^r} \leqslant \mathop{\mathbb{E}}_{(\vec{x},\vec{y},\vec{z})\sim\mathcal{D}_r} \left[F(\vec{x}) g(y^1) h(z^{2^{r-1}}) \right],$$

where $F(\vec{x}) = \prod_{i=1}^{2^{r-1}} f(x^i) \cdot \prod_{i=1}^{2^{r-1}-1} \overline{f(x'^i)}$.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Proof. The statement follows from applying the previous claim and noting that fixing $z^{2^{r-1}}$, the distribution of $(x^{2^{r-1}\prime}, y^{2^{r-1}+1})$ is $\mu_{x,y}|z = z^{2^{r-1}}$, hence

$$\overline{\mathbb{E}_{x^{2^{r-1}\prime}, y^{2^{r-1}+1}}\left[f(x^{2^{r-1}\prime})g(y^{2^{r-1}+1}) \,\middle|\, z^{2^{r-1}}\right]} = h(z^{2^{r-1}}).$$

□

Claim 8.4. *The graph H is connected.*

Proof. Otherwise, we could write $\Gamma = \Gamma_0 \cup \Gamma_1$, $\Phi = \Phi_0 \cup \Phi_1$ non-trivial partitions and have that there are no edges between Γ_0, Φ_1 as well as between Γ_1, Φ_0. We could then define the embedding over $\mathbb{F}_2$ defined as $\gamma(y) = i$ if $y \in \Gamma_i$, $\phi(z) = i$ if $z \in \Phi_i$ and have that $\gamma(y) + \phi(z) = 0$ in the support of μ, hence μ is linearly embeddable, and contradiction. □

Thus, it follows that taking $r = \lceil 2 + \log m \rceil$, the distribution of $(y^1, z^{2^{r-1}})$ in $\mathcal{D}_r$ has full support over $\Gamma \times \Phi$, and each element in $\Gamma \times \Phi$ has probability at least α^{2^r} in $\mathcal{D}_r$. We define the distribution of

$$\nu = ((x^1, x^{1\prime}, x^2, x^{2\prime} \ldots, x^{2^{r-1}-1}, x^{2^{r-1}-1\prime}, x^{2^{r-1}}), y^1, z^{2^{r-1}})$$

as distribution over $\Sigma^{2^r-1} \times \Gamma \times \Phi$.

Claim 8.5. *The distribution ν is not linearly embeddable.*

Proof. Otherwise, we would have a non-trivial embedding $\sigma\colon \Sigma^{2^r-1} \to G$, $\gamma\colon \Gamma \to G$ and $\phi\colon \Phi \to G$. Hence, at least two of these functions are not constant. Note that for $(x, y, z) \in \mathsf{supp}(\mu)$ we have that $(\vec{x}, y, z) \in \mathsf{supp}(\nu)$ for $\vec{x} = (x, x, \ldots, x)$, so $\sigma'\colon \Sigma \to G$, defined by $\sigma'(x) = \sigma(\vec{x})$ forms an embedding, together with γ and ϕ, of $\mathsf{supp}(\mu)$ to G, and contradiction. □

Thus, moving from μ and f, g, h to ν and F, g, h, we get that in order to prove Lemma 2.1 it suffices to prove the following lemma.

Lemma 8.6. *For all $\alpha > 0$, $m \in \mathbb{N}$ there exists $\xi > 0$ such that for all $\varepsilon > 0$ there is $\delta > 0$ such that the following holds.*

Suppose $|\Sigma|, |\Gamma|, |\Phi| \leqslant m$ and μ is a distribution over $\Sigma \times \Gamma \times \Phi$ such that (a) the support of μ cannot be linearly embedded, (b) $\mu(x, y, z) \geqslant \alpha$ for all $(x, y, z) \in \mathsf{supp}(\mu)$, (c) the support of $\mu_{y,z}$ is $\Gamma \times \Phi$. Then for all 1-bounded functions $f\colon \Sigma^n \to \mathbb{C}$, $g\colon \Gamma^n \to \mathbb{C}$ and $h\colon \Phi^n \to \mathbb{C}$ satisfying that $\mathsf{Stab}_{1-\xi}(g; \mu_y) \leqslant \delta$ we have that

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

Next, we reduce proving the above lemma to proving Lemma 8.1 in which we get to assume that the distribution $\mu_{y,z}$ is uniform over $\Gamma \times \Phi$.

Lemma 8.1 implies Lemma 8.6. Let f, g, h and μ be as in Lemma 8.6. We may write $\mu = (1 - s)\nu_1 + s\nu_2$, where $s \gtrsim_1 \alpha$ and ν_1, ν_2 have the same support as μ, the probability of each atom in them is at least $\alpha'(\alpha) > 0$ and $(\nu_2)_{y,z}$ is uniform over $\Gamma \times \Phi$. Thus, we choose a subset $I \subseteq [n]$ by including each element in it with probability $1 - s$, and sampling $(\mathbf{x}_I, \mathbf{y}_I, \mathbf{z}_I) \sim \nu_1^I$, $(\mathbf{x}_{\bar{I}}, \mathbf{y}_{\bar{I}}, \mathbf{z}_{\bar{I}}) \sim \nu_2^I$ yields a sample of $\mu^{\otimes n}$. We may thus write

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z}) \sim \mu^{\otimes n}} \left[f(\mathbf{x}) g(\mathbf{y}) h(\mathbf{z}) \right] = \mathop{\mathbb{E}}_{(\mathbf{x}_I,\mathbf{y}_I,\mathbf{z}_I) \sim \nu_1^{\otimes n}} \left[\mathop{\mathbb{E}}_{(\mathbf{x}',\mathbf{y}',\mathbf{z}') \sim \nu_2^{\bar{I}}} \left[f'(\mathbf{x}') g'(\mathbf{y}') h'(\mathbf{z}') \right] \right], \tag{27}$$

where

$$f' = f_{I \to \mathbf{x}_I}, \quad g' = g_{I \to \mathbf{y}_I}, \quad h' = h_{I \to \mathbf{z}_I}.$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

We note that by Lemma 3.15

$$\mathop{\mathbb{E}}_{\mathbf{y}_I}\left[\mathsf{Stab}_{1-\xi'}(g';\nu_1^{\bar{I}})\right] \leqslant \mathsf{Stab}_{1-c\xi'}(g)$$

for $c = c(s,m,\alpha) > 0$. Hence, taking $\xi' = \xi/c$ we get that $\mathbb{E}_{\mathbf{y}_I}\left[\mathsf{Stab}_{1-\xi'}(g';\nu_1^{\bar{I}})\right] \leqslant \delta$, and by Markov we get that

$$\mathbb{P}_{\mathbf{y}_I}\left[\mathsf{Stab}_{1-\xi'}(g';\nu_1^{\bar{I}}) \geqslant \sqrt{\delta}\right] \leqslant \sqrt{\delta}.$$

Using that in (27), along with the trivial bound $\left|\mathbb{E}_{(\mathbf{x}',\mathbf{y}',\mathbf{z}')\sim\nu_2^{\bar{I}}}\left[f'(\mathbf{x}')g'(\mathbf{y}')h'(\mathbf{z}')\right]\right| \leqslant 1$ for all x_I, y_I, z_I yields that

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \sqrt{\delta} + \mathop{\mathbb{E}}_{(\mathbf{x}_I,\mathbf{y}_I,\mathbf{z}_I)\sim\nu_1^{\otimes n}}\left[\left|\mathop{\mathbb{E}}_{(\mathbf{x}',\mathbf{y}',\mathbf{z}')\sim\nu_2^{\bar{I}}}\left[f'(\mathbf{x}')g'(\mathbf{y}')h'(\mathbf{z}')\right]1_{\mathsf{Stab}_{1-\xi'}(g';\nu_1^{\bar{I}})\leqslant\sqrt{\delta}}\right|\right]$$

Using Lemma 8.1, for sufficiently small δ the inner expectation is at most $\varepsilon/2$, hence restricting to $\delta \leqslant \varepsilon^2/4$ gives $\left|\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \varepsilon$, and the proof is complete. □

8.2 Lemma 2.5 implies Lemma 8.1

The remainder of this section is devoted to showing that Lemma 2.5 implies Lemma 8.1, and we begin with a remark. Note that in the statement of Lemma 8.1, the roles of x and y are exchangeable. Namely, to prove it, it is enough to prove that the conclusion of it holds for a distribution μ that satisfies (a), (b), and that $\mu_{x,y}$ is uniform; we assume this henceforth. We fix functions f, g, h as in the statement of Lemma 8.1, and we prove that the conclusion holds.

8.2.1 Defining the distribution $\tilde{\nu}$

We take r a large enough constant, and consider the distribution $\mathcal{D}_r$ resulting from μ by performing the path-trick, so that by Claim 8.3 we have that for some bounded functions $F, \tilde{h}$ it holds that

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right|^{2^r} \leqslant \left|\mathop{\mathbb{E}}_{(\vec{\mathbf{x}},\mathbf{y},\mathbf{z})\sim\mathcal{D}_r}\left[F(\vec{\mathbf{x}})g(\mathbf{y})\tilde{h}(\mathbf{z})\right]\right|; \tag{28}$$

here, we think of $\mathcal{D}_r$ as only recording the sequence of x's used, as well as y^1, $z^{2^{r-1}}$. Thus, as observed above, provided that r is large enough with respect to m, we get that in $\mathcal{D}_r$ the support of (y,z) is full; we fix such r.

Denote $\tilde{\Sigma} = \Sigma^{2^r}$, and define a graph X on $\tilde{\Sigma}$ where $\vec{x}$ and $\vec{x}'$ are adjacent if there is $(y,z) \in \Gamma \times \Phi$ such that $(\vec{x},y,z)$ and $(\vec{x}',y,z)$ are both in the support of $\mathcal{D}_r$. We look at the connected components of X, and from each one of them we pick a representative arbitrarily; denote by $\mathsf{rep}(\vec{x})$ the chosen representative form the connected component of $\vec{x}$, and let

$$\Sigma_{\mathsf{fin}} = \left\{\mathsf{rep}(\vec{x}) \mid \vec{x} \in \tilde{\Sigma}\right\} \subseteq \tilde{\Sigma}$$

denote the set of these representatives.

Next, we define a distribution $\tilde{\nu}$. The distribution $\tilde{\nu}$ first samples $(y,z) \in \Gamma \times \Phi$ uniformly, and then takes the unique $\vec{x} \in \Sigma_{\mathsf{fin}}$ such that $(\vec{x},y,z) \in \mathsf{supp}(\mathcal{D}_r)$, and outputs $(\vec{x},y,z)$. It is clear that the probability of each atom is at least $\alpha'(\alpha,m) > 0$, that $\tilde{\nu}_{y,z}$ is uniform and that in $\tilde{\nu}$, y, z implies x. It is also easy to observe that as by Claim 8.5 the support of ν cannot be linearly embedded, the support of $\tilde{\nu}$ can also not be linearly embedded.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

In the remainder of the proof, we will show that $\tilde{\nu}$ satisfies the relaxed base case, and therefore we may apply Lemma 2.5 to it. We then use a random-restriction argument to deduce Lemma 8.1.

Lemma 8.7. *The distribution $\tilde{\nu}$ satisfies the relaxed base case.*

Proof. Since $\mathsf{supp}(\tilde{\nu}) = \mathsf{supp}(\nu)$, Claim 8.5 implies that $\tilde{\nu}$ cannot be linearly embedded. The rest of the proof is deferred to Sections 8.3, 8.4. □

Together from Claim 8.13 and Lemma 8.7 we get that $\tilde{\nu}$ satisfies the relaxed base case. Therefore, we can apply Lemma 2.5 on it. In the next section, we show how to do so and establish Lemma 8.1.

8.2.2 Proving lemma 8.1: the merging and the random restriction arguments

Let

$$0 \ll \delta \ll \delta' \ll \varepsilon' \ll \gamma \ll \varepsilon \ll \xi \ll \xi' \ll \alpha, m^{-1} \leqslant 1;$$

we show that $\left|\mathbb{E}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \varepsilon$. Indeed, towards contradiction to Lemma 8.1, assume that

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| > \varepsilon. \tag{29}$$

Thus, by (28) we get that

$$\left|\mathop{\mathbb{E}}_{(\vec{\mathbf{x}},\mathbf{y},\mathbf{z})\sim\mathcal{D}_r^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \geqslant \varepsilon^{2^r}.$$

Consider the distribution $\mathcal{D}'_r$ over $\Sigma_{\mathsf{fin}} \times \Gamma \times \Phi$, defined by taking a sample $(\vec{x}, y, z) \sim \mathcal{D}_r$ and outputting $(\mathsf{rep}(\vec{x}), y, z)$.

Claim 8.8. *There is $n' \geqslant \gamma n$ and functions 1-bounded functions $f' \colon \Sigma_{fin}^{n'} \to \mathbb{C}$, $g' \colon \Gamma^{n'} \to \mathbb{C}$, $h' \colon \Phi^{n'} \to \mathbb{C}$ such that:*

1. $\left|\mathbb{E}_{(\vec{\boldsymbol{x}},\boldsymbol{y},\boldsymbol{z})\sim\mathcal{D}_r'^{\otimes n'}}\left[f'(\vec{\boldsymbol{x}})g(\boldsymbol{y})h(\boldsymbol{z})\right]\right| \geqslant \varepsilon'$.
2. $\mathit{Stab}_{1-\xi'}(g') \leqslant \delta'$.

Proof. This is a direct consequence of [[5], Lemma 3.10] adapted to the case of complex-valued functions in the obvious way. In the notation therein, Σ' represents the set of connected components of the graph X we defined, and noting the resulting distribution there (μ' in the notation therein) is exactly $\mathcal{D}'_r$. □

We fix f', g', h' from the above claim henceforth.

Note that for some $s \gg_{\alpha,m} 1$ we may write $\mathcal{D}'_r = (1-s)\mathcal{D} + s\tilde{\nu}$, where $\mathcal{D}$ is a distribution and $\tilde{\nu}$ is the distribution we have defined earlier; this is true since the probability of each atom in $\mathcal{D}'_r$ is at least $\alpha'(\alpha, m) > 0$ and $\mathsf{supp}(\tilde{\nu}) = \mathsf{supp}(\mathcal{D}'_r)$, so $\mathcal{D}'_r - s\nu$ is non-negative for $s = \alpha'/2$, and re-normalising it we get the distribution $\mathcal{D}$.

Consider a random restriction on f', g', h' as follows: sample $I \subseteq [n]$ by including each element with probability $(1-s)$, sample $(x, y, z) \sim \mathcal{D}^I$, and define $f'' \colon \Sigma^{[n]\setminus I} \to [-1, 1]$, $g'' \colon \Gamma^{[n]\setminus I} \to [-1, 1]$, $h'' \colon \Phi^{[n]\setminus I} \to [-1, 1]$ as

$$f''(\mathbf{x}') = f'_{I\to\mathbf{x}}(\mathbf{x}'), \quad g''(\mathbf{y}') = g'_{I\to\mathbf{y}}(\mathbf{y}'), \quad h''(\mathbf{z}') = h'_{I\to\mathbf{z}}(\mathbf{z}').$$

Consider the following events:

1. E_1: $\mathsf{Stab}_{1-\xi'/cs}(g''; \tilde{\nu}_y) \leqslant \sqrt{\delta}$ for some $c = c(m, \alpha) > 0$;

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

2. E_2: $\left|\mathbb{E}_{(\mathbf{x}',\mathbf{y}',\mathbf{z}')\sim\tilde{\nu}^{[n]\setminus I}}\left[f''(\mathbf{x}')g''(\mathbf{y}')h''(\mathbf{z}')\right]\right| \geqslant \frac{\varepsilon}{2}$;
3. E_3: $|[n]\setminus I| \geqslant \frac{s}{2}n$.

Claim 8.9. $\mathbb{P}[E_1] \geqslant 1-\sqrt{\delta}$.

Proof. Noting that by Lemma 3.15 we have that $\mathbb{E}_y\left[\mathsf{Stab}_{1-\xi'/sc}(g'')\right] \leqslant \mathsf{Stab}_{1-\xi'}(g') \leqslant \delta$, the result follows by Markvo's inequality. □

Claim 8.10. $\mathbb{P}[E_2] \geqslant \frac{\varepsilon}{2}$.

Proof. Setting $T = \left|\mathbb{E}_{(\mathbf{x}',\mathbf{y}',\mathbf{z}')\sim\tilde{\nu}^{[n]\setminus I}}\left[f''(\mathbf{x}')g''(\mathbf{y}')h''(\mathbf{z}')\right]\right|$, the expected value of T is at least ε, and $0 \leqslant T \leqslant 1$ always, hence we get that $T \geqslant \varepsilon/2$ with probability at least $\varepsilon/2$. □

Claim 8.11. $\mathbb{P}[E_3] \geqslant 1-o_n(1) \geqslant 1-\frac{\varepsilon}{4}$.

Proof. This is an immediate consequence of Chernoff's bound. □

Thus, $\mathbb{P}[E_1 \cap E_2 \cap E_3] > 0$, so we may fix x, y, z that satisfy these events. Then

$$\left|\mathbb{E}_{(\mathbf{x}',\mathbf{y}',\mathbf{z}')\sim\tilde{\nu}^{[n]\setminus I}}\left[f''(\mathbf{x}')g''(\mathbf{y}')h''(\mathbf{z}')\right]\right| \geqslant \frac{\varepsilon}{2},$$

however by Lemma 2.5 the last expression tends to 0 as δ' goes to 0, and as $\delta' \ll \varepsilon$ we get a contradiction to the assumption 29. This shows that Lemma 8.1 follows from Lemma 2.5.

8.3 Proof of Lemma 8.7: set up and the compactness argument

Consider $\mathcal{F} \subseteq P(\Sigma_{\mathsf{fin}})$ defined as follows: we have $F \in \mathcal{F}$ if there are functions $f\colon \Sigma_{\mathsf{fin}} \to \mathbb{C}$, $g\colon \Gamma \to \mathbb{C}, h\colon \Phi \to \mathbb{C}$ not all constant such that $f(\vec{x}) = g(y)h(z)$ in the support of $\tilde{\nu}$, and $\mathsf{supp}(f) = F$. Note that the collection $\mathcal{F}$ is closed under intersection, since if we have $F_1 \in \mathcal{F}$ using the functions f_1, g_1, h_1 and $F_2 \in \mathcal{F}$ using the functions f_2, g_2, h_2, then we have $F_1 \cap F_2 \in \mathcal{F}$ using the functions $f_1 f_2, g_1 g_2, h_1 h_2$.

Thus, we may consider minimal sets in $\mathcal{F}$, namely sets $\emptyset \neq F \in \mathcal{F}$ that do not strictly contain any set from $\mathcal{F}$. Let $F_1, \ldots, F_s$ be all minimal sets in $\mathcal{F}$. For $x \in \Sigma$, define $a(x) = (x, x, \ldots, x) \in \tilde{\Sigma}$, and let $b(x) \in \Sigma_{\mathsf{fin}}$ be the representative symbol from the connected component of $a(x)$, that is, $\mathsf{rep}(a(x))$.

Claim 8.12. *For all $x \in \Sigma$, $b(x) \notin F_1 \cup \ldots \cup F_s$.*

Proof. Suppose towards contradiction that $b(x^\star) \in F_1 \cup \ldots \cup F_s$ for some $x^\star \in \Sigma$, so that we may find functions f, g, h such that $f(\vec{x}) = g(y)h(z)$ on $\mathsf{supp}(\tilde{\nu})$, and additionally $f(b(x^\star)) \neq 0$. Note that since f, g, h cannot be all constant, it follows that either g or h must be non-constant.

Define $f'\colon \Sigma \to \mathbb{C}$ by $f'(x) = f(b(x))$, and note that it follows that $f'(x) = g(y)h(z)$ on $\mathsf{supp}(\mu)$. Also, $f'(x^\star) \neq 0$, so f' is not identically 0. Note that since the support of (x, y) in $\mathsf{supp}(\mu)$ is full, it follows that the function g can never vanish on Γ (since for all y, there is some z such that $(x^\star, y, z) \in \mathsf{supp}(\mu)$). This implies that $f'(x) \neq 0$ iff $h(z) \neq 0$ for every $(x, y, z) \in \mathsf{supp}(\mu)$. We now consider two cases:

1. If the function f' vanishes sometimes, that is, $\mathsf{supp}(f') \subsetneq \Sigma$, then we define $\sigma(x) = 1_{f'(x)=0}$, $\phi(z) = 1_{h(z)=0}$, and note that in $\mathsf{supp}(\mu)$ we have that $\sigma(x) + \phi(z) = 0 \pmod 2$ for all $(x, y, z) \in \mathsf{supp}(\mu)$, and that moreover this is a non-trivial embedding in $(\mathbb{F}_2, +)$. This is a contradiction to the fact $\mathsf{supp}(\mu)$ is not linearly embeddable.
2. Else, $\mathsf{supp}(f') = \Sigma$, and it follows that g and h also never vanish. If the argument of any of f', g or h is not constant, then we get that $\sigma(x) = \mathsf{arg}(f(x))$, $\gamma(y) = \mathsf{arg}(g(y))$, $\phi(z) = \mathsf{arg}(h(z))$ form a linear embedding of $\mathsf{supp}(\mu)$ into $([0, 2\pi), + \pmod{2\pi})$ and by Claim

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

3.16 it follows that $\mathsf{supp}(\mu)$ can be linearly embedded, in contradiction. Else, we may assume that the argument of each one of f', g and h is always the same, and by multiplying them by an appropriate complex number if necessary we may assume that they are all positive. By multiplying them by a large enough constant, we may assume that their range is $(1, 2^M)$ for some absolute constant M, hence

$$\sigma(x) = \log(f(x)), \quad \gamma(y) = \log(g(y)), \quad \phi(z) = \log(h(z))$$

forms a non-trivial embedding of $\mathsf{supp}(\mu)$ in $((0, M), \pmod M)$, and by Claim 3.16 it follows that $\mathsf{supp}(\mu)$ can be linearly embedded, in contradiction. □

We take $\Sigma' = \left\{ b(x) \,\middle|\, x \in \Sigma \right\}$, which is disjoint from $F_1 \cup \ldots \cup F_s$, and observe that if f, g, h satisfy that $f(x) = g(y)h(z)$, then $f|_{\Sigma'} \equiv 0$; indeed, otherwise the collection $\mathcal{F}$ would contain $\mathsf{supp}(f)$ which intersects Σ', hence there would be some minimal set in $\mathcal{F}$ intersecting Σ', and contradiction to the choice of Σ'.

Claim 8.13. *The set* $S = \left\{(\vec{x}, y, z) \in \mathsf{supp}(\tilde{\nu}) \,\middle|\, \vec{x} \in \Sigma'\right\}$ *cannot be linearly embedded.*

Proof. Otherwise there would be an Abelian group $(G, +)$ and embeddings $\sigma\colon \Sigma' \to G$, $\gamma\colon \Gamma \to G$ and $\phi\colon \Phi \to G$ not all constant such that $\sigma(\vec{x}) + \gamma(y) + \phi(z) = 0$ in S. Note that γ, ϕ cannot both be constant, and that defining $\sigma'(x) = \sigma(b(x))$ we get that σ', γ, ϕ form a non-trivial embedding of $\mathsf{supp}(\mu)$ to G, and contradiction. □

We move on to prove the heart of the relaxed base case, asserting that if $\mathbb{E}_{x,x' \in_R \Sigma'}\left[\left|f(x) - f(x')\right|^2\right] \geqslant \tau$, and f, g, h have 2-norm equal to 1, then $\left|\mathbb{E}_{(x,y,z)\sim\tilde{\nu}}\left[f(x)g(y)h(z)\right]\right|$ is bounded away from 1; moreover, the gap from 1 is at least polynomial in τ. We begin the proof of this assertion with the following claim, that employs a compactness argument and handles the case that $\tau > c$, for c which is an absolute constant depending on the alphabet sizes. After that, we will handle the case that τ is small by a closer inspection of the compactness argument.

Claim 8.14. *For all* $\tau > 0$, *there exists* $\lambda \in (0, 1)$ *such that if* $\mathbb{E}_{x,x' \in_R \Sigma'}\left[\left|f(x) - f(x')\right|^2\right] \geqslant \tau \|f\|_2^2$, *then*

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

Proof. Assume that the statement is false. Thus, we may find sequences of functions f_ℓ, g_ℓ, h_ℓ of 2-norm 1 such that $\mathbb{E}_{x,x' \in_R \Sigma'}\left[\left|f_\ell(x) - f_\ell(x')\right|^2\right] \geqslant \tau$ for all ℓ, and

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\tilde{\nu}}\left[f_\ell(x)g_\ell(y)h_\ell(z)\right]\right| \geqslant 1 - \frac{1}{\ell}.$$

Passing to limit, we find f, g, h of 2-norm 1 such that

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\tilde{\nu}}\left[f_\ell(x)g_\ell(y)h_\ell(z)\right]\right| \geqslant 1$$

and $\mathbb{E}_{x,x' \in_R \Sigma'}\left[\left|f(x) - f(x')\right|^2\right] \geqslant \tau$, and we next show that $f(x) = g(y)h(z)$ (or $-f(x) = g(y)h(z)$) in $\mathsf{supp}(\tilde{\nu})$, thereby reaching a contradiction (as $f|_{\Sigma'} \not\equiv 0$). Indeed, by Cauchy-Schwarz we have

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\tilde{\nu}}\left[f(x)g(y)h(z)\right]\right|^2 \leqslant \mathop{\mathbb{E}}_{(x,y,z)\sim\tilde{\nu}}\left[\left|f(x)\right|^2\right] \mathop{\mathbb{E}}_{(x,y,z)\sim\tilde{\nu}}\left[\left|g(y)\right|^2 \left|h(z)\right|^2\right] = \|f\|_2^2 \|g\|_2^2 \|h\|_2^2 = 1$$

where we used the fact that y, z are independent in $\tilde{\nu}$. Thus, we get that the above Cauchy-Schwarz inequality is tight, and so f is proportional to $\overline{gh}$. Considering the 2-norms of these functions, we

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

get that $f(x) = \theta\overline{g(y)h(z)}$ for some complex number θ of absolute value 1; without loss of generality we assume that $\theta = 1$ hence $f(x) = g(y)h(z)$. Note that as $\mathbb{E}_{x,x' \in_R \Sigma'}\left[\left|f(x) - f(x')\right|^2\right] \geqslant \tau$, f is not constant, so by definition we get that $\mathsf{supp}(f) \in \mathcal{F}$, and hence $f|_{\Sigma'} \equiv 0$, and contradiction to $\mathbb{E}_{x,x' \in_R \Sigma'}\left[\left|f(x) - f(x')\right|^2\right] \geqslant \tau$. □

8.4 Proof of Lemma 8.7: unravelling compactness

We now move on to handle the case τ is sufficiently small. We prove:

Lemma 8.15. *There is $c = c(m, \alpha) > 0$, such that for all $0 < \tau \leqslant c$, if $\mathbb{E}_{x,x' \in_R \Sigma'}\left[\left|f(x) - f(x')\right|^2\right] \geqslant \tau \|f\|_2^2$, then*

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\tilde{\nu}}\left[f(x)g(y)h(z)\right]\right| \leqslant (1 - \tau^{50m})\|f\|_2\|g\|_2\|h\|_2.$$

The rest of this section is devoted to the proof of Lemma 8.15.

Fix f, g and h as in the statement of the claim, and assume without loss of generality that their 2-norms are all 1. Let ν' be the distribution of $(x, y, z) \sim \tilde{\nu}$ conditioned on $x \in \Sigma'$. We write

$$(I) = \mathop{\mathbb{E}}_{(x,y,z)\sim\tilde{\nu}}\left[f(x)g(y)h(z)\right],$$

and assume towards contradiction that $|(I)| \geqslant 1 - \tau^{50m}$. Multiplying f by a suitable complex number of absolute value 1 we may assume that (I) is a non-negative real number, and hence $(I) \geqslant 1 - \tau^{50m}$.

Claim 8.16. $\mathbb{E}_{x \in_R \Sigma'}\left[\left|f(x)\right|^2\right] \gtrsim_m \tau$.

Proof. Follows as $\mathbb{E}_{x,x' \in_R \Sigma'}\left[\left|f(x) - f(x')\right|^2\right] \geqslant \tau$. □

Claim 8.17. *We have that $\left|f(x) - \overline{g(y)h(z)}\right| \leqslant \tau^{20m}$ for all $(x, y, z) \in \mathsf{supp}(\nu)$.*

Proof. Note that

$$(I) = 1 - \frac{1}{2}\mathop{\mathbb{E}}_{(x,y,z)\sim\tilde{\nu}}\left[\left|f(x) - \overline{g(y)h(z)}\right|^2\right].$$

Thus, if the conclusion of the claim fails, then $|(I)| \leqslant 1 - \Omega_{m,\alpha}(\tau^{40m}) < 1 - \tau^{50m}$, and contradiction. □

Claim 8.18. *We have that $\left|g(y)\right| \geqslant \tau$ for all $y \in \Gamma$.*

Proof. Assume otherwise, that is, that $\left|g(y^\star)\right| \leqslant \tau$ for some $y^\star$. Note that $\|h\|_\infty \leqslant \sqrt{m}\|h\|_2$ as the probability of each atom in h is at least $1/m$, so we get that $\left|g(y^\star)h(z)\right| \leqslant \sqrt{m}\tau$ for all z. Note that the support of the distribution $\nu'_{x,z}$ is $\Sigma' \times \Gamma$; this is because it contains $\left\{(b(x), y) \,\middle|\, (x, y) \in \mathsf{supp}(\mu_{x,y})\right\}$ and $\mu_{x,y}$ is full, so this is the same as $\Sigma' \times \Gamma$. Thus, for every $x \in \Sigma'$ we may find $z \in \Phi$ such that $(x, y^\star, z) \in \mathsf{supp}(\tilde{\nu})$, hence by Claim 8.17

$$\left|f(x)\right| \leqslant \left|g(y^\star)h(z)\right| + \tau^{20m} \leqslant O_m(\tau).$$

On the other hand, by Claim 8.16 we have $\mathbb{E}_{x \in \Sigma'}\left[\left|f(x)\right|^2\right] \gtrsim_{m,\alpha} \tau$, so we may find x such that the left hand side is at least $\Omega_{m,\alpha}(\sqrt{\tau})$. We thus get a contradiction to the fact that $\tau \leqslant c$, for small enough c depending only on m and α. □

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Consider the interval $(0, 1)$, and in it define the intervals $I_j = [\tau^{3(j+1)}, \tau^{3j})$ for $j = 0, 1, \ldots$. We say an interval I_j is free if it doesn't contain any point from either $\mathsf{Image}(|f|)$ or $\mathsf{Image}(|h|)$. Note that as the intervals I_j are disjoint and each one of these sets has size at most m, we may find $j \in \{0, 1, \ldots, 2m\}$ such that I_j is free, and we fix such j henceforth.

Claim 8.19. *We have that* $|f(x)| \geqslant \tau^{3j+1.5}$ *for all* $x \in \Sigma_{fin}$ *and* $|h(z)| \geqslant \tau^{3j+1.5}$ *for all* $z \in \Phi$.

Proof. Assume otherwise, and define

$$\sigma(x) = 1_{|f(x)| \geqslant \tau^{3j+1.5}}, \quad \phi(z) = 1_{|h(z)| \geqslant \tau^{3j+1.5}}.$$

Then by our assumption, at least one of σ, ϕ are not identically 1, say σ without loss of generality. Note that since $\|f\|_{2,\tilde{\nu}_x} = 1$, there is some x such that $|f(x)| \geqslant 1 > \tau$, so σ is also not constantly 0, hence σ is not constant. We next show that $\sigma(x) + \phi(z) = 0 \pmod 2$ for all $(x, y, z) \in \mathsf{supp}(\tilde{\nu})$, hence we conclude that $\mathsf{supp}(\nu)$ has a non-trivial Abelian embedding in $(\mathbb{F}_2, +)$, and contradiction to the fact $\tilde{\nu}$ cannot be linearly embedded.

The case that $\sigma(x) = 1$.

Suppose that $(x, y, z) \in \mathsf{supp}(\tilde{\nu})$ are such that $\sigma(x) = 1$. It follows from Claim 8.17 that

$$|g(y)h(z)| \geqslant |f(x)| - \tau^{10m} \geqslant \tau^{3j+1.5} - \tau^{10m},$$

and as $|g(y)| \leqslant \sqrt{m}\|g\|_{2,\tilde{\nu}_y} = \sqrt{m}$, it follows that

$$|h(z)| \geqslant \frac{\tau^{3j+1.5} - \tau^{10m}}{\sqrt{m}} \geqslant \tau^{3j+2},$$

where we used the fact that $\tau < c$ is small enough, and $j \leqslant 2m$. Thus, as I_j is free, it follows that $|h(z)| \geqslant \tau^{3j}$, and so $\phi(z) = 1$.

The case that $\sigma(x) = 0$.

Suppose that $(x, y, z) \in \mathsf{supp}(\tilde{\nu})$ are such that $\sigma(x) = 0$. It follows from Claim 8.17 that

$$|g(y)h(z)| \leqslant |f(x)| + \tau^{10m} \leqslant \tau^{3j+1.5} + \tau^{10m},$$

and as $|g(y)| \geqslant \tau$ by Claim 8.18, it follows that

$$|h(z)| \leqslant \frac{\tau^{3j+1.5} + \tau^{10m}}{\tau} \leqslant \tau^{3j+0.4},$$

where we used the fact that $\tau < c$ is small enough, and $j \leqslant 2m$. Thus, as I_j is free, it follows that $|h(z)| < \tau^{3(j+1)}$, and so $\phi(z) = 0$. □

We get that for all $(x, y, z) \in \mathsf{supp}(\tilde{\nu})$ it holds that

$$\left|\frac{\overline{f(x)}}{g(y)h(z)} - 1\right| = \frac{\left|\overline{f(x)} - g(y)h(z)\right|}{|g(y)h(z)|} \leqslant \frac{\tau^{20m}}{\Omega_{\alpha,m}(\sqrt{\tau})\tau^{3j+1.5}} \leqslant \tau^{10m}.$$

As the function f, g, and h do not vanish, we may choose a branch of the logarithm function and define

$$f'(x) = \log(\overline{f(x)}), \quad g'(y) = \log(g(y)), \quad h'(z) = \log(h(z)).$$

Define $d(a, b) = \min_{m \in \mathbb{Z}} |a - b - 2\pi \mathbf{i} m|$. Then

$$|d(f'(x), g'(y) + h'(z))| = d(\log\left(\frac{\overline{f(x)}}{g(y)h(z)}\right), 0) \leqslant \left|\frac{\overline{f(x)}}{g(y)h(z)} - 1\right| \leqslant \tau^{10m} \tag{30}$$

for all $(x, y, z) \in \mathsf{supp}(\tilde{\nu})$. Thus, f', g', and h' are approximate embeddings, and we next show how to extract proper embeddings from them in an Abelian group. First though, we argue that at least

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

one of them is far from constant; indeed, as there are $x, x' \in \Sigma'$ such that $|f(x) - f(x')| \geqslant \sqrt{\tau}$ it follows that $\left|e^{f'(x)} - e^{f'(x')}\right| \geqslant \sqrt{\tau}$. As all values of f' are at most $O_{\alpha,m}(1)$ in absolute value and $s \to e^s$ is $O_{\alpha,m}(1)$ Lipschitz in that range, we get that $d(f'(x), f'(x')) \geqslant \Omega_{\alpha,m}(\sqrt{\tau})$. We now extract the proper embeddings σ, γ, ϕ, and for that we work with the real part and the imaginary part separately depending on whether $\mathsf{Re}(f')$ is at least $\Omega_{\alpha,m}(\sqrt{\tau})$ far from constant, or $\mathsf{Im}(f')$ is at least $\Omega_{\alpha,m}(\sqrt{\tau})$ far from constant.

8.4.1 The case that the real part of f is far from constant

Looking at $S = \mathsf{Image}(\mathsf{Re}(f')) \cup \mathsf{Image}(\mathsf{Re}(g')) \cup \mathsf{Image}(\mathsf{Re}(h'))$, we have that $|S| \leqslant 3m$. We take $N = (\alpha\tau)^{-9m}$; from Dirichlet's approximation theorem it follows that there are $\sigma\colon \mathsf{Image}(f') \to \mathbb{Z}$, $\gamma\colon \mathsf{Image}(g') \to \mathbb{Z}$ and $\phi\colon \mathsf{Image}(h') \to \mathbb{Z}$ such that for some integer $1 \leqslant q \leqslant N$ we have

$$\left|\mathsf{Re}(f'(x)) - \frac{\sigma(x)}{q}\right| \leqslant \frac{1}{qN^{1/|S|}}, \quad \left|\mathsf{Re}(g'(y)) - \frac{\gamma(y)}{q}\right| \leqslant \frac{1}{qN^{1/|S|}}, \quad \left|\mathsf{Re}(h'(z)) - \frac{\phi(z)}{q}\right| \leqslant \frac{1}{qN^{1/|S|}}.$$

Claim 8.20. *σ is not constant.*

Proof. Consider x, x' such that $\left|\mathsf{Re}(f'(x)) - \mathsf{Re}(f'(x'))\right| \geqslant \Omega_{\alpha,m}(\sqrt{\tau})$. We get that

$$\left|\frac{\sigma(x)}{q} - \frac{\sigma(x')}{q}\right| \geqslant \Omega_{\alpha,m}(\sqrt{\tau}) - \frac{1}{qN^{1/|S|}}, \geqslant \Omega_{\alpha,m}(\sqrt{\tau}) - (\alpha\tau)^3 > 0,$$

so $\sigma(x) \neq \sigma(x')$. □

Claim 8.21. *$\sigma(x) - \gamma(y) - \phi(z) = 0$ for all $(x, y, z) \in \mathsf{supp}(\tilde{\nu})$.*

Proof. From the choice of σ, γ, ϕ and (30) it follows that for all $(x, y, z) \in \mathsf{supp}(\tilde{\nu})$ it holds that

$$\left|\frac{\sigma(x) - \gamma(y) - \phi(z)}{q}\right| \leqslant \frac{3}{qN^{1/|S|}} + \tau^{10m},$$

so

$$\left|\sigma(x) - \gamma(y) - \phi(z)\right| \leqslant \frac{3}{N^{1/|S|}} + \tau^{10m}q \leqslant \frac{3}{N^{1/|S|}} + \tau^{10m}N < 1$$

from the choice of N. It follows that $\sigma(x) - \gamma(y) - \phi(z)$ is an integer smaller than 1 in absolute value, hence it is 0. □

Combining Claims 8.22, 8.23 we get an Abelian embedding of $\mathsf{supp}(\tilde{\nu})$. Indeed, we can pick a large enough Q so that images of σ, γ, ϕ are all contained in the interval $[-Q, Q]$, and then consider them as maps from $\Sigma_{\mathsf{fin}}, \Gamma, \Phi$ to $([0, 3Q], +\pmod{3Q})$. Then by Claim 8.22 we conclude that σ is not constant (as we get that there are x, x' such that $\sigma(x) \neq \sigma(x')$ as they are both in $[-Q, Q]$ so they are different mod $3Q$), and from Claim 8.23 we have that $\sigma(x) - \gamma(y) - \phi(z) = 0 \pmod{3Q}$ in $\mathsf{supp}(\tilde{\nu})$. This is a contradiction to the fact $\tilde{\nu}$ has no linear embedding, and we are done.

8.4.2 The case that the imaginary part of f is far from constant

This case very similar to the previous one, and therefore our description will be briefer.

Looking at $S = \mathsf{Image}(\mathsf{Im}(f')) \cup \mathsf{Image}(\mathsf{Im}(g')) \cup \mathsf{Image}(\mathsf{Im}(h'))$, we have that $|S| \leqslant 3m$. We take $N = (\alpha\tau)^{-9m}$; from Dirichlet's approximation theorem it follows that there are $\sigma\colon \mathsf{Image}(f') \to \mathbb{Z}$, $\gamma\colon \mathsf{Image}(g') \to \mathbb{Z}$ and $\phi\colon \mathsf{Image}(h') \to \mathbb{Z}$ such that for some integer $1 \leqslant$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

$q \leqslant N$ we have

$$\left|\operatorname{Im}(f'(x)) - \frac{\sigma(x)}{q}\right| \leqslant \frac{1}{qN^{1/|S|}}, \quad \left|\operatorname{Im}(g'(y)) - \frac{\gamma(y)}{q}\right| \leqslant \frac{1}{qN^{1/|S|}}, \quad \left|\operatorname{Im}(h'(z)) - \frac{\phi(z)}{q}\right| \leqslant \frac{1}{qN^{1/|S|}}.$$

Claim 8.22. *σ is not constant.*

Proof. Consider x, x' such that $d(\mathbf{i}\operatorname{Im}(f'(x)), \mathbf{i}\operatorname{Im}(f'(x')) \geqslant \Omega_{\alpha,m}(\sqrt{\tau})$. We get that

$$d(\mathbf{i}\frac{\sigma(x)}{q}, \mathbf{i}\frac{\sigma(x')}{q}) \geqslant \Omega_{\alpha,m}(\sqrt{\tau}) - \frac{1}{qN^{1/|S|}}, \geqslant \Omega_{\alpha,m}(\sqrt{\tau}) - (\alpha\tau)^3 > 0,$$

so $\sigma(x) \neq \sigma(x')$. □

Claim 8.23. *$\sigma(x) - \gamma(y) - \phi(z) = 0$ for all $(x, y, z) \in \operatorname{supp}(\tilde{\nu})$.*

Proof. From the choice of σ, γ, ϕ and (30) it follows that for all $(x, y, z) \in \operatorname{supp}(\tilde{\nu})$ it holds that

$$d(\mathbf{i}\frac{\sigma(x)}{q}, \mathbf{i}\frac{\gamma(y) - \phi(z)}{q}) \leqslant \frac{3}{qN^{1/|S|}} + \tau^{10m},$$

so

$$d(\mathbf{i}\sigma(x), \mathbf{i}(\gamma(y) - \phi(z))) \leqslant \frac{3}{N^{1/|S|}} + \tau^{10m}q \leqslant \frac{3}{N^{1/|S|}} + \tau^{10m}N < 1$$

from the choice of N. It follows that $\sigma(x) - \gamma(y) - \phi(z)$ is an integer smaller than 1 in absolute value, hence it is 0. □

Combining Claims 8.22, 8.23 we get an Abelian embedding of $\operatorname{supp}(\tilde{\nu})$ as before.

9. Applications

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

9.1 Hardness of approximation of CSP's

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 9.1. *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 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 Fig. 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.[13] 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.

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 satisfying assignments to P and therefore every column

[13] We refer the readers to [5, 21] for detailed information on the semidefinite programme, its value and the local distributions.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

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*.

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 2.1, to analyse the above dictatorship test, then we have the following theorem on the soundness of the above test.

Theorem 9.2 (Restatement of Theorem 1.5). *Let* $P\colon \Sigma^3 \to \{0,1\}$ *be any predicate that satisfies the following conditions. (1) P does not satisfy any linear embedding, and (2) there exists an instance of 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],Theorem 1.1]. The only difference is that in the proof of [[5],Theorem 1.1], Lemma 2.1 with the added condition that the distribution μ is semi-rich was used. As we get off the semi-richness condition in our main analytical lemma, we get the above improved theorem that applies for a rather large class of 3-ary predicates. As the proof is identical to the proof of [[5],Theorem 1.1], we skip the proof of Theorem 9.2 in this version.

9.2 Counting lemmas

Theorem 9.3. *Suppose* μ *is a distribution over* $\Sigma \times \Gamma \times \Phi$ *such that* $\mathsf{supp}(\mu)$ *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*

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

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

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

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

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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

Indeed, note that

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

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 2.1.

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

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

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 9.4. *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 $(\boldsymbol{I}, \boldsymbol{x}')$ yields*

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

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 9.5 (Restatement of Theorem 1.6). *Suppose μ is a distribution over Σ^3 such that (1) the marginal distributions μ_x, μ_y, μ_z are identical, (2) $\{(x,x,x) \mid x \in \Sigma\} \subseteq \mathsf{supp}(\mu)$, and (3) $\mathsf{supp}(\mu)$ 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 $|S| \geqslant \delta|\Sigma|^n$,*

$$\mathbb{P}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})\sim\mu^{\otimes n}}\left[\boldsymbol{x} \in S, \boldsymbol{y} \in S, \boldsymbol{z} \in S\right] \geqslant \varepsilon.$$

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 9.4 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, that is, 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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

$\delta/2$; we refer to this event by E. Thus, we get that

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

where the last inequality is by Theorem 9.3. □

For example, Theorem 1.6 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.

Acknowledgements

We thank Yang Liu for spotting an error in Claim 7.9 in an earlier version of this paper.

Funding statement

Subhash Khot supported by the NSF Award CCF-1422159, NSF Award CCF-2130816, and the Simons Investigator Award. Dor Minzer supported by a Sloan Research Fellowship, NSF CCF award 2227876 and NSF CAREER award 2239160.

References

[1] Bateman, M. and Katz, N. (2012) New bounds on cap sets. *J Am Math Soc* **25**(2) 585–613.
[2] Bergelson, V. and Tao, T. (2014) Multiple recurrence in quasirandom groups. *Geom Funct Anal* **24**(1) 1–48.
[3] Bhangale, A., Harsha, P. and Roy, S. (2022) Mixing of 3-Term Progressions in Quasirandom Groups. In *13th Innovations in Theoretical Computer Science Conference (ITCS)*, vol. 215, pp. 20:1- 20:9.
[4] Bhangale, A. and Khot, S. (2021) Optimal inapproximability of satisfiable k-lin over non-abelian groups. In *Proceedings of the 53rd annual Symposium on Theory of Computing (STOC)*, pp. 1615–1628. ACM.
[5] Bhangale, A., Khot, S. and Minzer, D. (2022) On approximability of satisfiable k-csps: I. In *Proceedings of the 54th annual Symposium on Theory of Computing (STOC)*, pp. 976–988. ACM.
[6] Braverman, M., Khot, S., Lifshitz, N. and Minzer, D. (2021) An invariance principle for the multi-slice, with applications. In *Proceedings of the 62nd annual symposium on Foundations of Computer Science (FOCS)*, pp. 228–236. IEEE.
[7] Braverman, M., Khot, S. and Minzer, D. (2021) On Rich 2-to-1 Games. In *12th Innovations in Theoretical Computer Science Conference (ITCS)*, vol. **185**, pp. 27:1- 27:20.
[8] Brown, T. C. and Buhler, J. P. (1982) A density version of a geometric ramsey theorem. *J Combinatorial Theory, Se A* **32**(1) 20–34.
[9] Chase, G., Filmus, Y., Minzer, D., Mossel, E. and Saurabh, N. (2022) Approximate polymorphisms. In *Proceedings of the 54th annual Symposium on Theory of Computing (STOC)*, pp. 195–202. ACM.
[10] Croot, E., Lev, V. F. and Pach, P. P. (2017) Progression-free sets in $\mathcal{F}_4^n$ are exponentially small. *Ann Math* **185**(1) 331–337.
[11] Ellenberg, J. S. and Gijswijt, D. (2017) On large subsets of with no three-term arithmetic progression. *Ann Math* 185(1) 339–343.
[12] Gowers, W. T. (2008) Quasirandom groups. *Combinatorics, Probability and Computing* **17**(3) 363–387.
[13] Hazła, J., Holenstein, T. and Mossel, E. (2018) Product space models of correlation: Between noise stability and additive combinatorics. *Discrete Analysis* (2018). https://discreteanalysisjournal.com/article/6513-product-space-models-of-correlation-between-noise-stability-and-additive-combinatorics
[14] Jones, C. (2016) A noisy-influence regularity lemma for boolean functions. *CoRR*
[15] Khot, S. (2002) On the power of unique 2-prover 1-round games. In *Proceedings of the 34th annual Symposium on Theory of Computing (STOC)*, pp. 767–775. ACM.
[16] Meshulam, R. (1995) On subsets of finite abelian groups with no 3-term arithmetic progressions. *J Combinatorial Theory, Ser A* **71**(1) 168–172.

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

[17] Mossel, E. (2010) Gaussian bounds for noise correlation of functions. *Geom Funct Anal* **19**(6) 1713–1756.
[18] Mossel, E., O'Donnell, R. and Oleszkiewicz, K. (2005) Noise stability of functions with low influences: invariance and optimality. In *Proceedings of the 46th annual symposium on Foundations of Computer Science (FOCS)*, pp. 21–30. IEEE.
[19] O'Donnell, R. (2014) *Analysis of boolean functions*. Cambridge University Press.
[20] Peluse, S. (2018) Mixing for three-term progressions in finite simple groups. *Math Proc Cambridge* **165**(2) 279–286.
[21] Raghavendra, P. (2008) Optimal algorithms and inapproximability results for every csp? In *Proceedings of the 14th annual Symposium on Theory of Computing (STOC)*, pp. 245–254.
[22] Raghavendra, P. (2009) *Approximating NP-hard problems efficient algorithms and their limits*. University of Washington.
[23] Roth, K. F. (1953) On certain sets of integers. *J London Math Soc* **s1-28**(1) 104–109.
[24] Szemerédi, E. (1975) On sets of integers containing k elements in arithmetic progression. *Acta Arith* **27**(1) 199–245.
[25] Tao, T. (2013) Mixing for progressions in nonabelian groups, *Forum of Mathematics, Sigma*, vol. 1, pp. e2. Cambridge University Press.

A. Missing proofs from Section 4

In this section, we prove the implications from (8).

A.1 Proof that Lemma 4.2 implies Lemma 2.5

We use

$$0 \leqslant \delta \ll \xi \ll \varepsilon \ll c,\ M^{-1} \ll \alpha,\ m^{-1} \leqslant 1.$$

Note that

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})\mathrm{T}_{1-\xi}g(\mathbf{y})h(\mathbf{z})\right]\right| + \sum_{j=0}^{\infty}\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})(\mathrm{T}_{1-2^{-j-1}\xi}-\mathrm{T}_{1-2^{-j}\xi})g(\mathbf{y})h(\mathbf{z})\right]\right|.$$

For the first term, we have

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})\mathrm{T}_{1-\xi}g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \|\mathrm{T}_{1-\xi}g\|_2 = \sqrt{\mathsf{Stab}_{(1-\xi)^2}(g)} \leqslant \sqrt{\delta} \leqslant \frac{\varepsilon}{10},$$

so we focus on bounding the second term; fix some j, and denote $\xi' = 2^{-j}\xi$, $\mathrm{T}' = \mathrm{T}_{1-\xi'/2} - \mathrm{T}_{1-\xi'}$, and $\xi'' = M\xi'/\log(\xi')^3$. Clearly,

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f(\mathbf{x})\mathrm{T}'g(\mathbf{y})h(\mathbf{z})\right]\right| \leqslant \underbrace{\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[(\mathrm{I}-\mathrm{T}_{1-\xi''})f(\mathbf{x})\mathrm{T}'g(\mathbf{y})h(\mathbf{z})\right]\right|}_{(I)} + \underbrace{\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[\mathrm{T}_{1-\xi''}f(\mathbf{x})\mathrm{T}'g(\mathbf{y})h(\mathbf{z})\right]\right|}_{(II)}.$$

Upper bounding (II). From Lemma 4.2 we get that

$$(II) \lesssim_{M,m,\alpha} \frac{1}{\log(1/\xi')^6}.$$

Upper bounding (I). Clearly, we may bound it by

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[(\mathrm{I}-\mathrm{T}_{1-\xi''})f(\mathbf{x})\mathrm{T}_{1-\xi'}g(\mathbf{y})h(\mathbf{z})\right]\right| + \left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[(\mathrm{I}-\mathrm{T}_{1-\xi''})f(\mathbf{x})\mathrm{T}_{1-\xi'/2}g(\mathbf{y})h(\mathbf{z})\right]\right|,$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

and as the bound for each one of these is similar, we only explain the upper bound for the first. We can write

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

where a sample in the distribution μ' is drawn by first sampling $(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}$, and then re-sampling each coordinate of $\mathbf{y}$ independently with probability ξ'. Write $f=f^{\leqslant d}+f^{>d}$ where $d=\sqrt{M}\log(1/\xi')/\xi'$.

Claim A.1. $\left|\mathbb{E}_{(x,y,z)\sim\mu'^{\otimes n}}\left[(I-T_{1-\xi''})(f^{\leqslant d})(x)g(y)h(z)\right]\right|\leqslant\frac{M^{3/2}}{\log(1/\xi')^2}$.

Proof. By boundedness we have

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu'^{\otimes n}}\left[(\mathrm{I}-\mathrm{T}_{1-\xi''})(f^{\leqslant d})(\mathbf{x})g(\mathbf{y})h(\mathbf{z})\right]\right|&\leqslant\|(\mathrm{I}-\mathrm{T}_{1-\xi''})(f^{\leqslant d})\|_1\\&\leqslant\|(\mathrm{I}-\mathrm{T}_{1-\xi''})(f^{\leqslant d})\|_2\\&\leqslant(1-(1-\xi'')^d)\\&\leqslant d\xi''\\&\leqslant\frac{M^{3/2}}{\log(1/\xi')^2}.\end{aligned}$$

□

Claim A.2. $\left|\mathbb{E}_{(x,y,z)\sim\mu'^{\otimes n}}\left[(I-T_{1-\xi''})(f^{>d})(x)g(y)h(z)\right]\right|\leqslant\xi'^{100}$

Proof. The left hand side is at most

$$\mathop{\mathbb{E}}_{\mathbf{y}\sim\mu'_y}\left[\left|\mathbb{E}_{(\mathbf{x},\mathbf{y}',\mathbf{z})\sim\mu'^{\otimes n}}\left[(\mathrm{I}-\mathrm{T}_{1-\xi''})f^{>d}(\mathbf{x})h(\mathbf{z})\,\middle|\,\mathbf{y}'=\mathbf{y}\right]\right|\right]$$

Note that for each $i\in[n]$, the marginal distribution of $\mu'\mid\mathbf{y}'_i=y_i$ on x_i,z_i is $\xi'\mu_{x,z}+(1-\xi')\mu_{x,z}\mid y_i$. We consider the following alternative way of sampling from μ'; first, choose $I\subseteq[n]$ randomly by including each $i\in[n]$ in it with probability ξ'. On $i\in I$, sample $(x_i,z_i)\sim\mu_{x,z}$ and $y_i\sim\mu_y$ independently. On $i\notin I$, sample $(x_i,y_i,z_i)\sim\mu$. We thus have that the above expression can be written as

$$\begin{aligned}&\mathop{\mathbb{E}}_{\mathbf{y}\sim\mu'_y}\left[\left|\mathbb{E}_{I,(\mathbf{x}_{\bar I},\mathbf{y}'_{\bar I},\mathbf{z}_{\bar I})}\left[\mathop{\mathbb{E}}_{I,x_I,z_I}\left[(\mathrm{I}-\mathrm{T}_{1-\xi''})f^{>d}(\mathbf{x})h(\mathbf{z})\right]\,\middle|\,\mathbf{y}_I=\mathbf{y}'_{\bar I}\right]\right|\right]\\&\leqslant\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar I},\mathbf{y}_{\bar I},\mathbf{z}_{\bar I}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{x}_I,\mathbf{z}_I}\left[(\mathrm{I}-\mathrm{T}_{1-\xi''})f^{>d}(\mathbf{x})h(\mathbf{z})\right]\right|\right]\\&\leqslant\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar I},\mathbf{y}_{\bar I},\mathbf{z}_{\bar I}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{x}_I,\mathbf{z}_I}\left[f^{>d}(\mathbf{x})h(\mathbf{z})\right]\right|\right]+\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar I},\mathbf{y}_{\bar I},\mathbf{z}_{\bar I}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{x}_I,\mathbf{z}_I}\left[\mathrm{T}_{1-\xi''}(f^{>d})(\mathbf{x})h(\mathbf{z})\right]\right|\right]\\&=\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar I},\mathbf{y}_{\bar I},\mathbf{z}_{\bar I}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{x}_I,\mathbf{z}_I}\left[f^{>d}(\mathbf{x})h(\mathbf{z})\right]\right|\right]+\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar I},\mathbf{y}_{\bar I},\mathbf{z}_{\bar I}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{x}_I,\mathbf{z}_I}\left[(\mathrm{T}_{1-\xi''}f^{>d})(\mathbf{x})h(\mathbf{z})\right]\right|\right].\end{aligned}$$

We show that if f' is a function of 2-norm at most 1 such that $(f')^{\leqslant d}=0$, then

$$\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar I},\mathbf{y}_{\bar I},\mathbf{z}_{\bar I}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{x}_I,\mathbf{z}_I}\left[f'(\mathbf{x})h(\mathbf{z})\right]\right|\right]\leqslant\frac{1}{2}\xi'^{100}.$$

Since both $f^{>d}$ and $\mathrm{T}_{1-\xi''}f^{>d}$ are such functions, this would yield the statement of the claim. Indeed, if f' are such function, then considering the operator $\mathrm{S}_I\colon L_2(\Sigma^I,\mu_x^I)\to L_2(\Phi^I,\mu_z^I)$ we

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

have

$$\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar{I}},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{x}_I,\mathbf{z}_I}\left[f'(\mathbf{x})h(\mathbf{z})\right]\right|\right]=\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar{I}},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[\left|\langle \mathrm{S}_I(f'_{x_{\bar{I}}\to\mathbf{x}_{\bar{I}}}),h_{z_{\bar{I}}\to\mathbf{z}_{\bar{I}}}\rangle\right|\right]\leqslant\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar{I}},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[\|\mathrm{S}_I^*\mathrm{S}_I(f'_{x_{\bar{I}}\to\mathbf{x}_{\bar{I}}})\|_2\right].$$

By Jensen's inequality, we have that this is at most

$$\begin{aligned}\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar{I}},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[\|\mathrm{S}_I^*\mathrm{S}_I(f'_{x_{\bar{I}}\to\mathbf{x}_{\bar{I}}})\|_2^2\right]^{1/2}&\leqslant\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar{I}},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[\left(W_{\leqslant d'}(f'_{x_{\bar{I}}\to\mathbf{x}_{\bar{I}}})+(1-c(m,\alpha))^{2d'}\right)^{1/2}\right]\\&\leqslant(1-c)^{d'}+\sqrt{\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar{I}},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[W_{\leqslant d'}(f'_{x_{\bar{I}}\to\mathbf{x}_{\bar{I}}})\right]},\end{aligned}$$

where $d'=\xi' d/2$. In the second inequality, we used the fact that the operator $\mathrm{S}_I^*\mathrm{S}_I$ is connected and the probability of each transition is at least $\Omega_{\alpha,m}(1)$, so $\|\mathrm{S}_I^*\mathrm{S}_I f''\|_2\leqslant(1-c)^{d'}\|f''\|_2$ if $(f'')^{<d'}\equiv 0$. Note that

$$\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar{I}},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[W_{\leqslant d'}(f'_{x_{\bar{I}}\to\mathbf{x}_{\bar{I}}})\right]\leqslant\mathbb{P}_I\left[\left|[d]\cap I\right|\leqslant d'\right]\leqslant 2^{-\Omega(d')}$$

by Chernoff's bound. Overall, we get that

$$\mathop{\mathbb{E}}_{I,\mathbf{x}_{\bar{I}},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[\left|\mathop{\mathbb{E}}_{\mathbf{x}_I,\mathbf{z}_I}\left[f'(\mathbf{x})h(\mathbf{z})\right]\right|\right]\leqslant 2^{-\Omega_{m,\alpha}(d')}+2^{-\Omega(d')}\leqslant 2^{-\Omega_{m,\alpha}(\sqrt{M}\log(1/\xi'))}\leqslant\frac{1}{2}\xi'^{100}$$

□

We thus get that

$$(I)+(II)\lesssim_{M,m,\alpha}\frac{1}{\log(1/\xi')^2}+\xi'^{100}+\frac{1}{\log(1/\xi')^6}\lesssim\frac{1}{j^2+\log(1/\xi)^2}+2^{-j/2}\sqrt{\xi},$$

so

$$\sum_j(I)+(II)\lesssim_{M,m,\alpha}\sqrt{\xi}+\sum_{j=\log(1/\xi)}^{\infty}\frac{1}{j^2}\lesssim\sqrt{\xi}+\frac{1}{\log(1/\xi)},$$

hence $\sum_j(I)+(II)\leqslant\frac{\varepsilon}{2}$ provided ξ is small enough. This finishes the deduction of Lemma 2.5 from Lemma 4.2.

A.2 Softly truncating the effective degree of f from below: Lemma 4.3 implies Lemma 4.2

We choose

$$0\leqslant\xi_0\ll c,\ M^{-1}\ll\alpha,\ m^{-1}\leqslant 1,$$

and show that Lemma 4.3 implies Lemma 4.2. Letting $f'=T_{1-M\xi/\log(1/\xi)^3}f$, $g'=(\mathrm{T}_{1-\xi/2}-\mathrm{T}_{1-\xi})g$ and $\xi'=M\xi\log(1/\xi)^{100}$, we write

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

Using Lemma 4.3, we have that $(I)\lesssim_{m,\alpha,M}\frac{1}{\log^6(1/\xi)}$, and we next bound (II). We may write

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

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

where the distribution μ' is defined by first sampling $(x, y, z) \sim \mu^{\otimes n}$, and then for each i such that $x_i \in \Sigma'$, with probability ξ' re-sampling it according to $\mu_x \mid \Sigma'$. We consider the following alternative way of sampling according to μ'. Let $p = \mathbb{P}_{x\sim\mu_x}\left[x \in \Sigma'\right]$, and note that $p \geqslant \alpha$. We sample $R \subseteq [n]$ by independently including each $i \in [n]$ in R with probability p, and then sub-sample $I \subseteq R$ by including each $i \in R$ in I with probability ξ'. We then sample $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ as: for $i \in I$, we sample $\mathbf{x}_i \sim \mu_x \mid \Sigma'$, $y_i, z_i \sim \mu_{y,z} \mid x \in \Sigma'$ independently; for $i \in R \setminus I$ we sample $(x_i, y_i, z_i) \sim \mu \mid x_i \in \Sigma'$; for $i \notin R$, we sample $(x_i, y_i, z_i) \sim \mu \mid x_i \notin \Sigma'$. Thus, denoting $\nu = \mu_{y,z} \mid x \in \Sigma'$, we have

$$(II) \leqslant \mathop{\mathbb{E}}_{I,R,\mathbf{x},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[\left|f'(\mathbf{x})\right|\left|\mathop{\mathbb{E}}_{(\mathbf{y}_I,\mathbf{z}_I)\sim\nu^I}\left[g'(\mathbf{y})h(\mathbf{z})\right]\right|\right] \leqslant \sqrt{\mathop{\mathbb{E}}_{\substack{I,R,\mathbf{x},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}\\(\mathbf{y}_I,\mathbf{z}_I)\sim\nu^I,(\mathbf{y}'_I,\mathbf{z}'_I)\sim\nu^I}}\left[g'(\mathbf{y})h(\mathbf{z})\overline{g'(\mathbf{y}')h(\mathbf{z}')}\right]}$$
$$= \sqrt{\mathop{\mathbb{E}}_{\substack{I,R,\mathbf{x},\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}\\(\mathbf{y}_I,\mathbf{z}_I)\sim\nu^I,(\mathbf{y}'_I,\mathbf{z}'_I)\sim\nu^I}}\left[g'(\mathbf{y})L(\mathbf{y}',\mathbf{z},\mathbf{z}')\right]}.$$

Let

$$\tilde{g} = (g')_{y_{\bar{I}}\to\mathbf{y}_{\bar{I}}}, \quad \tilde{L} = L_{y'_{\bar{I}},z_{\bar{I}},z'_{\bar{I}}\to\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}},\mathbf{z}'_{\bar{I}}},$$

so that we may write the last expression as $\sqrt{\mathbb{E}_{I,R,\mathbf{y}_{\bar{I}},\mathbf{z}_{\bar{I}}}\left[\langle S_I\tilde{g}, \tilde{L}\rangle\right]}$, where $S_I\colon L_2(\Gamma^n, \nu_y) \to L_2(\Gamma^n, \nu_z)$ is defined as

$$S_I\tilde{g}(z) = \mathbb{E}_{(\mathbf{y},\mathbf{z})\sim\nu}\left[\tilde{g}(y) \mid \mathbf{z} = z\right].$$

We may bound

$$\left|\langle S_I\tilde{g}, \tilde{L}\rangle\right| \leqslant \|S_I\tilde{g}\|_2 \leqslant \|S_I^*S_I\tilde{g}\|_2^{1/2}\|\tilde{g}\|_2^{1/2} \lesssim \|S_I^*S_I\tilde{g}\|_2^{1/2},$$

so combining everything so far and using Jensen, we get that

$$(II) \lesssim \mathop{\mathbb{E}}_{I,y_{\bar{I}}}\left[\|S_I^*S_I\tilde{g}\|_{2,\nu_y}^2\right]^{1/8}.$$

Next, note that as $S_I^*S_I$ is connected and the probability of each atom is at least $\Omega_{\alpha,m}(1)$, we get that choosing $d = M\log(1/\xi)$ we have from Lemma 3.10 that

$$\|S_I^*S_I\tilde{g}\|_{2,\nu_y}^2 \leqslant W_{\leqslant d}[\tilde{g};\nu_y] + (1-c)^d.$$

We next bound the expectation of the first term on the right hand side using Lemma 3.15, and argue that it follows from the lemma that

$$\mathop{\mathbb{E}}_{I,y_{\bar{I}}}\left[W_{\leqslant d}[\tilde{g};\nu_y]\right] \lesssim \mathop{\mathbb{E}}_{I,y_{\bar{I}}}\left[\mathsf{Stab}_{1-1/d}(\tilde{g};\nu_y)\right] \lesssim \mathsf{Stab}_{1-c\xi'/d}(g';\mu_y),$$

for some $c = c(m, \alpha) > 0$. Indeed, in the setting of the lemma we have $1 - s = \xi'$; the distribution ν_2 is ν_y, and the distribution ν_1 is $\frac{p'}{1-\xi'}\nu'_1 + \frac{p''}{1-\xi'}\nu''_1$ where $p' = 1 - p$ and $p'' = 1 - \xi' - p' = p - \xi'$, ν'_1 is the distribution of $\mathbf{y}$ where $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \sim \mu \mid \mathbf{x} \notin \Sigma'$, and ν''_1 is the distribution of $\mathbf{y}$ where $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \sim \mu \mid \mathbf{x} \in \Sigma'$. It follows that

$$\mathop{\mathbb{E}}_{I,y_{\bar{I}}}\left[W_{\leqslant d}[\tilde{g};\nu_y]\right] \leqslant \sum_S\left(1 - \frac{c\xi'}{d}\right)^{|S|}\|(g')^{=S}\|_2^2 \leqslant \sum_S\left(1 - \frac{c\xi'}{d}\right)^{|S|}\left(\left(1 - \frac{\xi}{2}\right)^{|S|} - (1-\xi)^{|S|}\right)\|g^{=S}\|_2^2$$
$$\leqslant \max_{k\in\mathbb{N}}\left(1 - \frac{c\xi'}{d}\right)^k\left(\left(1 - \frac{\xi}{2}\right)^k - (1-\xi)^k\right).$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

If $k \leqslant \frac{1}{\xi \log(1/\xi)^{50}}$, the second factor is at most $\lesssim \xi k \lesssim \frac{1}{\log(1/\xi)^{50}}$. If $k > \frac{1}{\xi \log(1/\xi)^{50}}$, the first factor is at most

$$2^{-\Omega_{m,\alpha}(\frac{\xi' k}{d})} \leqslant 2^{-\Omega_{M,m,\alpha}(\log(1/\xi)^{50})} \leqslant \xi.$$

It follows that

$$(II) \lesssim \mathop{\mathbb{E}}_{I, y_{\bar{I}}} \left[\|S_I^* S_I \tilde{g}\|_{2,\nu_y}^2 \right]^{1/8} \lesssim (1-c)^d + \xi^{1/8} + \frac{1}{\log(1/\xi)^6} \lesssim \frac{1}{\log(1/\xi)^6}.$$

A.3 Getting the functions to be homogeneous: Lemma 4.1 implies Lemma 4.3

In this section, we prove that Lemma 4.1 implies Lemma 4.3. We begin by stating a few basic properties of the operators T and E.

Claim A.3. *Let $\chi \in B_1 \cup B_2$. Then*

1. *If $\chi \neq \chi_{const}$, then $T_{1-\xi'}\chi = (1-\xi')\chi$, and if $\chi = \chi_{const}$, then $T_{1-\xi'}\chi = \chi$.*
2. *If $\chi \in B_1$ then $E_{1-\xi'}\chi = \chi$.*
3. *If $\chi \in B_2$, then $E_{1-\xi'}\chi = (1-\xi')\chi$.*

Proof. For the first item, if $\chi \neq \chi_{\text{const}}$, then the average of it is 0, hence

$$\mathrm{T}_{1-\xi'}\chi(x) = (1-\xi')\chi(x) + \xi' \cdot 0 = (1-\xi')\chi(x).$$

The second part of the first item is clear. The second item is also clear, since $\mathrm{E}_{1-\xi'}$ may only change x if $x \in \Sigma'$, and $\chi \in B_1$ gets the same value on all elements of Σ'. For the third item, note that if $\chi \in B_2$, then $\mathbb{E}_{x\sim\mu_x}[1_{x\in\Sigma'}\chi(x)] = 0$, so

$$\mathrm{E}_{1-\xi'}\chi(x) = (1-\xi')\chi(x) + \xi' \frac{\mathbb{E}_{x\sim\mu_x}[1_{x\in\Sigma'}\chi(x)]}{\mathbb{E}_{x\sim\mu_x}[1_{x\in\Sigma'}]} = (1-\xi')\chi(x).$$

□

Claim A.4. *Let $f\colon \Sigma^n \to \mathbb{C}$.*

1. *If f is homogeneous of degree d, then $\mathrm{T}_{1-\xi'}f = (1-\xi')^d f$.*
2. *If f is effectively homogeneous of degree d, then $\mathrm{E}_{1-\xi'}f = (1-\xi')^d f$.*

□

Proof. Both items are immediate from Claim A.3 by writing f as linear combination of monomials of degree d, or effectively degree d in the case of the second item. □

Proof that Lemma 4.1 implies Lemma 4.3. Towards this end, fix f, g, h and α, m, M, ξ_0 and ξ as in Lemma 4.1; we shall assume that f, g and h have 2-norms equal to 1. Denote

$$f' = (\mathrm{I} - \mathrm{E}_{1-M\xi\log(1/\xi)^{100}})\mathrm{T}_{1-M\xi/\log(1/\xi)^3} f, \quad g' = (\mathrm{T}_{1-\xi/2} - \mathrm{T}_{1-\xi})g.$$

We write $f' = f_1' + f_2' + f_3'$ where f_1' is the part of f of effective degree of at most $d_1 = \frac{1}{\xi \log(1/\xi)^{200}}$, f_3' is the part of f of degree at least $d_2 = \frac{\log(1/\xi)^6}{\xi}$ and effective degree more than d_1, and f_2' is the part of f of effective degree more than d_1 and degree less than d_2. We write

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}} \left[f'(\mathbf{x}) g'(\mathbf{y}) h(\mathbf{z}) \right] = \underbrace{\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}} \left[f_1'(\mathbf{x}) g'(\mathbf{y}) h(\mathbf{z}) \right]}_{(I)} + \underbrace{\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}} \left[f_2'(\mathbf{x}) g'(\mathbf{y}) h(\mathbf{z}) \right]}_{(II)}$$
$$+ \underbrace{\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}} \left[f_3'(\mathbf{x}) g'(\mathbf{y}) h(\mathbf{z}) \right]}_{(III)},$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

and bound each one of these separately. First note that

$$|(I)|^2 \leqslant \|f_1'\|_2^2 \|g\|_2^2 \|h\|_2^2 \leqslant \|f_1'\|_2^2 = \sum_{\chi\colon \mathsf{effdeg}(\chi)\leqslant d_1} \left|\widehat{f_1'}(\chi)\right|^2.$$

Using Claim A.4, we get that the last sum is equal to

$$\sum_{\chi\colon \mathsf{effdeg}(\chi)\leqslant d_1} \left(1-\left(1-M\xi\log(1/\xi)^{100}\right)^{\mathsf{effdeg}(\chi)}\right)\left(1-\frac{M\xi}{\log(1/\xi)^3}\right)^{\deg(\chi)}\left|\widehat{f}(\chi)\right|^2$$
$$\leqslant \left(1-\left(1-M\xi\log(1/\xi)^{100}\right)^{d_1}\right)\sum_{\chi}\left|\widehat{f}(\chi)\right|^2,$$

which is at most $M\xi\log(1/\xi)^{100}d_1 \leqslant \frac{1}{\log(1/\xi)^{50}}$. Similarly, we have that

$$|(III)|^2 \leqslant \|f_3'\|_2^2 \|g\|_2^2 \|h\|_2^2 \leqslant \|f_3'\|_2^2 = \sum_{\chi\colon \deg(\chi)\geqslant d_2} \left|\widehat{f_3'}(\chi)\right|^2.$$

Using Claim A.4, we get that the last sum is equal to

$$\sum_{\chi\colon \deg(\chi)\geqslant d_2} \left(1-\left(1-M\xi\log(1/\xi)^{100}\right)^{\mathsf{effdeg}(\chi)}\right)\left(1-\frac{M\xi}{\log(1/\xi)^3}\right)^{\deg(\chi)}\left|\widehat{f}(\chi)\right|^2$$
$$\leqslant \left(1-\frac{M\xi}{\log(1/\xi)^3}\right)^{d_2}\sum_{\chi}\left|\widehat{f}(\chi)\right|^2,$$

which is at most $2^{-\frac{M\xi}{\log(1/\xi)^3}d_2} \leqslant \frac{1}{\log(1/\xi)^{50}}$.

And we next bound (II). Let g_1' be the part of g' of degree at most $d_1' = \frac{1}{\xi\log(1/\xi)^{10}}$, g_3' be the part of g' of degree at least $d_2' = \frac{\log(1/\xi)^{10}}{\xi}$, and g_2' be the part of g' of degree more than d_1' and less than d_2'. We write

$$(II) = \underbrace{\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f_2'(\mathbf{x})g_1'(\mathbf{y})h(\mathbf{z})\right]}_{(IV)} + \underbrace{\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f_2'(\mathbf{x})g_2'(\mathbf{y})h(\mathbf{z})\right]}_{(V)} + \underbrace{\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f_2'(\mathbf{x})g_3'(\mathbf{y})h(\mathbf{z})\right]}_{(VI)}$$

Similar arguments to before show that $|(IV)|, |(VI)| \leqslant \frac{1}{\log(1/\xi)^8}$, and we next bound (V). Write $f_2' = \sum_{i,j=d_1}^{d_2} f_{i,j}'$ where $f_{i,j}'$ is the part of f' homogeneous of degree i and effectively homogeneous j, and $g_2' = \sum_{i'=d_1'}^{d_2'} g_{2,i'}'$ where $g_{2,i'}'$ is the part of g' homogeneous of degree i', and $h = \sum_{d=0}^{n} h^{=d}$ where $h^{=d}$ is the homogeneous degree d part of h. We have

$$|(V)| \leqslant \sum_{i,j,i',d}\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f_{i,j}'(\mathbf{x})g_{2,i'}'(\mathbf{y})h^{=d}(\mathbf{z})\right]\right|.$$

We note that if $d > i + i'$, then the expectation is 0. For other d, we have by Lemma 4.1 that

$$\left|\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}}\left[f_{i,j}'(\mathbf{x})g_{i',j'}'(\mathbf{y})h^{=d}(\mathbf{z})\right]\right| \leqslant (1-\delta)^{\frac{i}{\log^C(i)}}$$

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

for some $C = C(m, \alpha) > 0$ and $\delta = \delta(m, \alpha) > 0$. Thus, the above sum is at most

$$(d_2')^4(1-\delta)^{\frac{1}{\xi \log^C(1/\xi)}} \leqslant \frac{1}{\log(1/\xi)^6}$$

where the last inequality holds as $0 < \xi \leqslant \xi_0$ and ξ_0 is sufficiently small. We conclude that $|(II)| \lesssim \frac{1}{\log(1/\xi)^6}$. Combining the bounds on $|(I)|$, $|(II)|$ and $|(III)|$ yields that

$$\left| \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{y},\mathbf{z})\sim\mu^{\otimes n}} \left[f'(\mathbf{x}) g'(\mathbf{y}) h(\mathbf{z}) \right] \right| \lesssim \frac{1}{\log(1/\xi)^6}.$$

thereby establishing the assertion of Lemma 4.3. □

B. Missing proofs from Section 5

In this section, we prove the claims establishing the SVD decompositions for homogeneous and non-homogeneous functions.

B.1 Proof of Claim 5.8

Proof. We think of g as a matrix M in $\mathbb{C}^{\Gamma^I \times \Gamma^J}$, whose $(\mathbf{a}, \mathbf{b})$ entry is $g(\mathbf{y}_I = \mathbf{a}, \mathbf{y}_J = \mathbf{b})$. The decomposition stated by the claim is an appropriately chosen singular-value decomposition of M; below are the details for completeness

Looking at $M^*M \in \mathbb{C}^{\Gamma^J \times \Gamma^J}$, we see that it is an $m \times m$ Hermitian positive semi-definite matrix, hence we may find an eigenbasis $g_1', \ldots, g_m'$ of $\mathbb{C}^{\Gamma^J}$ with non-negative eigenvalues $\lambda_1, \ldots, \lambda_m$. We note that the all 1 vector is an eigenvector of M^*M; indeed:

$$\begin{aligned}(M^*M\vec{1})_{\mathbf{a}} = \sum_{\mathbf{b}} (M^tM)_{\mathbf{a},\mathbf{b}} = \sum_{\mathbf{b}} \sum_{\mathbf{y}_I \in \Gamma^I} M^*[\mathbf{a}, \mathbf{y}_I] M[\mathbf{y}_I, \mathbf{b}] = \sum_{\mathbf{b}} \sum_{\mathbf{y}_I \in \Gamma^I} g(\mathbf{y}_I, \mathbf{b}) \overline{g(\mathbf{y}_I, \mathbf{a})} \\ = \sum_{\mathbf{y}_I \in \Gamma^I} \overline{g(\mathbf{y}_I, \mathbf{a})} \sum_{\mathbf{b}} g(\mathbf{y}_I, \mathbf{b}).\end{aligned}$$

Consider the function $A\colon \Gamma^I \to \mathbb{C}$ defined by $A(\mathbf{y}_I) = \sum_{\mathbf{b}} g(\mathbf{y}_I, \mathbf{b})$. Note that for each monomial in g that contains the variable from J sums up to 0 when we sum over $\mathbf{b}$, and any other monomial is multiplied by m. Thus, $A(\mathbf{y}_I) = m \cdot \tilde{g}$, where $\tilde{g}$ is the part of g which does not include a variable from J, and in particular it is a homogeneous function of degree d_1.

Thus, the sum we look at is proportional to $\langle \tilde{g}, g_{J\to\mathbf{a}} \rangle = \langle \tilde{g}, \tilde{g}_{J\to\mathbf{a}} \rangle$. The last equality holds since any monomial in $(g - \tilde{g})_{J\to\mathbf{a}}$ has degree at most $d-1$ and hence has inner product 0 with $\tilde{g}$. In particular, as $\langle \tilde{g}, \tilde{g}_{J\to\mathbf{a}} \rangle$ does not depend on $\mathbf{a}$, we get that the quantity $(M^*M\vec{1})_{\mathbf{a}}$ is the same for all $\mathbf{a}$, and hence $\vec{1}$ is an eigenvector of M^*M.

This means that in choosing the eigenbasis $g_1', \ldots, g_m'$, we can ensure that $g_1' \equiv 1$. This gives us the third and fourth bullets, and next we choose $g_1, \ldots, g_m$. Define $\tilde{g}_r = Mg_r'$; first we note that $\tilde{g}_r$ are orthogonal. Indeed,

$$\langle \tilde{g}_r, \tilde{g}_{r'} \rangle = \langle Mg_r', Mg_{r'}' \rangle = \langle M^*Mg_r', g_{r'}' \rangle = \lambda_r \langle g_r', g_{r'}' \rangle = \lambda_r 1_{r=r'}$$

This means that if we look only at the set R of i's such that $\lambda_i \neq 0$, then we get that $\{\tilde{g}_i\}_{i\in R}$ is orthogonal, and we choose $g_r = \tilde{g}_r / \sqrt{\lambda_r}$ (which has 2-norm equal to 1)

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

We prove that

$$M = \sum_{r \in R} \sqrt{\lambda_r} g_r(\mathbf{y}_I) \overline{g'_r(\mathbf{y}_J)}.$$

To see that, define $M' = \sum_{r \in R} \lambda_r g_r(\mathbf{y}_I) \overline{g'_r(\mathbf{y}_J)}$, and note that for all $r \in R$ we have $Mg'_r = M'g'_r = \sqrt{\lambda_r} g_r \neq 0$, and for $r \notin R$ we have $Mg'_r = 0$ as well as $M'g'_r = 0$, and hence $Mg'_r = M'g'_r$ for all r. This implies that $M = M'$, and we have thus establishes the decomposition of g as well as the first item (we can freely define g_r for $r \notin R$ by noting that as $\lambda_r = 0$ there, it doesn't change anything).

For the fifth item, we observe that

$$\sum_r \sqrt{\lambda_r}^2 = \sum_r \lambda_r = \frac{\mathrm{Tr}(M^*M)}{|\Gamma|^n} = \frac{1}{|\Gamma|^n} \sum_b (M^*M)_{b,b} = \frac{1}{|\Gamma|^n} \sum_{b,y} \left|g(y,b)\right|^2 = \|g\|_2^2 = 1.$$

Finally, we argue for second item. If $1 \in R$, then we get that $g_1(\mathbf{y}_I) = \lambda_1^{-1/2} \mathbb{E}_{\mathbf{y}_J}\left[g(\mathbf{y})\right]$, from which it is clear that g_1 is homogeneous of degree d. For $r \neq 1$ in R, we get that

$$g_r(\mathbf{y}_I) = \lambda_r^{-1/2} \mathop{\mathbb{E}}_{\mathbf{y}_J}\left[g(\mathbf{y}) g'_r(\mathbf{y}_J)\right],$$

and we note that on the right hand side only monomials in g that contain the variable from J can contribute (the rest give 0), hence we get a homogeneous function of degree $d - 1$. □

B.2 Proof of Claim 5.10

Proof. We run the same argument as in the proof of Claim 5.8, replacing λ_r's with γ_t's, and get that

$$\sum_{t \in T} \sqrt{\gamma_t} F_t(\mathbf{y}_I, \mathbf{z}_I) \overline{F'_t(\mathbf{y}_J, \mathbf{z}_J)}.$$

This gives the above decomposition and shows that it satisfies all items except for the third and fourth. For $t = 1$, we note as in the proof of Claim 5.8 that

$$F_1(\mathbf{y}_I, \mathbf{z}_I) = \gamma_1^{-1/2} \mathop{\mathbb{E}}_{\mathbf{y}_J, \mathbf{z}_J}\left[F(\mathbf{y}, \mathbf{z})\right],$$

from which it is clear that F_1 has effective degree at least d' and that it is constant on connected components.

For $t \geqslant 2$ we have that

$$F_t(\mathbf{y}_I, \mathbf{z}_I) = \gamma_t^{-1/2} \mathop{\mathbb{E}}_{\mathbf{y}_J, \mathbf{z}_J}\left[F(\mathbf{y}, \mathbf{z}) F'_t(\mathbf{y}_J, \mathbf{z}_J)\right],$$

and expanding F in as linear combination of monomials, we see that monomials from F not involving the variable from J retain their effective degree, while those monomials that involve the variable from J may drop their effective degree by at most 1 (that happens if the monomial has a character from WB_2 on that coordinate). It also follows from this representation that as F is constant on connected components, F_t is also constant on connected components.

Finally, we also note that

$$F'_t(\mathbf{y}_J, \mathbf{z}_J) = \gamma_t^{-1/2} \mathop{\mathbb{E}}_{\mathbf{y}_I, \mathbf{z}_I}\left[\overline{F(\mathbf{y}, \mathbf{z})} F_t(\mathbf{y}_I, \mathbf{z}_I)\right],$$

from which it is clear that F'_t is constant on connected components. This completes the proof of the third item, and therefore of Claim 5.10. □

https://doi.org/10.1017/S0963548325100114 Published online by Cambridge University Press

B.3 Proof of Claim 5.11

Proof. We think of g as a matrix M in $\mathbb{C}^{\Gamma^I \times \Gamma^J}$, whose $(\mathbf{a}, \mathbf{b})$ entry is $g(\mathbf{y}_I = \mathbf{a}, \mathbf{y}_J = \mathbf{b})$. The decomposition stated by the claim is an appropriately chosen singular-value decomposition of M; below are the details for completeness.

Looking at $M^*M \in \mathbb{C}^{\Gamma^J \times \Gamma^J}$, we see that it is an $m \times m$ Hermitian positive semi-definite matrix, hence we may find an eigenbasis $g'_1, \ldots, g'_m$ of $\mathbb{C}^{\Gamma^J}$ with non-negative eigenvalues $\lambda_1, \ldots, \lambda_m$. We define $\tilde{g}_r = Mg'_r$, set $R = \{r \mid \lambda_r \neq 0\}$ and normalise $g_r = \tilde{g}_r/\sqrt{\lambda_r}$ for $r \in R$, and as in Claim 5.8 we get that

$$M = \sum_{r \in R} \sqrt{\lambda_r} g_r(\mathbf{y}_I) \overline{g'_r(\mathbf{y}_J)},$$

as well as the first, third, and fourth items.

Finally, we argue for second item. For $r \in R$, then we get that $g_r(\mathbf{y}_I) = \lambda_r^{-1/2} \mathbb{E}_{\mathbf{y}_J}\left[g(\mathbf{y}) g'_r(\mathbf{y}_J)\right]$, and as for each $\mathbf{y}_J$, $g(\mathbf{y}) g'_r(\mathbf{y}_J)$ contains only monomials of degree $d - 1$ onwards, we get that the same is true for g_r. □

Cite this article: Bhangale A, Khot S, and Minzer D (2025). On approximability of satisfiable ***k***-CSPs: II. *Combinatorics, Probability and Computing*. https://doi.org/10.1017/S0963548325100114