

A Counting Lemma for Somewhat Restricted 3-APs

Amey Bhangale*

Subhash Khot†

Yang P. Liu‡

Dor Minzer§

Abstract

For a prime $p \geq 3$, a somewhat restricted 3-AP in $\mathbb{F}_p^n$ is a triplet $(x, x + a, x + 2a)$, where $x \in \mathbb{F}_p^n$ and $a \in \{0, 1, 2\}^n$. We prove a counting lemma for somewhat restricted 3-APs in dense sets in $\mathbb{F}_p^n$. More precisely, we prove that for all $\alpha > 0$, there exists $\beta > 0$, such that for sufficiently large n , if a set $A \subseteq \mathbb{F}_p^n$ has density at least α , then it contains at least β fraction of all somewhat restricted 3-APs.

Our proof builds on recently developed machinery from [Bhangale, Khot, Minzer, 2026]. Our main new ingredient is an arithmetic regularity lemma for patterns such as somewhat restricted 3-APs. This result is in the spirit of arithmetic regularity lemmas from the theory of Gowers uniformity norms [Green, Tao, 2010] and may be of independent interest.

1 Introduction

The emergence of arithmetic patterns in dense sets, first studied by Roth [Rot53], has proven to be a fundamental problem relating distinct areas in mathematics. Roth's theorem asserts that for all $\varepsilon > 0$, if $A \subseteq \{1, \dots, N\}$ has density at least ε and N is sufficiently large, then A must contain a triplet of the form $x, x + a, x + 2a$ with $a \neq 0$. Variants and extensions of Roth's theorem have received much attention over the years, and their study led to the development of important tools in combinatorics, analysis, ergodic theory, probability and more. Below are a few of these results:

1. **Quantitative bounds:** the precise density needed for the conclusion to hold has been the subject of intense study [Bou99, Bou08, San11, BS19, BS20, KM23]. By now, it is known that the answer is quasi-polynomial [KM23, BS23], matching the shape of the best known construction [Beh46, EHPS24].
2. **Longer progressions:** Szemerédi's [Sze75] extended Roth's theorem to progressions of arbitrarily large constant length. Further study of this problem led to the development of higher-order Fourier analysis [Gow98, Gow01] and of graph and hypergraph regularity [Sze78, RS78, Gow07].
3. **More general patterns:** the study of arithmetic progressions has motivated the broader question of the occurrence of general patterns in sufficiently dense sets. A concrete class of examples are polynomial progressions [BL96], which are patterns of the form $x + P_1(a), \dots, x + P_k(a)$ where $P_1, \dots, P_k$ are polynomials. Other examples include multi-dimensional patterns [FK78, Gow07] (such as $x + ia_1 + ja_2$ for $i, j = 0, 1, 2$), and combinatorial lines [FK91, Pol12, BKLM25].

*Department of Computer Science and Engineering, University of California, Riverside. Supported by the Hellman Fellowship award and NSF CAREER award 2440882.

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

‡School of Mathematics, Institute for Advanced Study, Princeton, NJ. This material is based upon work supported by the National Science Foundation under Grant No. DMS-1926686

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

The current work fits in the third item above, and we study restricted forms of 3-APs in the finite field setting, in which the common difference must come from a set much smaller than the domain.

1.1 The Finite Field Setting

In the finite field setting, one considers the problem of finding arithmetic patterns in a dense set $A \subseteq \mathbb{F}_p^n$. Here and throughout, p is thought of as a prime of constant size and n is thought of as tending to infinity. The theory in this setting is very close to the theory in the integer setting, and it avoids some of the technical difficulties (such as dealing with Bohr sets). In fact, the tools in the two settings are close analogs of one another, with the exception of the polynomial method [CLP17, EG17], which only seems applicable for some problems in the finite field setting.

Meshulam [Mes95] proved an analog of Roth's theorem in the finite field setting, following a similar density increment, Fourier-analytic strategy. Our work is mostly concerned with studying more restricted patterns than the ones addressed in Meshulam's work, defined as follows:

Definition 1.1. Fix a prime $p \geq 3$.

1. A somewhat restricted 3-AP in $\mathbb{F}_p^n$ is a triplet of the form $(x, x + a, x + 2a)$ such that $x \in \mathbb{F}_p^n$ and $a \in \{0, 1, 2\}^n \setminus \{\vec{0}\}$.
2. A restricted 3-AP in $\mathbb{F}_p^n$ is a triplet of the form $(x, x + a, x + 2a)$ such that $x \in \mathbb{F}_p^n$ and $a \in \{0, 1\}^n \setminus \{\vec{0}\}$.

One may consider two distinct problems with respect to restricted 3-APs and somewhat restricted 3-APs:

1. **Existence:** what density guarantees that A must contain at least one pattern?
2. **Counting:** if A has density $\alpha > 0$, does it have to contain $\Omega_\alpha(1)$ fraction the patterns (provided n is sufficiently large)?

The Fourier analytic machinery employed by Meshulam [Mes95] and subsequent works does not seem to apply well to this setting, and these problems were highlighted in [Gre25, HHM18]. For the first problem, the density Hales Jewett theorem implies that any constant density must contain at least one pattern. The quantitative bounds this argument gives were quite weak, and only recently effective bounds were given. For both problems, it is now known [BKM24a, BKLM25] that if $A \subseteq \mathbb{F}_p^n$ contains no restricted 3-AP, then it has density at most $1/(\log \log \log n)^{c_p}$ where $c_p > 0$ is a constant. As far as we know it may be the case that A must have exponentially small measure, and a recent result [CFP26] shows this is indeed the case for the richer class of progressions of the form $(x, x + a, x + 2a)$ where $a \in \{0, 1, \dots, (p+1)/2\}^n$.

The main objective of the current paper is to answer the counting problem above for somewhat restricted 3-APs. Note that for standard 3-APs, an existence result immediately yields a counting result, and we recall this argument next. Suppose that we know that for all $\alpha > 0$ there is $n_0(p, \alpha)$ such that if $A \subseteq \mathbb{F}_p^n$ has density at least α , then it has a 3-AP. Now fix $A \subseteq \mathbb{F}_p^n$ of density at least α , denote

$$3\text{AP}(A, W) = |\{x, a \in W \mid x, x + a, x + 2a \in A\}|,$$

and take $n_0 = n_0(\alpha/2, p)$. Taking an affine subspace $W \subset \mathbb{F}_p^n$ of dimension n_0 randomly, by an averaging argument we have that $|A \cap W| \geq \frac{1}{2}\alpha |W|$ with probability at least $\alpha/2$, and then by assumption $A \cap W$ must contain a 3-AP. Thus, $3\text{AP}(A, W) \geq 1$ with probability at least $\alpha/2$. On the other hand, we have that $\mathbb{E}_W [3\text{AP}(A, W)] = 3\text{AP}(A, \mathbb{F}_p^n) \cdot p^{2(n_0-n)}$, hence we conclude that $3\text{AP}(A, \mathbb{F}_p^n) \geq p^{2n} \frac{\alpha}{2p^{2n_0}} = \Omega_{\alpha,p}(p^{2n})$. Attempting to run such an argument in the setting of somewhat restricted 3-APs or restricted 3-APs fails

because these patterns are basis dependent. This raises the question of whether a counting lemma for such patterns still holds. Our main result, stated in the next section, asserts that the answer is positive in the case of somewhat restricted 3-APs.

1.2 Main Results

The main result of this paper is the following statement.

Theorem 1.2. *For all prime $p \geq 3$ and $\alpha > 0$, there exists $\beta > 0$, such that for large enough n , if $A \subseteq \mathbb{F}_p^n$ has density at least α , then*

$$\Pr_{\substack{x \in \mathbb{F}_p^n \\ a \in \{0,1,2\}^n}} [x, x+a, x+2a \in A] \geq \beta.$$

We do not give explicit quantitative bounds for β in Theorem 1.2. At best, the argument given here yields a bound for β that is inverse-tower-type in $1/\alpha$. This is because, as we discuss in Section 1.3, our argument uses an arithmetic regularity lemma in the spirit of [Gre05, GT10], which often leads to weak quantitative bounds. We believe that the statement of Theorem 1.2 should be true for restricted 3-APs, namely where $\{0, 1, 2\}^n$ is replaced with $\{0, 1\}^n$, but our methods cannot handle this yet. We discuss one of the key obstructions for this in Section 6.

Theorem 1.2 follows from our general result, which we now state. We need the following definition of embeddability of a distribution into an abelian group.

Definition 1.3. *Let Σ, Γ, Φ be finite alphabets, let μ be a distribution over $\Sigma \times \Gamma \times \Phi$, and let $(G, +)$ be an Abelian group. We say that μ has a $(G, +)$ -embedding if there are maps $\sigma: \Sigma \rightarrow G$, $\gamma: \Gamma \rightarrow G$, $\phi: \Phi \rightarrow G$ not all constant such that for all $(x, y, z) \in \text{supp}(\mu)$ we have*

$$\sigma(x) + \gamma(y) + \phi(z) = 0_G.$$

With this notion in mind, we state the general theorem below. Theorem 1.2 follows easily from this theorem (it is a routine exercise to check that somewhat restricted 3-APs do not admit $(\mathbb{Z}, +)$ -embeddings, see [BKM24a], for instance).

Theorem 1.4. *Let Σ be an alphabet of size at most m , and let μ be a distribution over Σ^3 and ν be a distribution over Σ such that*

1. *The marginal of μ on each coordinate is equal to ν , and $(x, x, x) \in \text{supp}(\mu)$.*
2. *The probability of each atom in μ is at least $s > 0$.*
3. *The distribution μ has no $(\mathbb{Z}, +)$ -embedding.*

Then for all $\alpha > 0$, there exists $\beta = \beta(\alpha, m, s) > 0$ such that for sufficiently large n , if $A \subseteq \Sigma^n$ has $\nu^{\otimes n}(A) \geq \alpha$, then

$$\Pr_{(x,y,z) \sim \mu^{\otimes n}} [x, y, z \in A] \geq \beta.$$

We remark that the first two conditions in Theorem 1.4 are necessary for the result to hold, but we believe that the third condition can be relaxed.

1.3 Techniques

In this section we give an overview of the proof of Theorem 1.2. We first outline a direct argument for the counting lemma for standard 3-APs, and then explain how to generalize it to our setting.

1.3.1 A Counting Lemma for Standard 3-APs

Let $A \subseteq \mathbb{F}_p^n$ have measure $\alpha > 0$ which we think of as a small constant, and denote for ease of notation $f(x) = 1_A(x)$. A standard Fourier analytic computation shows that if all non-trivial Fourier coefficients of f are at most ε in absolute value, then $\mathbb{E}_{x,a \in \mathbb{F}_p^n} [f(x)f(x+a)f(x+2a)] \geq \alpha^3 - \varepsilon\alpha$. Therefore, if all Fourier coefficients of f are at most $\alpha^2/2$ in absolute value, the conclusion immediately follows. Otherwise, one could pick a heavy Fourier character $0 \neq \gamma \in \mathbb{F}_p^n$ and show that the density of f increases on one of the hyperplanes defined by γ , and iterate the argument there. This works in the case of standard 3-APs (because a co-dimension 1 subspace contains a constant fraction of 3-APs), but fails in the case of somewhat restricted 3-APs because the density increment argument we know [BKM24a] passes down to a subspace of super-constant co-dimension. Thus, we wish to come up with an argument that avoids this type of iterative approach.

To do so, upon finding γ , one can define the averaging operator $T_\gamma: L_2(\mathbb{F}_p^n) \rightarrow L_2(\mathbb{F}_p^n)$ by

$$T_\gamma f(x) = \mathbb{E}_{y \in \mathbb{F}_p^n} [f(y) \mid \chi_\gamma(y) = \chi_\gamma(x)].$$

In words, T_γ averages over the cosets of the linear hyperplane defined by γ . For a single γ , one can observe that $T_\gamma f(x) = \sum_{t \in \mathbb{F}_p} \widehat{f}(t\gamma) \chi_{t\gamma}(x)$, so $\widehat{f}(\gamma)$ being large in absolute value means that the 2-norm of $(I - T_\gamma)f$ is noticeably smaller than that of f . Iterating, one finds a collection of characters $\mathcal{C} = \{\gamma_1, \dots, \gamma_r\}$, such that defining

$$T_{\mathcal{C}} f(x) = \mathbb{E}_{y \in \mathbb{F}_p^n} [f(y) \mid \chi_\gamma(y) = \chi_\gamma(x) \forall \gamma \in \mathcal{C}],$$

all Fourier coefficients of $(I - T_{\mathcal{C}})f$ are at most $O(1/\sqrt{r})$ in absolute value. Thus, one can write

$$\mathbb{E}_{x,a \in \mathbb{F}_p^n} [f(x)f(x+a)f(x+2a)] \geq \mathbb{E}_{x,a \in \mathbb{F}_p^n} [T_{\mathcal{C}} f(x)T_{\mathcal{C}} f(x+a)T_{\mathcal{C}} f(x+2a)] - O\left(\frac{1}{\sqrt{r}}\right), \quad (1)$$

so the task now is to analyze the right-hand side and show it is noticeable. Because the function $T_{\mathcal{C}} f$ is very structured and has only $O_r(1)$ many non-zero Fourier coefficients, this expectation seems more feasible to analyze. Indeed, taking W to be the linear subspace of codimension at most r defined by the vectors in $\mathcal{C}$, it is easy to observe that:

- Taking a random shift $v + W$ of W , with probability at least $\alpha/2$, the average of f inside $v + W$ is at least $\alpha/2$. Noting that $T_{\mathcal{C}} f$ is constant on each such affine shift, we get that with probability at least $\alpha/2$ over the choice of v , the value of $T_{\mathcal{C}} f$ on all points of $v + W$ is at least $\alpha/2$.
- Sampling $x, a \in \mathbb{F}_p^n$, the probability that $x, x+a, x+2a$ are all in $v + W$ is $p^{-2\text{codim}(W)}$, which is at least p^{-2r} .

With these two items in mind, one can get that

$$\mathbb{E}_{x,a \in \mathbb{F}_p^n} [T_{\mathcal{C}} f(x)T_{\mathcal{C}} f(x+a)T_{\mathcal{C}} f(x+2a)] \geq p^{-2r} \left(\frac{\alpha}{2}\right)^4,$$

which initially seems good, but upon further inspection is too weak.¹ Indeed, this bound is quantitatively weaker than the error term on the right hand side of (1), so it does not give any lower bound on $\mathbb{E}_{x,a \in \mathbb{F}_p^n} [f(x)f(x+a)f(x+2a)]$.

At a high-level, the attempt above can be thought of as decomposing f into a structure function $T_C f$, and a Fourier-quasi random function $(I - T_C)f$, and ultimately the issue is that there is an unfavorable tradeoff between the quasi-randomness parameter of $(I - T_C)f$ (which governs the error term) and the size of C (which governs the codimension of W). It turns out that one can break this dependency using stronger decompositions known as arithmetic regularity lemmas, appearing implicitly in [Bou86] for U^2 -norm and explicitly in [Gre05, GT10]. The idea is that by iterating the above process several times, fixing an arbitrary decay function $w: (0, 1)^2 \rightarrow (0, 1)$, one can find two collections C and C' such that:

- All Fourier coefficients of $(I - T_C)f$ are at most ε in absolute value.
- All Fourier coefficients of $(I - T_{C'})f$ are at most $\varepsilon' = w(\varepsilon, |C|^{-1})$ in absolute value.
- The functions $(I - T_C)f$ and $(I - T_{C'})f$ are ε -close in L_2 -norm.

With this in mind, one can replace (1) with

$$\mathbb{E}_{x,a \in \mathbb{F}_p^n} [f(x)f(x+a)f(x+2a)] \geq \mathbb{E}_{x,a \in \mathbb{F}_p^n} [T_{C'} f(x)T_{C'} f(x+a)T_{C'} f(x+2a)] - O(\varepsilon'). \quad (2)$$

The point now is that one can still take the linear subspace W as above with respect to C (and not to C'). While it is no longer the case that $T_{C'} f$ is constant on cosets of W , we know it is close to being constant on a typical coset, thanks to the L_2 -closeness to $T_C f$. This is sufficient to repair the argument above.

1.3.2 A Counting Lemma for Somewhat Restricted 3-APs

We now explain the way our argument for somewhat restricted 3-APs fits the above template.

The first complication is that the direct Fourier-analytic computation showing that small Fourier coefficients imply counting no longer works. This issue was already studied in [BKM24a, BKM24b], and morally speaking, it is known that if f has correlation at most η with all functions of the form $\chi_\gamma(x) \cdot L(x)$ where $\gamma \in \mathbb{F}_p^n$ and $L: \mathbb{F}_p^n \rightarrow \mathbb{C}$ has $\|L\|_2 = 1$ and Fourier analytic degree at most d , then

$$\mathbb{E}_{\substack{x \in \mathbb{F}_p^n \\ a \in \{0,1,2\}^n}} [f(x)f(x+a)f(x+2a)] \geq \alpha^3 - \delta(\eta, d),$$

where $\delta(\eta, d)$ goes to 0 as $\eta \rightarrow 0$ and $d \rightarrow \infty$; see Theorem 2.20 for a more precise statement.

The second complication that arises is that one has to come up with an analog of the averaging operator T_C above. This was done in [BKM26], and they introduce noise operators $T_{C,1-\varepsilon}$. Morally speaking, for $x \in \mathbb{F}_p^n$, the value of $T_{C,1-\varepsilon} f(x)$ can be thought of as the average of $f(y)$ over points y such that $\chi_\gamma(y) = \chi_\gamma(x)$ for all $\gamma \in C$, and additionally x and y agree on roughly $(1-\varepsilon)$ -fraction of the coordinates. A key feature of this operator is that it acts like the identity operator on functions of the form $\chi'_\gamma(x)L'(x)$ with $\gamma \in \text{span}_{\mathbb{N}}(C)$ and $\deg(L') \ll 1/\varepsilon$, which suggests it is the correct analog of T_C above. Indeed, [BKM26] prove a regularity lemma with respect to this operator, showing that for all $\eta > 0$, given f , one can find C of size $O_\eta(1)$ and $\varepsilon = \Omega_\eta(1)$, such that $(I - T_{C,1-\varepsilon})f$ has correlation at most η with all functions of the form $\chi_\gamma(x)L(x)$. We remark that technically, the operator $T_{C,1-\varepsilon}$ is significantly harder to work with

¹A more careful argument gives $p^{-r}\alpha^3$, but this does not make a difference.

compared to T_C , and this is partly because it does not behave as nicely as T_C with respect to the Fourier decomposition.

With this in mind, attempting to reproduce the initial failed attempt, one could start with the inequality

$$\mathbb{E}_{\substack{x \in \mathbb{F}_p^n \\ a \in \{0,1,2\}^n}} [f(x)f(x+a)f(x+2a)] \geq \mathbb{E}_{\substack{x \in \mathbb{F}_p^n \\ a \in \{0,1,2\}^n}} [T_{C,1-\varepsilon}f(x)T_{C,1-\varepsilon}f(x+a)T_{C,1-\varepsilon}f(x+2a)] - \delta(\eta, 1/\varepsilon),$$

where $\lim_{\varepsilon, \eta \rightarrow 0} \delta(\eta, 1/\varepsilon) = 0$, and then analyze the expectation on the right hand side. Towards this end, one may use the fact from [BKM26] that the function $T_{C,1-\varepsilon}f$ can be ξ -approximated in L_2 -distance by a function of the form

$$\sum_{\gamma \in \text{span}(C)} \chi_\gamma(x) L_\gamma(x),$$

where L_γ are all of degree and 2-norm at most $O_{\xi, \eta}(1)$. Intuitively, this says that after we restrict $T_{C,1-\varepsilon}f$ down to a coset of W , we will get a low-degree function (instead of a constant function). This is of course more complicated to analyze, but luckily (after appropriate preprocessing), one could use the invariance principle of [MOO10] to relate the resulting expectation to a similar looking expectation over Gaussian space, which can be lower bounded via results of [CDP15]. We remark that to actually make such an argument go through, one needs to ensure that the collection C has a large rank (roughly asserting that no non-trivial combination of vectors from C has small Hamming weight). This leads to significant complications which ignore in this overview. We also remark that the ability to solve the Gaussian problem we get in the end is the primary reason we are only able to handle somewhat restricted 3-APs at the moment (as opposed to restricted 3-APs), and we discuss this further in Section 6.

With the initial failed attempt recovered, the main contribution of this paper is to show an appropriate analog of the arithmetic regularity lemma for operators of the form $T_{C,1-\varepsilon}$, and this is achieved in Lemma 4.1. We then boost it to Lemma 4.4, which makes the same assertion with additional quasirandomness properties of the collection C in the spirit of rank discussed above (see Definitions 2.13 and 2.14). Finally, in Section 5 we show how to use our arithmetic regularity lemma in conjunction with the invariance principle to prove Theorem 1.4.

Statement of AI use: we used ChatGPT 5.5 during the polishing of this write-up, and to write the proofs from Section A. All other mathematical content in this paper is due to the authors.

2 Preliminaries

Notations. For a vector $x \in \Sigma^n$ and a subset $I \subseteq [n]$ of coordinates, we denote by x_I the vector in Σ^I which results by dropping from x all coordinates outside I . We denote by $\mathbf{i}$ the complex root of -1 , by $\bar{a}$ the complex conjugate of the number $a \in \mathbb{C}$, and by $\mathbb{D} \subseteq \mathbb{C}$ the set of complex numbers of absolute value 1. We denote by $I \subseteq_\rho [n]$ a random subset of $[n]$ in which we include each $i \in [n]$ independently with probability ρ . We will use standard big- O notations, and write $A = O(B)$ if $A \leq C \cdot B$ for an absolute constant C , and $A = \Omega(B)$ if $A \geq C \cdot B$ for an absolute constant $C > 0$. If there is a dependency of the hidden constant on some auxiliary parameter, say m , we denote $A = O_m(B)$ and $A = \Omega_m(B)$.

Our arguments will use sufficiently rapid decay functions $w: (0, 1) \rightarrow (0, 1)$, which by default will always be monotonically increasing and satisfying $w(\varepsilon) \leq \varepsilon$. We say that a tuple of scales $\rho_0, \dots, \rho_\ell$ w -separated, and denote

$$\rho_\ell \ll_w \rho_{\ell-1} \ll_w \dots \ll_w \rho_0,$$

if $\rho_{j+1} \leq w(\rho_j)$ for all $0 \leq j < \ell$.

2.1 Product Functions

Definition 2.1. Let Σ be a finite alphabet and $n \in \mathbb{N}$. A function $P: \Sigma^n \rightarrow \mathbb{D}$ is called a product function if there are functions $P_1, \dots, P_n: \Sigma \rightarrow \mathbb{D}$ such that $P(x) = \prod_{i=1}^n P_i(x_i)$.

Given a product function $P: \Sigma^n \rightarrow \mathbb{D}$, the decomposition as $P_1(x_1) \cdots P_n(x_n)$ is only unique up to multiplying the P_i 's by constants of absolute value 1, and this will be inconvenient for us at times. To remedy that we fix a distinguished element $\sigma^* \in \Sigma$, and by default normalize so that $P(\sigma^*, \dots, \sigma^*) = 1$ and $P_1(\sigma^*) = \dots = P_n(\sigma^*) = 1$. Unless specified otherwise, throughout this paper we will consider normalized product functions P and use their normalized decompositions.

We next define the projections of a product function.

Definition 2.2. Given a product function $P(x) = \prod_{i=1}^n P_i(x_i)$ and $I \subseteq [n]$, we define the function $P|_I: \Sigma^n \rightarrow \mathbb{D}$ as $P|_I(x) = \prod_{i \in I} P_i(x_i)$. For a collection of product functions $\mathcal{P} \subseteq \{P: \Sigma^n \rightarrow \mathbb{D}\}$ and $I \subseteq [n]$, we denote by $\mathcal{P}|_I$ the collection $\{P|_I \mid P \in \mathcal{P}\}$.

Next, we define the order of a product function and the notion of cyclic and discrete product functions.

Definition 2.3. Let $P: \Sigma^n \rightarrow \mathbb{D}$. We define the order of P , denoted by $\text{ord}(P)$, as the smallest integer k such that $P(x)^k = 1$ for all $x \in \Sigma^n$. If no such k exists we denote $\text{ord}(P) = \infty$.

Definition 2.4. Let $P: \Sigma^n \rightarrow \mathbb{D}$ be a product function.

1. We say that P is a cyclic product function if there is a prime p and an integer r such that $\text{ord}(P) = p^r$.
2. We say that P is M -discrete if $\text{ord}(P) \leq M$.

We have the following basic fact.

Fact 2.5. Let $P: \Sigma^n \rightarrow \mathbb{D}$ be a product function.

1. If $\text{ord}(P) = M < \infty$, we may find $\sigma_1, \dots, \sigma_n: \Sigma \rightarrow \mathbb{Z}_M$ such that $P(x) = e^{\frac{1}{M} 2\pi i \sum_{i=1}^n \sigma_i(x_i)}$.
2. If P is M -discrete, then for all $x, x' \in \Sigma^n$, we either have $P(x) = P(x')$, or $|P(x) - P(x')| \geq \frac{1}{M}$.

Proof. For the first item, writing $P(x) = \prod_{i=1}^n P_i(x_i)$ we get that P_i^M is constant for all i , and by normalization it must be the constant 1 function. Thus, it follows that there is $\sigma_i: \Sigma \rightarrow \mathbb{Z}_M$ such that $P_i(x_i) = e^{\frac{1}{M} 2\pi i \sigma_i(x_i)}$ for all $x_i \in \Sigma$, completing the proof of the first item.

For the second item, note that if $P(x) \neq P(x')$, then as they are both roots of unity of order $\text{ord}(P)$ we get that

$$|P(x) - P(x')| = \left| e^{\frac{1}{\text{ord}(P)} 2\pi i r} - 1 \right|$$

for some $0 < r < \text{ord}(P)$. We assume $r \leq \text{ord}(P)/2$, otherwise we can replace r with $\text{ord}(P) - r$ and the above equality is still valid. If $r \geq \text{ord}(P)/4$, then the real part of $e^{\frac{1}{\text{ord}(P)} 2\pi i r}$ is $\cos\left(\frac{r}{\text{ord}(P)} 2\pi\right) \leq 0$, we get $|P(x) - P(x')| \geq 1$. Otherwise, we get

$$\left| e^{\frac{1}{\text{ord}(P)} 2\pi i r} - 1 \right| \geq \sin\left(\frac{r}{\text{ord}(P)} 2\pi\right) \geq \frac{2}{\pi} \frac{r}{\text{ord}(P)} 2\pi \geq \frac{1}{\text{ord}(P)},$$

where we used the inequality $\sin(\theta) \geq \frac{2}{\pi}\theta$ for $0 \leq \theta \leq \frac{\pi}{2}$. This proves the second item. $\square$

The next claim shows that discrete product functions far from constants have small expectation and are very high degree. To state it, we define the standard noise operator over product spaces.

Definition 2.6. Let Σ be a finite alphabet and let ν be a distribution over Σ . For $\rho \in [0, 1]$ and $x \in \Sigma$, we define the distribution $y \sim T_{\rho, \nu} x$ as outputting x with probability ρ , and else outputting $y \sim \nu$. For $n \geq 2$ and $x \in \Sigma^n$, we define the distribution $T_{\rho, \nu}^{\otimes n} x$ by outputting y sampled by taking $y_i \sim T_{\rho, \nu} x_i$ independently for each i .

We also think of $T_{\rho, \nu}^{\otimes n}$ as a linear operator on $L_2(\Sigma^n, \nu^{\otimes n})$, defined as

$$T_{\rho, \nu}^{\otimes n} f(x) = \mathbb{E}_{y \sim T_{\rho, \nu}^{\otimes n} x} [f(y)]$$

for each $f: \Sigma^n \rightarrow \mathbb{C}$ and $x \in \Sigma^n$. When n and ν are clear from context, we often omit them from notation.

The next fact allows us to decompose a discrete product function as a product of cyclic product functions.

Fact 2.7. For all $M \in \mathbb{N}$ there exists $T \in \mathbb{N}$ such that the following holds. Suppose $P: \Sigma^n \rightarrow \mathbb{D}$ is an M -discrete product function. Then we may find cyclic product functions $Q_1, \dots, Q_T$ with $\text{ord}(Q_i) \leq M$ such that $P(x) = Q_1(x) \cdots Q_T(x)$.

Proof. Working coordinate by coordinate, it suffices to prove the statement for $n = 1$. By assumption we know that $P(x) = e^{\frac{1}{M'} 2\pi i k(x)}$ for some $M' \leq M$ and $k: \Sigma \rightarrow \mathbb{Z}_{M'}$. Factorize $M' = p_1^{r_1} \cdots p_T^{r_T}$, and define $M'_{-i} = \prod_{j \neq i} p_j^{r_j}$. As $\gcd(M'_{-1}, \dots, M'_{-T}) = 1$, we get by Bezout's identity that there are integers $a_1, \dots, a_T$ such that

$$1 = \sum_{i=1}^T a_i M'_{-i}.$$

Plugging this in, we get that

$$P(x) = e^{\frac{1}{M'} 2\pi i k(x)} = \prod_{i=1}^T e^{\frac{a_i M'_{-i}}{M'} 2\pi i k(x)} = \prod_{i=1}^T e^{\frac{a_i}{p_i^{r_i}} 2\pi i k(x)}.$$

Defining $Q_i(x) = e^{\frac{a_i}{p_i^{r_i}} 2\pi i k(x)}$ gives the result. $\square$

We will often consider combinations of functions from a collection $\mathcal{P}$, and we make the following definition.

Definition 2.8. Let $\mathcal{P}$ be a collection of product functions over Σ^n . We define

$$\text{span}_{\mathbb{N}}(\mathcal{P}) = \left\{ \prod_{P \in \mathcal{P}} P(x)^{\alpha_P} \mid \alpha_P \in \mathbb{N} \forall P \in \mathcal{P} \right\}.$$

We note that if each function in $\mathcal{P}$ has order at most $M < \infty$, then $\text{span}_{\mathbb{N}}(\mathcal{P})$ is finite and has size at most $M^{|\mathcal{P}|}$.

2.2 Two Rank Notions

In this section we define the two notions of rank used in our argument. Both use the symbolic distance between product functions, defined as follows.

Definition 2.9. Let $P, P': \Sigma^n \rightarrow \mathbb{D}$ be product functions. We define $\Delta_{\text{symbolic}}(P, P')$ to be the smallest integer k such that there are $u_1, u'_1, \dots, u_n, u'_n: \Sigma \rightarrow \mathbb{D}$ such that

$$P(x) = \prod_{i=1}^n u_i(x_i), \quad P'(x) = \prod_{i=1}^n u'_i(x_i)$$

and $u_i = u'_i$ for all but k of $i \in [n]$.

Definition 2.10. For $P: \Sigma^n \rightarrow \mathbb{D}$ and a collection of product functions $\mathcal{P}$ over Σ^n , we define

$$\Delta_{\text{symbolic}}(P, \text{span}_{\mathbb{N}}(\mathcal{P})) = \min_{Q \in \text{span}_{\mathbb{N}}(\mathcal{P})} \Delta_{\text{symbolic}}(P, Q).$$

The following simple fact will be used throughout the paper.

Fact 2.11. Let $P, Q, R: \Sigma^n \rightarrow \mathbb{D}$ be product functions. Then:

1. $\Delta_{\text{symbolic}}(P, R) \leq \Delta_{\text{symbolic}}(P, Q) + \Delta_{\text{symbolic}}(Q, R)$.
2. For all integers k , $\Delta_{\text{symbolic}}(P^k, Q^k) \leq \Delta_{\text{symbolic}}(P, Q)$.

Proof. The proof of both items is immediate by definition. □

Fact 2.12. Let $P: \Sigma^n \rightarrow \mathbb{D}$ be an M -discrete product function, and let ν be a distribution over Σ in which the probability of each atom is at least α . Suppose that $\Delta_{\text{symbolic}}(P, 1) \geq D$. Then:

1. $|\mathbb{E}_{x \sim \nu^{\otimes n}} [P(x)]| \leq 2^{-\Omega_{\alpha, M}(D)}$.
2. For all $\tau > 0$, $\|\text{T}_{1-\tau}^{\otimes n} P\|_2 \leq 2^{-\Omega_{\alpha, \tau, M}(D)}$.

Proof. Write $P(x) = \prod_{i=1}^n P_i(x_i)$. By independence, it follows that

$$\left| \mathbb{E}_{x \sim \nu^{\otimes n}} [P(x)] \right| = \prod_{i=1}^n \left| \mathbb{E}_{x_i \sim \nu} [P_i(x_i)] \right|. \tag{3}$$

By assumption at least D of the P_i are not constant, and without loss of generality we assume these are $P_1, \dots, P_D$. Fix $1 \leq i \leq D$. Then

$$\left| \mathbb{E}_{x_i \sim \nu} [P_i(x_i)] \right|^2 = \mathbb{E}_{x_i, x'_i \sim \nu} [P_i(x_i) \overline{P_i(x'_i)}] = 1 - \frac{1}{2} \mathbb{E}_{x_i, x'_i \sim \nu} [|P_i(x_i) - P_i(x'_i)|^2]$$

Since P_i is not constant there are $a, b \in \Sigma$ such that $P_i(a) \neq P_i(b)$, so by Fact 2.5 $|P_i(a) - P_i(b)| \geq \frac{1}{M}$. It follows that $\frac{1}{2} \mathbb{E}_{x_i, x'_i \sim \nu} [|P_i(x_i) - P_i(x'_i)|^2] \geq \alpha^2/M$, and so $|\mathbb{E}_{x_i \sim \nu} [P_i(x_i)]|^2 \leq 1 - \frac{\alpha^2}{2M}$. Plugging into (3) finishes the proof of the first item.

For the second item, by independence again $\|T_{1-\tau}^{\otimes n} P\|_2 = \prod_{i=1}^n \|T_{1-\tau} P_i\|_2$. Fixing $1 \leq i \leq D$, we get that

$$\|T_{1-\tau} P_i\|_2^2 = \mathbb{E}_{\substack{x_i \sim \nu \\ x'_i \sim T_{1-\tau} x_i}} [P_i(x_i) \overline{P_i(x'_i)}] = 1 - \frac{1}{2} \mathbb{E}_{\substack{x_i \sim \nu \\ x'_i \sim T_{1-\tau} x_i}} [|P_i(x_i) - P_i(x'_i)|^2].$$

By the same argument as before, it follows that $\mathbb{E}_{\substack{x_i \sim \nu \\ x'_i \sim T_{1-\tau} x_i}} [|P_i(x_i) - P_i(x'_i)|^2] \geq \frac{\alpha^2 \tau}{M}$, and the second item follows. $\square$

The following definition is the stronger notion of rank we consider, and roughly speaking it asserts that non-trivial combinations of functions from $\mathcal{P}$ are always high degree.

Definition 2.13. Let $\mathcal{P}$ be a collection of product functions. We say that it has rank at least D , and denote $\text{rk}(\mathcal{P}) \geq D$, if for all integers $\{\alpha_P\}_{P \in \mathcal{P}}$ such that $0 \leq \alpha_P < \text{ord}(P)$ and at least one is non zero, the function $S(x) = \prod_P P(x)^{\alpha_P}$ satisfies that $\Delta_{\text{symbolic}}(S, 1) \geq D$.

The following definition is a weaker notion of rank, and roughly speaking it asserts that any combination of functions from $\mathcal{P}$ either gives a constant function, or else a function far from constant in symbolic distance.

Definition 2.14. Let $\mathcal{P}$ be a collection of product functions. We say that it has soft-rank at least D , and denote $\text{soft-rk}(\mathcal{P}) \geq D$, if for all integers $\{\alpha_P\}_{P \in \mathcal{P}}$, defining $S(x) = \prod_P P(x)^{\alpha_P}$, we either have that S is constant or that $\Delta_{\text{symbolic}}(S, 1) \geq D$.

The contexts in which rank and soft-rank are used is somewhat different. The notion of rank is used in scenarios where we are only dealing with a single function defined over a probability space. In that scenario, rank governs certain pseudorandomness properties, such as the probability the function attains a certain value. The weaker soft-rank is used in scenarios we have multiple functions defined over correlated probability spaces. Therein, it governs some form of weak independence between them. To formally explain this multi-function scenario we require the following definition.

Definition 2.15. Suppose that Σ, Γ, Φ are finite alphabets, μ is a distribution over $\Sigma \times \Gamma \times \Phi$, and $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ are collections of product functions over Σ^n, Γ^n and Φ^n respectively. The collection of product functions $\mathcal{D}_\mu$ over $\text{supp}(\mu)^n$ is defined by including in it, for each $P \in \mathcal{P}$, (and similarly for each $Q \in \mathcal{Q}$ and $R \in \mathcal{R}$) the function $P'(x, y, z) = P(x)$.

In words, the collection $\mathcal{D}_\mu$ can be thought of as a joint lifting of the collections $\mathcal{P}, \mathcal{Q}$ and $\mathcal{R}$ to a common domain. Thus, the rank and soft-rank of $\mathcal{D}_\mu$ can be thought of as pseudorandomness notions measuring the

correlations/independence of $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ over their correlated probability spaces. Though much of what we say holds for general $\mathcal{P}, \mathcal{Q}, \mathcal{R}$, the case where they are all the same will be sufficient for us, and we make the following definition.

Definition 2.16. *Let $\mathcal{P}$ be a collection of product functions, and let μ be a distribution over Σ^3 . We define $\text{softrk}_\mu(\mathcal{P}) = \text{softrk}(\mathcal{D}_\mu(\mathcal{P}, \mathcal{P}, \mathcal{P}))$*

We now elaborate on the motivation behind the notions of rank and softrank, and on how one goes about achieving large rank/softrank.

1. **The single function setting:** suppose that we have a single collection of product functions $\mathcal{P}$, which we want to “process” into a high-rank collection. Intuitively, a violation to the high-rank condition means that there is a non-trivial dependency between functions in $\mathcal{P}$. If this dependency took a very simple form, such as being able to express $P_1 \in \mathcal{P}$ as a product of other $P \in \mathcal{P}$, then one could move to a smaller collection $\mathcal{P}' \subseteq \mathcal{P}$ in a way such that $T_{1-\varepsilon, \mathcal{P}} \approx T_{1-\varepsilon, \mathcal{P}'}$.² Indeed, we show that there is a way to modify $\mathcal{P}$ to achieve high rank, so that the noise operator does not change by too much; in fact, the rank we are able to guarantee is arbitrarily large compared to ε and the size of the collection.³

Having high rank is very helpful in analyzing events such as $\{x \mid P(x) = c_P \forall P \in \mathcal{P}\}$, for fixed values C_P . In particular, it allows one to give strong estimates of their probabilities, show strong mixing results on random walks starting in them, and so on.

2. **The multi-function setting:** suppose we have multiple collections of product functions $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ over $\Sigma^n, \Gamma^n, \Phi^n$ respectively, and we have a distribution μ over $\Sigma \times \Gamma \times \Phi$. Then $\mathcal{D} = \mathcal{D}_\mu(\mathcal{P}, \mathcal{Q}, \mathcal{R})$ is a lifting of $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ to the common domain $\text{supp}(\mu)^n$. Thus, if $\text{rk}(\mathcal{D})$ is large, then as discussed above, we would be able to analyze correlations between the values of $\mathcal{P}, \mathcal{Q}$ and $\mathcal{R}$ when the input is sampled from $\mu^{\otimes n}$ very well; in fact, their values would be nearly independently distributed. How can we ensure that $\mathcal{D}$ has a high rank, though?

Applying a modification procedure as in the last time is possible, but it results in modifications to $\mathcal{D}$, instead of to $\mathcal{P}, \mathcal{Q}, \mathcal{R}$. This presents issues in our analysis, the most serious of them is that we are unable to guarantee that the noise operators corresponding to each one of them do not change. To circumvent this issue, we instead use the weaker notion of softrank in such contexts, which turns out to be useful in our setting.

A procedure to gain high softrank is as follows. If $\mathcal{D}$ has small softrank, then we can combine in $\mathcal{D}$ to get a function that is close to constant in symbolic distance, which in our case is possible only when it depends on a small subset $I \subseteq [n]$ of the coordinates. By projecting each function in $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ on $\bar{I}$, the aforementioned combination becomes completely constant. Iterating this procedure, we get modifications of $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ whose joint lifting has a large softrank.⁴

2.3 The μ -norm and the CSP Stability Result

We need the following notions and results from [BKM26]. The first of which is a definition of a certain semi-norm.

²We remark that, in general the dependencies one has to deal with need not take this form, and then a different modification of $\mathcal{P}$ is necessary.

³Strictly speaking, this involves flexibility in choosing the noise rate, which we are able to achieve.

⁴Crucially, by taking the noise rate of the averaging operators to be sufficiently small, we can ensure that this modification does not change by much the noise operators of $\mathcal{P}, \mathcal{Q}, \mathcal{R}$.

Definition 2.17. For a distribution μ over $\Sigma \times \Gamma \times \Phi$ and a function $f: (\Sigma^n, \mu_x^{\otimes n}) \rightarrow \mathbb{C}$, we define the μ semi-norm of f as

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

For a distribution ν over Σ , we define a collection of distributions over Σ^3 in which one of the marginals is ν .

Definition 2.18. For an alphabet Σ , a distribution ν over Σ and a parameter $\alpha > 0$, define the collections

$$M_\nu = \left\{ \mu \mid \begin{array}{l} \mu \text{ is a pairwise connected distribution over } \Sigma^3 \text{ with no } (\mathbb{Z}, +)\text{-embedding} \\ \mu_x = \nu \text{ or } \mu_y = \nu \text{ or } \mu_z = \nu \end{array} \right\},$$

$$M_\alpha = \{ \mu \mid \mu(x, y, z) \geq \alpha \ \forall (x, y, z) \in \text{supp}(\mu) \},$$

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

Definition 2.19. Let ν be a distribution over Σ , let $\alpha > 0$, and let $f: (\Sigma^n, \nu^{\otimes n}) \rightarrow \mathbb{C}$ be a function. We define

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

When α is clear from context, we often omit it from the notation and write $\|f\|_\nu$ instead of $\|f\|_{\nu, \alpha}$.

The last result in this section is the inverse theorem from [BKM24b].

Theorem 2.20. For all $m \in \mathbb{N}$ there is a constant S_m and a finite Abelian group G of size at most S_m such that for all $\varepsilon, \alpha > 0$ there exists $\delta > 0$ and $d \in \mathbb{N}$ such that the following holds. Suppose that Σ, Γ, Φ are alphabets of size at most m , and μ is a distribution over $\Sigma \times \Gamma \times \Phi$ with no $\mathbb{Z}$ embeddings in which the probability of each atom is at least α . Then there exists an embedding (σ, γ, ϕ) of $\text{supp}(\mu)$ into G , such that if $f: \Sigma^n \rightarrow \mathbb{C}$ is a 1-bounded function with $\|f\|_\mu \geq \varepsilon$, then there is $\chi \in \hat{G}^{\otimes n}$ and $L: \Sigma^n \rightarrow \mathbb{C}$ of degree at most d and $\|L\|_2 \leq 1$ such that

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

2.4 Noise Operators

We will use the following noise operator from [BKM26].

Definition 2.21. Let Σ be a finite alphabet, let ν be a distribution over Σ , let $\mathcal{P} = \{P_1, \dots, P_r: \Sigma^n \rightarrow \mathbb{D}\}$ be a collection of functions and let $I \subseteq [n]$. For each $x \in \Sigma^n$ we define the distribution $T_{\nu, \mathcal{P}, I}x$ as:

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

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

As usual, we will associate with $T_{\nu, \mathcal{P}, I}, T_{\nu, \mathcal{P}, 1-\varepsilon}$ averaging operators over functions, which by abuse of notation we denote as $T_{\nu, \mathcal{P}, I}, T_{\nu, \mathcal{P}, 1-\varepsilon}: L_2(\Sigma^n, \nu^{\otimes n}) \rightarrow L_2(\Sigma^n, \nu^{\otimes n})$. We define

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

When the distribution ν is clear from context, we will often drop ν from notation and simply write $T_{\mathcal{P}, I}$ and $T_{\mathcal{P}, 1-\varepsilon}$ in place of $T_{\nu, \mathcal{P}, I}$ and $T_{\nu, \mathcal{P}, 1-\varepsilon}$.

2.4.1 Properties of the Noise Operator

We now list a few properties of the noise operator established in [BKM26]. The first one is [BKM26, Fact 4.2].

Fact 2.22. *For all $I \subseteq [n]$ and $\mathcal{P}$, $\nu^{\otimes n}$ is a stationary distribution of $T_{\nu, \mathcal{P}, I}$. Consequently $\nu^{\otimes n}$ is a stationary distribution of $T_{\nu, \mathcal{P}, 1-\varepsilon}$.*

We will need the following variant of [BKM26, Fact 4.3].

Fact 2.23. *Suppose that $\mathcal{P}, \mathcal{P}'$ are collections of product functions such that:*

1. *For all $P \in \mathcal{P}$, $\Delta_{\text{symbolic}}(P, \text{span}_{\mathbb{N}}(\mathcal{P}')) \leq k$.*
2. *For all $P' \in \mathcal{P}'$, $\Delta_{\text{symbolic}}(P', \text{span}_{\mathbb{N}}(\mathcal{P})) \leq k$.*

Then for all $\varepsilon > 0$:

1. *There is a coupling of (x, y, y') such that (x, y) is distributed according to $(x, T_{\nu, \mathcal{P}, 1-\varepsilon}x)$, (x, y') is distributed according to $(x, T_{\nu, \mathcal{P}', 1-\varepsilon}x)$ and $\Pr[y \neq y'] \leq k(|\mathcal{P}| + |\mathcal{P}'|)\varepsilon$.*
2. *For any 1-bounded function $f: \Sigma^n \rightarrow \mathbb{C}$, $\|T_{\nu, \mathcal{P}', 1-\varepsilon}f - T_{\nu, \mathcal{P}, 1-\varepsilon}f\|_2 \leq 2\sqrt{k(|\mathcal{P}| + |\mathcal{P}'|)\varepsilon}$.*

Proof. For each $P \in \text{span}_{\mathbb{N}}(\mathcal{P})$, let $\tilde{P} \in \text{span}_{\mathbb{N}}(\mathcal{P}')$ be the minimizer of $\Delta(P, \tilde{P})$, and let $K(P) = \{i \mid P_i \neq \tilde{P}_i\}$, so that $|K(P)| \leq k$. Similarly, define $K'(P')$ for $P' \in \mathcal{P}'$, and set $K = \bigcup_{P \in \mathcal{P}} K(P) \cup \bigcup_{P' \in \mathcal{P}'} K'(P')$. Then we get that $|K| \leq (|\mathcal{P}| + |\mathcal{P}'|)k$. Note that if $I \subseteq [n] \setminus K$, then $T_{\nu, \mathcal{P}, I}$ and $T_{\nu, \mathcal{P}', I}$ are identical as they are both equal to $T_{\nu, \mathcal{P}|_I, I}$. We have $\Pr_{I \subseteq [n]} [I \subseteq [n] \setminus K] \geq 1 - |K|\varepsilon$, so it follows that the statistical distance between $(x, T_{\nu, \mathcal{P}', 1-\varepsilon}x)$ and $(x, T_{\nu, \mathcal{P}, 1-\varepsilon}x)$ is at most $|K|\varepsilon$, and the first item follows.

For the second item, fix a coupling (x, y, y') as in the first item. Then

$$\|T_{\nu, \mathcal{P}', 1-\varepsilon}f - T_{\nu, \mathcal{P}, 1-\varepsilon}f\|_2^2 = \mathbb{E}_x \left[|T_{\nu, \mathcal{P}, 1-\varepsilon}f(x) - T_{\nu, \mathcal{P}', 1-\varepsilon}f(x)|^2 \right] \leq \mathbb{E}_{x, y, y'} \left[|f(y) - f(y')|^2 \right],$$

which is at most $4\mathbb{E}_{x, y, y'} [1_{y \neq y'}]$ by the 1-boundedness of f , and the second item follows. $\square$

The following result is essentially [BKM26, Claim 4.5], except that we replaced the assumption that $\mathcal{P}$ arises from some group ($\mathcal{P} \subseteq \mathcal{P}(\Sigma, G, \sigma)$ in the notations therein) with the assumption that each function in $\mathcal{P}$ is discrete; this is in fact the property used in the proof therein. In words, the result says that for sufficiently small noise rate, the operator $T_{\nu, \mathcal{P}, 1-\varepsilon}$ acts like the identity on low-degree functions.

Lemma 2.24. *Let $d, r, m \in \mathbb{N}$ and $\alpha, \varepsilon > 0$. Suppose that Σ has size at most m , ν is a distribution over Σ in which the probability of each atom is at least α , and $\mathcal{P}$ is a collection of $O_m(1)$ -discrete product functions of size at most r . Then for any $L: \Sigma^n \rightarrow \mathbb{C}$ of degree at most d we have that*

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

The following result is essentially [BKM26, Lemma 4.9], where again we replace the assumption about $\mathcal{P}$ as before. The proof is the same as therein. In words, the result says that if f is 1-bounded, then $T_{\nu, \mathcal{P}, 1-\varepsilon}f$ can be approximated arbitrarily well in L_2 -distance by a function of the form $\sum_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} P(x) L_P(x)$, where each L_P is a low-degree function with bounded 2-norm.

Lemma 2.25. *For all $m, r \in \mathbb{N}$, $\alpha, \varepsilon > 0$ and $\xi > 0$ there exist $C, D \in \mathbb{N}$ such that the following holds. Suppose that Σ has size at most m , ν is a distribution over Σ in which the probability of each atom is at least α , and $\mathcal{P}$ is a collection of $O_m(1)$ -discrete product functions of size at most r . Then for any 1-bounded function $f: \Sigma^n \rightarrow \mathbb{C}$, there exists a function $f': \Sigma^n \rightarrow \mathbb{C}$ such that:*

1. *The function f' approximates $T_{\nu, \mathcal{P}, 1-\varepsilon} f$: $\|T_{\nu, \mathcal{P}, 1-\varepsilon} f - f'\|_2 \leq \xi$.*
2. *The function f' can be written as*

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

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

3 High and Low Rank Collections

In this section we give a few results concerning the notions of rank in Definitions 2.13 and 2.14. First in Sections 3.1, 3.2 and 3.3 we show processes that turn a given low-rank/low-sofrank collection $\mathcal{P}$ into a high-rank/high-sofrank collection. Then, in Section 3.4 we prove some quasirandomness properties of high-rank collections.

3.1 Achieving a High Rank

In this section we show how to process a given collection of product functions $\mathcal{P}$ into a high-rank collection $\mathcal{P}'$. The following claim captures the main sub-procedure in that process.

Claim 3.1. *Let $\mathcal{P} \subseteq \{P: \Sigma^n \rightarrow \mathbb{D}\}$ be a collection of M -discrete cyclic product functions. If $\text{rk}(\mathcal{P}) \leq D$, then there exists $P \in \mathcal{P}$ such that one of the following holds:*

1. *Setting $\mathcal{P}' = \mathcal{P} \setminus \{P\}$, we have $\Delta_{\text{symbolic}}(P, \text{span}_{\mathbb{N}}(\mathcal{P}')) \leq D$.*
2. *There exists an M -discrete cyclic product function $P': \Sigma^n \rightarrow \mathbb{C}$ such that $\text{ord}(P') < \text{ord}(P)$, $\Delta_{\text{symbolic}}(P', \text{span}_{\mathbb{N}}(\mathcal{P})) \leq D$, and setting $\mathcal{P}' = (\mathcal{P} \setminus \{P\}) \cup \{P'\}$, we have*

$$\Delta_{\text{symbolic}}(P, \text{span}_{\mathbb{N}}(\mathcal{P}')) \leq D.$$

Proof. By definition, since $\text{rk}(\mathcal{P}) \leq D$ we may find not-all-zero integers $\{\alpha_P\}_{P \in \mathcal{P}}$ such that $0 \leq \alpha_P \leq \text{ord}(P) - 1$ and the function $R(x) = \prod_{P \in \mathcal{P}} P(x)^{\alpha_P}$ satisfies that $\Delta_{\text{symbolic}}(R, 1) \leq D$. For each prime p , let $\mathcal{P}_p \subseteq \mathcal{P}$ be the subset of product functions whose order is a power of p , and fix p such that for some $P \in \mathcal{P}_p$ we have $\alpha_P \neq 0$. For each p' , let $\beta_{p'} = \max_{P' \in \mathcal{P}_{p'}} \text{ord}(P')$, and denote

$$s = \prod_{p' \neq p: \mathcal{P}_{p'} \neq \emptyset} \beta_{p'}.$$

Then by Fact 2.11 we have that

$$\Delta_{\text{symbolic}}(R^s, 1) \leq \Delta_{\text{symbolic}}(R, 1) \leq D.$$

For $p' \neq p$ and $P' \in \mathcal{P}_{p'}$, as the order of P' divides s , we get that $P'(x)^s \equiv 1$. For $P \in \mathcal{P}_p$, as s is coprime to p we may find integers i and j such that $is + j \text{ord}(P) = 1$, so $is\alpha_P \equiv \alpha_P \pmod{\text{ord}(P)}$, and therefore $\alpha'_P := s\alpha_P \not\equiv 0 \pmod{\text{ord}(P)}$ for some $P \in \mathcal{P}_p$. Thus,

$$R(x)^s = \prod_{P \in \mathcal{P}_p} P(x)^{\alpha'_P},$$

where $0 \leq \alpha'_P \leq \text{ord}(P) - 1$ for all $P \in \mathcal{P}_p$ and at least one of them is non-zero. There are two cases now.

Case 1: If some α'_P is coprime to p , then we fix such P and find β such that $\beta \cdot \alpha'_P \equiv 1 \pmod{\text{ord}(P)}$. Using Fact 2.11 we get that $\Delta(R^{s\beta}, 1) \leq D$, and a direct computation gives

$$R(x)^{s\beta} = P(x) \prod_{P' \in \mathcal{P} \setminus \{P\}} P'(x)^{\alpha'_{P'}\beta}.$$

Thus, $\Delta(P, \prod_{P' \in \mathcal{P} \setminus \{P\}} P'^{-\alpha'_{P'}\beta}) \leq D$, and the first item holds.

Case 2: Else, we may find an integer $k \geq 1$ and integers β_P , at least one of which, say β_{P^*} , is coprime to p , such that $\alpha'_P = p^k \beta_P$. Thus, we may write

$$R(x)^s = \left(\prod_{P \in \mathcal{P}_p} P(x)^{\beta_P} \right)^{p^k} = R'(x)^{p^k},$$

where we define $R'(x) = \prod_{P \in \mathcal{P}_p} P(x)^{\beta_P}$. Since $\Delta_{\text{symbolic}}(R, 1) \leq D$, we may find $J \subseteq [n]$ of size at most D , such that $R|_{\bar{J}} \equiv 1$. We get that for $P' = R'|_{\bar{J}}$, we have $(P')^{p^k} \equiv 1$. We now note that:

1. P' is a cyclic product function with $\text{ord}(P') \leq p^k \leq \alpha'_{P^*} < \text{ord}(P^*)$.
2. We have that

$$\Delta \left((P^*)^{-\beta_{P^*}}, \overline{P'} \prod_{P \in \mathcal{P}_p \setminus \{P^*\}} P^{\beta_P} \right) = \Delta \left(P', (P^*)^{\beta_{P^*}} \prod_{P \in \mathcal{P}_p \setminus \{P^*\}} P^{\beta_P} \right) \leq |J| \leq D.$$

As β_{P^*} is coprime to p , we may find s' such that $s' \cdot \beta_{P^*} \equiv 1 \pmod{\text{ord}(P^*)}$, and get from the above and Fact 2.11 that

$$\Delta \left(P^*, (P')^{s'} \prod_{P \in \mathcal{P}_p \setminus \{P^*\}} P^{-s' \cdot \beta_P} \right) = \Delta \left((P^*)^{s' \beta_{P^*}}, (P')^{s'} \prod_{P \in \mathcal{P}_p \setminus \{P^*\}} P^{-s' \beta_P} \right) \leq D.$$

Together, we get that the second item holds for $\mathcal{P}' = (\mathcal{P} \setminus \{P^*\}) \cup \{P'\}$. □

Iterating Claim 3.1, we get the following lemma, which will be used in subsequent sections.

Lemma 3.2. *Let $\mathcal{P}$ be a collection of M -discrete cyclic product functions, let $k = M|\mathcal{P}| + 1$ and let $D_1 \leq D_2 \leq \dots \leq D_k$ be integers. Then we may find a collection $\mathcal{P}'$ of M -discrete cyclic product functions with $|\mathcal{P}'| \leq |\mathcal{P}|$ such that for some $1 \leq i \leq k$, the following properties hold:*

1. $\text{rk}(\mathcal{P}') \geq D_i$.
2. For each $P \in \mathcal{P}$, we have that $\Delta_{\text{symbolic}}(P, \text{span}_{\mathbb{N}}(\mathcal{P}')) \leq D_1 + \dots + D_{i-1}$.
3. For each $P' \in \mathcal{P}'$, we have that $\Delta_{\text{symbolic}}(P', \text{span}_{\mathbb{N}}(\mathcal{P})) \leq D_1 + \dots + D_{i-1}$.

Proof. Starting with $\mathcal{P}'_0 = \mathcal{P}$ and $i = 1$, we iterate the following process:

- If $\text{rk}(\mathcal{P}'_{i-1}) \geq D_i$, halt.
- Else, apply Claim 3.1 on $\mathcal{P}'_{i-1}$ to find $\mathcal{P}'_i$ as therein. Thus, we get that for each $P \in \text{span}_{\mathbb{N}}(\mathcal{P}'_{i-1})$ and $P' \in \text{span}_{\mathbb{N}}(\mathcal{P}'_i)$ we have

$$\Delta_{\text{symbolic}}(P, \text{span}_{\mathbb{N}}(\mathcal{P}'_i)) \leq D_i, \quad \Delta_{\text{symbolic}}(P', \text{span}_{\mathbb{N}}(\mathcal{P}'_{i-1})) \leq D_i.$$

Increment i and iterate.

Note that as $\sum_{P \in \mathcal{P}'_i} \text{ord}(P) < \sum_{P \in \mathcal{P}'_{i-1}} \text{ord}(P)$, the process terminates at some $i \leq k$, and we take $\mathcal{P}' = \mathcal{P}'_{i-1}$.

The first item automatically holds. For the second item fix $P \in \text{span}_{\mathbb{N}}(\mathcal{P})$. Applying the second bullet of the process we get functions $P_j \in \text{span}_{\mathbb{N}}(\mathcal{P}'_j)$ such that

$$\Delta_{\text{symbolic}}(P, P'_1) \leq D_1, \quad \Delta_{\text{symbolic}}(P'_j, P'_{j+1}) \leq D_j \quad \forall j \leq i-2.$$

Combining via Fact 2.11 gives that $\Delta_{\text{symbolic}}(P, P'_{i-1}) \leq D_1 + \dots + D_{i-1}$, so the second item holds. The proof of the third item is identical. $\square$

3.2 Achieving a High Softrank

In this section we show how to process a given collection of product functions $\mathcal{P}$ into a high sofrank collection. This is captured by the following lemma.

Lemma 3.3. *Let Σ be a finite alphabet, let μ be a distribution over Σ^3 , let $\mathcal{P}$ be a collection of M -discrete cyclic product functions and set $k = M^{3|\mathcal{P}|}$. Then for all integers $D_1 \leq D_2 \leq \dots \leq D_k$ we may find $1 \leq i \leq k$ and a subset $J \subseteq [n]$ such that the following holds:*

1. $|J| \leq D_1 + \dots + D_{i-1}$.
2. $\text{soft}\text{rk}_{\mu}(\mathcal{P}|_J) \geq D_i$.

Proof. Start with $i = 1$ and $\mathcal{P}_1 = \mathcal{P}$. As long as $\text{soft}\text{rk}_{\mu}(\mathcal{P}_i) < D_i$, by definition we may find integers $\{\alpha_P, \beta_P, \gamma_P\}_{P \in \mathcal{P}}$ such that the function $R: \text{supp}(\mu) \rightarrow \mathbb{D}$ defined as

$$R(x, y, z) = \prod_{P \in \mathcal{P}} P(x)^{\alpha_P} P(y)^{\beta_P} P(z)^{\gamma_P}$$

satisfies that $R \not\equiv 1$ but $\Delta_{\text{symbolic}}(R, 1) \leq D_i$. We may find $J_i \subseteq [n]$ of size at most D_i such that $R|_{J_i} \equiv 1$, define $\mathcal{P}_{i+1} = \mathcal{P}|_{J_i}$, increment i by 1 and iterate.

We note that if for integers $\{\alpha'_P, \beta'_P, \gamma'_P\}_{P \in \mathcal{P}}$ we have that

$$\prod_{P \in \mathcal{P}} P(x)^{\alpha'_P} P(y)^{\beta'_P} P(z)^{\gamma'_P} \equiv 1,$$

in $\mathcal{P}_i$, then this remains true for the corresponding functions in $\mathcal{P}_{i+1}$. Thus, each step of the iterative process increases the number of $\{\alpha'_P, \beta'_P, \gamma'_P\}_{P \in \mathcal{P}}$ as above by at least 1, so the process terminates within k steps. Let i be the index we stop with, and set $J = J_1 \cup \dots \cup J_{i-1}$. Then $|J| \leq D_1 + \dots + D_{i-1}$ so the first item holds, and $\mathcal{P}_i = \mathcal{P}|_J$ so the second item holds. $\square$

3.3 Achieving High Rank and High Soft Rank

We now combine Lemmas 3.2 and 3.3 to get the following result:

Lemma 3.4. *Let Σ be a finite alphabet, let μ be a distribution over Σ^3 , let $\mathcal{P}$ be a collection of M -discrete cyclic product functions. Then there exists $k = k(|\mathcal{P}|, M)$ such that for all integers $D_1 \leq D_2 \leq \dots \leq D_k$ we may find $1 \leq i \leq k$ and a collection $\mathcal{P}'$ of M -discrete cyclic product functions with $|\mathcal{P}'| \leq |\mathcal{P}|$ such that the following holds:*

1. $\text{softrk}_\mu(\mathcal{P}')$ and $\text{rk}(\mathcal{P}')$ are both at least $D_i - (D_1 + \dots + D_{i-1})$.
2. For each $P \in \mathcal{P}$, we have that $\Delta_{\text{symbolic}}(P, \text{span}_{\mathbb{N}}(\mathcal{P}')) \leq 2(D_1 + \dots + D_{i-1})$.
3. For each $P' \in \mathcal{P}'$, we have that $\Delta_{\text{symbolic}}(P', \text{span}_{\mathbb{N}}(\mathcal{P})) \leq 2(D_1 + \dots + D_{i-1})$.

Proof. Take k_1 and k_2 from Lemmas 3.2 and 3.3 and set $k = k_1 \cdot k_2$. Consider the sequence $E_1 \leq E_2 \leq \dots \leq E_{k_1}$ given as $E_j = D_{jk_2}$. Then by Lemma 3.2 we may find a collection $\tilde{\mathcal{P}}$ and j as therein. In particular, $\text{rk}(\tilde{\mathcal{P}}) \geq E_j = D_{j \cdot k_2}$ and the second and third item there hold with the right hand side being $E_1 + \dots + E_{j-1} \leq D_1 + \dots + D_{(j-1) \cdot k_2}$. Consider the sequence $E'_1 \leq E'_2 \leq \dots \leq E'_{k_2}$ given by $E'_\ell = D_{(j-1)k_2 + \ell}$. Then applying Lemma 3.3 we may find an ℓ and a subset J of size at most $E'_1 + \dots + E'_{\ell-1}$ such that $\text{softrk}_\mu(\tilde{\mathcal{P}}|_{\bar{J}}) \geq E'_\ell$.

We take $\mathcal{P}' = \tilde{\mathcal{P}}|_{\bar{J}}$ and $i = (j-1)k_2 + \ell$. Then $\text{softrk}_\mu(\mathcal{P}') = \text{softrk}_\mu(\tilde{\mathcal{P}}|_{\bar{J}}) \geq E'_\ell = D_i$ and

$$\text{rk}(\mathcal{P}') \geq \text{rk}(\tilde{\mathcal{P}}) - |J| \geq D_{jk_2} - (D_1 + \dots + D_{i-1}) \geq D_i - (D_1 + \dots + D_{i-1}),$$

so the first item of the lemma holds. For the second item, for $P \in \mathcal{P}$ we have

$$\begin{aligned} \Delta_{\text{symbolic}}(P, \text{span}_{\mathbb{N}}(\mathcal{P}')) &\leq \Delta_{\text{symbolic}}(P, \text{span}_{\mathbb{N}}(\tilde{\mathcal{P}})) + |J| \leq E_1 + \dots + E_{j-1} + E'_1 + \dots + E'_{\ell-1} \\ &\leq 2(D_1 + \dots + D_{i-1}). \end{aligned}$$

The proof of the third item is identical. □

3.4 Quasirandomness of High-rank Collections

In this section, we prove a few lemmas that show the utility of having a collection of product functions $\mathcal{P}$ with high rank.

The first lemma shows that if $\mathcal{P}$ has a high rank, then the distribution of $\{P(x)\}_{P \in \mathcal{P}}$ where $x \sim \nu^{\otimes n}$ is close to the uniform distribution on $\prod_{P \in \mathcal{P}} \text{Image}(P)$.

Lemma 3.5. *Let Σ be an alphabet of size m and let ν be a distribution over Σ in which the probability of each atom is at least α . Suppose that $\mathcal{P}$ is a collection of M -discrete, n -variate product functions with $|\mathcal{P}| = r$, and suppose that $\text{rk}(\mathcal{P}) \geq D$. Then for any $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$ we have that*

$$\left| \Pr_{x \sim \nu^{\otimes n}} [\forall P \in \mathcal{P}, P(x) = b_P] - \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)} \right| \leq 2^{-\Omega_{M,r,\alpha}(D)}.$$

Proof. Denote $\mathcal{P} = \{P_1, P_2, \dots, P_r\}$. Then we may expand:

$$\begin{aligned} \Pr_{x \sim \nu^{\otimes n}} [\forall P \in \mathcal{P}, P(x) = b_P] &= \mathbb{E}_{x \sim \nu^{\otimes n}} \left[\prod_{P \in \mathcal{P}} \left(\frac{1}{\text{ord}(P)} \sum_{j=0}^{\text{ord}(P)-1} (\overline{b_P} P(x))^j \right) \right] \\ &= \sum_{\vec{j}=(j_1, \dots, j_r)} \left(\prod_{i=1}^r \frac{\overline{b_P}^{j_i}}{\text{ord}(P_i)} \right) \mathbb{E}_{x \sim \nu^{\otimes n}} \left[\prod_{i=1}^r P_i(x)^{j_i} \right] \end{aligned}$$

Denote $S_{\vec{j}}(x) := \prod_{i=1}^r P_i(x)^{j_i}$. We now analyze the expectation for the following two different scenarios.

- (a) $\vec{j} = \vec{0}$: In this case, we have that the expectation is 1.
- (b) $\vec{j} \neq \vec{0}$: In this case, $\Delta_{\text{symbolic}}(S_{\vec{j}}, 1) \geq D$ since $\text{rk}(\tilde{\mathcal{P}}) \geq D$. Using Fact 2.12 (1), we have

$$\left| \mathbb{E}_{x \sim \nu^{\otimes n}} \left[\prod_{i=1}^r P_i(x)^{j_i} \right] \right| \leq 2^{-\Omega_{\alpha, M}(D)}.$$

Using the above, summing over all $\vec{j}$, we get that

$$\left| \Pr_{x \sim \nu^{\otimes n}} [\forall P \in \mathcal{P}, P(x) = b_P] - \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)} \right| \leq 2^{-\Omega_{M, r, \alpha}(D)}. \quad \square$$

The next lemma shows that if the collection of product functions $\mathcal{P}$ has high rank and softrank with respect to a distribution μ over Σ^3 , then for any $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$, the probability that the product functions evaluate to $\vec{b}$ on each one of x, y, z , where $(x, y, z) \sim \mu^{\otimes n}$, is non-negligible.

Lemma 3.6. *Let Σ be an alphabet of size m and let μ be a distribution over Σ^3 such that $(x, x, x) \in \text{supp}(\mu)$ for all $x \in \Sigma$, and in which the probability of each atom is at least α . Suppose that $\mathcal{P}$ is a collection of M -discrete, n -variate product functions, and suppose that $\text{rk}(\mathcal{P}), \text{softrk}_{\mu}(\mathcal{P}) \geq D$. For any $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$ we have that*

$$\Pr_{(x, y, z) \sim \mu^{\otimes n}} [P(x) = P(y) = P(z) = b_P] \geq \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)^3} - 2^{-\Omega_{M, m, \alpha}(D)}.$$

We prove a more general version of the above lemma in which the conclusion holds even after conditioning on a small subset of coordinates taking values from the support of μ .

Lemma 3.7. *Let Σ be an alphabet of size m and let μ be a distribution over Σ^3 such that $(x, x, x) \in \text{supp}(\mu)$ for all $x \in \Sigma$, and in which the probability of each atom is at least α . Suppose that $\mathcal{P}$ is a collection of M -discrete, n -variate product functions with $|\mathcal{P}| = r$, and suppose that $\text{rk}(\mathcal{P}), \text{softrk}_{\mu}(\mathcal{P}) \geq D$. Suppose $J \subseteq [n]$ of size $|J| \leq \frac{D}{2}$. Then for every $(w, w', w'') \in \text{supp}(\mu|_J)$ and for any $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$ we have that*

$$\Pr_{(x, y, z) \sim \mu^{\vec{J}}} [\forall P \in \mathcal{P}, P(w, x) = P(w', y) = P(w'', z) = b_P] \geq \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)^3} - 2^{-\Omega_{M, r, \alpha}(D)}.$$

Proof. Fix any $(w, w', w'') \in \text{supp}(\mu|_J)$ and $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$. We begin by expressing the probability as follows. For the sake of ease of notation, in what follows, we let $o_P := \text{ord}(P)$ and $\mathcal{P} = \{P_1, P_2, \dots, P_r\}$.

$$\begin{aligned}
& \Pr_{(x,y,z) \sim \mu^{\vec{J}}} [\forall P \in \mathcal{P}, P(w, x) = P(w', y) = P(w'', z) = b_P] \\
&= \mathbb{E}_{(x,y,z) \sim \mu^{\vec{J}}} \left[\prod_{P \in \mathcal{P}} \left(\frac{1}{o_P} \sum_{j=0}^{o_P-1} (\overline{b_P} P(w, x))^j \right) \cdot \left(\frac{1}{o_P} \sum_{j=0}^{o_P-1} (\overline{b_P} P(w', y))^j \right) \cdot \left(\frac{1}{o_P} \sum_{j=0}^{o_P-1} (\overline{b_P} P(w'', z))^j \right) \right] \\
&= \mathbb{E}_{(x,y,z) \sim \mu^{\vec{J}}} \left[\prod_{P \in \mathcal{P}} \frac{1}{o_P^3} \left(\sum_{j,k,\ell=0}^{o_P-1} \left(\overline{b_P}^{(j+k+\ell)} P(w, x)^j P(w', y)^k P(w'', z)^\ell \right) \right) \right] \\
&= \sum_{\substack{\vec{j}=(j_1, \dots, j_r) \\ \vec{k}=(k_1, \dots, k_r) \\ \vec{\ell}=(\ell_1, \dots, \ell_r)}} \mathbb{E}_{(x,y,z) \sim \mu^{\vec{J}}} \left[\left(\prod_{i=1}^r \frac{1}{o_{P_i}^3} \left(\overline{b_{P_i}}^{(j_i+k_i+\ell_i)} P_i(w, x)^{j_i} P_i(w', y)^{k_i} P_i(w'', z)^{\ell_i} \right) \right) \right]
\end{aligned}$$

We let $\alpha_i = \prod_{i' \in J} P_{i,i'}(w_{i'})$ be the evaluation of P_i at w restricted to the coordinates in J . Similarly, let $\beta_i = \prod_{i' \in J} P_{i,i'}(w'_{i'})$ and $\gamma_i = \prod_{i' \in J} P_{i,i'}(w''_{i'})$. Also, let $\tilde{P} = P|_J$ and $\tilde{\mathcal{P}} = \{\tilde{P}_1, \tilde{P}_2, \dots, \tilde{P}_r\}$. With these notations, we can rewrite the above expression as:

$$\begin{aligned}
& \Pr_{(x,y,z) \sim \mu^{\vec{J}}} [\forall P \in \mathcal{P}, P(w, x) = P(w', y) = P(w'', z) = b_P] \\
&= \sum_{\substack{\vec{j}=(j_1, \dots, j_r) \\ \vec{k}=(k_1, \dots, k_r) \\ \vec{\ell}=(\ell_1, \dots, \ell_r)}} \underbrace{\left(\prod_{i=1}^r \frac{1}{o_{P_i}^3} \right) \left(\prod_{i=1}^r \left(\overline{b_{P_i}}^{(j_i+k_i+\ell_i)} \right) \alpha_i^{j_i} \beta_i^{k_i} \gamma_i^{\ell_i} \right)}_{=: \Theta(\vec{j}, \vec{k}, \vec{\ell})} \mathbb{E}_{(x,y,z) \sim \mu^{\vec{J}}} \left[\prod_{i=1}^r \tilde{P}_i(x)^{j_i} \tilde{P}_i(y)^{k_i} \tilde{P}_i(z)^{\ell_i} \right]
\end{aligned}$$

Let us denote the product function within the expectation by $S_{(\vec{j}, \vec{k}, \vec{\ell})}$, i.e.,

$$S_{(\vec{j}, \vec{k}, \vec{\ell})}(x, y, z) := \prod_{i=1}^r \tilde{P}_i(x)^{j_i} \tilde{P}_i(y)^{k_i} \tilde{P}_i(z)^{\ell_i}.$$

As $\text{rk}(\mathcal{P}), \text{softrk}_\mu(\mathcal{P}) \geq D$ and $|J| \leq D/2$, we have $\text{rk}(\tilde{\mathcal{P}}), \text{softrk}_\mu(\tilde{\mathcal{P}}) \geq D/2$. We now analyze the quantity $\Theta(\vec{j}, \vec{k}, \vec{\ell})$ for the following three different scenarios.

- (a) $\vec{j} = \vec{k} = \vec{\ell} = \vec{0}$: In this case, we have $\Theta(\vec{j}, \vec{k}, \vec{\ell}) = 1$.
- (b) $\Delta_{\text{symbolic}}(S_{(\vec{j}, \vec{k}, \vec{\ell})}, 1) \geq D/2$: In this case, using Fact 2.12 (1), we have $\left| \Theta(\vec{j}, \vec{k}, \vec{\ell}) \right| \leq 2^{-\Omega_{\alpha, M}(D)}$.
- (c) $S_{(\vec{j}, \vec{k}, \vec{\ell})}$ is a constant function: As $\text{softrk}_\mu(\tilde{\mathcal{P}}) \geq D/2$, this is the only remaining case. We now use the fact that $\text{rk}(\tilde{\mathcal{P}})$ is also large. Since (x, x, x) is in the support of μ and $S_{(\vec{j}, \vec{k}, \vec{\ell})}$ is a constant function, $S_{(\vec{j}, \vec{k}, \vec{\ell})}(x, x, x)$ is also a constant function. As $\text{rk}(\tilde{\mathcal{P}}) \geq D/2$, this can only happen if $j_i + k_i + \ell_i = 0$

mod o_{P_i} for all $1 \leq i \leq r$. Thus, $S_{(\vec{j}, \vec{k}, \vec{\ell})}(x, x, x)$ is the constant 1 function and hence $S_{(\vec{j}, \vec{k}, \vec{\ell})}$ is also the constant 1 function. Furthermore, as $b_{P_i} \in \text{Image}(P_i)$, we get $\overline{b_{P_i}}^{(j_i + k_i + \ell_i)} \equiv 1$ for all $1 \leq i \leq r$. Thus,

$$\Theta(\vec{j}, \vec{k}, \vec{\ell}) = \left(\prod_{i=1}^r \alpha_i^{j_i} \beta_i^{k_i} \gamma_i^{\ell_i} \right).$$

As we are in the case that $\Delta_{\text{symbolic}}(S_{(\vec{j}, \vec{k}, \vec{\ell})}, 1) < D/2$, it must be the case that the following function is a constant function for every $(w, w', w'') \in \text{supp}(\mu|_J)$ and $(x, y, z) \in \text{supp}(\mu^{\vec{J}})$.

$$\begin{aligned} \prod_{i=1}^r \left(P_i(w, x)^{j_i} P_i(w', y)^{k_i} P_i(w'', z)^{\ell_i} \right) &= \prod_{i=1}^r \alpha_i^{j_i} \beta_i^{k_i} \gamma_i^{\ell_i} \prod_{i=1}^r \tilde{P}_i(x)^{j_i} \tilde{P}_i(y)^{k_i} \tilde{P}_i(z)^{\ell_i} \\ &= \prod_{i=1}^r \alpha_i^{j_i} \beta_i^{k_i} \gamma_i^{\ell_i}. \end{aligned}$$

This is because in this case, the function on the left hand side will have symbolic distance from 1 strictly less than D and we know that $\text{softrk}_{\mu}(\mathcal{P}) \geq D$. Note that in the last step above, we used the fact that $S_{(\vec{j}, \vec{k}, \vec{\ell})}$ is the constant 1 function. As $(x, x, x) \in \text{supp}(\mu)$ for all $x \in \Sigma$, we must that that above function, when evaluated at (w, w, w) for $w \in \Sigma^J$ is equal to

$$\prod_{i=1}^r \alpha_i^{j_i} \alpha_i^{k_i} \alpha_i^{\ell_i} = \prod_{i=1}^r \alpha_i^{j_i + k_i + \ell_i} = 1,$$

where $\alpha_i = \prod_{i' \in J} P_{i, i'}(w_{i'})$ and again we used the fact that $\alpha_i \in \text{Image}(P_i)$ and $j_i + k_i + \ell_i = 0 \pmod{o_{P_i}}$. Thus, we have

$$\Theta(\vec{j}, \vec{k}, \vec{\ell}) = 1$$

If we denote the number of tuples $(\vec{j}, \vec{k}, \vec{\ell})$ that fall into cases (a) and (c) above by C_0 , then we have $C_0 \geq 1$.

As the total number of tuples is upper bounded by $\left(\prod_{P \in \mathcal{P}} \text{ord}(P) \right)^{3r} = O_{M,r}(1)$, we get

$$\Pr_{(x,y,z) \sim \mu^{\vec{J}}} [\forall P \in \mathcal{P}, P(w, x) = P(w', y) = P(w'', z) = b_P] \geq C_0 \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)^3} - 2^{-\Omega_{M,r,\alpha}(D)},$$

as required. $\square$

Consider the noise operator $T_{1-\kappa, \mu}^{\otimes n}$ on $(\Sigma^n)^3$ as per Definition 2.6: on a tuple (x, y, z) it outputs (x', y', z') where for each coordinate $i \in [n]$ we have $(x'_i, y'_i, z'_i) = (x_i, y_i, z_i)$ with probability $1 - \kappa$ and we take $(x'_i, y'_i, z'_i) \sim \mu$ with probability κ independently. As usual, when clear from context we omit the n and μ from notation. The following lemma shows that for a typical tuple (x, y, z) the evaluations of the product function on $(x', y', z') \sim T_{1-\kappa, \mu}^{\otimes n}(x, y, z)$ is essentially independent of (x, y, z) if $\text{softrk}_{\mu}(\mathcal{P})$ is large.

Lemma 3.8. *Let Σ be an alphabet of size m and let μ be a distribution over Σ^3 such that $(x, x, x) \in \text{supp}(\mu)$ for all $x \in \Sigma$, and in which the probability of each atom is at least α . Suppose that $\mathcal{P}$ is a collection of M -discrete, n -variate product functions with $|\mathcal{P}| = r$, and suppose that $\text{softrk}_{\mu}(\mathcal{P}) \geq D$.*

Suppose $J \subseteq [n]$ has size $|J| \leq \frac{D}{2}$. Then for every $(w, w', w'') \in \text{supp}(\mu|_J)$, $\vec{a}, \vec{b}, \vec{c} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$ with

$$\begin{aligned} S_{\vec{a}} &= \{x \in \Sigma^{\bar{J}} \mid P(w, x) = a_P \ \forall P \in \mathcal{P}\}, \\ S_{\vec{b}} &= \{y \in \Sigma^{\bar{J}} \mid P(w', y) = b_P \ \forall P \in \mathcal{P}\}, \\ S_{\vec{c}} &= \{z \in \Sigma^{\bar{J}} \mid P(w'', z) = c_P \ \forall P \in \mathcal{P}\}. \end{aligned}$$

and

$$S = \{(x, y, z) \in (\Sigma^{\bar{J}})^3 \mid x \in S_{\vec{a}}, y \in S_{\vec{b}}, z \in S_{\vec{c}}\},$$

we have that

$$\|\mathbf{T}_{1-\kappa} \mathbf{1}_S - \mu(S)\|_{L_2(\mu)} \leq 2^{-\Omega_{\alpha, \kappa, r, M}(D)}.$$

Proof. The proof is mostly identical to the proof of Lemma 3.7. Let $o_P := \text{ord}(P)$ and $\mathcal{P} = \{P_1, P_2, \dots, P_r\}$. We start by writing $\mathbf{1}_S$ as follows.

$$\begin{aligned} \mathbf{1}_S(x, y, z) &= \prod_{P \in \mathcal{P}} \left(\frac{1}{o_P} \sum_{j=0}^{o_P-1} (\overline{a_P} P(w, x))^j \right) \cdot \left(\frac{1}{o_P} \sum_{j=0}^{o_P-1} (\overline{b_P} P(w', y))^j \right) \cdot \left(\frac{1}{o_P} \sum_{j=0}^{o_P-1} (\overline{c_P} P(w'', z))^j \right) \\ &= \prod_{P \in \mathcal{P}} \frac{1}{o_P^3} \left(\sum_{j,k,\ell=0}^{o_P-1} \overline{a_P}^j \overline{b_P}^k \overline{c_P}^\ell P(w, x)^j P(w', y)^k P(w'', z)^\ell \right) \\ &= \sum_{\substack{\vec{j}=(j_1, \dots, j_r) \\ \vec{k}=(k_1, \dots, k_r) \\ \vec{\ell}=(\ell_1, \dots, \ell_r)}} \left(\prod_{i=1}^r \frac{1}{o_{P_i}^3} \overline{a_{P_i}}^{j_i} \overline{b_{P_i}}^{k_i} \overline{c_{P_i}}^{\ell_i} P_i(w, x)^{j_i} P_i(w', y)^{k_i} P_i(w'', z)^{\ell_i} \right) \end{aligned}$$

Similar to the above proof, we let $\alpha_i = \prod_{i' \in J} P_{i,i'}(w_{i'})$ be the evaluation of P_i at w restricted to the coordinates in J . Similarly, let $\beta_i = \prod_{i' \in J} P_{i,i'}(w'_{i'})$ and $\gamma_i = \prod_{i' \in J} P_{i,i'}(w''_{i'})$. Also, let $\tilde{P} = P|_{\bar{J}}$ and $\tilde{\mathcal{P}} = \{\tilde{P}_1, \tilde{P}_2, \dots, \tilde{P}_r\}$. With these notations, we can rewrite the above expression as:

$$\begin{aligned} \mathbf{1}_S(x, y, z) &= \sum_{\substack{\vec{j}=(j_1, \dots, j_r) \\ \vec{k}=(k_1, \dots, k_r) \\ \vec{\ell}=(\ell_1, \dots, \ell_r)}} \left(\prod_{i=1}^r \frac{1}{o_{P_i}^3} \overline{a_{P_i}}^{j_i} \overline{b_{P_i}}^{k_i} \overline{c_{P_i}}^{\ell_i} P_i(w, x)^{j_i} P_i(w', y)^{k_i} P_i(w'', z)^{\ell_i} \right) \\ &= \sum_{\substack{\vec{j}=(j_1, \dots, j_r) \\ \vec{k}=(k_1, \dots, k_r) \\ \vec{\ell}=(\ell_1, \dots, \ell_r)}} \left(\prod_{i=1}^r \frac{1}{o_{P_i}^3} \overline{a_{P_i}}^{j_i} \overline{b_{P_i}}^{k_i} \overline{c_{P_i}}^{\ell_i} \alpha_i^{j_i} \beta_i^{k_i} \gamma_i^{\ell_i} \tilde{P}_i(x)^{j_i} \tilde{P}_i(y)^{k_i} \tilde{P}_i(z)^{\ell_i} \right). \end{aligned}$$

Let us denote the product function within the summand by $S_{(\vec{j}, \vec{k}, \vec{\ell})}$, i.e.,

$$S_{(\vec{j}, \vec{k}, \vec{\ell})}(x, y, z) := \prod_{i=1}^r \tilde{P}_i(x)^{j_i} \tilde{P}_i(y)^{k_i} \tilde{P}_i(z)^{\ell_i}.$$

We have

$$1_S = \sum_{\substack{\vec{j}=(j_1,\dots,j_r) \\ \vec{k}=(k_1,\dots,k_r) \\ \vec{\ell}=(\ell_1,\dots,\ell_r)}} \left(\prod_{i=1}^r \frac{1}{o_{P_i}^3} \overline{a_{P_i}}^{j_i} \overline{b_{P_i}}^{k_i} \overline{c_{P_i}}^{\ell_i} \alpha_i^{j_i} \beta_i^{k_i} \gamma_i^{\ell_i} \right) S_{(\vec{j}, \vec{k}, \vec{\ell})}$$

As $\text{softrk}_\mu(\mathcal{P}) \geq D$ and $|J| \leq D/2$, we have $\text{softrk}_\mu(\tilde{\mathcal{P}}) \geq D/2$. Let η be the sum of all the terms in the summand when the function $S_{(\vec{j}, \vec{k}, \vec{\ell})}$ is constant on the support of $\mu^{\otimes n'}$. We claim that $\mu(S)$ is close to η . To see this, as $\text{softrk}_\mu(\mathcal{P}) \geq D/2$, the only other terms in the summation are those with $\Delta_{\text{symbolic}}(S_{(\vec{j}, \vec{k}, \vec{\ell})}, 1) \geq D/2$. Thus,

$$\begin{aligned} & \left| \mathbb{E}_{(x,y,z) \sim \mu^{\vec{J}}} [1_S(x, y, z)] - \eta \right| \\ & \leq \sum_{\substack{\vec{j}, \vec{k}, \vec{\ell} \\ \Delta_{\text{symbolic}}(S_{(\vec{j}, \vec{k}, \vec{\ell})}, 1) \geq D/2}} \left(\prod_{i=1}^r \frac{1}{o_{P_i}^3} \overline{a_{P_i}}^{j_i} \overline{b_{P_i}}^{k_i} \overline{c_{P_i}}^{\ell_i} \alpha_i^{j_i} \beta_i^{k_i} \gamma_i^{\ell_i} \right) \left| \mathbb{E}_{(x,y,z) \sim \mu^{\vec{J}}} [S_{(\vec{j}, \vec{k}, \vec{\ell})}(x, y, z)] \right| \end{aligned}$$

Using Fact 2.12 (1), for $S_{(\vec{j}, \vec{k}, \vec{\ell})}$ with $\Delta_{\text{symbolic}}(S_{(\vec{j}, \vec{k}, \vec{\ell})}, 1) \geq D/2$, we have

$$\left| \mathbb{E}_{(x,y,z) \sim \mu^{\vec{J}}} [S_{(\vec{j}, \vec{k}, \vec{\ell})}(x, y, z)] \right| \leq 2^{-\Omega_{\alpha, M}(D)}.$$

Thus,

$$|\mu(S) - \eta| \leq 2^{-\Omega_{\alpha, r, M}(D)}.$$

Applying the noise operator $T_{1-\kappa}$ to the function 1_S , we have

$$T_{1-\kappa} 1_S - \eta = \sum_{\substack{\vec{j}, \vec{k}, \vec{\ell} \\ \Delta_{\text{symbolic}}(S_{(\vec{j}, \vec{k}, \vec{\ell})}, 1) \geq D/2}} \left(\prod_{i=1}^r \frac{1}{o_{P_i}^3} \overline{a_{P_i}}^{j_i} \overline{b_{P_i}}^{k_i} \overline{c_{P_i}}^{\ell_i} \alpha_i^{j_i} \beta_i^{k_i} \gamma_i^{\ell_i} \right) T_{1-\kappa} S_{(\vec{j}, \vec{k}, \vec{\ell})}.$$

Observing the fact that the constant in front of $T_{1-\kappa} S_{(\vec{j}, \vec{k}, \vec{\ell})}$ in the above summation is at most 1 in absolute value, by the triangle inequality,

$$\|T_{1-\kappa} 1_S - \eta\|_{L^2(\mu)} \leq \sum_{\substack{\vec{j}, \vec{k}, \vec{\ell} \\ \Delta_{\text{symbolic}}(S_{(\vec{j}, \vec{k}, \vec{\ell})}, 1) \geq D/2}} \|T_{1-\kappa} S_{(\vec{j}, \vec{k}, \vec{\ell})}\|_{L^2(\mu)}.$$

Using Fact 2.12 (2), we have $\|T_{1-\kappa} S_{(\vec{j}, \vec{k}, \vec{\ell})}\|_{L^2(\mu)} \leq 2^{-\Omega_{\alpha, \kappa, M}(D)}$. Using this and the above established fact that $|\mu(S) - \eta| \leq 2^{-\Omega_{\alpha, r, M}(D)}$, we have

$$\|T_{1-\kappa} 1_S - \mu(S)\|_{L^2(\mu)} \leq 2^{-\Omega_{\alpha, \kappa, rM}(D)},$$

as required. □

4 The Arithmetic Regularity Lemma

In this section we state and prove our arithmetic regularity lemmas.

4.1 The Basic Version

Below is the basic version of our arithmetic regularity lemma. At first reading, the lemma should be interpreted as saying that f can be approximated by $T_{\mathcal{P},1-\varepsilon}f$ in the $\|\cdot\|_\nu$ semi norm, for an appropriate choice of $\mathcal{P}$ and ε . At this resolution, the result is identical to [BKM26, Lemma 5.3], which is insufficient for our purposes. The key additional feature in the statement below is that f is η -close to a different function $T_{\mathcal{P}',1-\varepsilon'}f$ in the $\|\cdot\|_\nu$ semi-norm, where:

- η could be taken to be arbitrarily small compared to $|\mathcal{P}|$ and ε of the original crude estimator $T_{\mathcal{P},\varepsilon}f$;
- the crude estimator $T_{\mathcal{P},\varepsilon}f$ is close to the strong estimator $T_{\mathcal{P}',\varepsilon'}f$ in ℓ_2 distance.

For technical reasons we also allow flexibility in the choice of the noise parameter, but the reader should ignore this and think of ε' as being fixed to ε_{i+2} below.

Lemma 4.1. *Let Σ be a finite alphabet of size at most m , let ν be a probability distribution over Σ in which the probability of each atom is at least $\alpha > 0$, and let $w: (0, 1) \times \mathbb{N} \rightarrow (0, 1)$ be a sufficiently rapid decay function. Then for all $\xi > 0$, there exists $k \in \mathbb{N}$, such that taking the sequence defined inductively as $\varepsilon_0 = 1/2$ and $\varepsilon_{i+1} = w(\varepsilon_i, i)$, the following holds. For any 1-bounded function $f: \Sigma^n \rightarrow \mathbb{C}$, for any $i_0 \in \mathbb{N}$, there exists $i_0 \leq i \leq k + i_0$ and collections of cyclic product functions $\mathcal{P}, \mathcal{P}'$ such that:*

1. *The collection $\mathcal{P}$ has size at most $1/\varepsilon_i$, is $O_m(1)$ -discrete, and for all $\varepsilon' \in [\varepsilon_{i+2}, \varepsilon_{i+1})$ we have*

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

2. *The collection $\mathcal{P}'$ has size at most $1/\varepsilon_{i+4}$, is $O_m(1)$ -discrete, and we have*

$$\|f - T_{\mathcal{P}',1-\varepsilon_{i+5}}f\|_\nu \leq \varepsilon_{i+3}.$$

3. *For all $\varepsilon' \in [\varepsilon_{i+2}, \varepsilon_{i+1})$ we have that*

$$\|T_{\mathcal{P}',1-\varepsilon_{i+5}}f - T_{\mathcal{P},1-\varepsilon'}f\|_2 \leq \xi.$$

Proof. Fix ξ and α and take d and $\delta > 0$ from Theorem 2.20 for ξ and α . The proof is by two similar iterative processes, and we now describe them. Let $\eta_0 = 1$, $\mathcal{P}_0 = \emptyset$, $g_0 = T_{\mathcal{P}_0,1-\eta_0}f = \mathbb{E}[f]$, $f_0 = f - g_0$ and set $i = 0$ and $u = i_0$. If there is $\varepsilon' \in (\varepsilon_{u+2}, \varepsilon_{u+1})$ such that $f - T_{\mathcal{P}_i,1-\varepsilon'}f$ is δ -correlated with a function of the form $L_i \cdot P_i$ where $L_i: \Sigma^n \rightarrow \mathbb{C}$ has degree at most d and 2-norm 1 and P_i is an $O_m(1)$ -discrete product function, do the following:

1. Note that $\text{ord}(P_i) \leq O_m(1)$, thus by Fact 2.7 we may write it as $Q_1 \cdots Q_T$ where each Q_i is a cyclic product function with order at most $O_m(1)$ and $T \leq O_m(1)$. Define $\mathcal{P}_{i+1} = \mathcal{P}_i \cup \{Q_1, \dots, Q_T\}$, $\eta_{i+1} = \varepsilon'$, $g_{i+1} = T_{\mathcal{P}_{i+1},1-\eta_{i+1}}f$ and $f_{i+1} = f - g_{i+1}$.
2. Increase i by 1, increase u by 2 and repeat.

We claim that this process terminates within $O(1/\delta^2)$ steps. This is a direct consequence of the following claim.

Claim 4.2. *If the process has not terminated by step R , then $\|g_{R+1}\|_2^2 \geq R\delta_{10}^2$.*

Proof. The proof is identical to the proof of [BKM26, Claim 5.2], and we give it for completeness. For each $i \leq R$ write $T_i = T_{\mathcal{P}_i, 1-\eta_i}$ for simplicity, and choose complex numbers θ_i of absolute value 1 such that $\langle f_i, \theta_i L_i P_i \rangle$ is real-valued and is at least δ . Recalling that $f_i = (I - T_i)f$ and noting that $(I - T_i)$ is self adjoint, we conclude that

$$\langle f, \theta_i (I - T_i) L_i P_i \rangle \geq \delta.$$

We now study the quantity $(I) = \sum_{i=1}^R \langle T_{R+1} f, \theta_i (I - T_i) L_i P_i \rangle$, and prove an upper bound and a lower bound on its absolute value.

The lower bound: since T_{R+1} is self adjoint, we can write

$$(I) = \langle f, \sum_{i=1}^R \theta_i T_{R+1} (I - T_i) L_i P_i \rangle = \langle f, \sum_{i=1}^R \theta_i (T_{R+1} - T_{R+1} T_i) L_i P_i \rangle.$$

Fixing $1 \leq i \leq R$, we analyze the inner product on the right hand side. First, we have the pointwise equality $T_{R+1} L_i P_i(x) = P_i(x) T_{R+1} L_i$ (as the function P_i is constant over $y \sim T_{R+1} x$). Next, by Lemma 2.24 we have that $\|(I - T_{R+1}) L_i\| \lesssim_{d,m,|\mathcal{P}_{R+1}|,\alpha} \eta_{R+1}^{1/3}$. Altogether, it follows that $\|T_{R+1} L_i P_i - L_i P_i\|_2 \leq \eta_{R+1}^{1/6}$. We now claim in a similar fashion that $\|T_{R+1} T_i L_i P_i - T_i L_i P_i\|_2 \leq \eta_R^{1/6}$. First, note that by Lemma 2.25 we may find a function h_i of the form $\sum_{P' \in \text{span}_{\mathbb{N}}(\mathcal{P}_i)} L_{P'} P'$ such that $\|T_i L_i P_i - h_i\| \leq \eta_R$, and each $L_{P'}$ has degree and 2-norm at most $O_{\eta_R}(1)$. Noting that by Lemma 2.24 we have $\|T_{R+1} L_{P'} P' - L_{P'} P'\|_2 \lesssim_{|\mathcal{P}_{R+1}|,\eta_R} \eta_{R+1}^{1/3}$, we conclude that

$$\|T_{R+1} T_i L_i P_i - T_i L_i P_i\|_2 \leq \|T_{R+1} h_i - h_i\| + 2\|L_i P_i - h_i\|_2 \leq \sum_{P'} \|T_{R+1} L_{P'} P' - L_{P'} P'\|_2 + 2\eta_R,$$

which is at most $O_{|\mathcal{P}_{R+1}|,\eta_R}(\eta_{R+1}^{1/3}) + 2\eta_R \leq \eta_R^{1/6}$. We conclude that

$$\begin{aligned} |(I)| &\geq \langle f, \sum_{i=1}^R \theta_i (I - T_i) L_i P_i \rangle - \sum_{i=1}^R \|(T_{R+1} - T_{R+1} T_i) L_i P_i - (I - T_i) L_i P_i\|_2 \\ &\geq R\delta - \sum_{i=1}^R \eta_{R+1}^{1/6} + \eta_R^{1/6} \\ &\geq 0.99R\delta. \end{aligned}$$

The upper bound: by Cauchy-Schwarz

$$|(I)| \leq \|T_{R+1} f\|_2 \left\| \sum_{i=1}^R \theta_i (I - T_i) L_i P_i \right\|_2.$$

Looking at the second term, taking square and expanding, we get that it is at most

$$\sum_{i=1}^R \|(I - T_i)L_i P_i\|_2^2 + 2 \sum_{i' < i} |\langle (I - T_i)L_i P_i, (I - T_{i'})L_{i'} P_{i'} \rangle|. \quad (4)$$

The first sum is at most $4R$ as $\|(I - T_i)L_i P_i\|_2 \leq 2$ for each i . For the second term, fix $i' < i$. Then by self adjointness of $(I - T_i)$ and Cauchy-Schwarz

$$|\langle (I - T_i)L_i P_i, (I - T_{i'})L_{i'} P_{i'} \rangle| \leq \|L_i P_i\|_2 \|(I - T_i)(I - T_{i'})L_{i'} P_{i'}\|_2 \leq \|(I - T_i)(I - T_{i'})L_{i'} P_{i'}\|_2.$$

An argument identical to the one in the “upper bound” shows that $\|T_i(I - T_{i'})L_{i'} P_{i'} - (I - T_{i'})L_{i'} P_{i'}\|_2 \leq \eta_{i-1}^{1/6}$. Plugging this into (4) and then back up gives that

$$|(I)| \leq \|T_{R+1}f\|_2 \sqrt{4R + \sum_i \eta_{i-1}^{1/6}} \leq \|T_{R+1}f\|_2 \sqrt{5R}.$$

Combining the bounds: putting the upper bound and the lower bound on $|(I)|$ together gives that $0.99R\delta \leq |(I)| \leq \|T_{R+1}f\|_2 \sqrt{5R}$, and simplifying yields the statement of the claim. $\square$

As $\|g_i\|_2 \leq \|f\|_2 \leq 1$ always, the process terminates after $i^* \leq O(1/\delta^2)$ steps and let u be the index of ε in the end. At the end of the process we have found $\mathcal{P}_{i^*}$ which satisfies the first item (with $\varepsilon_{u+2}, \varepsilon_{u+1}$). It is unlikely to be a stronger approximation as required in the second item, though. To fix that, the idea is to run more rounds of the iterative process, but with different parameters so as to find a stronger approximation. We call this the secondary process.

Each run of the secondary process has a parameter s , which in the beginning is set to $s = u$. We then take $\xi' = \varepsilon_{s+3}$ and choose d' and δ' from Theorem 2.20 for ξ' and δ' . Now run the process above with ξ' and δ' from the point we stopped it. Then the process would terminate after at most $O(1/\delta'^2)$ steps, and we would reach j such that have that $\|f_j\|_\nu \leq \xi'$. We thus get the candidate decomposition as in the statement with $\mathcal{P} = \mathcal{P}_{i^*}$ with η_{i^*} , and $\mathcal{P}' = \mathcal{P}_j$ with η_j , and it is clear that the first two items hold. If the third item holds we terminate, and otherwise we take $s = j + 2$ and run the secondary process again.

If this iterative process terminates within $100/\xi^2$ steps, then it is clear we get a decomposition as needed. The key claim now is that if the secondary process runs too many times, we can still find a decomposition as per the following claim.

Claim 4.3. *Suppose that the secondary process has not terminated within S runs, and that $S \geq \frac{100}{\xi^4}$. Then we can find $j \leq S - 2$ such that taking $\mathcal{P} = \mathcal{P}_j$ and $\mathcal{P}' = \mathcal{P}_{j+2}$ satisfies the claim.*

Proof. Fix $j_1, \dots, j_S$, where j_t is the index j at the end of the execution of the secondary process when it is run for the t th time. Consider the functions $g_{j_1}, \dots, g_{j_S}$, and for simplicity we denote them by $G_1, \dots, G_S$ and the corresponding noise operators by $T_1, \dots, T_S$.

Inspect the quantity $(I) = \sum_{\ell < S/2} \|(T_{2\ell} - T_{2(\ell-1)})f\|_2^2$. Then by self adjointness we may write it as

$$\begin{aligned} (I) &= \sum_{\ell < S/2} \langle (T_{2\ell} - T_{2(\ell-1)})^2 f, f \rangle = \langle \sum_{\ell < S/2} (T_{2\ell} - T_{2(\ell-1)})^2 f, f \rangle \\ &\leq \left\| \sum_{\ell < S/2} (T_{2\ell} - T_{2(\ell-1)})^2 f \right\|, \end{aligned}$$

where the last inequality is by Cauchy-Schwarz. Squaring the last quantity and expanding, we get it is at most

$$\sum_{\ell} \|(T_{2\ell} - T_{2(\ell-1)})^2 f\|_2^2 + 2 \sum_{\ell' < \ell} |\langle (T_{2\ell} - T_{2(\ell-1)})^2 f, (T_{2\ell'} - T_{2(\ell'-1)})^2 f \rangle|. \quad (5)$$

For the first sum, each summand is at most 16. As per the second sum, fix $\ell' < \ell$ and note that

$$|\langle (T_{2\ell} - T_{2(\ell-1)})^2 f, (T_{2\ell'} - T_{2(\ell'-1)})^2 f \rangle| = |\langle (T_{2\ell} - T_{2(\ell-1)}) f, (T_{2\ell} - T_{2(\ell-1)})(T_{2\ell'} - T_{2(\ell'-1)})^2 f \rangle|,$$

which by Cauchy-Schwarz and the crude bound $\|(T_{2\ell} - T_{2(\ell-1)}) f\|_2 \leq 2$, is at most

$$2\|(T_{2\ell} - T_{2(\ell-1)})(T_{2\ell'} - T_{2(\ell'-1)})^2 f\|_2. \quad (6)$$

Repeating the argument as in the “lower bound” of Claim 4.2, we get that

$$\begin{aligned} \|T_{2\ell}(T_{2\ell'} - T_{2(\ell'-1)})^2 f - (T_{2\ell'} - T_{2(\ell'-1)})^2 f\|_2 &\leq \eta_{2\ell-1}^{1/6}, \\ \|T_{2(\ell-1)}(T_{2\ell'} - T_{2(\ell'-1)})^2 f - (T_{2\ell'} - T_{2(\ell'-1)})^2 f\|_2 &\leq \eta_{2\ell-2}^{1/6}, \end{aligned}$$

which by the triangle inequality gives that (6) $\leq 2\eta_{2\ell-2}^{1/6}$. Plugging this into (5) gives that

$$(5) \leq 16 \left(\frac{S}{2} - 1 \right) + \sum_{\ell} 2\ell \eta_{2\ell-2}^{1/12} \leq 8S,$$

and plugging this above gives that $(I) \leq \sqrt{8S}$. By definition of (I) we conclude that there is ℓ such that

$$\|T_{2\ell} f - T_{2\ell-2} f\|_2^2 \leq \frac{2\sqrt{8}}{\sqrt{S}}.$$

Pick $\mathcal{P} = \mathcal{P}_{2\ell-2}$ and $\mathcal{P}' = \mathcal{P}_{2\ell}$. The first two items of the claim hold by the fact that the secondary process terminated at times $j_{2\ell-2}$ and $j_{2\ell}$. The third property follows as

$$\|T_{2\ell} f - T_{2\ell-2} f\|_2^2 \leq \frac{2\sqrt{8}}{\sqrt{S}} \leq \frac{2\sqrt{8}}{10/\xi^2} \leq \xi^2. \quad \square$$

Thus, if the secondary process was run for at least $100/\xi^4$ steps we are done by Claim 4.3, and the proof is concluded. $\square$

4.2 The Version with High Rank

We now state and prove a version of Lemma 4.1, in which the collection $\mathcal{P}$ is guaranteed to have high rank and high softrank.

Lemma 4.4. *Let Σ be a finite alphabet of size at most m , let ν be a probability distribution over Σ in which the probability of each atom is at least $\alpha > 0$, let μ be a distribution over Σ^3 with marginals equal to ν , and let $w: (0, 1) \rightarrow (0, 1)$ be a sufficiently rapid decay function. Then for every $\xi > 0$, and for every 1-bounded function $f: \Sigma^n \rightarrow \mathbb{C}$, there exist w -separated scales*

$$\rho_5 \ll_w \rho_4 \ll_w \rho_3 \ll_w \rho_2 \ll_w \rho_1 \ll_w \rho_0$$

satisfying $\rho_0 \leq \xi$ and $\rho_5 \geq \Omega_{w,m,\alpha,\xi}(1)$, and collections of cyclic product functions $\mathcal{P}, \mathcal{P}'$ such that

1. The collection $\mathcal{P}$ satisfies:

- (a) $|\mathcal{P}| \leq 1/\rho_0$.
- (b) $\mathcal{P}$ is $O_m(1)$ -discrete.
- (c) $\|f - T_{\mathcal{P}, 1-\rho_1} f\|_\nu \leq \xi$.
- (d) $\text{rk}(\mathcal{P}) \geq \frac{1}{\rho_2}$.
- (e) $\text{softrk}_\mu(\mathcal{P}) \geq \frac{1}{\rho_2}$.

2. The collection $\mathcal{P}'$ has size at most $1/\rho_4$, is $O_m(1)$ -discrete and we have $\|f - T_{\mathcal{P}', 1-\rho_5} f\|_\nu \leq \rho_3$.

3. We have that $\|T_{\mathcal{P}', 1-\rho_5} f - T_{\mathcal{P}, 1-\rho_1} f\|_2 \leq \xi$.

Proof. We first define a very rapidly decaying function $w' : (0, 1) \times \mathbb{N} \rightarrow (0, 1)$ with which we will want to apply Lemma 4.1. Set $\tilde{w}(\varepsilon, i) = w(\varepsilon)$. On input (ε, i) , first compute ε_i and $M = O_m(1)$ from Lemma 4.1 for $\tilde{w}$. Then, take k_i from Lemma 3.4 for M -discrete collections of size at most $1/\varepsilon_i$. Finally, take $w'(\varepsilon, i) = 100^{2(k_1 + \dots + k_i)} \cdot w^{2(k_1 + \dots + k_i)}(\varepsilon)$. For ease of notation, throughout we denote η_0 and $\eta_{i+1} = w'(\eta_i, i)$ as used in Lemma 4.1 for the decay function w' and $\xi/4$.

Applying Lemma 4.1 we find i , $\mathcal{P}$ and $\mathcal{P}'$ satisfying the properties therein with the decay function w' . Set $k = k_1 + \dots + k_i$ and consider the sequence $D_0 \leq D_1 \leq \dots \leq D_k$ given as

$$D_0 = \frac{1}{\eta_{i+1}}, \quad D_1 = 2w^{(2)}(D_0^{-1})^{-1}, \quad \dots, \quad D_k = 2w^{(2)}(D_{k-1}^{-1})^{-1}.$$

We note that $D_k \leq \eta_{i+2}^{-1}$. By Lemma 3.4 we may find $1 \leq j \leq k$ and a collection $\tilde{\mathcal{P}}$ of $O_m(1)$ -discrete cyclic product functions such that $\text{rk}(\tilde{\mathcal{P}}), \text{softrk}(\tilde{\mathcal{P}}) \geq D_j - (D_1 + \dots + D_{j-1})$, each function in $\mathcal{P}$ is $3D_{j-1}$ -close to a function in $\text{span}_{\mathbb{N}}(\tilde{\mathcal{P}})$, and each function in $\tilde{\mathcal{P}}$ is $3D_{j-1}$ -close to a function in $\text{span}_{\mathbb{N}}(\mathcal{P})$. We prove that the collections $\tilde{\mathcal{P}}$ and $\mathcal{P}'$ satisfy the properties of the lemma with the scales

$$\rho_0 = \eta_i, \quad \rho_1 = w(D_{j-1}^{-1}), \quad \rho_2 = w(\rho_1), \quad \rho_3 = \eta_{i+3}, \quad \rho_4 = \eta_{i+4}, \quad \rho_5 = \eta_{i+5}.$$

Item 1(a) follows because $|\tilde{\mathcal{P}}| \leq |\mathcal{P}| \leq \eta_i^{-1}$ by Lemma 4.1. Item 1(b) follows by Lemma 3.4. For item 1(c), by Fact 2.23 we get that

$$\|T_{\tilde{\mathcal{P}}, 1-\rho_1} f - T_{\mathcal{P}, 1-\rho_1} f\|_2 \leq \sqrt{2(|\mathcal{P}| + |\tilde{\mathcal{P}}|) \cdot 3D_{j-1} \cdot \rho_1} \leq \rho_1^{1/4}, \quad (7)$$

and by the first item of Lemma 4.1, the triangle inequality and the trivial $\|g\|_\nu \leq \|g\|_2$, we get that

$$\|f - T_{\mathcal{P}, 1-\rho_1} f\|_\nu \leq \frac{\xi}{4} + \rho_1^{1/4} \leq \xi.$$

Item 3 follows by combining the third item in Lemma 4.1 and (7) with the triangle inequality. Item 2 follows from the second item in Lemma 4.1. Lastly, for items 1(d) and 1(e), since w is sufficiently rapidly decaying we get that $\text{rk}(\tilde{\mathcal{P}}), \text{softrk}(\tilde{\mathcal{P}}) \geq D_j/2 \geq \rho_2^{-1}$. $\square$

5 Proof of the Counting Lemma

Apart from the arithmetic regularity lemma that we proved in the previous section, we also need the invariance principle from [Mos10] and a multidimensional Gaussian reverse hypercontractivity theorem due to [CDP15]. To state these results, we first give some preliminaries.

5.1 Preliminaries

5.1.1 Fourier Analysis

It will be convenient to work with the Fourier basis for the space of functions $L^2(\Sigma^n, \nu^{\otimes n})$. The definitions below and the proofs of the propositions can be found in [O'D14].

Definition 5.1. An n -dimensional multi-index is a tuple $\alpha \in \mathbb{N}^n$. We write

$$\text{supp}(\alpha) = \{i : \alpha_i \neq 0\}, \quad \#\alpha = |\text{supp}(\alpha)|, \quad |\alpha| = \sum_{i=1}^n \alpha_i.$$

We write $\alpha \in \mathbb{N}_{< m}^n$ when we want to emphasize that each $\alpha_i \in \{0, 1, \dots, m-1\}$.

Definition 5.2. Given functions $\phi_0, \dots, \phi_{m-1} \in L^2(\Sigma, \nu)$ and a multi-index $\alpha \in \mathbb{N}_{< m}^n$, we define $\phi_\alpha \in L^2(\Sigma^n, \nu^{\otimes n})$ by

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

Proposition 5.3. Let $\phi_0, \dots, \phi_{m-1}$ be an orthonormal basis for $L^2(\Sigma, \nu)$ where ϕ_0 is the constant 1 function. Then the collection $(\phi_\alpha)_{\alpha \in \mathbb{N}_{< m}^n}$ is an orthonormal basis for $L^2(\Sigma^n, \nu^{\otimes n})$.

5.1.2 Influences

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

Definition 5.4. For a function $f : (\Sigma^n, \nu^{\otimes n}) \rightarrow \mathbb{C}$ and a coordinate $i \in [n]$, the influence of i is defined as

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

Here, $(x_{-i} = y, x_i = a)$ is the point in Σ^n whose i th coordinate is a , and the rest of its coordinates are filled according to y . The total influence of a function is defined as $I[f, \nu^{\otimes n}] = \sum_{i=1}^n I_i[f, \nu^{\otimes n}]$.

We have the following proposition.

Proposition 5.5. For a function $f : (\Sigma^n, \nu^{\otimes n}) \rightarrow \mathbb{C}$, we have

$$I[f, \nu^{\otimes n}] = \sum_{\alpha \in \mathbb{N}_{< m}^n} |\alpha| |\hat{f}(\alpha)|^2, \quad I_i[f, \nu^{\otimes n}] = \sum_{\substack{\alpha \in \mathbb{N}_{< m}^n \\ \alpha_i \neq 0}} |\hat{f}(\alpha)|^2,$$

where $\hat{f}(\alpha) := \langle f, \phi_\alpha \rangle$ and $|\alpha|$ is the number of non-zero coordinates in α .

For a function f that is a sum of small functions with small influences, the following proposition upper bounds the influences of the function f .

Proposition 5.6. Let $f : (\Sigma^n, \nu^{\otimes n}) \rightarrow \mathbb{C}$ be such that

$$f(x) = \sum_{j=1}^R \theta_j \cdot f_j(x)$$

where $f_j : (\Sigma^n, \nu^{\otimes n}) \rightarrow \mathbb{C}$ and $|\theta_j| \leq 1$ for all $1 \leq j \leq R$, then for all $i \in [n]$, we have $I_i[f, \nu^{\otimes n}] \lesssim_R \sum_j I_i[f_j, \nu^{\otimes n}]$.

Proof. We expand $I_i(f)$ as follows.

$$\begin{aligned}
I_i(f) &= \mathbb{E}_{\substack{y \sim \nu^{\otimes(n-1)} \\ a, b \sim \nu}} \left[|f(x_{-i} = y, x_i = a) - f(x_{-i} = y, x_i = b)|^2 \right] \\
&= \mathbb{E}_{\substack{y \sim \nu^{\otimes(n-1)} \\ a, b \sim \nu}} \left[\left| \sum_{j=1}^R \theta_j (f_j(x_{-i} = y, x_i = a) - f_j(x_{-i} = y, x_i = b)) \right|^2 \right] \\
&\lesssim_R \mathbb{E}_{\substack{y \sim \nu^{\otimes(n-1)} \\ a, b \sim \nu}} \left[\sum_{j=1}^R |\theta_j (f_j(x_{-i} = y, x_i = a) - f_j(x_{-i} = y, x_i = b))|^2 \right],
\end{aligned}$$

where the last inequality is by the Cauchy-Schwarz inequality. Continuing, we get that

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

as required. $\square$

5.1.3 The Invariance Principle

In this section, we present the invariance principle of [MOO10], and we begin with some set-up. Suppose that Σ, Γ, Φ are finite alphabets of sizes m_1, m_2, m_3 respectively, and μ is a probability measure over $\Sigma \times \Gamma \times \Phi$ in which the probability of each atom is at least $\alpha > 0$. We set up orthonormal bases for (Σ, μ_x) , (Γ, μ_y) and (Φ, μ_z) given by $v_0, \dots, v_{m_1-1}$, $u_0, \dots, u_{m_2-1}$ and $w_0, \dots, w_{m_3-1}$. Consider the ensemble of random variables

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

where $(x, y, z) \sim \mu$. We define the covariance matrix $P \in \mathbb{C}^{(m_1+m_2+m_3-3) \times (m_1+m_2+m_3-3)}$ whose rows and columns correspond to the functions in $\mathcal{X}$, and the entry corresponding to two random variables in $\mathcal{X}$. For example, for v_i and u_j , the corresponding entry is

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

Let $L^2(\mathbb{C}^n, \gamma^n)$ be the inner product space of functions with standard complex $\mathcal{CN}(0, 1)$ Gaussian measure.⁵ We define an ensemble of standard *proper*⁶ complex Gaussian variables having the same covariance matrix P as follows. Let $Z \sim \mathcal{CN}(0, I_{m_1+m_2+m_3-3})$ be $(m_1+m_2+m_3-3)$ independent standard complex Gaussian variables. We define

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

as $\mathcal{G} := P^{1/2}Z$ to be an ensemble of centered proper Gaussian random variables with covariance matrix P . Note that $\mathcal{G}$ satisfies (i) $\mathbb{E}[\mathcal{G}\mathcal{G}^T] = 0$ and (ii) $\mathbb{E}[\mathcal{G}\mathcal{G}^*] = P$. The invariance principle relates the behavior

⁵A complex random variable Z whose real and imaginary parts are independent normally distributed random variables with mean zero and variance $\frac{1}{2}$.

⁶An ensemble $\mathcal{G}$ of standard complex Gaussian variables is proper if it satisfies $\mathbb{E}[\mathcal{G}\mathcal{G}^T] = 0$.

of low-influence, multi-linear polynomials over $\mathcal{X}$ and over $\mathcal{G}$. Below, we state the version that we need from [Mos10] specialized to our case of interest, but before that, we need a few definitions.

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

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

Definition 5.7. *The degree of M as above is defined as*

$$\deg(M) = \max_T \{|T| \mid m_T \neq 0\}.$$

Definition 5.8. *The influence of variable (i, j) on M as above is defined as*

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

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

Define $\text{trunc}_{[0,1]}: \mathbb{C} \rightarrow [0, 1]$ as follows:

$$\text{trunc}_{[0,1]}(a) = \begin{cases} 0 & \text{if } \text{Re}(a) < 0, \\ \text{Re}(a) & \text{if } 0 \leq \text{Re}(a) \leq 1, \\ 1 & \text{if } \text{Re}(a) > 1. \end{cases}$$

The invariance principle theorem from Mossel [Mos10] is as follows.

Theorem 5.9. *For all $\alpha > 0$, $k, m \in \mathbb{N}$, $d \in \mathbb{N}$, $C > 0$ and $\varepsilon > 0$, there exists $\tau > 0$ such that the following holds. Suppose that $|\Sigma| = m_1$, $|\Gamma| = m_2$, $|\Phi| = m_3$, that μ is a distribution over $\Sigma \times \Gamma \times \Phi$ in which the probability of each atom is at least α , and let $\mathcal{X}$ and $\mathcal{G}$ be the ensembles of random variables as above with the same covariance matrix. Let $M: \mathbb{C}^{qn} \rightarrow \mathbb{C}^k$ is a multi-linear polynomial with $\max_{i,j} I_{i,j}[M] \leq \tau$, $\deg(M) \leq d$ and $\|M\|_2 \leq C$.*

1. *If $\Psi: \mathbb{C}^k \rightarrow \mathbb{C}$ is differentiable three times and its third order derivatives are at most C in absolute value, then*

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

2. *Define $\zeta: \mathbb{C}^k \rightarrow \mathbb{R}$ by $\zeta(a_1, \dots, a_k) = \sqrt{\sum_{i=1}^k |\text{trunc}_{[0,1]}(a_i) - a_i|^2}$. Then*

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

We will also need the following fact.

Fact 5.10. *Consider the function $\zeta: \mathbb{C} \rightarrow [0, \infty)$ defined as $\zeta(a) = |\text{trunc}_{[0,1]}(a) - a|$. Then ζ is 2-Lipshitz function.*

Proof. Let $a, b \in \mathbb{C}$; we show that $|\zeta(a) - \zeta(b)| \leq 2|a - b|$. It is enough to show this for the real numbers a and b . If a, b are both on $[0, 1]$, then the left-hand side is 0 and the claim is trivial. If exactly one of a and b is in $[0, 1]$, say a then

$$|\zeta(a) - \zeta(b)| = \zeta(b)$$

If $b > 1$, then $\zeta(b) = b - 1 \leq |b - a|$. Otherwise, $b < 0$, and $\zeta(b) = |b| \leq |b - a|$. It remains to consider the case that both a and b are not in $[0, 1]$. When $a, b < 0$, then $|\zeta(a) - \zeta(b)| = |a| - |b| = |a - b|$. If $a, b > 1$, then $|\zeta(a) - \zeta(b)| = |a - b|$. Finally, if $a < 0$ and $b > 1$, we have $\zeta(a) - \zeta(b) = -a + (b - 1) \leq 2|a - b|$, and the proof is concluded. $\square$

5.1.4 Gaussian Reverse Hypercontractivity

Finally, we will need the multidimensional version of the Gaussian reverse hypercontractivity theorem from [CDP15]. We use the following special case of the theorem from [HHM18].

Theorem 5.11. *Let $\eta > 0$ and let $\mathcal{G} = (\mathcal{G}_1, \dots, \mathcal{G}_\ell)$ be a jointly proper Gaussian collection of ℓ random vectors such that:*

1. *For each $i \in [\ell]$, $\mathcal{G}_i = (G_{i,1}, \dots, G_{i,n})$ is a random vector distributed as n independent $\mathcal{CN}(0, 1)$ Gaussians.*
2. *The $n\ell \times n\ell$ covariance matrix $P = \mathbb{E}[\mathcal{G}\mathcal{G}^*]$ of the Gaussian ensemble satisfies $P - \eta I \succeq 0$, for some $\eta > 0$.*

Then, for all functions $f_1, \dots, f_\ell \in L^2(\mathbb{C}^n, \gamma^n)$ such that $f_1, \dots, f_\ell : \mathbb{C}^n \rightarrow [0, 1]$ and $\mathbb{E}[f_j(\mathcal{G}_j)] = \mu_j$, we have:

$$\mathbb{E} \left[\prod_{j=1}^{\ell} f_j(\mathcal{G}_j) \right] \geq \left(\prod_{j=1}^{\ell} \mu_j \right)^{1/\eta}.$$

Remark 5.12. *The statement of the theorem from [CDP15] (and [HHM18]) is for real Gaussian random variables (See Theorem A.1). In Appendix A, we show how to derive the above theorem from their theorem statement.*

5.2 Proof of the Main Theorem

Let $A \subseteq \Sigma^n$ such that $\frac{|A|}{|\Sigma|^n} \geq \delta$ and let $f = 1_A$ be the indicator function of the set A . The goal is to estimate the following quantity

$$\Delta_1(f) := \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)f(y)f(z)]$$

where the minimum non-zero probability of the distribution μ is $\alpha > 0$. Note that $\mu|_x = \mu|_y = \mu|_z = \nu$ are the marginals of μ . Hence, $\mu \in M_{\nu, \alpha}$ as per Definition 2.18. Therefore, the conditions in the arithmetic regularity lemma, Lemma 4.4, are satisfied for this distribution and we will use it henceforth.

5.2.1 Making the Function Resilient

As the first step of this process, we show that it suffices to bound Δ_1 for resilient functions, defined as follows:

Definition 5.13. For $B \in \mathbb{N}$ and $0 < \gamma \leq 1$, we say a function $g: \Sigma^n \rightarrow [0, 1]$ is (B, γ) -resilient if for every $I \subseteq [n]$ of size at most B and every $v \in \Sigma^I$ we have that $\nu^{\bar{I}}(g_{I \rightarrow v}) \geq \gamma \cdot \nu^{\otimes n}(g)$.

We have the following claim.

Claim 5.14. Suppose $\Delta_1(g) \geq \delta'(B, \eta)$ for every $(B, 1/2)$ -resilient function $g: \Sigma^n \rightarrow \{0, 1\}$ with $\eta = \nu^{\otimes n}(g)$. Then, for every $f: \Sigma^n \rightarrow \{0, 1\}$ with $\nu^{\otimes n}(f) = \delta$, we have

$$\Delta_1(f) \geq \alpha^{\frac{2B}{\delta\alpha^B}} \delta'(B, \delta).$$

Proof. If f is $(B, 1/2)$ -resilient, then $\Delta_1(f) \geq \delta'(B, \delta)$ and we are done. Otherwise, we have that f is not $(B, 1/2)$ -resilient. This means there exists $I_1 \subseteq [n]$ and $w \in \Sigma^{I_1}$ such that $|I_1| \leq B$ and $\nu^{\bar{I}_1}(f_{I_1 \rightarrow w}) \leq \delta/2$. As $\nu^{\otimes n}(f) = \delta$, there exists v_1 such that $\nu^{\bar{I}_1}(f_{I_1 \rightarrow v_1}) \geq \delta + \frac{\delta}{2}\alpha^B$. Now, if $f_{I_1 \rightarrow v_1}$ is $(B, 1/2)$ -resilient then we get that

$$\begin{aligned} \Delta_1(f) &= \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)f(y)f(z)] \\ &\geq \alpha^{|I_1|} \cdot \mathbb{E}_{(x,y,z) \sim \mu^{\bar{I}_1}} [f_{I_1 \rightarrow v_1}(x)f_{I_1 \rightarrow v_1}(y)f_{I_1 \rightarrow v_1}(z)] \\ &\geq \alpha^B \cdot \delta'(B, \delta), \end{aligned}$$

and hence the conclusion holds. If $f_{I_1 \rightarrow v_1}$ is not $(B, 1/2)$ -resilient, then we can find I_2 and v_2 and continue the process. Each time, the measure of the restricted function increases by an additive factor of at least $\frac{\delta}{2}\alpha^B$. Thus, the process stops after at most $t \leq \frac{2}{\delta\alpha^B}$ iterations, as otherwise the measure of the restricted function becomes greater than 1, which cannot happen as $f(x) \in \{0, 1\}$ for all x . Let $I = I_1 \cup I_2 \cup \dots \cup I_t$ and $v = v_1 \circ v_2 \circ \dots \circ v_t$ be the union of the subsets and the concatenation of the strings found during the process. We have $|I| \leq t \cdot B \leq \frac{2B}{\delta\alpha^B}$. The function $f_{I \rightarrow v}$ is now $(B, 1/2)$ -resilient with measure at least $\frac{\delta}{2}$, and hence

$$\begin{aligned} \Delta_1(f) &= \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f(x)f(x+a)f(x+2a)] \\ &\geq \alpha^{|I|} \cdot \mathbb{E}_{(x,y,z) \sim \mu^{\bar{I}}} [f_{I \rightarrow v}(x)f_{I \rightarrow v}(y)f_{I \rightarrow v}(z)] \\ &\geq \alpha^{\frac{2B}{\delta\alpha^B}} \cdot \delta'(B, \delta). \end{aligned} \quad \square$$

Thus, it suffices to establish a lower bound $\Delta_1(f)$ under the assumption that f is $(B, 1/2)$ -resilient, and focus on this case henceforth.

5.2.2 Applying the Arithmetic Regularity Lemma

We now apply Lemma 4.4. For better readability, it will be convenient to denote various functions of the lemma using simpler notation. We will use the following notation.

$$\begin{aligned}
f &= T_{\mathcal{P}, 1-\rho_1} f, & f' &= T_{\mathcal{P}', 1-\rho_5} f \\
F(x) &= \sum_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} P(x) \cdot L_P(x), & \tilde{F}(x) &= \text{trunc}_{[0,1]}(F(x)) \\
F_d(x, y) &= \sum_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} P(x) \cdot L_P(y), & \tilde{F}_d(x', x'') &:= \text{trunc}_{[0,1]}(F_d(x', x'')).
\end{aligned} \tag{8}$$

Here, f and f' are from Lemma 4.4, and F is from Lemma 2.25 applied to the function f with $\|F - f\|_2 \leq \xi$. The function F_d is the same as the function F except that it uses two separate inputs for the product functions and the low-degree functions in the decomposition of F . We will also use the following notation throughout.

$$\begin{aligned}
m &:= |\Sigma| \\
M &:= \max_{P \in \mathcal{P}} \{\text{ord}(P)\} \text{ where } M = O_m(1). \\
r &:= |\mathcal{P}| \leq \frac{1}{\rho_0} \\
\|L_P\|_2 &\leq C, \quad \deg(L_P) \leq D \text{ where } C \text{ and } D \text{ depend on } \xi \text{ and } \rho_1.
\end{aligned}$$

We set the parameters as follows (the parameters τ , η and κ would appear later in the proof). Let $w: (0, 1) \rightarrow (0, 1)$ be a sufficiently rapid decay function and set⁷

$$\rho_5 \ll_w \rho_4 \ll_w \rho_3 \ll_w \tau \ll_w \rho_2 \ll_w \kappa \ll_w \rho_1 \ll_w \rho_0 \leq \eta \leq \xi \ll \delta \ll m^{-1}, \alpha. \tag{9}$$

We now replace f with f' in the expression $\Delta_1(f)$ by paying only a factor of $O(\rho_3)$ as follows.

$$\begin{aligned}
\Delta_1(f) &:= \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f'(x)f(y)f(z)] + \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [(f - f')(x)f(y)f(z)] \\
&\geq \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f'(x)f(y)f(z)] - \|f - f'\|_{\mu} \\
&\geq \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f'(x)f(y)f(z)] - \rho_3.
\end{aligned}$$

Repeating a similar process for the second and third occurrences of f , we get

$$\Delta_1(f) \geq \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f'(x)f'(y)f'(z)] - 3\rho_3. \tag{10}$$

The following claim shows that the resilience of f implies resilience of f .

Claim 5.15. *The function f is $(B, 1/4)$ -resilient.*

Proof. To show that f is $(B, 1/4)$ -resilient, fix a subset $J \subseteq [n]$ of size at most B and a string $v \in \Sigma^J$. Let $n' = n - |J|$. We start by looking at the function $f_{J \rightarrow v}$.

$$f_{J \rightarrow v}(x) = \mathbb{E}_{y \sim T_{\mathcal{P}, 1-\rho_1}((v, x))} [f(y)].$$

Our goal is to show that the following quantity is at least $\frac{\delta}{4}$.

$$\nu^{\otimes n'}(f_{J \rightarrow v}) = \mathbb{E}_{x \sim \nu^{\otimes n'}} \left[\mathbb{E}_{y \sim T_{\mathcal{P}, 1-\rho_1}((v, x))} [f(y)] \right].$$

⁷Note that this setting corresponds to applying Lemma 4.4 with the decay function $w(\cdot)^2$ as $\rho_3 \ll_{w^2} \rho_2$ and $\rho_2 \ll_{w^2} \rho_1$.

Recall the operator $T_{\nu, \mathcal{P}, I}: L_2(\Sigma^n, \nu^{\otimes n}) \rightarrow L_2(\Sigma^n, \nu^{\otimes n})$ from Definition 2.21. We have

$$\nu^{\otimes n'}(f_{J \rightarrow v}) = \mathbb{E}_{x \sim \nu^{\otimes n'}} \left[\mathbb{E}_{I \subseteq_{\rho_1} [n]} \left[\mathbb{E}_{y \sim T_{\mathcal{P}, I}((v, x))} [f(y)] \right] \right].$$

We will show that with probability at least 0.99 over the choice of I , the statistical distance between the distribution on y in the above expectation and the distribution on $z \sim \nu^{\otimes n}$ conditioned on $z_{J'} = v'$ for $J' = J \cap I$ and $v' = v|_{J'}$ is upper bounded by $\frac{\delta}{10}$. Once we show that, as f takes value in $\{0, 1\}$, we will get that

$$\nu^{\otimes n'}(f_{J \rightarrow v}) \geq 0.99 \mathbb{E}_{I \subseteq_{1-\rho_1} [n]} \left[\nu^{\otimes n-|J'|}(f_{J' \rightarrow v'}) \right] - \frac{\delta}{10} \geq 0.99 \cdot \frac{\delta}{2} - \frac{\delta}{10} \geq \frac{\delta}{4},$$

where the second inequality used the fact that f is $(B, 1/2)$ -resilient and $\nu^{\otimes n}(f) \geq \delta$.

Towards showing the closeness in the statistical distance, define the following two distributions:

- $\mathcal{D}_{I,v}$: Sample $x \sim \nu^{\otimes n'}$, and output $y \sim T_{\mathcal{P}, I}((v, x))$.
- $\mathcal{D}'_{J',v'}$: Sample $z \sim \nu^{\otimes n}$ conditioned on $z_{J'} = v'$.

Let $\mathcal{P} = \{P_1, P_2, \dots, P_r\}$. Thus, the proof will be complete if we show the following two statements.

- (a) With probability at least 0.99 over $I \subseteq_{\rho_1} [n]$, $\text{rk}(\mathcal{P}|_I) \geq \frac{1}{4\rho_2}$.
- (b) If $\text{rk}(\mathcal{P}|_I) \geq \frac{\rho_1}{2\rho_2}$, then the distribution $\mathcal{D}_{I,v}$ is $\frac{\delta}{10}$ -close to the distribution $\mathcal{D}'_{J',v'}$.

Proof of (a). Fix a vector $\vec{\alpha} \in (\mathbb{N} \cup \{0\})^{|\mathcal{P}|}$ where $0 \leq \alpha_P < \text{ord}(P)$ for all $P \in \mathcal{P}$ that are not all zero. Let $S_{\vec{\alpha}}(x) = \prod_{P \in \mathcal{P}} P(x)^{\alpha_P}$. As $\text{rk}(\mathcal{P}) \geq \frac{1}{\rho_2}$, we have $\Delta_{\text{symbolic}}(S_{\vec{\alpha}}, 1) \geq \frac{1}{\rho_2}$. Let $A(\vec{\alpha}) \subseteq [n]$ be the set of coordinates where $(S_{\vec{\alpha}})_i \neq 1$. Thus, we have $|A(\vec{\alpha})| \geq \frac{1}{\rho_2}$. For $I \subseteq_{\rho_1} [n]$, we have that $\mathbb{E}_I[|I \cap A(\vec{\alpha})|] \geq \frac{\rho_1}{\rho_2}$. By Chernoff bound, the probability that $|I \cap A(\vec{\alpha})| \leq \frac{\rho_1}{2\rho_2}$ is at most $e^{-\frac{\rho_1}{8\rho_2}}$. Finally, taking the union bound over all the admissible $\vec{\alpha}$ (the number of admissible $\vec{\alpha}$ s is at most $e^{O_m(1/\rho_0)}$ as $\text{ord}(P) = O_m(1)$ for all $P \in \mathcal{P}$ and $|\mathcal{P}| \leq \frac{1}{\rho_0}$), we have that there exists an $\vec{\alpha}$ such that $|I \cap A(\vec{\alpha})| \leq \frac{\rho_1}{2\rho_2}$ is at most $e^{O_m(1/\rho_0)} \cdot e^{-\frac{\rho_1}{8\rho_2}} \leq 0.01$. Therefore, with probability at least 0.99 over $I \subseteq_{\rho_1} [n]$, we have $\text{rk}(\mathcal{P}|_I) \geq \frac{\rho_1}{2\rho_2}$ as required.

Proof of (b). Consider the following two sets for $v' \in \Sigma^{J'}$ and $w \in \Sigma^n$:

$$S_{v'} = \{z \mid z_{J'} = v'\},$$

$$T_w = \{z \mid P(x) = P(w) \ \forall P \in \mathcal{P} \text{ and } z_{\bar{I}} = w_{\bar{I}}\}.$$

For any fixed $w \in \Sigma^n$, we will now compute $\Pr_{y \sim \mathcal{D}'_{J',v'}}[y = w]$ and $\Pr_{y \sim \mathcal{D}_{I,v}}[y = w]$. The first one is easy:

$$\Pr_{y \sim \mathcal{D}'_{J',v'}}[y = w] = \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(S_{v'})}.$$

We now estimate $\Pr_{y \sim \mathcal{D}_{I,v}}[y = w]$. Recall the distribution $\mathcal{D}_{I,v}$: Sample $x \sim \nu^{\otimes n'}$, and output $y \sim T_{\mathcal{P}, I}((v, x))$. Thus, the following conditions must be met for $y = w$

1. $x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}$, as the noise operator $T_{\mathcal{P}, I}$ does not change $x|_{\bar{I} \cap \bar{J}}$.
2. $P((v, x)) = P(w) \forall P \in \mathcal{P}$. Let this event be E .
3. Finally, $y_I = w_I$.

Thus,

$$\begin{aligned} & \Pr_{\substack{x \sim \nu^{\otimes n'} \\ y \sim T_{\mathcal{P}, I}((v, x))}} [y = w] \\ &= \Pr_{x \sim \nu^{\otimes n'}} [x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}] \cdot \Pr_{x \sim \nu^{\otimes n'}} [E \mid x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}] \cdot \Pr_{\substack{x \sim \nu^{\otimes n'} \\ y \sim T_{\mathcal{P}, I}((v, x))}} [y = w \mid E \wedge x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}]. \end{aligned}$$

We now estimate each of these probabilities. Note that

$$\Pr_{x \sim \nu^{\otimes n'}} [x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}] = \prod_{\ell \in \bar{I} \cap \bar{J}} \nu(w_\ell) = \prod_{\ell \in \bar{I} \cap \bar{J}} \nu(w_\ell) \cdot \frac{\prod_{\ell \in J'} \nu(w_\ell)}{\prod_{\ell \in J'} \nu(w_\ell)} = \frac{\nu^{\bar{I}}(w|_{\bar{I}})}{\nu^{\otimes n}(S_{v'})}.$$

Also,

$$\Pr_{\substack{x \sim \nu^{\otimes n'} \\ y \sim T_{\mathcal{P}, I}((v, x))}} [y = w \mid E \wedge x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}] = \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(T_w)}.$$

Therefore,

$$\begin{aligned} \left| \Pr_{\substack{x \sim \nu^{\otimes n'} \\ y \sim T_{\mathcal{P}, I}((v, x))}} [y = w] - \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(S_{v'})} \right| &\leq \left| \frac{\nu^{\bar{I}}(w|_{\bar{I}})}{\nu^{\otimes n}(S_{v'})} \Pr_{x \sim \nu^{\otimes n'}} [E \mid x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}] \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(T_w)} - \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(S_{v'})} \right| \\ &\leq \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(S_{v'})} \left| \frac{\nu^{\bar{I}}(w|_{\bar{I}}) \Pr_{x \sim \nu^{\otimes n'}} [E \mid x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}] - \nu^{\otimes n}(T_w)}{\nu^{\otimes n}(T_w)} \right|. \end{aligned}$$

Let $s = \prod_{P \in \mathcal{P}} \text{ord}(P)^{-1}$. Then, using the fact that $\text{rk}(\mathcal{P}|_I) \geq \frac{\rho_1}{2\rho_2}$ and Lemma 3.5, we have

$$\left| \Pr_{x \sim \nu^{\otimes n'}} [E \mid x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}] - s \right| \leq 2^{-\Omega_{M, r, \alpha}(\frac{\rho_1}{\rho_2})},$$

and

$$\left| \nu^{\otimes n}(T_w) - s \cdot \nu^{\bar{I}}(w|_{\bar{I}}) \right| \leq 2^{-\Omega_{M, r, \alpha}(\frac{\rho_1}{\rho_2})} \nu^{\bar{I}}(w|_{\bar{I}}).$$

Although the latter statement does not follow directly from Lemma 3.5, observe that the same proof works if we replace the distribution $\nu^{\otimes n}$ with the conditional distribution $\nu^{\otimes n}|_{x|_{\bar{I}} = w_{\bar{I}}}$ and using the fact that $\text{rk}(\mathcal{P}|_I) \geq \frac{\rho_1}{2\rho_2}$. Adding the conditioning back gives the required bound as above.

If we let $\eta = 2^{-\Omega_{M,r,\alpha}(\frac{\rho_1}{\rho_2})}$, then using the fact that $\rho_2 \ll \rho_1 \ll s$, we have

$$\begin{aligned}
\left| \Pr_{\substack{x \sim \nu^{\otimes n'} \\ y \sim T_{\mathcal{P},I}((v,x))}} [y = w] - \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(S_{v'})} \right| &\leq \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(S_{v'})} \left| \frac{\nu^{\bar{I}}(w|_{\bar{I}}) \Pr_{x \sim \nu^{\otimes n'}}[E \mid x_{\bar{I} \cap \bar{J}} = w_{\bar{I} \cap \bar{J}}] - \nu^{\otimes n}(T_w)}{\nu^{\otimes n}(T_w)} \right| \\
&\leq \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(S_{v'})} \left| \frac{\nu^{\bar{I}}(w|_{\bar{I}}) \cdot (s + \eta) - (s \cdot \nu^{\bar{I}}(w|_{\bar{I}}) - \eta \cdot \nu^{\bar{I}}(w|_{\bar{I}}))}{(s/2) \cdot \nu^{\bar{I}}(w|_{\bar{I}})} \right| \\
&\leq \frac{4\eta}{s} \frac{\nu^{\otimes n}(w)}{\nu^{\otimes n}(S_{v'})}.
\end{aligned}$$

Summing over all of w , we get that the the distribution $\mathcal{D}_{I,v}$ is $\frac{4\eta}{s}$ -close to the distribution $\mathcal{D}'_{J',v'}$. As $\rho_2 \ll \rho_1 \ll s$, we have $\frac{4\eta}{s} \leq \frac{\delta}{10}$ as required. $\square$

Define the following quantity,

$$\Delta_2(f') := \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f'(x)f'(y)f'(z)].$$

Using (10), as we have

$$\Delta_1(f) \geq \Delta_2(f') - 3\rho_3, \quad (11)$$

and our next goal is to establish a lower bound on $\Delta_2(f')$.

5.2.3 Removing influential variables

In the next step, we apply a random restriction to a small subset of coordinates. The primary goal of this restriction is to ensure that in the approximate decomposition of f , given by F ,

$$F(x) = \sum_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} P(x) \cdot L_P(x),$$

the restricted low-degree functions L_P for every $P \in \text{span}_{\mathbb{N}}(\mathcal{P})$ have small influences.

We formally define the influence of a variable in the decomposed function.

Definition 5.16. We say that a function $F : (\Sigma^n, \nu^{\otimes n}) \rightarrow \mathbb{C}$ given by

$$F(x) = \sum_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} P(x) \cdot L_P(x).$$

has τ -small shifted low-degree influences if for every $i \in [n]$ and every $P \in \text{span}_{\mathbb{N}}(\mathcal{P})$ it holds that $I_i[L_P, \nu^{\otimes n}] \leq \tau$.

The following claim shows the existence of a small set such that most restrictions to the variables in the set entail low influences for the restricted function F . The proof follows the strategy of the regularity lemma from [Jon16]. We prove the claim in the appendix, focusing on the bounds of the parameters.

Claim 5.17. Consider the function F from above. Then for any $\tau > 0$, there exists $J \subseteq [n]$ of size $O_{|\mathcal{P}|, |\Sigma|, D, C, \eta, \tau}(1)$ such that with probability at least $1 - \eta$ over $w \sim \nu^{\otimes |J|}$, we have

$$\max_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} \left\{ \max_i I_i[(L_P)_{J \rightarrow w}] \right\} \leq \tau.$$

We take the set J from the previous claim. We next restrict the function f' to inputs x such that $x|_J = w$ for some $w \in \Sigma^J$. Denote the restricted function by $f'_{J \rightarrow w}$. We consider the following four events regarding w :

1. E_1 : all the influences of $L_P|_{J \rightarrow w}$ are at most τ for every $P \in \text{span}_{\mathbb{N}}(\mathcal{P})$.
2. E_2 : the ℓ_2 norms of the restricted low-degree functions $L_P|_{J \rightarrow w}$ are bounded by C' where $C' = 100C2^{O_m(r)}$.
3. E_3 : $\|f'|_{J \rightarrow w} - f|_{J \rightarrow w}\|_2 \leq \sqrt{\xi}$.
4. E_4 : $\|F_{J \rightarrow w} - f|_{J \rightarrow w}\|_2 \leq \sqrt{\xi}$.

The following claim shows that with high probability over $w \sim \Sigma^J$, all the events E_1, E_2, E_3 , and E_4 hold.

Claim 5.18. $\Pr_{w \sim \Sigma^J} [E_1 \wedge E_2 \wedge E_3 \wedge E_4] \geq 0.98$.

Proof. Consider a function L_P where $P \in \text{span}_{\mathbb{N}}(\mathcal{P})$. As $\|L_P\| \leq C$, by Markov's inequality the probability over the choice of w that $\|L_P|_{J \rightarrow w}\|_2 \geq C'$ is at most $\frac{C^2}{C'^2}$. As the number of the functions L_P in the expansion of F is upper bounded by $|\text{span}_{\mathbb{N}}(\mathcal{P})| \leq 2^{O_m(r)}$, by union bound, the event E_2 holds with probability at least $1 - \frac{3C^2 2^{O_m(r)}}{C'^2} \geq 0.01$. Using Claim 5.17, E_1 holds with probability at least $1 - \eta$. As $\|f' - f\|_2^2 \leq \xi^2$, Markov's inequality gives that the probability that $\|f'|_{J \rightarrow w} - f|_{J \rightarrow w}\|_2^2 \geq \xi$ is at most ξ . Similarly, $\Pr[\neg E_4] \leq \xi$ as $\|F - f\|_2 \leq \xi$. Therefore, by the union bound E_1, E_2, E_3 and E_4 hold together with probability at least $1 - \eta - 2\xi - 0.01 \geq 0.98$. $\square$

Fix J from above and let E_w be the indicator that the events E_1, E_2, E_3 and E_4 happen for the given w . Recall, $\Delta_2(f')$,

$$\Delta_2(f') = \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n}} [f'(x)f'(y)f'(z)].$$

We can rewrite the above expectations as follows.

$$\begin{aligned} \Delta_2(f') &= \mathbb{E}_{(w,w',w'') \sim \mu^J} \left[\mathbb{E}_{(x,y,z) \sim \mu^{\bar{J}}} [f'|_{J \rightarrow w}(x)f'|_{J \rightarrow w'}(y)f'|_{J \rightarrow w''}(z)] \right] \\ &\geq 0.98 \cdot \alpha^{|J|} \mathbb{E}_{w \sim \nu^J} \left[\mathbb{E}_{(x,y,z) \sim \mu^{\bar{J}}} [f'|_{J \rightarrow w}(x)f'|_{J \rightarrow w}(y)f'|_{J \rightarrow w}(z)] \middle| E_w \right] \\ &\geq \alpha^{|J|+1} \mathbb{E}_{w \sim \nu^J} \left[\underbrace{\mathbb{E}_{(x,y,z) \sim \mu^{\bar{J}}} [f'|_{J \rightarrow w}(x)f'|_{J \rightarrow w}(y)f'|_{J \rightarrow w}(z)]}_{\Delta_2(f', E_w)} \middle| E_w \right] \end{aligned}$$

where in the second step, we use Claim 5.18 and the fact that $w = w' = w''$ happens with probability at least $\alpha^{|J|}$. Next, we estimate $\Delta_2(f', E_w)$ defined above.

5.2.4 Moving to a Decoupled Function

We will fix the set J and w for the remainder of this section, and assume that E_w holds. Let us denote $n' = n - |J|$. For succinctness, we denote $f'_{J \rightarrow w}$, $f_{J \rightarrow w}$ and $F_{|J \rightarrow w}$ simply by f'_w , f_w and F_w , respectively. The next step is to decouple the input from F_w to the product functions and the low-degree functions. Towards this, for a given function $F_w : (\Sigma^{n'}, \nu^{\otimes n'}) \rightarrow \mathbb{C}$, define a function $F_{d,w} : (\Sigma^{n'} \times \Sigma^{n'}, \nu^{\otimes n'} \times \nu^{\otimes n'}) \rightarrow \mathbb{C}$

$$F_{d,w}(x', x'') = \sum_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} P(w, x') \cdot L_P(w, x'').$$

We also truncate the functions so that they remain bounded. Towards this, consider the following functions,

$$\tilde{F}_{d,w}(x', x'') := \text{trunc}_{[0,1]}(F_{d,w}(x', x'')),$$

and

$$\tilde{F}_w(x) = \text{trunc}_{[0,1]}(F_w(x)).$$

The following lemma from [BKM26] allows one to replace the function F_w with $F_{d,w}$ by changing the distribution $X \sim \nu^{\otimes n'}$ to a coupled distribution $(X', X'') \sim \nu^{\otimes n'} \times \nu^{\otimes n'}$ as long as $\mathcal{P}$ has high rank. The decoupled distribution $\mathcal{D}$ is as follows (recall from (9) that we choose κ such that $\rho_2 \ll \kappa \ll \rho_1$):

1. Sample $(X, Y, Z) \sim \mu^{\otimes n'}$.
2. Sample $(X', Y', Z') \sim T_{\mathcal{P},0}(X, Y, Z)$.
3. Sample $(X'', Y'', Z'') \sim T_{1-\kappa}(X, Y, Z)$. By that, we mean that for each coordinate $i \in \bar{J}$ we have $(X''_i, Y''_i, Z''_i) = (X_i, Y_i, Z_i)$ with probability $1 - \kappa$ and we take $(X''_i, Y''_i, Z''_i) \sim \mu$ with probability κ independently.

We have the following lemma from [BKM26] adapted to our setting above.

Lemma 5.19. ([BKM26, Lemma 6.7]) *Let $D > 0$ and $\mathcal{P} = \{P_1, \dots, P_r : \Sigma^{n'} \rightarrow \mathbb{C}\}$ be a collection of product functions from the decomposed function F_w such that $\text{rk}(\mathcal{P}) \geq D$. Then for every $\kappa > 0$, the coupling $\mathcal{D}$ is a coupling between (X', X'') and X such that*

$$\mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[|F_{d,w}(X', X'') - F_w(X)|^2 \right] \leq O_{\rho_0, \rho_1, |\Sigma|}(\kappa) + 2^{-\Omega_{\rho_1, \rho_1, |\Sigma|}(\kappa(D))}.$$

Towards estimating $\Delta_2(f', E_w)$, we change the function f'_w to $F_{d,w}$. However, for what comes next, after fixing the input to the product functions in $F_{d,w}$, we require that the expectation involves the same function on the second input and that the function has non-negligible measure. Therefore, we do the transformation more carefully by conditioning on certain events defined below.

For a fixed $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$, define the following sets

$$S_{w, \vec{b}} = \{x \in \Sigma^{n'} \mid P(w, x) = b_P \text{ for every } P \in \mathcal{P}\}.$$

Also, for a given $x \in \Sigma^{\bar{J}}$, define the vector $\vec{b}(x)$ where $b_P(x) = P(w, x)$ for all $P \in \mathcal{P}$.

Abusing the notation, for a vector $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$, we use the following function in the subsequent analysis.

$$F_{d,w}(\vec{b}, x'') := \sum_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} b_P \cdot L_P(w, x'').$$

where for $P' \in \text{span}_{\mathbb{N}}(\mathcal{P})$ with $P'(x) = \prod_{P \in \mathcal{P}} P(x)^{\alpha_P}$, we have $b_{P'} = \prod_{P \in \mathcal{P}} b_P^{\alpha_P}$. Also, let $\tilde{F}_{d,w}(\vec{b}, x'') = \text{trunc}_{[0,1]}(F_{d,w}(\vec{b}, x''))$.

Suppose we sample (X, Y, Z) , (X', Y', Z') and (X'', Y'', Z'') from the decoupled distribution $\mathcal{D}$ defined above. Consider the following event $E' = E'_1 \wedge E'_2$ on the string X where events E'_i are defined as follows for $\vec{b} = \vec{b}(X)$.

1. E'_1 : $\mathbb{E}_{x \sim \nu^{\otimes n'}} [f'(w, x) \mid x \in S_{w, \vec{b}}] \geq \frac{\delta}{32}$.
2. E'_2 : $\mathbb{E}_{(x, x', x'') \sim \mathcal{D}} [|\tilde{F}_{d,w}(\vec{b}, x'') - f'_w(x)|^2 \mid x \in S_{w, \vec{b}}] \leq \xi^{1/4}$.

The event E'_1 says that the measure of f'_w remains large when restricted to inputs x such that $x \in S_{w, \vec{b}}$. The event E'_2 shows the ℓ_2 closeness of the decoupled function $\tilde{F}_{d,w}(\vec{b}, x'')$ and $f'_w(x)$, again conditioned on $x \in S_{w, \vec{b}}$.

The following claim shows that E' occurs with non-negligible probability.

Claim 5.20. *If we sample (X, Y, Z) , (X', Y', Z') and (X'', Y'', Z'') from the decoupled distribution $\mathcal{D}$, then the probability that E' happens is at least $\frac{\delta}{32} - \xi^{1/4} \geq \frac{\delta}{64}$.*

Proof. As $\|f - f\|_{\nu} \leq \xi$, we have $\mathbb{E}_{x \sim \nu^{\otimes n}} [f(x)] \geq \frac{\delta}{2}$. Since f is $(B, 1/4)$ -resilient (Claim 5.15), we have $\mathbb{E}_{x \sim \nu^{\otimes n'}} [f(w, x)] \geq \frac{\delta}{8}$. Furthermore, since $\|f'_w - f_w\|_2 \leq \sqrt{\xi}$ (because of w satisfying event E_3), we get $\mathbb{E}_{x \sim \nu^{\otimes n'}} [f'(w, x)] \geq \frac{\delta}{8} - \xi \geq \frac{\delta}{16}$. Note that f' is a 1-bounded function. Therefore, if we sample $(X, Y, Z) \sim \mu^{\otimes n'}$, and letting $\vec{b} = \vec{b}(X)$, then with probability at least $\frac{\delta}{32}$, we have $\mathbb{E}_{x \sim \nu^{\otimes n'}} [f'(w, x) \mid x \in S_{w, \vec{b}}] \geq \frac{\delta}{32}$.

In the rest of the proof, we estimate $\Pr[E'_2]$. Before doing that, let us first upper bound the quantity $\mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(X', X'') - f'_w(X) \right|^2 \right]$. We have,

$$\begin{aligned} & \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(X', X'') - f'_w(X) \right|^2 \right] \\ &= \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(X', X'') - f_w(X) + f_w(X) - f'_w(X) \right|^2 \right] \\ &\lesssim \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(X', X'') - f_w(X) \right|^2 \right] + \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| f_w(X) - f'_w(X) \right|^2 \right]. \end{aligned}$$

Since w satisfies the event E_3 , the second expectation is upper bounded by ξ . We now bound the first expectation,

$$\begin{aligned} & \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(X', X'') - f_w(X) \right|^2 \right] \\ &= \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(X', X'') - F_w(X) + F_w(X) - f_w(X) \right|^2 \right] \\ &\lesssim \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(X', X'') - F_w(X) \right|^2 \right] + \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| F_w(X) - f_w(X) \right|^2 \right]. \end{aligned}$$

Again, since w satisfies the event E_4 , the second expectation is upper bounded by ξ . As for the first expectation,

$$\begin{aligned}
& \mathbb{E}_{X, X', X''} \left[\left| \tilde{F}_{d,w}(X', X'') - F_w(X) \right|^2 \right] \\
&= \mathbb{E}_{X, X', X''} \left[\left| \tilde{F}_{d,w}(X', X'') - F_{d,w}(X', X'') + F_{d,w}(X', X'') - F_w(X) \right|^2 \right] \\
&\lesssim \mathbb{E}_{X, X', X''} \left[\left| \tilde{F}_{d,w}(X', X'') - F_{d,w}(X', X'') \right|^2 \right] + \mathbb{E}_{X, X', X''} \left[\left| F_{d,w}(X', X'') - F_w(X) \right|^2 \right]. \quad (12)
\end{aligned}$$

As for the first term, using the property that $\eta(a) = |a - \text{trunc}_{[0,1]}(a)|$ is 2-Lipshitz from Fact 5.10,

$$\begin{aligned}
& \mathbb{E}_{X, X', X''} \left[\left| \tilde{F}_{d,w}(X', X'') - F_{d,w}(X', X'') \right|^2 \right] \\
&= \mathbb{E}_{X, X', X''} \left[\eta(F_{d,w}(X', X''))^2 \right] \\
&\lesssim \mathbb{E}_{X, X', X''} \left[\eta(f_w(X))^2 + |F_{d,w}(X', X'') - f_w(X)|^2 \right] \\
&= \mathbb{E}_{X, X', X''} \left[|F_{d,w}(X', X'') - f_w(X)|^2 \right] \quad (\text{Using } \eta(f_w(X)) = 0) \\
&\leq \mathbb{E}_{X, X', X''} \left[|F_{d,w}(X', X'') - F_w(X) + F_w(X) - f_w(X)|^2 \right] \\
&\lesssim \mathbb{E}_{X, X', X''} \left[|F_{d,w}(X', X'') - F_w(X)|^2 \right] + \mathbb{E}_{X, X', X''} \left[|F_w(X) - f_w(X)|^2 \right] \\
&\leq \rho'_2 + \xi,
\end{aligned}$$

where $\rho'_2 \ll \xi$ if $\rho_2 \ll \rho_1, \rho_0$. Here, we used Lemma 5.19 to bound the first expectation by $O_{\rho_0, \rho_1, |\Sigma|}(\kappa) + 2^{-\Omega_{\rho_0, \rho_1, |\Sigma|, \kappa}(1/\rho_2)}$ which is at most ρ'_2 (by choosing small enough κ and $\rho_2 \ll \kappa$ and noting that $\text{rk}(\mathcal{P}|_{\bar{J}}) \geq \text{rk}(\mathcal{P})/2$), and the second expectation is bounded by ξ because of the event E_4 . The second expectation from (12) is at most ρ'_2 as we already bounded the same expectation above. Therefore, we get,

$$\mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(X', X'') - f'_w(X) \right|^2 \right] \lesssim \xi + \rho'_2. \quad (13)$$

Now suppose that for with probability at least $\xi^{1/4}$ over $X \sim \nu^{\otimes n'}$ we have

$$\mathbb{E}_{(x, x', x'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(\vec{b}, x'') - f'_w(x) \right|^2 \mid x \in S_{w, \vec{b}(X)} \right] \geq \xi^{1/4}.$$

This would imply

$$\mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \tilde{F}_{d,w}(X', X'') - f'_w(X) \right|^2 \right] \geq \xi^{1/2} \gg \xi + \rho'_2,$$

contradicting (13). Thus, $\Pr[E'_2] \geq 1 - \xi^{1/4}$.

Therefore, by the union bound, we have $\Pr[E'_1 \wedge E'_2] \geq \frac{\delta}{32} - \xi^{1/4}$, as required. $\square$

Let $W_{\vec{b}}$ be the event that $(X, Y, Z) \sim \mu^{\otimes n'}$ are such that $X, Y, Z \in S_{w, \vec{b}}$. We have the following claim that shows the closeness of the distribution $(X'', Y'', Z'')|W_{\vec{b}}$ to the distribution $\mu^{\otimes n'}$, provided $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$.

Claim 5.21. Fix any $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$. In the decoupled distribution, the conditional distribution on $(X'', Y'', Z'') | W_{\vec{b}}$ is $\tilde{\rho}_2$ -close to the distribution $\mu^{\otimes n'}$ where $\tilde{\rho}_2 = 2^{-\Omega_{\alpha, \kappa, r, M}(1/\rho_2)}$.

Proof. Fix any subset $A \subseteq (\Sigma^{n'})^3$. The measure of A under $\mu^{\otimes n'}$ is simply $\mu^{\otimes n'}(A)$. We now compute the probability that $(X'', Y'', Z'') \in A$ conditioned on the event $W_{\vec{b}}$. Let $S = S_{w, \vec{b}} \times S_{w, \vec{b}} \times S_{w, \vec{b}}$.

$$\Pr_{\substack{(X, Y, Z) \sim \mu^{\otimes n'} \\ (X'', Y'', Z'') \sim T_{1-\kappa}(X, Y, Z)}} [(X'', Y'', Z'') \in A \mid (X, Y, Z) \in S] = \frac{1}{\mu^{\otimes n'}(S)} \langle T_{1-\kappa} 1_S, 1_A \rangle_{L^2(\mu)}.$$

Subtracting $\mu^{\otimes n'}(A) = \frac{\mu^{\otimes n'}(S)}{\mu^{\otimes n'}(S)} \langle 1, 1_A \rangle_{L^2(\mu)}$ and taking absolute value, we get

$$\begin{aligned} \left| \frac{1}{\mu^{\otimes n'}(S)} \langle T_{1-\kappa} 1_S, 1_A \rangle_{L^2(\mu)} - \mu^{\otimes n'}(A) \right| &= \frac{1}{\mu^{\otimes n'}(S)} \langle T_{1-\kappa} 1_S - \mu^{\otimes n'}(S), 1_A \rangle_{L^2(\mu)} \\ &\leq \frac{1}{\mu^{\otimes n'}(S)} \|T_{1-\kappa} 1_S - \mu^{\otimes n'}(S)\|_{L^2(\mu)}. \end{aligned}$$

Using Lemma 3.8, we have $\|T_{1-\kappa} 1_S - \mu^{\otimes n'}(S)\|_{L^2(\mu)} \leq 2^{-\Omega_{\alpha, \kappa, r, M}(1/\rho_2)}$. For the denominator, since $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$, using Lemma 3.7, we have $\mu^{\otimes n'}(S) \geq \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)^3} - 2^{-\Omega_{M, r, \alpha}(1/\rho_2)} = \Omega_{r, M}(1)$.

Thus,

$$\left| \Pr_{\substack{(X, Y, Z) \sim \mu^{\otimes n'} \\ (X'', Y'', Z'') \sim T_{1-\kappa}(X, Y, Z)}} [(X'', Y'', Z'') \in A \mid (X, Y, Z) \in S] - \mu^{\otimes n'}(A) \right| \leq 2^{-\Omega_{\alpha, \kappa, r, M}(1/\rho_2)}.$$

□

Consider the distribution $\mathcal{B}$ on the vectors $\vec{b}$ where we first sample $X \sim \nu^{\otimes n'}$ and then output $\vec{b}(X)$. We have

$$\begin{aligned} \Delta_2(f', E_w) &= \mathbb{E}_{(x, y, z) \sim \mu^{n'}} [f'_w(x) f'_w(y) f'_w(z)] \\ &\geq \mathbb{E}_{\vec{b} \sim \mathcal{B}} \left[\Pr[W_{\vec{b}}] \cdot 1_{\vec{b} \text{ satisfies } E'} \cdot \mathbb{E}_{(x, y, z) \sim \mu^{n'}} [f'_w(x) f'_w(y) f'_w(z) \mid W_{\vec{b}}] \right] \quad (f' \text{ in non-negative}) \\ &\geq \mathbb{E}_{\vec{b} \sim \mathcal{B}} \left[\Pr[W_{\vec{b}}] \cdot 1_{\vec{b} \text{ satisfies } E'} \cdot \left(\mathbb{E}_{\substack{(x, y, z) \\ (x', y', z') \sim \mathcal{D} \\ (x'', y'', z'')}} [\tilde{F}_{d, w}(\vec{b}, x'') \tilde{F}_{d, w}(\vec{b}, y'') \tilde{F}_{d, w}(\vec{b}, z'') \mid W_{\vec{b}}] - 3\xi^{1/8} \right) \right]. \end{aligned}$$

In the last inequality, for $\vec{b}$ satisfying the event E' , we used

$$\begin{aligned}
& \mathbb{E}_{(x,y,z) \sim \mu^{n'}} [f'_w(x)f'_w(y)f'_w(z) \mid W_{\vec{b}}] - \mathbb{E}_{\substack{(x,y,z) \\ (x',y',z') \sim \mathcal{D} \\ (x'',y'',z'')}} [\tilde{F}_{d,w}(\vec{b}, x'')f'_w(y)f'_w(z) \mid W_{\vec{b}}] \\
&= \mathbb{E}_{\substack{(x,y,z) \\ (x',y',z') \sim \mathcal{D} \\ (x'',y'',z'')}} [(f'_w(x) - \tilde{F}_{d,w}(\vec{b}, x''))f'_w(y)f'_w(z) \mid W_{\vec{b}}] \\
&\leq \mathbb{E}_{\substack{(x,y,z) \\ (x',y',z') \sim \mathcal{D} \\ (x'',y'',z'')}} \left[|f'_w(x) - \tilde{F}_{d,w}(\vec{b}, x'')| \mid W_{\vec{b}} \right] \quad (f' \text{ is } [0, 1] \text{ bounded}) \\
&= \mathbb{E}_{\substack{(x,y,z) \\ (x',y',z') \sim \mathcal{D} \\ (x'',y'',z'')}} \left[|f'_w(x) - \tilde{F}_{d,w}(\vec{b}, x'')| \mid x \in S_{w,\vec{b}} \right] \\
&\leq \mathbb{E}_{(x,x',x'')} \left[|f'_w(x) - \tilde{F}_{d,w}(\vec{b}, x'')|^2 \mid x \in S_{w,\vec{b}} \right]^{1/2}. \quad (\text{Cauchy-Schwarz})
\end{aligned}$$

As $\vec{b}$ under consideration satisfies the event E'_2 , the last expectation is upper bounded by $\xi^{1/4}$. Repeating the same thing for the y and the z variables, and using $[0, 1]$ -boundedness of f'_w and $\tilde{F}_{d,w}$, we get the required bound.

Using Lemma 3.7, for every $\vec{b} \in \prod_{P \in \mathcal{P}} \text{Image}(P)$, $\Pr[W_{\vec{b}}]$ is lower bounded by

$$\Pr[W_{\vec{b}}] \geq \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)^3} - 2^{-\Omega_{\rho_0, m, \alpha}(1/\rho_2)} \geq \frac{1}{2} \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)^3}.$$

Thus, we fix an arbitrary $\vec{b}_\star$ for which the event E' occurs. Note that, such a $\vec{b}_\star$ is in $\prod_{P \in \mathcal{P}} \text{Image}(P)$ by definition. Using Lemma 3.5, we have

$$\Pr_{\vec{b} \sim \mathcal{B}} [\vec{b} = \vec{b}_\star] \geq \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)} - 2^{-\Omega_{\rho_0, m, \alpha}(1/\rho_2)} \geq \frac{1}{2} \prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)}.$$

We get

$$\begin{aligned}
\Delta_2(f', E_w) &\geq \mathbb{E}_{\vec{b} \sim \mathcal{B}} \left[\Pr[W_{\vec{b}}] \cdot \mathbf{1}_{\vec{b} \text{ satisfies } E'} \cdot \left(\mathbb{E}_{\substack{(x,y,z) \\ (x',y',z') \sim \mathcal{D} \\ (x'',y'',z'')}} [\tilde{F}_{d,w}(\vec{b}, x'')\tilde{F}_{d,w}(\vec{b}, y'')\tilde{F}_{d,w}(\vec{b}, z'') \mid W_{\vec{b}}] - 3\xi^{1/8} \right) \right] \\
&\geq \frac{1}{4} \left(\prod_{P \in \mathcal{P}} \frac{1}{\text{ord}(P)} \right)^4 \cdot \left(\mathbb{E}_{\substack{(x,y,z) \\ (x',y',z') \sim \mathcal{D} \\ (x'',y'',z'')}} [\tilde{F}_{d,w}(\vec{b}_\star, x'')\tilde{F}_{d,w}(\vec{b}_\star, y'')\tilde{F}_{d,w}(\vec{b}_\star, z'') \mid W_{\vec{b}_\star}] - 3\xi^{1/8} \right).
\end{aligned}$$

We now give a lower bound on the above expectation in the next section.

5.2.5 Applying the Invariance Principle

From the last section, our goal is to estimate the following quantity:

$$\Delta_3(\tilde{F}_{d,w}, \vec{b}_\star) := \mathbb{E}_{\substack{(x,y,z) \\ (x',y',z') \sim \mathcal{D} \\ (x'',y'',z'')}} \left[\tilde{F}_{d,w}(\vec{b}_\star, x'') \tilde{F}_{d,w}(\vec{b}_\star, y'') \tilde{F}_{d,w}(\vec{b}_\star, z'') \mid W_{\vec{b}_\star} \right],$$

where $\vec{b}_\star$ is such that it satisfies the event E' . Using Claim 5.21 and 1-boundedness of the function $\tilde{F}_{d,w}$, we have

$$\Delta_3(\tilde{F}_{d,w}, \vec{b}_\star) \geq \mathbb{E}_{(x'',y'',z'') \sim \mu^{\otimes n'}} \left[\tilde{F}_{d,w}(\vec{b}_\star, x'') \tilde{F}_{d,w}(\vec{b}_\star, y'') \tilde{F}_{d,w}(\vec{b}_\star, z'') \right] - \tilde{\rho}_2. \quad (14)$$

Furthermore, since $\vec{b}_\star$ satisfies the events E'_1 and E'_2 , we have

$$\mathbb{E}_{(x,x') \sim \nu^{\otimes n'}} \left[\tilde{F}_{d,w}(\vec{b}_\star, x'') \mid x \in S_{w,\vec{b}_\star} \right] \geq \frac{\delta}{64} - \xi^{1/4}.$$

Again, using Claim 5.21 and the 1-boundedness of the function $\tilde{F}_{d,w}$, we get

$$\mathbb{E}_{x'' \sim \nu^{\otimes n'}} \left[\tilde{F}_{d,w}(\vec{b}_\star, x'') \right] \geq \frac{\delta}{64} - \xi^{1/4} - \tilde{\rho}_2 \geq \frac{\delta}{128}. \quad (15)$$

In the following claim, we use the invariance principle to lower bound the expectation from (14).

Claim 5.22. *For any $\vec{b}$ that satisfies the event E' ,*

$$\mathbb{E}_{(x'',y'',z'') \sim \mu^{\otimes n'}} \left[\tilde{F}_{d,w}(\vec{b}, x'') \tilde{F}_{d,w}(\vec{b}, y'') \tilde{F}_{d,w}(\vec{b}, z'') \right] \geq \delta^{\Omega_{m,\alpha}(1)}.$$

Proof. Consider the function $F_{d,w}(\vec{b}, \cdot) : (\Sigma^{n'}, \nu^{\otimes n'}) \rightarrow \mathbb{C}$ and its decomposition from (8). Every function L_P in the decomposition of the function F_w , we have $I_i(L_P) \leq \tau$ for every $i \in J$. As the function $F_{d,w}(\vec{b}, \cdot)$ is the weighted sum of these functions L_P for $P \in \mathcal{P}$, we can express $F_{d,w}(\vec{b}, \cdot) = \sum_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} \theta_P \cdot L_P$ where $|\theta_P| = 1$. Applying Proposition 5.6, as $|\text{span}_{\mathbb{N}}(\mathcal{P})| = O_r(1)$, we get that $I_i(F_{d,w}(\vec{b}, \cdot)) \leq O_r(\tau) = O_{\rho_1}(\tau)$ for all $i \in J$.

Let $\Psi : \mathbb{C}^3 \rightarrow \mathbb{C}$ be any smooth extension of the function $\Psi(a, b, c) = abc$ for a, b, c that have absolute value at most 1, by which we mean it has third order derivatives bounded by $O(1)$ and additionally it satisfies

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

for all complex numbers $a, b, c, a', b', c' \in \mathbb{C}$. Let $\mathcal{X} = \{X_1, \dots, X_{m-1}, Y_1, \dots, Y_{m-1}, Z_1, \dots, Z_{m-1}\}$ and $\mathcal{G} = \{G_{1,x}, \dots, G_{m-1,x}, \dots, G_{1,z}, \dots, G_{m-1,z}\}$ be the ensemble of random variables and the Gaussian random variables for the distribution μ , as defined in Section 5.1.3. Let $X = (\{X_1, \dots, X_{m-1}\}, Y = (Y_1, \dots, Y_{m-1}), Z = (Z_1, \dots, Z_{m-1})$ and $G_x = (G_{1,x}, \dots, G_{m-1,x}), G_y = (G_{1,y}, \dots, G_{m-1,y}), G_z = (G_{1,z}, \dots, G_{m-1,z})$. We now apply the invariance principle, Theorem 5.9, to the expectation under consideration to move to the Gaussian space. Towards this, we work with the multi-linear representation $M_{d,w} : \mathbb{C}^{(m-1)n'} \rightarrow \mathbb{C}$ of the function $F_{d,w}$. Also, let $\tilde{M}_{d,w} = \text{trunc}_{[0,1]}(M_{d,w})$ be its truncated version.

Consider the following quantities

$$\begin{aligned}
\Lambda_1 &= \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n'}} \left[\Psi \left(\tilde{F}_{d,w}(\vec{b}, x), \tilde{F}_{d,w}(\vec{b}, y), \tilde{F}_{d,w}(\vec{b}, z) \right) \right] \\
\Lambda'_1 &= \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n'}} \left[\Psi \left(F_{d,w}(\vec{b}, x), F_{d,w}(\vec{b}, y), F_{d,w}(\vec{b}, z) \right) \right] \\
\Lambda_2 &= \mathbb{E}_{\mathcal{X}} \left[\Psi \left(M_{d,w}(\vec{b}, X), M_{d,w}(\vec{b}, Y), M_{d,w}(\vec{b}, Z) \right) \right] \\
\Lambda_3 &= \mathbb{E}_{\mathcal{G}} \left[\Psi \left(M_{d,w}(\vec{b}, G_x), M_{d,w}(\vec{b}, G_y), M_{d,w}(\vec{b}, G_z) \right) \right] \\
\Lambda'_3 &= \mathbb{E}_{\mathcal{G}} \left[\Psi \left(\tilde{M}_{d,w}(\vec{b}, G_x), \tilde{M}_{d,w}(\vec{b}, G_y), \tilde{M}_{d,w}(\vec{b}, G_z) \right) \right]
\end{aligned}$$

Note that $\Lambda'_1 = \Lambda_2$, as in Λ_2 , we just moved to the multi-linear representation of $F_{d,w}$. We have the following claims.

Claim 5.23. $|\Lambda_1 - \Lambda'_1| \leq O_{\rho_1}(\tilde{\rho}_2) + O(\xi^{1/4})$.

Proof. Using the smoothness property of Ψ , we have

$$\begin{aligned}
&|\Lambda_1 - \Lambda'_1| \\
&\lesssim \mathbb{E}_{(x,y,z) \sim \mu^{\otimes n'}} \left[\sqrt{\left| \tilde{F}_{d,w}(\vec{b}, x) - F_{d,w}(\vec{b}, x) \right|^2 + \left| \tilde{F}_{d,w}(\vec{b}, y) - F_{d,w}(\vec{b}, y) \right|^2 + \left| \tilde{F}_{d,w}(\vec{b}, z) - F_{d,w}(\vec{b}, z) \right|^2} \right].
\end{aligned}$$

Now, using the fact that the marginal distribution of μ on each coordinate is μ , by the Cauchy-Schwarz inequality,

$$|\Lambda_1 - \Lambda'_1| \lesssim \sqrt{\mathbb{E}_{x \sim \nu^{\otimes n'}} \left[\left| \tilde{F}_{d,w}(\vec{b}, x) - F_{d,w}(\vec{b}, x) \right|^2 \right]}. \quad (16)$$

Recall the decoupled distribution $\mathcal{D}$ from Lemma 5.19. We move from the distribution $\nu^{\otimes n'}$ to the marginal of the decoupled distribution as follows. Using Claim 5.21, the statistical distance between the distribution on X'' in $(X, X', X'') \sim \mathcal{D}|W_{\vec{b}}$ and the distribution $x \sim \nu^{\otimes n'}$ is at most $\tilde{\rho}_2$. Thus, we may couple X'' in (X, X', X'') with x distributed according to $\nu^{\otimes n'}$ such that $\Pr[X'' \neq x] \leq \tilde{\rho}_2$. Let E be the event that $x \neq X''$ in the coupling. We have,

$$\begin{aligned}
&\left| \mathbb{E}_{x \sim \nu^{\otimes n'}} \left[\left| \tilde{F}_{d,w}(\vec{b}, x) - F_{d,w}(\vec{b}, x) \right|^2 \right] - \mathbb{E}_{(X, X', X'') \sim \mathcal{D}|W_{\vec{b}}} \left[\left| \tilde{F}_{d,w}(\vec{b}, X'') - F_{d,w}(\vec{b}, X'') \right|^2 \right] \right|^2 \\
&= \mathbb{E}_{(X, X', X'') \sim \mathcal{D}|W_{\vec{b}}} \left[\left(\left| \tilde{F}_{d,w}(\vec{b}, x) - F_{d,w}(\vec{b}, x) \right|^2 - \left| \tilde{F}_{d,w}(\vec{b}, X'') - F_{d,w}(\vec{b}, X'') \right|^2 \right) 1_E \right]^2 \\
&\leq \Pr[E] \cdot \mathbb{E}_{(X, X', X'') \sim \mathcal{D}|W_{\vec{b}}} \left[\left(\left| \tilde{F}_{d,w}(\vec{b}, x) - F_{d,w}(\vec{b}, x) \right|^2 - \left| \tilde{F}_{d,w}(\vec{b}, X'') - F_{d,w}(\vec{b}, X'') \right|^2 \right)^2 \right] \\
&\leq \Pr[E] \cdot \mathbb{E}_{(X, X', X'') \sim \mathcal{D}|W_{\vec{b}}} \left[\left(\left| \tilde{F}_{d,w}(\vec{b}, x) - F_{d,w}(\vec{b}, x) \right|^2 + \left| \tilde{F}_{d,w}(\vec{b}, X'') - F_{d,w}(\vec{b}, X'') \right|^2 \right)^2 \right] \\
&\lesssim \Pr[E] \cdot \left(\mathbb{E} \left[\left| \tilde{F}_{d,w}(\vec{b}, x) \right|^4 \right] + \mathbb{E} \left[\left| F_{d,w}(\vec{b}, x) \right|^4 \right] + \mathbb{E} \left[\left| \tilde{F}_{d,w}(\vec{b}, X'') \right|^4 \right] + \mathbb{E} \left[\left| F_{d,w}(\vec{b}, X'') \right|^4 \right] \right)
\end{aligned}$$

The first and the third expectations are bounded by 1. We bound the second expectation, and the last expectation can be bounded analogously. Note that $|\mathbf{F}_{d,w}(\vec{b}, x)| \leq \sum_P |L_P(w, x)|$. We now use the fact that the $\deg(L_P(w, \cdot))$ and $\|L_P(w, \cdot)\|_2$ are bounded by $O_{\rho_1}(1)$. By Hölder's inequality and hypercontractivity, we get

$$\mathbb{E} \left[|\mathbf{F}_{d,w}(\vec{b}, x)|^4 \right] \lesssim_r \sum_P \|L_P(w, \cdot)\|_4^4 \lesssim_{r, \rho_1} \|L_P(w, \cdot)\|_2^4 = O_{\rho_1}(1).$$

Thus, we get

$$\left| \mathbb{E}_{x \sim \nu^{\otimes n'}} \left[\left| \tilde{\mathbf{F}}_{d,w}(\vec{b}, x) - \mathbf{F}_{d,w}(\vec{b}, x) \right|^2 \right] - \mathbb{E}_{(X, X', X'') \sim \mathcal{D} | W_{\vec{b}}} \left[\left| \tilde{\mathbf{F}}_{d,w}(\vec{b}, X'') - \mathbf{F}_{d,w}(\vec{b}, X'') \right|^2 \right] \right| = O_{\rho_1}(\tilde{\rho}_2).$$

Hence,

$$|\Lambda_1 - \Lambda'_1| \lesssim \sqrt{\mathbb{E}_{(X, X', X'') \sim \mathcal{D} | W_{\vec{b}}} \left[\left| \tilde{\mathbf{F}}_{d,w}(\vec{b}, X'') - \mathbf{F}_{d,w}(\vec{b}, X'') \right|^2 \right]} + O_{\rho_1}(\tilde{\rho}_2).$$

Now, using the property that $\eta(a) = |a - \text{trunc}_{[0,1]}(a)|$ is 2-Lipshitz from Fact 5.10,

$$\begin{aligned} & \mathbb{E}_{(X, X', X'') \sim \mathcal{D} | W_{\vec{b}}} \left[\left| \tilde{\mathbf{F}}_{d,w}(\vec{b}, X'') - \mathbf{F}_{d,w}(\vec{b}, X'') \right|^2 \right] \\ &= \mathbb{E}_{(X, X', X'') \sim \mathcal{D} | W_{\vec{b}}} \left[\eta(\mathbf{F}_{d,w}(\vec{b}, X''))^2 \right] \\ &= \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\eta(\mathbf{F}_{d,w}(\vec{b}, X''))^2 \mid X \in S_{w, \vec{b}} \right] \\ &= \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\eta(\mathbf{f}'_w(X))^2 + \left| \mathbf{f}'_w(X) - \mathbf{F}_{d,w}(\vec{b}, X'') \right|^2 \mid X \in S_{w, \vec{b}} \right] \\ &= \mathbb{E}_{(X, X', X'') \sim \mathcal{D}} \left[\left| \mathbf{f}'_w(X) - \mathbf{F}_{d,w}(\vec{b}, X'') \right|^2 \mid X \in S_{w, \vec{b}} \right] \quad (\eta(\mathbf{f}'_w(X)) = 0) \end{aligned}$$

As $\vec{b}$ satisfies the vent E' (more specifically, the event E'_2), the above expectation is at most $\xi^{1/4}$. $\square$

Claim 5.24. $|\Lambda_2 - \Lambda_3| = O_{\rho_1}(\tau')$ where $\tau' \rightarrow 0$ as $\tau \rightarrow 0$.

Proof. This follows from a direct application of Theorem 5.9 by noting that the degree and the ℓ_2 norm of the multilinear function $\mathbf{M}_{d,w}(\vec{b}, \cdot)$ is upper bounded by $O_{\rho_1}(1)$. $\square$

Claim 5.25. $\Lambda'_3 \geq \delta^{\Omega_{m, \alpha}(1)}$.

Proof. Recall that from (16) we derived the following in the proof of Claim 5.23.

$$\mathbb{E}_{x \sim \nu^{\otimes n'}} \left[\left| \tilde{\mathbf{F}}_{d,w}(\vec{b}, x) - \mathbf{F}_{d,w}(\vec{b}, x) \right|^2 \right] = O_{\rho_1}(\tilde{\rho}_2) + O(\xi^{1/4}).$$

This along with the lower bound from (15), we get

$$\mathbb{E}_{x \sim \nu^{\otimes n'}} \left[\mathbf{F}_{d,w}(\vec{b}_\star, x) \right] \geq \frac{\delta}{128} - O_{\rho_1}(\tilde{\rho}_2) + O(\xi^{1/4}) \geq \frac{\delta}{256}.$$

By applying Theorem 5.9 again, we have

$$\left| \mathbb{E}_{\mathcal{X}} \left[\mathbf{M}_{d,w}(\vec{b}, X) \right] - \mathbb{E}_{\mathcal{G}} \left[\mathbf{M}_{d,w}(\vec{b}, G_x) \right] \right| \leq O_{\rho_1}(\tau'),$$

and thus,

$$\mathbb{E}_{\mathcal{G}} \left[\mathbf{M}_{d,w}(\vec{b}, G_x) \right] \geq \frac{\delta}{256} - O_{\rho_1}(\tau') \geq \frac{\delta}{512}.$$

The same holds for $\mathbb{E}_{G_y} \left[\mathbf{M}_{d,w}(\vec{b}, G_y) \right]$ and $\mathbb{E}_{G_z} \left[\mathbf{M}_{d,w}(\vec{b}, G_z) \right]$. Applying Theorem 5.9 (item 2.), we get

$$\left| \mathbb{E}_{\mathcal{X}} \left[\eta(\mathbf{M}_{d,w}(\vec{b}, X)) \right] - \mathbb{E}_{\mathcal{G}} \left[\eta(\mathbf{M}_{d,w}(\vec{b}, G_x)) \right] \right| \leq O_{\rho_1}(\tau').$$

As by the Cauchy-Schwarz inequality,

$$\mathbb{E}_{\mathcal{X}} \left[\eta(\mathbf{M}_{d,w}(\vec{b}, X)) \right]^2 \leq \mathbb{E}_{x \sim \nu^{\otimes n'}} \left[\left| \tilde{\mathbf{F}}_{d,w}(\vec{b}, x) - \mathbf{F}_{d,w}(\vec{b}, x) \right|^2 \right] = O_{\rho_1}(\tilde{\rho}_2) + O(\xi^{1/4}),$$

we get,

$$\mathbb{E}_{\mathcal{G}} \left[\eta(\mathbf{M}_{d,w}(\vec{b}, G_x)) \right]^2 = O_{\rho_1}(\tilde{\rho}_2) + O(\xi^{1/4}) + O_{\rho_1}(\tau'). \quad (17)$$

Thus, using the fact that $\mathbb{E}_{\mathcal{G}} \left[\mathbf{M}_{d,w}(\vec{b}, G_x) \right] \geq \frac{\delta}{512}$ and the above bound, we get

$$\mathbb{E}_{\mathcal{G}} \left[\tilde{\mathbf{M}}_{d,w}(\vec{b}, G_x) \right] \geq \frac{\delta}{1024},$$

Let T be the covariance matrix of the ensemble $\mathcal{G}$. In Proposition 5.28 below, we show that the matrix T satisfies $T - \eta_{\alpha, |\Sigma|} I \succeq 0$, for some constant $\eta_{\alpha, |\Sigma|} > 0$ that depends on α and $|\Sigma|$. Hence, using Theorem 5.11, we have

$$\mathbb{E}_{\mathcal{G}^{n'}} \left[\tilde{\mathbf{M}}_{d,w}(\vec{b}, G_x), \tilde{\mathbf{M}}_{d,w}(\vec{b}, G_y), \tilde{\mathbf{M}}_{d,w}(\vec{b}, G_z) \right] \geq \left(\frac{\delta^3}{1024^3} \right)^{1/\eta_{m, \alpha}}.$$

□

Claim 5.26. $|\Lambda_3 - \Lambda'_3| = O_{\xi}(1)$.

Proof. Using the smoothness property of Ψ , we have

$$\begin{aligned} & |\Lambda_3 - \Lambda'_3| \\ & \lesssim \mathbb{E}_{\mathcal{G}} \left[\sqrt{\left| \mathbf{M}_{d,w}(\vec{b}, G_x) - \tilde{\mathbf{M}}_{d,w}(\vec{b}, G_x) \right|^2 + \left| \mathbf{M}_{d,w}(\vec{b}, G_y) - \tilde{\mathbf{M}}_{d,w}(\vec{b}, G_y) \right|^2 + \left| \mathbf{M}_{d,w}(\vec{b}, G_z) - \tilde{\mathbf{M}}_{d,w}(\vec{b}, G_z) \right|^2} \right]. \end{aligned}$$

Using the fact that $\sqrt{|a|^2 + |b|^2 + |c|^2} \leq (|a| + |b| + |c|)$ for all $a, b, c \in \mathbb{C}$, we get

$$\begin{aligned} & |\Lambda_3 - \Lambda'_3| \\ & \lesssim \mathbb{E}_{\mathcal{G}} \left[\left| \mathbf{M}_{d,w}(\vec{b}, G_x) - \tilde{\mathbf{M}}_{d,w}(\vec{b}, G_x) \right| + \left| \mathbf{M}_{d,w}(\vec{b}, G_y) - \tilde{\mathbf{M}}_{d,w}(\vec{b}, G_y) \right| + \left| \mathbf{M}_{d,w}(\vec{b}, G_z) - \tilde{\mathbf{M}}_{d,w}(\vec{b}, G_z) \right| \right]. \end{aligned}$$

From (17), we have

$$\mathbb{E}_{\mathcal{G}} \left[\left| \mathbf{M}_{d,w}(\vec{b}, G_x) - \tilde{\mathbf{M}}_{d,w}(\vec{b}, G_x) \right| \right]^2 = \mathbb{E}_{\mathcal{G}} \left[\eta(\mathbf{M}_{d,w}(\vec{b}, G_x)) \right]^2 = O_{\rho_1}(\tilde{\rho}_2) + O(\xi^{1/4}) + O_{\rho_1}(\tau') = O_{\xi}(1).$$

We have the same bound for the other terms in the upper bound of $|\mathbf{\Lambda}_3 - \mathbf{\Lambda}'_3|$ above, and hence

$$|\mathbf{\Lambda}_3 - \mathbf{\Lambda}'_3| = O_{\xi}(1). \quad \square$$

Using Claims 5.23, 5.24, 5.26, and 5.25, we have the following.

$$\begin{aligned} \mathbb{E}_{(x'', y'', z'') \sim \mu^{\otimes n'}} \left[\tilde{\mathbf{F}}_{d,w}(\vec{b}, x'') \tilde{\mathbf{F}}_{d,w'}(\vec{b}, y'') \tilde{\mathbf{F}}_{d,w''}(\vec{b}, z'') \right] &\geq \delta^{\Omega_{m,\alpha}(1)} - O_{\rho_1}(\tilde{\rho}_2) - O_{\rho_1}(\tau') - O_{\xi}(1) \\ &\geq \delta^{\Omega_{m,\alpha}(1)}, \end{aligned}$$

as required, where the last inequality follows from the relations (9). $\square$

We are now ready to prove the main theorem for resilient functions.

Lemma 5.27. *Suppose $f : \Sigma^n \rightarrow \{0, 1\}$ is a $(B, 1/2)$ resilient function for $B \gg \Omega_{\rho_1}(1)$, then $\Delta_1(f) \geq \alpha^{B+1} \Omega_{\rho_0}(\delta^{\Omega_{m,\alpha}(1)})$.*

Proof. Apply the arithmetic regularity lemma (Lemma 4.4) to the function f and the distribution μ for $\xi \ll \delta, \alpha, m^{-1}$. We use the notations from (8) and the parameters satisfying (9). From (11), we have

$$\Delta_1(f) \geq \Delta_2(f') - 3\rho_3.$$

Next, as in Section 5.2.3, we randomly restrict at most B coordinates to a string w so that in the restricted function F_w , all the influences of the restricted functions L_P in its decomposition are $\tau \ll O_{\rho_0}(1)$ (along with w satisfying the event E). For this restriction, we get,

$$\Delta_2(f') \geq \alpha^{B+1} \cdot \Delta_2(f', E_w).$$

From Section 5.2.4, for some vector $\vec{b}_\star$ satisfying the event E' , we get

$$\Delta_2(f', E_w) \geq \Omega_{\rho_0}(1) \left(\Delta_3(\tilde{\mathbf{F}}_{d,w}, \vec{b}_\star) - 3\xi^{1/8} \right)$$

Finally, using (14) and Claim 5.22, we have

$$\Delta_3(\tilde{\mathbf{F}}_{d,w}, \vec{b}_\star) \geq \delta^{\Omega_{m,\alpha}(1)} - \tilde{\rho}_2,$$

where $\tilde{\rho}_2$ is from Claim 5.21. Combining all these, we get

$$\begin{aligned} \Delta_1(f) &\geq \alpha^{B+1} \Omega_{\rho_0}(1) \left(\delta^{\Omega_{m,\alpha}(1)} - \tilde{\rho}_2 - 3\xi^{1/8} \right) - 3\rho_3 \\ &= \alpha^{B+1} \Omega_{\rho_0}(\delta^{\Omega_{m,\alpha}(1)}), \end{aligned}$$

where in the last equality, we use the relations (9) among different parameters. $\square$

The proof of the main theorem follows from Lemma 5.14 and Lemma 5.27.

Proof of Theorem 1.4: We start with $f : \Sigma^n \rightarrow \{0, 1\}$ with density $\delta > 0$. Using Claim 5.17, in order to move to a restricted function where all influences of the decomposed function are at most τ , and still have the average of the restricted function being $\Omega(\delta)$, we need f to be $(B, 1/2)$ -resilient where $B = O_{|\mathcal{P}|, |\Sigma|, D, C, \eta, \tau}(1) = O_{\rho_1}(1)$. Using Claim 5.14, if f is not $(B, 1/2)$ -resilient to begin with, we first make f resilient by losing a multiplicative factor of $2^{-\Omega_{B, \delta, \alpha}(1)} = 2^{-\Omega_{\rho_1}(1)}$ towards estimating $\Delta_1(f)$. Using the estimate for $(B, 1/2)$ resilient functions from Lemma 5.27, we have

$$\Delta_1(f) \geq 2^{-\Omega_{\rho_1}(1)} \Omega_{\rho_0}(\delta^{\Omega_{m, \alpha}(1)}).$$

thereby finishing the proof of the main theorem. $\square$

The following proposition shows that for μ such that $\text{supp}(\mu)$ has no $\mathbb{Z}$ -embedding, the corresponding covariance matrix from the proof of Claim 5.22 is positive definite.

Proposition 5.28. *Let μ be a distribution on Σ^3 such that $\text{supp}(\mu)$ has no $\mathbb{Z}$ -embedding and let $\alpha > 0$ be the minimum non-zero probability of an atom in μ . Let μ_1, μ_2, μ_3 denote the marginals of μ on the first, second, and third coordinates, respectively. Assume that each marginal μ_i has full support on Σ . For $i \in \{1, 2, 3\}$, let*

$$f_0^{(i)} = 1, f_1^{(i)}, \dots, f_{q-1}^{(i)}$$

be an orthonormal basis of $L^2(\Sigma, \mu_i)$ where $q = |\Sigma|$. Define the random vector

$$W = (f_1^{(1)}(X), \dots, f_{q-1}^{(1)}(X), f_1^{(2)}(Y), \dots, f_{q-1}^{(2)}(Y), f_1^{(3)}(Z), \dots, f_{q-1}^{(3)}(Z)) \in \mathbb{C}^{3(q-1)},$$

and let $M = \text{Cov}(W)$ be the covariance matrix of W , i.e., $M = \mathbb{E}_\mu[WW^]$ where $*$ denotes conjugate transpose. Then M is Hermitian positive definite, i.e., $M - \eta_{\alpha, |\Sigma|} I \succeq 0$ for some $\eta_{\alpha, |\Sigma|} > 0$.*

Proof. Let

$$\theta = (\alpha_1, \dots, \alpha_{q-1}, \beta_1, \dots, \beta_{q-1}, \gamma_1, \dots, \gamma_{q-1})^\top \in \mathbb{C}^{3(q-1)}.$$

Define complex-valued functions $g, h, k : \Sigma \rightarrow \mathbb{C}$ by

$$g(x) = \sum_{j=1}^{q-1} \overline{\alpha_j} f_j^{(1)}(x), \quad h(y) = \sum_{j=1}^{q-1} \overline{\beta_j} f_j^{(2)}(y), \quad \text{and} \quad k(z) = \sum_{j=1}^{q-1} \overline{\gamma_j} f_j^{(3)}(z).$$

Then

$$\theta^* W = g(X) + h(Y) + k(Z).$$

Consequently,

$$\begin{aligned} \theta^* M \theta &= \mathbb{E}_\mu \left[|\theta^* W|^2 \right] \\ &= \mathbb{E}_\mu \left[|g(X) + h(Y) + k(Z)|^2 \right]. \end{aligned}$$

It follows immediately that M is Hermitian positive semidefinite. We show that it is strictly positive definite. Suppose, toward a contradiction, that there exists $\theta \neq 0$ such that

$$\theta^* M \theta = 0.$$

Then $g(X) + h(Y) + k(Z)$ is 0. Hence $g(x) + h(y) + k(z) = 0$ for every (x, y, z) in the support of μ . Write

$$g = g_R + \mathbf{i}g_I, \quad h = h_R + \mathbf{i}h_I, \quad k = k_R + \mathbf{i}k_I,$$

where all real and imaginary parts are real-valued functions. Taking real and imaginary parts gives

$$g_R(x) + h_R(y) + k_R(z) = 0$$

and

$$g_I(x) + h_I(y) + k_I(z) = 0$$

for every $(x, y, z) \in \text{supp}(\mu)$. For every $j \geq 1$, the function $f_j^{(i)}$ is orthogonal to the constant function 1 in $L^2(\Sigma, \mu_i)$. Therefore,

$$\mathbb{E}_{\mu_1}[g] = 0, \quad \mathbb{E}_{\mu_2}[h] = 0, \quad \mathbb{E}_{\mu_3}[k] = 0.$$

In particular, if g, h, k were all constant, then they would all be identically zero. Since the basis functions are linearly independent and $\theta \neq 0$, this is impossible. Hence at least one of g, h, k is nonconstant. It follows that at least one of the two real triples

$$(g_R, h_R, k_R) \quad \text{or} \quad (g_I, h_I, k_I)$$

contains a nonconstant function. Choose such a triple and denote it by (G, H, K) . Then

$$G(x) + H(y) + K(z) = 0 \quad \text{for every } (x, y, z) \in \text{supp}(\mu),$$

and at least one of G, H, K is nonconstant. Consider the homogeneous linear system

$$A(x) + B(y) + C(z) = 0 \quad \text{for every } (x, y, z) \in \text{supp}(\mu).$$

Its coefficient matrix has integer entries. The tuple (G, H, K) is a real solution lying outside the linear subspace in which A, B, C are all constant. Since the kernel of a matrix with rational entries has a basis over $\mathbb{Q}$, the real solution space is the real span of its rational solutions. Therefore, if every rational solution had A, B, C all constant, then every real solution would also have A, B, C all constant. This contradicts the existence of (G, H, K) . Hence there exists a rational solution

$$(A, B, C)$$

such that at least one of A, B, C is nonconstant and

$$A(x) + B(y) + C(z) = 0 \quad \text{for every } (x, y, z) \in \text{supp}(\mu).$$

Multiplying by a common denominator, we may assume that A, B, C are integer-valued. Thus (A, B, C) is a nonconstant embedding of $\text{supp}(\mu)$ into $\mathbb{Z}$, contradicting the hypothesis. Therefore,

$$\theta^* M \theta > 0 \quad \text{for every nonzero } \theta \in \mathbb{C}^{3(q-1)}.$$

Hence M is Hermitian positive definite.

From this, standard compactness argument shows that $\theta^* M \theta$ is indeed bounded away from 0 as a function of $|\Sigma|$ and α for every nonzero $\theta \in \mathbb{C}^{3(q-1)}$ with $\|\theta\|_2 = 1$. $\square$

6 Discussion and Open Problems

A key conceptual challenge in extending Theorem 1.2 to the setting of restricted 3-APs is the corresponding problem in Gaussian space, which we now describe, and for simplicity we fix $p = 3$. Let ω_3 be the primitive root of unity of order 3, and let μ be the uniform distribution over $\{(x, x + a, x + 2a) \mid x \in \mathbb{F}_3, a \in \{0, 1\}\}$. Consider the 6 by 6 correlation matrix, M , whose rows and columns are indexed by $\{1, 2\} \times \{1, 2, 3\}$ and are given by

$$M((a, i), (b, j)) = \mathbb{E}_{(x_1, x_2, x_3) \sim \mu} \left[\omega^{ax_i} \omega^{-bx_j} \right].$$

Explicitly, using the lexicographic ordering on $\{1, 2\} \times \{1, 2, 3\}$, M can be written as

$$M = \begin{pmatrix} 1 & \frac{1+\omega}{2} & \frac{1+\omega^2}{2} & 0 & 0 & 0 \\ \frac{1+\omega^2}{2} & 1 & \frac{1+\omega}{2} & 0 & 0 & 0 \\ \frac{1+\omega}{2} & \frac{1+\omega^2}{2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1+\omega^2}{2} & \frac{1+\omega}{2} \\ 0 & 0 & 0 & \frac{1+\omega}{2} & 1 & \frac{1+\omega^2}{2} \\ 0 & 0 & 0 & \frac{1+\omega^2}{2} & \frac{1+\omega}{2} & 1 \end{pmatrix}. \quad (18)$$

It turns out that this matrix is singular, and this presents difficulties. The result of [CDP15] can be used to bound expectations of the form

$$\mathbb{E}_{(G_1, \dots, G_k)} \left[\prod_{i=1}^k f_i(G_i) \right],$$

where $G_1, \dots, G_k$ jointly distributed Gaussian random variables with non-singular covariance matrix, and each f_i is non-negative bounded function. Strictly speaking, their result is phrased for real-valued Gaussian distributions, but it also holds for complex Gaussians.

For this multi-function version, the non-singularity assumption is indeed necessary, and we give an example. Multiplying M by $(1, \omega, \omega^2, 1, \omega^2, \omega)$ shows that M is singular, and this implies that complex Gaussians $(G_1, \dots, G_6)$ with covariance matrix M satisfy the linear relation $G_1 + \omega^2 G_2 + \omega G_3 = 0$. Take a constant $\varepsilon > 0$ and define $f_1(G) = f_2(G) = 1_{|G| \leq \varepsilon}$ and $f_3(G) = 1_{|G| \geq 3\varepsilon}$. Then for G_1, G_2, G_3 as above, if $f_1(G_1) = f_2(G_2) = 1$, then we get that

$$|G_3| = |G_1 + \omega^2 G_2| \leq 2\varepsilon < 3\varepsilon,$$

so $f_3(G_3) = 0$. This shows that $\mathbb{E}_{G_1, G_2, G_3} [f_1(G_1) f_2(G_2) f_3(G_3)] = 0$.

We are not aware of any similar example where the functions f_1, f_2, f_3 are all the same. In fact, we believe the following statement should be true:

Conjecture 6.1. *For all $\alpha > 0$, there exists $\beta > 0$, such that for all $n \in \mathbb{N}$, if $f: \mathbb{C}^{2n} \rightarrow \{0, 1\}$ has measure at least α , then*

$$\mathbb{E}_{(G_1, \dots, G_6)} [f(G_1, G_4) f(G_2, G_5) f(G_3, G_6)] \geq \beta.$$

Here, $G_1, \dots, G_6$ is the n -dimensional Gaussian distribution, where each coordinate is sampled independently as jointly distributed centered complex Gaussian random variables with the covariance matrix M from (18).

We remark that we know how to prove dimension-dependent bounds for Conjecture 6.1, and the real challenge here is to establish bounds that hold for all dimensions.

Remark 6.2. *There may be a more general phenomenon underlying statements along the lines of Conjecture 6.1 for matrices M whose main diagonal are all 1's. If M is singular and there is $v \in \ker(M)$ not orthogonal to the all 1 vector, then even the single function can produce an expectation equal to 0, so the analogous statement to Conjecture 6.1 fails for M . It is feasible that this is the only obstruction, and that if a covariance matrix M has that $\ker(M) \subseteq \text{span}(\vec{1})^\perp$, then M admits a counting result in the spirit of Conjecture 6.1. If true, such a result would be very interesting in our opinion, as it gives a clear distinction between the single-function version and the multi-function version.*

References

- [Beh46] F. A. Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences of the United States of America*, 32(12):331–332, 1946.
- [BKLM25] Amey Bhangale, Subhash Khot, Yang P. Liu, and Dor Minzer. On inverse theorems and combinatorial lines. In *66th IEEE Annual Symposium on Foundations of Computer Science (FOCS 2025)*, pages 1672–1684. IEEE, 2025.
- [BKM24a] Amey Bhangale, Subhash Khot, and Dor Minzer. Effective bounds for restricted 3-arithmetic progressions in $\mathbb{F}_p^n$. *Discrete Analysis*, page 22, 2024.
- [BKM24b] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable k -CSPs: IV. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC 2024)*, pages 1423–1434. ACM, 2024.
- [BKM26] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable k -CSPs: V. *TheoretCS*, 5:9, 2026.
- [BL96] Vitaly Bergelson and Alexander Leibman. Polynomial extensions of van der Waerden’s and Szemerédi’s theorems. *Journal of the American Mathematical Society*, 9(3):725–753, 1996.
- [Bou86] Jean Bourgain. A Szemerédi-type theorem for sets of positive density in $\mathbb{R}^k$. *Israel Journal of Mathematics*, 54(3):307–316, 1986.
- [Bou99] Jean Bourgain. On triples in arithmetic progression. *Geometric and Functional Analysis*, 9:968–984, 1999.
- [Bou08] Jean Bourgain. Roth’s theorem on progressions revisited. *Journal d’Analyse Mathématique*, 104:155–192, 2008.
- [BS19] Thomas F. Bloom and Olof Sisask. Logarithmic bounds for Roth’s theorem via almost-periodicity. *Discrete Analysis*, page 20, 2019.
- [BS20] Thomas F. Bloom and Olof Sisask. Breaking the logarithmic barrier in Roth’s theorem on arithmetic progressions, 2020.
- [BS23] Thomas F. Bloom and Olof Sisask. An improvement to the Kelley–Meka bounds on three-term arithmetic progressions, 2023.

- [CDP15] Wei-Kuo Chen, Nikos Dafnis, and Grigoris Paouris. Improved Hölder and reverse Hölder inequalities for Gaussian random vectors. *Advances in Mathematics*, 280:643–689, 2015.
- [CFP26] David Conlon, Jacob Fox, and Huy Tuan Pham. A note on arithmetic progressions with restricted differences, 2026.
- [CLP17] Ernie Croot, Vsevolod F. Lev, and Péter Pál Pach. Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small. *Annals of Mathematics*, 185(1):331–337, 2017.
- [EG17] Jordan S. Ellenberg and Dion Gijswijt. On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression. *Annals of Mathematics*, 185(1):339–343, 2017.
- [EHPS24] Christian Elsholtz, Zach Hunter, Laura Proske, and Lisa Sauermann. Improving Behrend’s construction: sets without arithmetic progressions in integers and over finite fields, 2024.
- [FK78] Hillel Furstenberg and Yitzhak Katznelson. An ergodic Szemerédi theorem for commuting transformations. *Journal d’Analyse Mathématique*, 34(1):275–291, 1978.
- [FK91] Hillel Furstenberg and Yitzhak Katznelson. A density version of the Hales–Jewett theorem. *Journal d’Analyse Mathématique*, 57(1):64–119, 1991.
- [Gow98] W. T. Gowers. A new proof of Szemerédi’s theorem for arithmetic progressions of length four. *Geometric and Functional Analysis*, 8(3):529–551, 1998.
- [Gow01] W. T. Gowers. A new proof of Szemerédi’s theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001.
- [Gow07] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Annals of Mathematics*, 166(3):897–946, 2007.
- [Gre05] Ben Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geometric and Functional Analysis*, 15(2):340–376, 2005.
- [Gre25] Ben Green. 100 open problems. <https://people.maths.ox.ac.uk/greenbj/papers/open-problems.pdf>, 2025. Manuscript; most recent update: December 2025.
- [GT10] Ben Green and Terence Tao. An arithmetic regularity lemma, an associated counting lemma, and applications. In *An Irregular Mind*, volume 21 of *Bolyai Society Mathematical Studies*, pages 261–334. Springer, 2010.
- [HHM18] Jan Hązła, Thomas Holenstein, and Elchanan Mossel. Product Space Models of Correlation: Between Noise Stability and Additive Combinatorics. *Discrete Analysis*, dec 27 2018.
- [Jon16] Chris Jones. A noisy-influence regularity lemma for boolean functions. *CoRR*, abs/1610.06950, 2016.
- [KM23] Zander Kelley and Raghu Meka. Strong bounds for 3-progressions. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 933–973. IEEE, 2023.
- [Mes95] Roy Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *Journal of Combinatorial Theory, Series A*, 71(1):168–172, 1995.

- [MOO10] Elchanan Mossel, Ryan O’Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010.
- [Mos10] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010.
- [O’D14] Ryan O’Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.
- [Pol12] D. H. J. Polymath. A new proof of the density Hales–Jewett theorem. *Annals of Mathematics*, 175(3):1283–1327, 2012.
- [Rot53] K. F. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, s1-28(1):104–109, 1953.
- [RS78] I. Z. Ruzsa and Endre Szemerédi. Triple systems with no six points carrying three triangles. In *Combinatorics*, volume 18 of *Colloquia Mathematica Societatis János Bolyai*, pages 939–945, Amsterdam–New York, 1978. North-Holland. Proceedings of the Fifth Hungarian Colloquium, Keszthely, 1976.
- [San11] Tom Sanders. On Roth’s theorem on progressions. *Annals of Mathematics*, 174(1):619–636, 2011.
- [Sze75] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27:199–245, 1975.
- [Sze78] Endre Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes*, volume 260 of *Colloques Internationaux CNRS*, pages 399–401, Paris, 1978. CNRS.

A Proof of Theorem 5.11

We begin by stating the multidimensional version of the Gaussian reverse hypercontractivity theorem from [CDP15, HHM18].

Theorem A.1. ([CDP15, HHM18]) *Let $\eta > 0$ and let $\mathcal{G} = (\mathcal{G}_1, \dots, \mathcal{G}_\ell)$ be a jointly Gaussian collection of ℓ random vectors such that:*

1. *For each $i \in [\ell]$, $\mathcal{G}_i = (G_{i,1}, \dots, G_{i,n})$ is a random vector distributed as n independent $\mathcal{N}(0, 1)$ real Gaussians.*
2. *The $n\ell \times n\ell$ covariance matrix $P = \mathbb{E} [\mathcal{G}\mathcal{G}^T]$ of the Gaussian ensemble satisfies $P - \eta I \succeq 0$, for some $\eta > 0$.*

Then, for all functions $f_1, \dots, f_\ell \in L^2(\mathbb{R}^n, \gamma^n)$, where γ is the standard $\mathcal{N}(0, 1)$ Gaussian measure, such that $f_1, \dots, f_\ell : \mathbb{R}^n \rightarrow [0, 1]$ and $\mathbb{E} [f_j(\mathcal{G}_j)] = \mu_j$, we have:

$$\mathbb{E} \left[\prod_{j=1}^{\ell} f_j(\mathcal{G}_j) \right] \geq \left(\prod_{j=1}^{\ell} \mu_j \right)^{1/\eta}.$$

We show how to derive Theorem 5.11 from the above theorem by reducing the complex case to the real case. Towards this end, we use the following definition.

Definition A.2. For $H \in \mathbb{C}^{m \times m}$, define its realification by

$$\mathcal{R}(H) := \begin{pmatrix} \operatorname{Re} H & -\operatorname{Im} H \\ \operatorname{Im} H & \operatorname{Re} H \end{pmatrix} \in \mathbb{R}^{2m \times 2m}.$$

The key feature of $\mathcal{R}(H)$ is that it correctly maps the inner product over $\mathbb{C}^m$ to the inner product over $\mathbb{R}^{2m}$. In particular, we have the following lemma asserting that if a Hermitian matrix H is positive definite, then $\mathcal{R}(H)$ is also positive definite.

Lemma A.3. If $H \in \mathbb{C}^{m \times m}$ is Hermitian, then $\mathcal{R}(H) \in \mathbb{R}^{2m \times 2m}$ is real symmetric and for every $x, y \in \mathbb{R}^m$

$$(x + \mathbf{i}y)^* H (x + \mathbf{i}y) = \begin{pmatrix} x \\ y \end{pmatrix}^\top \mathcal{R}(H) \begin{pmatrix} x \\ y \end{pmatrix}.$$

Consequently, $H \succeq 0 \iff \mathcal{R}(H) \succeq 0$, and $H \succ 0 \iff \mathcal{R}(H) \succ 0$.

Proof. Write

$$H = M + \mathbf{i}N, \quad M = \operatorname{Re} H, \quad N = \operatorname{Im} H.$$

Since H is Hermitian,

$$M^\top = M \quad N^\top = -N, \quad \mathcal{R}(H) = \begin{pmatrix} M & -N \\ N & M \end{pmatrix}.$$

Let $z = x + \mathbf{i}y$, with $x, y \in \mathbb{R}^m$. A direct calculation gives

$$\begin{aligned} z^* H z &= (x^\top - \mathbf{i}y^\top)(M + \mathbf{i}N)(x + \mathbf{i}y) = x^\top M x - x^\top N y + y^\top N x + y^\top M y \\ &= \begin{pmatrix} x \\ y \end{pmatrix}^\top \begin{pmatrix} M & -N \\ N & M \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \end{aligned}$$

giving the first assertion of the lemma. The second assertion is immediate by definitions. $\square$

We are now ready to prove Theorem 5.11, which we restate here.

Theorem 5.11. Let $\eta > 0$ and let $\mathcal{G} = (\mathcal{G}_1, \dots, \mathcal{G}_\ell)$ be a jointly proper Gaussian collection of ℓ random vectors such that:

1. For each $i \in [\ell]$, $\mathcal{G}_i = (G_{i,1}, \dots, G_{i,n})$ is a random vector distributed as n independent $\mathcal{CN}(0, 1)$ Gaussians.
2. The $n\ell \times n\ell$ covariance matrix $P = \mathbb{E}[\mathcal{G}\mathcal{G}^*]$ of the Gaussian ensemble satisfies $P - \eta I \succeq 0$, for some $\eta > 0$.

Then, for all functions $f_1, \dots, f_\ell \in L^2(\mathbb{C}^n, \gamma^n)$ such that $f_1, \dots, f_\ell : \mathbb{C}^n \rightarrow [0, 1]$ and $\mathbb{E}[f_j(\mathcal{G}_j)] = \mu_j$, we have:

$$\mathbb{E} \left[\prod_{j=1}^{\ell} f_j(\mathcal{G}_j) \right] \geq \left(\prod_{j=1}^{\ell} \mu_j \right)^{1/\eta}.$$

Proof. Write

$$\mathcal{G}_j = X_j + \mathbf{i}Y_j, \quad X_j, Y_j \in \mathbb{R}^n,$$

and define the normalized realification

$$\tilde{\mathcal{G}}_j := \sqrt{2} \begin{pmatrix} X_j \\ Y_j \end{pmatrix} \in \mathbb{R}^{2n}, \quad \tilde{\mathcal{G}} = (\tilde{\mathcal{G}}_1, \dots, \tilde{\mathcal{G}}_\ell).$$

Since the coordinates of $\mathcal{G}_j$ are independent $\mathcal{CN}(0, 1)$ random variables, the coordinates of $\tilde{\mathcal{G}}_j$ are independent and mean 0. The normalization $\sqrt{2}$ makes them $\mathcal{N}(0, 1)$ random variables.

Write $P = (P_{jk})_{j,k=1}^\ell$, where each P_{jk} is an $n \times n$ complex matrix, and let $\tilde{P} := \text{Cov}(\tilde{\mathcal{G}})$ be the real covariance matrix of the realified collection. Then we have

$$\tilde{P}_{jk} = \begin{pmatrix} \text{Re } P_{jk} & -\text{Im } P_{jk} \\ \text{Im } P_{jk} & \text{Re } P_{jk} \end{pmatrix}. \quad (19)$$

Indeed, by definition we have

$$P_{jk} = \mathbb{E}[\mathcal{G}_j \mathcal{G}_k^*] = \mathbb{E}[(X_j + \mathbf{i}Y_j)(X_k^T - \mathbf{i}Y_k^T)] = \mathbb{E}[X_j X_k^T] + \mathbb{E}[Y_j Y_k^T] + \mathbf{i}\mathbb{E}[Y_j X_k^T] - \mathbf{i}\mathbb{E}[X_j Y_k^T].$$

On the other hand, the properness of the collection gives

$$0 = \mathbb{E}[\mathcal{G}_j \mathcal{G}_k^T] = \mathbb{E}[(X_j + \mathbf{i}Y_j)(X_k + \mathbf{i}Y_k)^T] = \mathbb{E}[X_j X_k^T] - \mathbb{E}[Y_j Y_k^T] + \mathbf{i}\mathbb{E}[Y_j X_k^T] + \mathbf{i}\mathbb{E}[X_j Y_k^T],$$

implying $\mathbb{E}[X_j X_k^T] = \mathbb{E}[Y_j Y_k^T]$ and $\mathbb{E}[Y_j X_k^T] = -\mathbb{E}[X_j Y_k^T]$. Plugging this above gives

$$\text{Re } P_{jk} = \mathbb{E}[X_j X_k^T] + \mathbb{E}[Y_j Y_k^T] = 2\mathbb{E}[X_j X_k^T] = 2\mathbb{E}[Y_j Y_k^T],$$

and

$$\text{Im } P_{jk} = \mathbb{E}[Y_j X_k^T] - \mathbb{E}[X_j Y_k^T] = -2\mathbb{E}[X_j Y_k^T] = 2\mathbb{E}[Y_j X_k^T].$$

This implies (19), and so $\tilde{P} = \mathcal{R}(P)$, where $\mathcal{R}$ denotes blockwise realification. From Lemma A.3, we conclude that hence

$$\tilde{P} - \eta I_{2n\ell} = \mathcal{R}(P - \eta I_{n\ell}) \succeq 0.$$

Therefore, $\tilde{\mathcal{G}}$ satisfies the covariance hypothesis of the real reverse Gaussian hypercontractivity theorem, Theorem A.1, with n replaced by $2n$.

For $x, y \in \mathbb{R}^n$, define

$$\tilde{f}_j(x, y) := f_j\left(\frac{x + \mathbf{i}y}{\sqrt{2}}\right).$$

Then

$$\tilde{f}_j(\tilde{\mathcal{G}}_j) = f_j(\mathcal{G}_j) \quad \text{and} \quad \mathbb{E}[\tilde{f}_j(\tilde{\mathcal{G}}_j)] = \mu_j.$$

Applying Theorem A.1 to $\tilde{\mathcal{G}}_1, \dots, \tilde{\mathcal{G}}_\ell$ therefore yields

$$\mathbb{E}_{\tilde{\mathcal{G}}} \left[\prod_{j=1}^\ell f_j(\mathcal{G}_j) \right] = \mathbb{E}_{\tilde{\mathcal{G}}} \left[\prod_{j=1}^\ell \tilde{f}_j(\tilde{\mathcal{G}}_j) \right] \geq \left(\prod_{j=1}^\ell \mu_j \right)^{1/\eta},$$

as required. □

B Proof of Claim 5.17

The following lemma will be crucial in proving the claim.

Lemma B.1. Fix $i \in [n]$ and a function $f : (\Sigma^n, \nu^{\otimes n}) \rightarrow \mathbb{C}$. Then,

$$\mathbb{E}_{a \sim \nu} [I(f_{x_i \rightarrow a})] = I(f) - I_i(f).$$

Proof. Let $|\Sigma| = m$ and consider the Fourier basis $\phi_0 = \mathbf{1}, \phi_1, \dots, \phi_{m-1}$ for the space $L^2(\Sigma, \nu)$. Using Proposition 5.3, f has the following expansion

$$f(x) = \sum_{\alpha \in \mathbb{N}_{< m}^n} \hat{f}(\alpha) \phi_\alpha(x).$$

Thus, for $x \in \Sigma^{n-1}$, the expansion of the function $f_{x_i \rightarrow a}(x)$ is

$$f_{x_i \rightarrow a}(x) = \sum_{\alpha \in \mathbb{N}_{< m}^{n-1}} \sum_{t=0}^{m-1} \phi_t(a) \hat{f}(\alpha^{i \rightarrow t}) \phi_\alpha(x),$$

where $\alpha^{i \rightarrow t} \in \mathbb{N}_{< m}^n$ with $(\alpha^{i \rightarrow t})_i = t$ and the other coordinates same as α . Now using Proposition 5.5,

$$\begin{aligned} I(f_{x_i \rightarrow a}) &= \sum_{\alpha \in \mathbb{N}_{< m}^{n-1}} |\alpha| \left| \sum_{t=0}^{m-1} \phi_t(a) \hat{f}(\alpha^{i \rightarrow t}) \right|^2 \\ &= \sum_{\alpha \in \mathbb{N}_{< m}^{n-1}} |\alpha| \left(\sum_t \left| \hat{f}(\alpha^{i \rightarrow t}) \right|^2 + \sum_{t \neq t'} \phi_t(a) \overline{\phi_{t'}(a)} \hat{f}(\alpha^{i \rightarrow t}) \overline{\hat{f}(\alpha^{i \rightarrow t'})} \right). \end{aligned}$$

Using the fact that the characters are orthogonal, and hence $\mathbb{E}_{a \sim \nu} [\phi_t(a) \overline{\phi_{t'}(a)}] = 0$ for $t \neq t'$, we get,

$$\begin{aligned} \mathbb{E}_{a \sim \nu} [I(f_{x_i \rightarrow a})] &= \sum_{\alpha \in \mathbb{N}_{< m}^{n-1}} |\alpha| \left(\sum_t \left| \hat{f}(\alpha^{i \rightarrow t}) \right|^2 \right) \\ &= \sum_{\alpha \in \mathbb{N}_{< m}^n} |\alpha| \left| \hat{f}(\alpha) \right|^2 - \sum_{\substack{\alpha \in \mathbb{N}_{< m}^n \\ \alpha_i \neq 0}} \left| \hat{f}(\alpha) \right|^2 \\ &= I(f) - I_i(f), \end{aligned}$$

as required. □

We restate the claim here.

Claim 5.17. Consider the function F from above. Then for any $\tau > 0$, there exists $J \subseteq [n]$ of size $O(|\mathcal{P}|, |\Sigma|, D, C, \eta, \tau)(1)$ such that with probability at least $1 - \eta$ over $w \sim \nu^{\otimes |J|}$, we have

$$\max_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} \left\{ \max_i I_i[(LP)_{J \rightarrow w}] \right\} \leq \tau.$$

Proof. We build a decision tree $\mathcal{D}$ as follows. Start with the root node and associate it with functions $\{L_P \mid P \in \text{span}_{\mathbb{N}}(\mathcal{P})\}$. At each node N in the tree $\mathcal{D}$, we pick a variable $i \in [n]$ and add $|\Sigma|$ many child nodes to it. Suppose node N is associated with functions $\{(L_P)|_N \mid P \in \text{span}_{\mathbb{N}}(\mathcal{P})\}$, then the children nodes will be associated with the restriction $x_i \rightarrow a$ of functions $\{(L_P)|_N \mid P \in \text{span}_{\mathbb{N}}(\mathcal{P})\}$ for $a \in \Sigma$.

We prove the statement by starting with the tree with just the root node. Each time, if for most leaf nodes T , at least one of the functions $\{(L_P)|_T\}$ has an influential variable, then we split the node T into $|\Sigma|$ nodes each with the restriction $x_i \rightarrow a$ for $a \in \Sigma$. We show that after $O(1)$ many steps, most of the leaf nodes in $\mathcal{D}$ are such that all functions $\{(L_P)|_T\}$ associated with T have all influences at most τ .

For a given decision tree, consider the following distribution, denoted by $\nu(\mathcal{D})$ on the leaf nodes of $\mathcal{D}$. Start with a root node of $\mathcal{D}$, at each step sample the a^{th} child of the current node according to the probability $\nu(a)$ and keep going down until the process hits a leaf node T . Consider the following energy function associated with a given decision tree:

$$\varphi(\mathcal{D}) = \mathbb{E}_{T \sim \nu(\mathcal{D})} \left[\sum_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} I[(L_P)|_T] \right],$$

where the expectation is over the leaf nodes T of $\mathcal{D}$, weighted according to the distribution ν , and $I[f]$ is the total influence of f . Note that if $\mathcal{D}_0$ is just the root node, then $\varphi(\mathcal{D}_0) = O_{|\mathcal{P}|, |\Sigma|, C, D}(1)$, as $\deg(L_P) \leq D$ and $\|L\|_2 \leq C$ and hence $I[L_P] \leq CD$.

Consider any decision tree $\mathcal{D}$. Suppose that in $\mathcal{D}$, for at least η fraction of leaf nodes T , weighed according to $\nu(\mathcal{D})$, at least one of the associated functions has some variable i with influence at least τ . In this case, we can split each of these leaf nodes according to the respective influential variable. Using Lemma B.1, the new tree $\mathcal{D}'$ has the property

$$\varphi(\mathcal{D}') \leq \varphi(\mathcal{D}) - \eta \cdot \tau.$$

From the above, we concluded that if for at least η fraction of leaf nodes T , at least one of the associated functions has a variable i with influence at least τ , then the energy of the next tree (generated by splitting each of these leaf nodes according to the respective influential variable) decreases by at least $\eta \cdot \tau$. Hence, after $k := O_{|\mathcal{P}|, |\Sigma|, C, D}(\frac{1}{\tau\eta})$ many steps, we get a tree $\mathcal{D}^{\text{final}}$ with the property that for at least $(1 - \eta)$ fraction of leaf nodes T in $\mathcal{D}^{\text{final}}$, the associated functions $\{(L_P)|_T\}$ have all the influences at most τ .

In order to prove the claim, it will be convenient to have a homogeneous tree where all nodes at a given level ℓ are split based on a fixed variable $i(\ell)$. We can modify the above process of getting a decision tree to ensure that the tree is always homogeneous: At each iteration j , suppose in the above process we split the nodes based on the variables $x_{j_1}, x_{j_2}, \dots, x_{j_t}$, then in the modified process we split all the nodes at last level based on all the values of $x_{j_1}, x_{j_2}, \dots, x_{j_t}$ variables (this will increase the depth of the homogeneous tree by t , assuming all x_{j_m} s are distinct). Similar to the above analysis, the energy of the new homogeneous tree is at most the energy of the original non-homogeneous tree. Hence, this process will also stop after $k = O_{|\mathcal{P}|, |\Sigma|, C, D}(\frac{1}{\tau\eta})$ iterations. However, note that in each iteration, the depth of the tree increases by up to the number of leaf nodes in the previous iteration. If we let $d(\ell)$ denote the depth of the homogeneous tree before the ℓ^{th} iteration, then we have

$$d(0) = 1, \quad d(\ell + 1) \leq d(\ell) + 2^{d(\ell)}.$$

Solving this recurrence gives $d(k) \leq 2 \uparrow k - 1$.

If we let the restricted coordinates in the above homogeneous tree by the set $J \subseteq [n]$, then $|J| = O_{|\mathcal{P}|, |\Sigma|, C, D, \eta, \tau}(1)$ and the guarantee on the decision tree implies that with probability at least $1 - \eta$ over $w \sim \nu^{\otimes |J|}$, we have

$$\max_{P \in \text{span}_{\mathbb{N}}(\mathcal{P})} \left\{ \max_i I_i[(L_P)_{J \rightarrow w}] \right\} \leq \tau,$$

as required □