Electronic Colloquium on Computational Complexity, Report No. 54 (2023)

On Approximability of Satisfiable k-CSPs: III

Amey Bhangale[*] Subhash Khot[†] Dor Minzer[‡]

Abstract

In this paper we study functions on the Boolean hypercube that have the property that after applying certain random restrictions, the restricted function is correlated to a linear function with non-negligible probability. If the given function is correlated with a linear function then this property clearly holds. Furthermore, the property also holds for low-degree functions as low-degree functions become a constant function under a random restriction with a non-negligible probability. We show that this essentially is the only possible reason. More specifically, we show that the function must be correlated to a product of a linear function and a low-degree function. One of the main motivations of studying this question comes from the recent work of the authors [BKM22b] towards understanding approximability of satisfiable Constraint Satisfaction Problems.

Towards proving our structural theorem, we analyze a 2-query direct product test for the table $F : \binom{[n]}{qn} \to \{0,1\}^{qn}$ where $q \in (0,1)$. We show that, for every constant $\varepsilon > 0$, if the test passes with probability $\varepsilon > 0$, then there is a global function $g : [n] \to \{0,1\}$ such that for at least $\delta(\varepsilon)$ fraction of sets, the global function g agrees with the given table on *all except* $\alpha(\varepsilon)$ *many locations*. The novelty of this result lies in the fact that $\alpha(\varepsilon)$ is independent of the set sizes. Prior to our work, such a conclusion (in fact, a stronger conclusion with $\alpha = 0$) was shown by Dinur, Filmus, and Harsha [DFH19] albeit when the test accepts with probability $1 - \varepsilon$ for a small constant $\varepsilon > 0$. The setting of parameters in our direct product tests is fundamentally different compared to [DG08, IKW12, DS14, DFH19] and hence our analysis involves new techniques, including the use of the small-set expansion property of graphs defined on multi-slices. Such expansion property was recently shown in [BKLM22].

As one application of our structural result, we give a 4-query linearity test under the p-biased distribution. More specifically, for any $p \in (\frac{1}{3}, \frac{2}{3})$, we give a test that queries a given function $f : \{0,1\}^n \to \{0,1\}$ at 4 locations, where the marginal distribution of each query is $\mu_p^{\otimes n}$. The test has perfect completeness and soundness $\frac{1}{2} + \varepsilon$ – in other words, for every constant $\varepsilon > 0$, if the test passes with probability at least $\frac{1}{2} + \varepsilon$, then the function f is correlated to a linear function under the $\mu_p^{\otimes n}$ measure. This qualitatively improves the results on the linearity testing under the p-biased distribution from the previous work [KS09, DFH19] in which the authors studied the test with soundness $1 - \varepsilon$, for ε close to 0.

1 Introduction

Analysis of Boolean functions plays a crucial role in many areas of mathematics and computer science, including complexity theory, hardness of approximation, coding theory, additive combinatorics, social choice, etc. Among the set of Boolean functions, linear functions are among the simplest class of functions and hence linearity testing, i.e., checking whether a given Boolean function is a linear function or far from it, is one of the most fundamental and well-studied problems in the analysis of Boolean functions. In this paper,

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

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

[‡]Department of Mathematics, Massachusetts Institute of Technology. Supported by a Sloan Research Fellowship, NSF CCF award 2227876 and NSF CAREER award 2239160. Email: dminzer@mit.edu.

ISSN 1433-8092

we study certain problems in the analysis of Boolean functions and problems in property testing, including linearity testing and agreement testing.

The main motivation for studying these set of problems comes from the recent work by the authors and this work can be thought of as a continuation of the line of research from the previous work by the authors [BKM22a, BKM22b]. The primary focus in this paper is to understand the structure of a boolean function under a random restriction. Fix a distribution ν on $\{0,1\}$ and a constant $\eta \in (0,1)$. Given a function $f : \{0,1\}^n \to \{0,1\}$, consider the process of randomly restricting a subset of the variables as follows. First choose a random subset $I \subseteq [n]$ by including $i \in I$ with probability η independently for each $i \in [n]$ and then select $z \in \{0,1\}^{|I|}$ from the distribution ν^I. The function f under the restriction (I, z) is defined as $f_{I\to z} : \{0,1\}^{n-|I|} \to \{0,1\}$ where $f_{I\to z}(x) = f(x, z|_I)$, i.e., we fix the variables from I according to z. In this work, we study the properties of f if $f_{I\to z}$ is correlated with a linear function with noticeable probability. In order prove the structural result, we also study the direct product testing under a different regime of parameters that was not studied before. Finally, we use our structural result to analyze linearity tests under a biased distribution.

We now formally describe these problems and the main results that we prove in this work.

1.1 Problem 1: Large Fourier coefficient after a random restriction

Let μ be a distribution over $\{0,1\}$ in which the probability of each atom is at least $\alpha > 0$, and write $\mu = \beta U + (1-\beta)\mu'$ where U is the uniform distribution over $\{0,1\}$, μ' is some distribution over $\{0,1\}$ with full support, and $0 < \beta < \alpha/2$ is thought of as a constant. We denote $I \sim_p [n]$ the choice a random subset of $[n]$ that results from including each element from $[n]$ in it with probability p. Suppose that $f\colon (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$ is a function with 2-norm at most 1 satisfying that

$$\Pr_{\substack{I\sim_{1-\beta}[n]\\ z\sim\mu'^I}}\left[\exists S \subseteq \overline{I}, \left|\widehat{f_{I\to z}}(S)\right| \geqslant \delta\right] \geqslant \eta. \tag{1}$$

In words, with noticeable probability, after a suitable random restriction and looking at the underlying measure of the restricted function as the uniform distribution, the restricted function has a significant Fourier coefficient. What can we say about the structure of the function f in that case?

The most natural guess would be that the function f itself has to be correlated with some linear function $\chi_S(x) = \prod_{i\in S}(-1)^{x_i}$. Inspecting, it is indeed clear that any function f that is correlated with some χ_S indeed satisfies (1), however it turns out that there are other examples. If f is a low-degree function, say a function of degree much smaller than $1/\beta$, then we expect the random restriction to fix the value of f with considerable probability, and hence we expect $\left|\widehat{f_{I\to z}}(\emptyset)\right|$ to be large with considerable probability. More generally, it is enough that f is correlated with a low-degree function for the above to occur with noticeable probability.

More generally, one could combine the two examples above and show that any function f that is correlated with a function of the form $\chi_S(x)\cdot g(x)$, where g is a low-degree function, satisfies (1) provided that deg(g) is significantly smaller than $1/\beta$. Indeed, after such random restriction, the restriction of χ_S is a different character (up to a sign), and the restriction of g is close to being a constant function with significant probability. Hence we would get that after random restriction f is correlated with a function of the form $a\chi_{S'}$ for some real number $a \in \mathbb{R}$, and in particular it has a significant Fourier coefficient.

Our first result asserts that this structure in fact captures all functions f satisfying (1).

Theorem 1.1. *For all $\alpha, \beta > 0$ and $\delta, \eta > 0$, there are $\delta'(\delta) > 0$, $d(\alpha, \delta) \in \mathbb{N}$ such that if a function $f\colon (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$ with 2-norm at most 1 as in the above set-up satisfies (1), then there is $S \subseteq [n]$ and a function $g\colon \{0,1\}^n \to \mathbb{R}$ of 2-norm at most 1 of degree at most d, such that*

$$\left| \mathop{\mathbb{E}}_{x \sim \mu^{\otimes n}} [f(x)\chi_S(x)g(x)] \right| \geqslant \delta'.$$

Moreover, the function g is given as $g = (\chi_S f)^{\leqslant d}$.

Motivation. Besides being a natural question to consider, we are motivated to study the above problem and prove Theorem 1.1 by the study of satisfiable CSPs. In particular, in [BKM22b] the authors proved an analytical lemma [BKM22b, Lemma 1] that plays a crucial role in classifying the complexity of approximation of satisfiable constraint satisfaction for the case of 3-ary CSPs. En route to extending this result to larger arity CSPs, the authors have been thinking about a stability version of this problem [BKM22c] which naturally leads to a structure as given in (1). While such structure, by its own, is already significant, it is hard to really call it a global structure, since it only asserts that f possesses some distinctive structure after a random restriction, which limits its applicability. Indeed, while we believe such structure to be sufficient for some applications (such as resolving the non-linear embedding hypothesis from [BKM22a]), we can see that to make further progress one needs a more "full-fledged" global characterization of a function f satisfying (1). This is where the current paper enters the picture, and the original motivation for us to prove Theorem 1.1.

Upon trying to think of Theorem 1.1, we have realized it is related to two other notable problems in TCS, namely the linearity testing problem over the biased cube, and the direct product testing problem. Below, we discuss these problems, and state our results about them.

1.2 Problem 2: The linearity testing problem over the biased cube

The next problem we consider is the biased version of the classical linearity testing problem. Let μ_q be the q-biased distribution over $\{0,1\}$, i.e. the distribution in which $\mu_q(1) = q$ and $\mu_q(0) = 1 - q$, and let ν be a distribution over $\left\{ (a,b,c,d) \in \{0,1\}^4 \,\middle|\, a+b+c+d = 0 \pmod 2 \right\}$ whose marginal on each coordinate is μ_q in which the probability of each element is at least $\alpha > 0$. In the linearity testing problem over the q-biased cube, we have a function $f\colon (\{0,1\}^n, \mu_q^{\otimes n}) \to \{-1,1\}$ satisfying that

$$\Pr_{(x,y,z,w) \sim \nu^{\otimes n}} [f(x)f(y)f(z)f(w) = 1] \geqslant \frac{1}{2} + \delta, \tag{2}$$

namely that $f(x)f(y)f(z)f(w) = 1$ with probability noticeably larger than $1/2$, and the goal is to prove that f must possess some special structure in this case. The classical version of this problem is concerned with the case that $q = 1/2$, in which case it was shown that f must have a heavy Fourier coefficient, i.e. must be correlated with a function of the form χ_S. Initially, this was shown for the so-called 99% regime [BLR90], in which $\delta \geqslant 1/2 - \varepsilon$ for some small ε, and later this was extended to the 1% regime, in which case δ is thought of as small [BCH+96, KLX10].

For any $q \neq 1/2$, one can recover the result for the 99% regime using the same local-correction techniques [KS09, DFH19] and show that f must be in fact close to a function of the form χ_S. However, the techniques in the more challenging 1% regime completely break down, and as far as we know the linearity testing question is open for any $q \neq \frac{1}{2}$ in this regime.

Theorem 1.1 already by itself gives some structural result for functions f satisfying (2), and to see that, we re-write (2) as

$$\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[f(x)f(y)f(z)f(w)\right] \geqslant 2\delta. \tag{3}$$

Inspecting (3), one may apply random-restrictions properly so as to transform inequality (3) to measuring the advantage certain restrictions of f have in the standard linearity testing problem over the uniform hypercube, which shows that with noticeable probability, a random restriction of f has a significant Fourier coefficient as in the setting of Theorem 1.1. Thus, f must be correlated with a function of the form $\chi_S g$ for a low-degree function g.

Ideally, one would expect that the answer to the linearity testing question over the q-biased cube to also be about correlations just with χ_S, which raises the question of whether the g part is necessary in the above result. In general, we do not know the answer to that, but we are able to show that it boils down to the following problem, for which we need the notion of resilient functions.

Definition 1.2. *Let μ be a probability measure over $\{0,1\}$. A function $g\colon (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$ is called (r,ε) resilient if for any $S \subseteq [n]$ of size at most r and any $s \in \{0,1\}^S$,*

$$\left|\mathbb{E}_{x\sim\mu^{\otimes n}}\left[f(x) \mid x_S = s\right] - \mathop{\mathbb{E}}_{x\sim\mu^{\otimes n}}\left[f(x)\right]\right| \leqslant \varepsilon.$$

In other words, restricting any set of at most r coordinates changes the average of g by at most ε.

It turns out that to "remove" the g part from the above structural result, it is sufficient (and also necessary, in a sense) to show that if $g_1, \dots, g_4$ are bounded, noise stable functions (which should be thought of as low-degree functions), that are resilient, then

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[g_1(x)g_2(y)g_3(z)g_4(w)\right]\right| \leqslant o(1). \tag{4}$$

In general, we do not know how to solve this problem, however in some cases of interest we are able to do so, namely in the case that ν is pairwise independent.

To spell it out, in this case, ν is the distribution in which (a) each one of $(1,1,0,0)$, $(1,0,1,0)$, $(1,0,0,1)$, $(0,1,1,0)$, $(0,1,0,1)$ and $(0,0,1,1)$ receives probability q_1, (b) the point $(1,1,1,1)$ receives probability q_2, and (c) the point $(0,0,0,0)$ receives probability q_3, where $q_1 = \frac{q(1-q)}{2}$, $q_2 = \frac{q(3q-1)}{2}$ and $q_3 = 1 - \frac{5q}{2} + \frac{3q^2}{2}$. In this case, we are able to resolve the above problem, thereby prove the following result:

Theorem 1.3. *Let $q \in (\frac{1}{3}, \frac{2}{3})$, and suppose that ν is a pairwise independent distribution over the set $\left\{(a,b,c,d) \in \{0,1\}^4 \;\middle|\; a+b+c+d = 0 \pmod 2\right\}$ in which the marginal of each coordinate is μ_q. Then for every $\delta > 0$, there is $\delta' > 0$ such that if $f\colon (\{0,1\}^n, \mu_q^{\otimes n}) \to \{-1,1\}$ satisfies* (3)*, then there is $S \subseteq [n]$ such that*

$$\left|\mathop{\mathbb{E}}_{x\sim\mu^{\otimes n}}\left[f(x)\chi_S(x)\right]\right| \geqslant \delta'.$$

We remark that our argument gives in fact a version of Theorem 1.3 for real-valued functions with bounded 12-norms, as well as a list-decoding version. See Section 4 for more details.

1.3 Problem 3: Direct product testing

The third and final problem considered in this paper is the direct product testing problem which is described as follows. Fix any $q \in (0, 1)$ and consider a table $F : \binom{[n]}{qn} \to \{0, 1\}^{qn}$. For a subset $S \subseteq [n]$ of size qn, the entry $F[S]$ can be thought of as a function $f_S : S \to \{0, 1\}$, by fixing an arbitrary ordering of the set $[n]$. F is called a direct product table if there is a function $g : [n] \to \{0, 1\}$ such that for all S, $F[S] = g|_S$. Here, $g|_S$ is the function g restricted to the coordinates in S. In direct product testing, one would like to check, by querying a few locations from the table F, if the table is coming from a global function $g : [n] \to \{0, 1\}$. In other words, is there a function $g : [n] \to \{0, 1\}$ such that for many subsets $S \subseteq [n]$, the entry $F[S]$ is equal to $g|_S$?

The direct product testing problem has been extensively studied in [DR06, DG08, IKW12, DS14, DFH19] and one of the main motivations of studying direct product testing is its application to constructing Probabilistically Checkable Proofs with small soundness (for instance, see [IKW12]). There is a natural test to check if the table F is a direct product and it is as follows: Select a random set A of size $q'n$ and two random subsets $B_1 \subseteq [n] \setminus A$ and $B_2 \subseteq [n] \setminus A$ each of size $(q - q')n$, for some $q' < q$, and check if $F[S_1]|_{S_1 \cap S_2} = F[S_2]|_{S_1 \cap S_2}$, where $S_i = A \cup B_i$ for $i = 1, 2$. Denote the distribution on the sets (S_1, S_2) by $\mathcal{D}_{q,q'}$. Clearly, if F is a direct product function, then the test passes with probability 1. The challenging task is to show that if the test passes with non-negligible probability, then F is *close* to being a direct product function.

Similar to linearity testing, the direct product testing has been studied in the 99% regime [DR06, DFH19] (in which one wants to draw the conclusion when the test passes with probability $1 - \varepsilon$) and in the 1% regime [DG08, IKW12, DS14] (in which one wants to draw the conclusion when the test passes with probability ε). Here, ε can be though of as a small quantity. In this work, we study the direct product test in the 1% regime when q, q' are constants independent of n. The regime of parameters we consider is tailored to our applications (i.e., proving Theorem 1.1, and hence proving Theorem 1.3), and to the best of our knowledge does not currently appear in the literature.

If the test passes with probability ε, then one possibility is that the table F could be obtained (probabilistically) by choosing some $g : [n] \to \{0, 1\}$, and defining $F[S]$ independently for each S as $g|_S$ with probability $\sqrt{\varepsilon}$, and otherwise to be a random element of $\{0, 1\}^{qn}$. More generally, one can take a list of functions $g_1, \dots, g_m : [n] \to \{0, 1\}$ such that for all $i \neq j$ we have that $\Delta(g_i, g_j) \leqslant O(1)$, and then for each S independently, with probability $\sqrt{\varepsilon}$ choosing $F[S] = g_i|_S$ for some random $i \in [m]$, and otherwise taking $F[S]$ to be uniformly chosen. Our direct product theorem asserts that the above examples essentially exhaust all possible F's that satisfy the direct product test.

Theorem 1.4. *For all $0 < q' < q < 1$ and $\varepsilon > 0$, there are $r \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose that $F : \binom{[n]}{qn} \to \{0, 1\}^{qn}$ satisfies*

$$\Pr_{(S_1, S_2) \sim \mathcal{D}_{q,q'}} [F[S_1]|_{S_1 \cap S_2} = F[S_2]|_{S_1 \cap S_2}] \geqslant \varepsilon.$$

Then there exists a function $g : [n] \to \{0, 1\}$ such that for at least a δ fraction of $S \in \binom{[n]}{qn}$, we have $|\{i \in S \mid F[S]_i \neq g(i)\}| \leqslant r$.

The novelty of this result lies in the fact that r is independent of n. Prior to our work, such a conclusion (in fact, a stronger conclusion with $r = 0$) was shown by Dinur, Filmus, and Harsha [DFH19] albeit when the test accepts with probability $1 - \varepsilon$ for small constants $\varepsilon > 0$. We cannot have $r = 0$ in our conclusion as the test passes with a small probability.[1] Furthermore, the setting of parameters in our direct product tests are fundamentally different compared to the previous work on direct product testing and hence our analysis involves new techniques, including the use of the small-set expansion property of graphs defined on multi-slices.[2] Such expansion property was recently shown in [BKLM22].

For our application, we need to apply the direct product theorem over a q-biased hypercube which is defined as follows. Consider the q-biased measure over $P([n])$, i.e. $\mu_q^{\otimes n}(A) = q^{|A|}(1-q)^{n-|A|}$, and let $G\colon (P[n], \mu_q^{\otimes n}) \to P([n])$ be an assignment that to each $A \in P([n])$ assigns a subset of it $G[A] \subseteq A$ in a *locally consistent manner*. Namely, for $\alpha \in (0,1)$, consider the distribution $\mathcal{D}_{q,\alpha}$ over $A, A' \subseteq [n]$ that results from by taking, for each $i \in [n]$ independently, i to be both in A, A' with probability αq, i to be in $A \setminus A'$ with probability $(1-\alpha)q$, i to be in $A' \setminus A$ with probability $(1-\alpha)q$. The function G is locally consistent if

$$\Pr_{(A,A')\sim\mathcal{D}_{q,\alpha}}\left[G[A]\cap(A\cap A') = G[A']\cap(A\cap A')\right] \geqslant \varepsilon.$$

The following corollary, that follows from Theorem 1.4, asserts that in this case, G must be correlated to a global subset $S \subseteq [n]$.

Corollary 1.5. *For all $\alpha, \varepsilon > 0$ and $0 < q < \frac{1}{2-\alpha}$, there are $r \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose that $G\colon (P[n], \mu_q^{\otimes n}) \to P([n])$ satisfies*

$$\Pr_{(A,A')\sim\mathcal{D}_{q,\alpha}}\left[G[A]\cap(A\cap A') = G[A']\cap(A\cap A')\right] \geqslant \varepsilon.$$

Then there exists $S \subseteq [n]$ such that $\Pr_{A\sim\mu_q^{\otimes n}}\left[|G[A]\Delta(S\cap A)| \leqslant r\right] \geqslant \delta$.

1.4 Related Work

As mentioned before, various kinds of linearity tests have been extensively studied. To begin with, Blum, Luby and Rubienfeld [BLR90] gave the 3-query lineary test under uniform distribution in the 99% regime. [BCH$^+$96, KLX10] improved this result by showing that if the function on $\{0,1\}^n$ passes the BLR test with probability $\frac{1}{2}+\varepsilon$, for any constant $\varepsilon > 0$, then it has a non-trivial correlation with some linear function. In the p-biased setting, Kopparty and Saraf [KS09] gives $O_p(1)$-query linearity test with soundness $1-\varepsilon$ for ε close to 0. David, Dinur, Goldenberg, Kindler and Shinkar [DDG$^+$15] gave a linearity testing in the 99% regime on a slice of the Boolean hypercube. Recently, in order to reduce the number of queries in the biased linearity testing, Dinur, Filmus and Harsha [DFH19] gave a 4-query linearity test (more generally, a 2^{d+1}-query degree-d test) with soundness $1-\varepsilon$, in the p-biased setting.

The direct product tests (also known as agreement tests) were first studied by Goldreich and Safra [GS00] in which they show that it can be testable with constantly many queries. Dinur and Reingold [DR06] gave a 2-query direct product test in the 99% accepting regime. Dinur and Goldenberg [DG08] improved this to the 1% regime. More specifically, given a table $F : \binom{[n]}{k} \to \{0,1\}^k$, if the test described in the introduction passes with probability at least $\varepsilon \geqslant 1/k^\alpha$ for some $\alpha < 1$, then there is a global function $g : [n] \to \{0,1\}$ such that for at least $\varepsilon^{O(1)}$ fraction of the sets S, $F[S] \overset{\leqslant k^\delta}{\neq} g(S)$ for some constant $0 < \delta < 1$. Here

[1] Consider a global function $g : [n] \to \{0,1\}$ and define $F[S] = g(S) + \eta$, where η is a random noise with hamming weight $\leqslant C$ for some constant C. It is easy to see that F will pass the test with a small constant probability and yet there is no global function that fully agrees with $F[S]$ on a constant fraction of $S \in \binom{[n]}{qn}$.

[2] Given an alphabet size $m \in \mathbb{N}$, thought of as a constant, and $\vec{k} = (k_1, k_2, \ldots, k_m)$ whose entries sum of up n, the $\vec{k}$-multi-slice is the set of vectors $x \in [m]^n$ in which each symbol $i \in [m]$ appears precisely k_i times.

the notation $\overset{\leqslant\beta}{\neq}$ means that the two strings agree on all except β many locations. They also show that one cannot get a meaningful conclusion of the test passes with probability less than $\frac{1}{k}$. More formally, there is a function F such that the test accepts F with probability at least $\Omega(k'/k)$, where k' is the intersection size of the two sets from the test distribution, for any function $g : [n] \to \{0,1\}$, the fraction of sets S on which $g(S) \overset{\leqslant 0.9k}{\neq} F[S]$ is at most $\frac{k}{n}$. Thus, for $k' = \Theta(k)$, and $k = n^{1-\varepsilon}$, the claim says there is no global structure even if the test passes with probability $\Omega(1)$. In our case, though, this claim does not give any meaningful conclusion, as the quantity $\frac{k}{n} = q$, a large constant.

In order to bring down the soundness of the test (compared to the quantity 2^k, which is the alphabet size), Impagliazzo, Kabanets, and Wigderson gave a 3-query test that has soundness $\exp(-k^\alpha)$ for some $\alpha > 0$. They also gave a different proof of the 2-query test from [DG08] and obtained similar results. Dinur and Livni Navon [DLN17] improved the soundness of the 3-query test to $\varepsilon = \exp(-\Omega(k))$ when $N \gg k$ ($N > 2^{\Omega(k)}$). In the latter result, the global function approximately agrees with F on at least $\varepsilon - 4\varepsilon^2$ fraction of the sets. Here, the approximate agreement can be taken as an all but arbitrary small constant fraction of the coordinates in S.

Recently, Dinur, Filmus and Harsha [DFH19] analyzed the 2-query test in the 99% regime to get a stronger conclusion. More specifically, they showed that if the test passes with probability at least $1 - \varepsilon$ for a sufficiently small constant $\varepsilon > 0$, then there is a global function g such that for at least $1 - O(\varepsilon)$ fraction of the sets S, $F[S] = g(S)$. Note that in the conclusion, they get a stronger agreement with the global function. They also gave a higher-dimension version of the direct product test where $F[S]$ represents a degree d functions (as opposed to linear functions) on the variables in S. In the same work [DFH19], the authors use this direct product test to get a 4-query linearity test over a biased measure on the hypercube.

1.5 Techniques

In this section, we give the proofs overview of the three theorems mentioned in the introduction.

1.5.1 Proof overview of Theorem 1.1

By the hypothesis of the theorem, we know that after a random restriction, the function f is correlated with a linear function with non-negligible probability. If we put a further restriction on the function, then the (further) restricted function stays correlated with the same linear function with non-negligible probability. We use this fact to conclude that the correlated linear function is independent of the actual restriction, i.e., it depends on the subset being restricted but independent of the settings to the variables in the subset. Once we establish this structure, we show that for different subsets I_1 and I_2 that intersect at many locations, the corresponding linear functions are similar on the domain $\{0,1\}^{I_1 \cap I_2}$. We exploit this structure further by using our direct product theorem to conclude that f is correlated to a global nearly-linear function. We now explain each of these parts in more detail.

We denote $I \sim_p [n]$ the choice a random subset of $[n]$ that results from including each element from $[n]$ in it with probability p. Let $\chi_S(x) := \prod_{i \in S} (-1)^{x_i}$ the multiplicative character over the uniform measure.

Step I: Local linear structure. Suppose we have a function f as in the statement of Theorem 1.1. By the premise, we know that choosing a random restriction $I \sim_{1-\beta} [n]$ and $z \sim {\mu'}^I$, the restricted function $f_{I \to z}$ has a significant Fourier coefficient $S_{I,z}$ with noticeable probability. A priori, it may be the case that even if we fix the set of restricted coordinates I, for each z we would get a completely different and unrelated character $S_{I,z}$, and the first step in our argument is to show that this cannot be the case over all I.

Towards showing that $S_{I,z}$ typically does not depend on z, we consider a heavier random restriction in which we first choose I as above, then $I' \sim_{1/2} \bar{I}$, and randomly restrict the coordinates of $I \cup I'$ according to a measure μ'', after which the underlying measure of $f_{I\to z, I'\to z'}$ is still the uniform measure; in other words, z' is chosen uniformly from $\{0,1\}^{I'}$. Since after the restriction $I \to z$ we already have a heavy Fourier coefficient $S_{I,z}$ with noticeable probability, it follows that $f_{I\to z, I'\to z'}$ also has a heavy Fourier coefficient, namely $S_{I,z} \cap \bar{I}'$, with noticeable probability. Note that the identity of this coefficient now does not depend on the setting of z'. At the same time, when we view the common random restriction $\tilde{I} \to \tilde{z}$ that combines I and I', there is no longer "separation" of what is the I-part and what is the I'-part, and this allows us to argue that the identity of $S_{I,z}$ does not really depend on z. Formally, for this step we use the small set expansion property of the hypercube.

Step II: Local consistency. Thus, we can think that for each I, we have a list of heavy coefficients, $\tilde{W}_I$ that capture all of the heavy coefficients that may occur when we randomly restrict the coordinates of I. Using a list-decoding type version of the argument above, we show that together, all $S \subseteq \bar{I}$ that are individually only rarely a heavy coefficient of a random restriction of f on I, even together do not contribute much to the probability that a restriction of f has a significant Fourier coefficient. Using this fact, we are able to establish that the lists $\tilde{W}_I$ must have certain local consistency properties. Roughly speaking, we show that if we choose I_1, I_2 randomly that intersect on $(1-\beta)$ of their elements (for suitably chosen $\beta > 0$), with significant probability we have a pair of compatible characters in the lists of I_1, I_2. That is, with significant probability we will be able to find $S_1 \in \tilde{W}_{I_1}$ and $S_2 \in \tilde{W}_{I_2}$ such that $S_1 \cap \overline{I_1 \cup I_2} = S_2 \cap \overline{I_1 \cup I_2}$. Clearly, such property would happen if there was a global character $S \subseteq [n]$ such that many of the lists $\tilde{W}_I$ contain $S \cap \bar{I}$, and the intuition suggests that this is the only way to create such a situation. In the next part of the argument, we use a direct product theorem, namely Corollary 1.5, to carry out such an argument.

Step III: Invoking the direct product testing theorem. We show that on top of being locally consistent, the lists $\tilde{W}_I$ are also bounded, hence we may define an assignment G to the I's that to each I selects randomly a character $F[I] \in \tilde{W}_I$. Having defined F, we observe that the local consistency of the lists translates to the fact that the assignment $G[A] = F[\overline{A}]$ passes the direct product test with significant probability. Thus, we may invoke Corollary 1.5 to deduce that there exists $S \subseteq [n]$ for which, for a significant fraction of the I's, $|F[I]\Delta(S \cap I)| = O(1)$.

Step IV: Deducing the correlation with a global nearly-linear function. Stated otherwise, the last conclusion asserts that after random restriction, with significant probability the function $f_{I\to z}$ is correlated with a function of the form $\chi_{S'}$ for S' such that $|S'\Delta(I \cap S)| = O(1)$. Thus, the function $\chi_{I\cap S} \cdot f_{I\to z}$ has significant mass on the low-degree part, and is hence not noise sensitive – i.e. it has stability bounded away from 0. Thus, the function $(\chi_S \cdot f)_{I\to z}$ (which up to a sign is the same as the previous function) is somewhat noise stable with significant probability over the choice of I and z, which allows us to deduce via Lemma 2.8 that the function $\chi_S \cdot f$ is somewhat noise stable, and hence is correlated with its low-degree part.

1.5.2 Proof overview of Theorem 1.3: linearity testing over a biased hypercube

We begin with an overview of the proof of Theorem 1.3 and the overall idea is as follows. We know that [BCH+96] if the function passes the linearity test with probability $1/2 + \varepsilon$ under the uniform measure, then the function is correlated with a linear function. In order to use this structure, we first do a certain random restriction on a subset of coordinates such that for the rest of the coordinates, our test queries are distributed uniformly. Now, using the linearity testing over the uniform measure, we can conclude that the restricted functions are correlated with a linear function. At this point, we use our Theorem 1.1 to conclude that the original function must be correlated with a product of a linear function and a low-degree polynomial.

In order to get rid of the low-degree polynomial from the conclusion, we design the test carefully so that its contribution in the final correlation is negligible. We now explain how to achieve these high-level ideas in more detail.

Step I: From linearity testing to large Fourier coefficients under random restrictions. Suppose that we are given a function $f\colon (\{0,1\}^n, \mu_q^{\otimes n}) \to \{-1,1\}$ satisfying the premise of Theorem 1.3, i.e. such that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}[f(x)f(y)f(z)f(w)]\right| \geqslant \delta. \tag{5}$$

Using standard averaging arguments, after choosing restrictions $f_{I\to a}$, $f_{I\to b}$, $f_{I\to c}$, $f_{I\to d}$ in a correlated manner that changes the underlying measure to be uniform, with significant probability we get that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu'^{[n]\setminus I}}[f_{I\to a}(x)f_{I\to b}(y)f_{I\to c}(z)f_{I\to d}(w)]\right| \geqslant \frac{\delta}{2},$$

where ν' is the uniform distribution over $(x,y,z,w) \in \{0,1\}^4$ such that $x+y+z+w=0$. Thus, using the standard Fourier analytic analysis of the test over the uniform measure, we conclude that with significant probability the function $f_{I\to a}$ has a heavy Fourier coefficient. Invoking Theorem 1.1 we conclude that f is correlated with a function of the form $\chi_S \cdot g$, where g is a low-degree function, and moreover g takes the form $g = (\chi_S f)^{\leqslant d}$ for some $d = O_\delta(1)$.

Step II: The list decoding argument. We would like to argue that since f is correlated with $\chi_S \cdot g$, we can "switch" one of the f's above with $\chi_S \cdot g$, and still get that the expectation in (5) is significant. To carry out such argument, we require a list-decoding version of the previous argument. Namely, we need to find a list of functions $\chi_{S_1} \cdot g_1, \ldots, \chi_{S_m} \cdot g_m$ that are all correlated with f and furthermore that "explain" all of the advantage of the expectation in (5), in the sense that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[(f - \sum_{i=1}^{m}\chi_{S_i}\cdot g_i)(x)f(y)f(z)f(w)\right]\right| = o(1). \tag{6}$$

Such arguments are rather easy to carry out in the uniform measure, however in our setting we are facing two additional challenges. First, since our decoding procedure above is not very simple, we are only able to apply it in a black-box way, so if we want to apply it iteratively we have to be careful so that the functions we work with satisfy the prerequisites of our basic decoding procedure. In our situation, this amounts to the functions not having too large 2-norm. Second, in contrast to the standard hypercube, the functions $\chi_{S_i} \cdot g_i$ need not be orthogonal hence there is no "natural" bound on the list size m. Indeed, such bound is simply false, so one cannot simply take all of the functions $\chi_{S_i} \cdot g_i$ that are correlated with f.

We overcome these challenges by allowing some flexibility in the degree of g_i's and in the level of correlation we require. Roughly speaking, the idea is that for $\chi_{S_1} g_1$ and $\chi_{S_2} g_2$ to be correlated, the characters S_1, S_2 must be close to each other (in the sense that $|S_1 \Delta S_2|$ is small). Thus, as g_1 is the low-degree part of $f \cdot \chi_{S_1}$ and g_2 is the low-degree part of $f \cdot \chi_{S_2}$, we expect these to overlap, and so if we "increase" the degree in which we truncate, we expect the function $\chi_{S_1} g_1$ to already include in it all of the mass of $\chi_{S_2} g_2$, and so we would be able to drop $\chi_{S_2} g_2$ from the list.

After carefully doing this argument, we are indeed able to find a bounded m and a list $\chi_{S_1} g_1, \ldots, \chi_{S_m} g_m$ as above so that (6) holds. This means that for some i, we get that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}[(\chi_{S_i}\cdot g_i)(x)f(y)f(z)f(w)]\right| \geqslant \Omega(\delta/m),$$

and we have effectively switched one of the f's into a function with the desired structure. Repeating this argument a few more times, we find $S_1, \ldots, S_4$ and $g_1, \ldots, g_4$ of low degree given as $g_i = (\chi_{S_i} f)^{\leqslant d_i}$ for some $d_i = O_\delta(1)$ such that

$$\left| \mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[(\chi_{S_1} \cdot g_1)(x)(\chi_{S_2} \cdot g_2)(y)(\chi_{S_3} \cdot g_3)(z)(\chi_{S_4} \cdot g_4)(w)\right] \right| \geqslant \delta'. \tag{7}$$

Step III: The invariance principle argument. Letting $T = S_1 \cap S_2 \cap S_3 \cap S_4$, we show that unless all of the S_i's are almost equal to T (in the sense that $|S_i \Delta T| = O(1)$), the above expectation is small. Hence, we get that each one of the S_i's is close to T, and for simplicity of presentation in this overview, we assume that $S_i = T$ for all i. Thus, as $\chi_T(x)\chi_T(y)\chi_T(z)\chi_T(w) = 1$ in the support of ν, it follows that

$$\left| \mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[g_1(x)g_2(y)g_3(z)g_4(w)\right] \right| \geqslant \delta'.$$

In other words, we have reduced the original problem of studying the structure of functions f that have an advantage in the linearity test over μ_q to the same problem, except that now the functions $g_1, \ldots, g_4$ are low-degree. The slight caveat here is that while f's were bounded (in fact, Boolean), the g_i's are not, however this is easy to fix, and we show that instead of using degree truncations, one can apply a suitable noise operator and still get an inequality as above. Thus, for the sake of this overview, we think of g_i's as low-degree bounded functions.

It can be shown that if f is not correlated with any χ_S, then the average of g_i is close to 0, even after restricting any set of $O(1)$ many coordinates. Thus, using standard regularity arguments, we can show that there is a set of coordinates T' of size $O(1)$ such that after restricting them, the restrictions of $g_1, \ldots, g_4$ all have small low-degree influence and still have averages close to 0. In this case, we are able to appeal to the invariance principle [MOO05], and more specifically to a version from [Mos10]. For the sake of simplicity of presentation, we ignore the restriction of T' for now, so that the invariance principle implies that the value of $\mathbb{E}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[g_1(x)g_2(y)g_3(z)g_4(w)\right]$ is close to the value of an expectation of the form $\mathbb{E}_{(z^1,z^2,z^3,z^4)\sim\tilde{\nu}^{\otimes n}}\left[P_1(z^1)P_2(z^2)P_3(z^3)P_4(z^4)\right]$, where $P_1, \ldots, P_4 \colon \mathbb{R}^n \to [-1, 1]$ are functions over Gaussian space with the same average as $g_1, \ldots, g_4$, and $\tilde{\nu}$ is a distribution of jointly distributed Gaussian random variables with the same pairwise correlations as of ν. However, ν is pairwise independent (this is the only place in which we use this fact), so the last Gaussian expectation is easy to compute and is just equal to the product of averages of $P_1, \ldots, P_4$, which is 0. This is a contradiction to (7), and so it is not possible that f is not correlated with any of χ_S, completing the overview of the proof.

1.5.3 Proof overview of Theorem 1.4: direct product testing

In the 99% regime, in order to come up with the global function that agrees with the given table F, in most cases, just taking the majority vote works. More formally, if we define the function $g : [n] \to \{0, 1\}$ by setting $g(i) = \mathsf{Majority}_{S, S\ni i} F[S]|_i$, then this g will have the property that it will approximately agree with F on almost all of the domain $\binom{[n]}{qn}$. Such a proof strategy was shown to work [DR06, DFH19] in the high acceptance regime of the direct product tests.

This above strategy, however, fails badly in the 1% regime. To see this, for every S, define $F[S]$ to be a random element from $\{0^{qn}, 1^{qn}\}$ with equal probability. It is easy to see that F will pass the test with probability $1/2$. On the other hand, the function g defined by taking the majority vote, looks like a random function and hence is very far from the table F.

Step I: Getting the local structure. One of the frameworks that was very successful in analyzing various direct product tests in the 1% regime is from the work of Impagliazzo, Kabanets, and Wigderson [IKW12]. This framework, that we will explain next, has been used in [DLN17, BDN17] to analyze various agreement tests. As seen before, although taking the majority vote among all the sets containing i does not work, we can define functions that have agreement with F *locally*. More specifically, given a subset S and an assignment $\sigma \in \{0,1\}^{qn}$, if we define a function $g_{S,\sigma} : [n] \to \{0,1\}$ by setting $g_{S,\sigma}(i) = \mathsf{Majority}_{S', S' \ni i, F[S']|_{S\cap S'} = \sigma|_{S'\cap S}} F[S']|_i$, then at least for the earlier example, one of the gs will end up being the all 0s function and will have agreement with the table F. In other words, we define the function by taking the majority vote *only* among the sets that are consistent with the given pair (S, σ).

This intuition can be made to work even when the test passes with probability $\varepsilon > 0$ where ε is a small constant, or even a sub-constant. However, in general, the functions $g_{S,\sigma}$ agree with the table F on only a $o(1)$-fraction of the domain. Recall, we are interested in finding a global function g that agrees with F on at least $\delta(\varepsilon)$ fraction of the domain for some fixed function δ independent of n.

Step II: Stitching different local functions. To remedy this, the next important component in the framework is to *stitch* these local functions $g_{S,\sigma}$ to come up with a global function g that has the required property. In our set-up, we differ from the previous work in this step of stitching different local functions. If we define the domain $\mathcal{C}_{S,\sigma} \subseteq \binom{[n]}{qn}$ as those sets of size qn on which the function $g_{S,\sigma}$ agrees with the table $F[.]$, then one way to show that these different functions $g_{S,\sigma}s$ are similar to each is to show that the families $\mathcal{C}_{S,\sigma}$ and $\mathcal{C}_{S',\sigma'}$ have many sets in common for a typical (S, σ) and (S', σ'). This would be enough to conclude that $g_{S,\sigma} \approx g_{S',\sigma'}$ and then get the final required global structure. This was shown to work in [DG08, IKW12] where the set sizes $qn = o(\sqrt{n})$, i.e., when $q = o(1/\sqrt{n})$.

The difficulty that arises in our setting of the parameters is that the sets S are of size $\Theta(n)$ and hence we cannot directly show that for a typical pair (S, σ) and (S', σ'), the corresponding functions agree with each other. We can, however, show that for a typical (S, σ), there are many $(\tilde{S}, \sigma')$, where $\tilde{S}$ is a slight perturbation of the set S resulting in changing a constant fraction of the coordinates in S, such that the families $\mathcal{C}_{S,\sigma}$ and $\mathcal{C}_{\tilde{S},\sigma'}$ have many sets in common. From this, we can conclude that the functions $g_{S,\sigma} \overset{\leqslant O(1)}{\neq} g_{\tilde{S},\sigma'}$ for a typical (S, σ). This still is not enough to guarantee an existence of the global function that agrees with the table F on $\delta(\varepsilon)$ fraction of the domain and the reason is that we could only show the approximate equality between $g_{S,\sigma}$ and $g_{\tilde{S},\sigma'}$ where $\tilde{S}$ is correlated to S.

Step III: Using the small-set expansion property. In order to break the correlation between the pairs (S, σ) and $(\tilde{S}, \sigma')$ for which we could show $g_{S,\sigma} \approx g_{\tilde{S},\sigma'}$, we use the small set expansion property of a certain graph defined on the multi-slice $\{0, 1, 2\}^n$. Note that from the approximate equality $g_{S,\sigma} \overset{\leqslant O(1)}{\neq} g_{\tilde{S},\sigma'}$, we have

$$\mathop{\mathbb{E}}_{(S,\sigma),(\tilde{S},\sigma')} \left[\mathop{\mathbb{E}}_{T\subseteq[n],|T|=n/C} \left[\mathbf{1}_{g_{S,\sigma}(T) = g_{\tilde{S},\sigma'}(T)} \right] \right] \geqslant \varepsilon^{O(1)},$$

where C is a large constant depending on the approximate equality of the functions $g_{S,\sigma}$ and $g_{\tilde{S},\sigma'}$. This gives,

$$\mathop{\mathbb{E}}_{T\subseteq[n],|T|=n/C} \left[\mathop{\mathbb{E}}_{(S,\sigma),(\tilde{S},\sigma')} \left[\mathbf{1}_{g_{S,\sigma}(T) = g_{\tilde{S},\sigma'}(T)} \right] \right] \geqslant \varepsilon^{O(1)},$$

Now for a typical subset T, we define a graph on (S, σ) where the edges are given by the distribution in the above expectation.[3] We partition the vertex set based on the values of $g_{S,\sigma}(T)$. Then it is possible that all the parts in the partition are small but still the above expectation is large, unless the graph is a small set expander. The graph in our case turns out to be a small-set expander and hence we can conclude that one of the parts in the partition is large and therefore, we can break the correlation to conclude that

$$\mathop{\mathbb{E}}_{T\subseteq[n],|T|=n/C}\left[\mathop{\mathbb{E}}_{(S,\sigma),(S'',\sigma'')}\left[\mathbf{1}_{g_{S,\sigma}(T)=g_{S'',\sigma''}(T)}\right]\right] \geqslant \delta(\varepsilon),$$

for some function δ of ε. From this, we conclude that $g_{S,\sigma} \overset{\leqslant O(1)}{\neq} g_{S'',\sigma''}$ happens with probability $\delta(\varepsilon)$ for a random pairs (S, σ) and (S'', σ''). This shows that a constant fraction of these local function $g_{S,\sigma}$ are close to each other and hence there is a global function that (approximately) agrees with the table F on a constantly many sets in the domain.

2 Preliminaries

In this section we introduce some basic tools used throughout the paper, mostly from analysis of Boolean functions. We refer the reader to [O'D14] for a more thorough introduction and discussion.

Notations. We denote $I \sim_p [n]$ the choice a random subset of $[n]$ that results from including each element from $[n]$ in it with probability p. Here and throughout, we denote by $\chi_S(x) = \prod_{i\in S}(-1)^{x_i}$ the multiplicative character over the uniform measure. Later on, when we discuss character over the q-biased measures we will denote it by $\chi_S^q(x) = \frac{x_i - q}{\sqrt{q(1-q)}}$. We use big-$O$ notations, meaning that the notation $f = O(g)$ says that $f \leqslant C \cdot g$ where $C > 0$ is an absolute constant, and $f = \Omega(g)$ says that $f \geqslant cg$ where $c > 0$ is an absolute constant. To simplify keeping track of various parameters, we shall use the notation $0 < a \ll b \ll c \leqslant 1$ to say that first c is chosen, then b is chosen sufficiently smaller compared to c, and then a is chosen sufficiently small with respect to a.

2.1 The Efron-Stein decomposition

Throughout the paper, we will be dealing with product probability measures over the Boolean hypercube, i.e. $(\{0,1\}^n, \mu = \mu_1 \times \ldots \times \mu_n)$, and mostly with the case that each one of the μ_i's is the q-biased distribution.

Given any product space $(\Omega = \Omega_1 \times \ldots \times \Omega_n, \mu = \mu_1 \times \ldots \times \mu_n)$, one may consider the space of real-valued functions $L_2(\Omega = \Omega_1 \times \ldots \times \Omega_n, \mu = \mu_1 \times \ldots \times \mu_n)$ equipped with the inner product

$$\langle f, g\rangle = \mathop{\mathbb{E}}_{x\sim\mu}[f(x)g(x)]$$

for all $f, g \colon \Omega \to \mathbb{R}$.

The Efron-Stein decomposition of a function $f \colon \Omega \to \mathbb{R}$ is a natural orthogonal decomposition of f that is often convenient to use. Here, for each $S \subseteq [n]$ we define the space $V^{\subseteq S}$ of functions over Ω that depend only on coordinates from S, and then $V^{=S} = V^{\subseteq S} \cap \bigcap_{S'\subsetneq S} V^{\subseteq S'\perp}$, which is the space of functions

[3] In the actual argument, we do not need σ and we view $S = A \cup B$ where $A \cap B = \emptyset$. Hence we use the multi-slice $\{0, 1, 2\}^n$ to represent the vertices. For instance, $S = A \cup B$ is represented by a string $\mathbf{x}$ where $x_i = 1$ if $i \in A$, $x_i = 2$ if $i \in B$ and $x_i = 0$ otherwise.

depending only on coordinates from S and orthogonal to any function that depends on less coordinates. With respect to this, we denote by $f^{=S} \in V^{=S}$ the projection of f to $V^{=S}$, so that

$$f = \sum_{S \subseteq [n]} f^{=S}.$$

Given this decomposition, one can verify that the Parseval and Plancherel identities hold, i.e. that

$$\langle f, g \rangle = \sum_{S \subseteq [n]} \langle f^{=S}, g^{=S} \rangle, \qquad \|f\|_2^2 = \sum_{S \subseteq [n]} \left\| f^{=S} \right\|_2^2.$$

The degree decomposition. Sometimes, it will be convenient for us to consider the coarser degree decomposition $f = \sum_{d=0}^{n} f^{=d}$, wherein we define $f^{=d} = \sum_{|S|=d} f^{=S}$. We also define $f^{\leqslant d} = \sum_{i=0}^{d} f^{=i}$, and refer to $f^{\leqslant d}$ as the degree d part of f. The degree of f, denoted by $\deg(f)$, is defined to be the largest d so that $f^{=d} \neq 0$.

Definition 2.1. *The degree d weight of a function $f\colon (\Omega, \mu) \to \mathbb{R}$ is defined as $W^{=d}[f] = \left\| f^{=d} \right\|_2^2$. The weight of f up to degree d is defined as $W^{\leqslant d}[f] = \left\| f^{\leqslant d} \right\|_2^2$.*

It is easy to see, by orthogonality of the $f^{=i}$'s, that $W^{\leqslant d}[f] = \sum_{i=0}^{d} W^{=i}[f]$.

2.2 Influences

Influences are a central notion in analysis of Boolean functions, and our arguments use the notions of influences as well as low-degree influences.

Definition 2.2. *For a function $f\colon (\Omega = \Omega_1 \times \ldots \times \Omega_n, \mu = \mu_1 \times \ldots \times \mu_n) \to \mathbb{R}$ and $i \in [n]$, the influence of the ith coordinate is defined to be as follows. Sample $x \sim \mu$, and then sample y by taking $y_j = x_j$ for all $j \neq i$ and sampling $y_i \sim \mu_i$ independently; we define*

$$I_i[f] = \mathop{\mathbb{E}}_{x,y} \left[(f(x) - f(y))^2 \right].$$

Subsequently, the low-degree influence of a function f is defined as

Definition 2.3. *For a function $f\colon (\Omega = \Omega_1 \times \ldots \times \Omega_n, \mu = \mu_1 \times \ldots \times \mu_n) \to \mathbb{R}$, $d \in \mathbb{N}$ and $i \in [n]$, the degree d influence of the ith coordinate is defined to be $I_i^{\leqslant d}[f] = I_i[f^{\leqslant d}]$.*

2.3 Fourier decomposition

The Fourier decomposition is a refinement of the Efron-Stein decomposition that is available in some settings, such as the q-biased probability measure.

Definition 2.4. *Let $q \in (0,1)$, and denote $\sigma = \sqrt{q(1-q)}$ the standard deviation of a q-biased random coin. We define $\chi_i^q\colon \{0,1\} \to \mathbb{R}$ as*

$$\chi_i^q(x_i) = \frac{x_i - q}{\sigma}.$$

For $S \subseteq [n]$, we define $\chi_S^q\colon \{0,1\}^n \to \mathbb{R}$ by $\chi_S^q(x) = \prod_{i \in S} \chi_i^q(x_i)$.

For the q-biased measure, one can show that for $f\colon (\{0,1\}^n, \mu_q^{\otimes n}) \to \mathbb{R}$, it holds that $f^{=S}(x) = \widehat{f}(S;\mu_q)\chi_S^q(x)$ where $\widehat{f}(S;\mu_q)$ is called the Fourier coefficient of f with respect to S and is given by

$$\widehat{f}(S;\mu_q) = \langle f, \chi_S^q \rangle.$$

2.4 Random restrictions

In this section, we define the notions of restrictions and of random restrictions that will be extensively used in the paper. Since the focus of current paper is on the Boolean hypercube with a biased measure, we restrict our discussion to this domain.

Given a function $f\colon (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$, a set of coordinates $I \subseteq [n]$ and a partial input $z \in \{0,1\}^I$, the restricted function $f_{I\to z}\colon \{0,1\}^{[n]\setminus I} \to \mathbb{R}$ is defined as

$$f_{I\to z}(y) = f(x_I = z, x_{\bar{I}} = y).$$

Here and throughout, we denote by $(x_I = z, x_{\bar{I}} = y)$ the point whose I-coordinates are set according to z, and whose $\bar{I}$ coordinates are set according to y.

A random restriction of a function $f\colon (\{0,1\}^n \mu^{\otimes n}) \to \mathbb{R}$ refers to a restriction in which either (or both) I and z are chosen randomly. Typically, when one says random restriction one has a parameter $\alpha \in (0,1)$, chooses $I \subseteq [n]$ by including each element $i \in [n]$ independently with probability α, choosing $z \sim \mu^I$ and then considering the function $f_{I\to z}$ as a function from $(\{0,1\}^{[n]\setminus I}, \mu^{[n]\setminus I})$ to $\mathbb{R}$. For us, however, it will be important to consider a more general notion of random restriction, in which the underlying measure of the restricted function changes.

Suppose that the measure μ can be written as $\mu = \beta\mathcal{D}_1 + (1-\beta)\mathcal{D}_2$, where $\mathcal{D}_1$ and $\mathcal{D}_2$ are distributions and $\beta \in (0,1)$. In such situations (that have already appeared in the introduction), we will often consider the following random restriction process: choose $I \subseteq [n]$ by including each element $i \in [n]$ in it with probability β, choose $z \sim \mathcal{D}_1^I$, and consider the function $f_{I\to z}$ as a function from $(\{0,1\}^{[n]\setminus I}, \mathcal{D}_2^{[n]\setminus I})$ to $\mathbb{R}$. Note that under these random choices, choosing $y \sim \mathcal{D}_2^{[n]\setminus I}$, the distribution of the point $(x_I = z, x_{\bar{I}} = y)$ is still μ, hence this restriction process still makes sense.

Indeed, this restriction process and some of its properties has already appeared in previous works in this series [BKM22a, BKM22b], and it will also play a crucial role in this work. In a sense, it allows us to change distributions to other distributions that are more favorable to work with, so long as the supports of the distributions are the same. Indeed, a typical scenario wherein we use this idea is to go from some distribution over a domain to the uniform distribution over the same domain.

2.5 Noise Stability

In this section, we define the standard notion of noise stability and prove several basic properties of it.

Definition 2.5. *Let μ be a distribution over $\{0,1\}$, and let $\rho \in [0,1]$. For $x \in \{0,1\}$, a ρ-correlated bit $y \in \{0,1\}$ is sampled by taking $y = x$ with probability ρ, and otherwise sampling $y \sim \mu$ independently. We denote this distribution by $y \sim_{\rho,\mu} x$.*

Given a distribution μ over $\{0,1\}$ and $\rho \in [0,1]$, we denote by $\mathrm{T}_{\mu,\rho}\colon L_2(\{0,1\},\mu) \to L_2(\{0,1\},\mu)$ the corresponding averaging operator defined as $\mathrm{T}_{\mu,\rho} f(x) = \mathbb{E}_{y\sim_{\rho,\mu} x}\left[f(x)\right]$.

For multi-variate functions $f\colon (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$, one similarly defines ρ-correlated inputs; given $x \in \{0,1\}^n$, the distribution over $y \sim_{\mu^{\otimes n},\rho} x$ is sampled by taking, for each $i \in [n]$ independently, $y_i = x_i$

with probability ρ, and otherwise sampling $y_i \sim \mu$. The corresponding averaging operator $\mathrm{T}_{\mu^{\otimes n},\rho}$ is easily seen then to be the same as $\mathrm{T}_{\mu,\rho}^{\otimes n}$. When the measure μ and n are clear from context, we often omit them from the notation.

Definition 2.6. *Let μ be a distribution over $\{0,1\}$, let $\rho \in [0,1]$ and let $f\colon (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$ be a function. The noise stability of f with correlation parameter ρ is defined as*

$$\mathsf{Stab}_\rho(f;\mu^{\otimes n}) = \langle f, \mathrm{T}_\rho f\rangle = \mathop{\mathbb{E}}_{x\sim\mu, y\sim_\rho x}[f(x)f(y)].$$

When the measure is clear from context, we often abbreviate the stability notation, and simply write $\mathsf{Stab}_\rho(f)$.

Intuitively, for a function f which is noise stable, the values of $f(x)$ and $f(y)$ are correlated if x and y are correlated inputs. One way to generate correlated inputs x and y is to choose a common random restriction on a subset of coordinates, and sample the rest of the coordinates independently; the correlation of $f(x)$ and $f(y)$, after the random restriction then, may be associated with the bias the function has after random restriction. Indeed, the following lemma expresses the noise stability of f as a function of the empty Fourier coefficient of a random restriction of f (which captures its bias).

Lemma 2.7. *Let μ be a distribution over $\{0,1\}$, and let $f\colon (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$ be a function. Then $\mathsf{Stab}_{1-\kappa}(f) = \mathbb{E}_{I\sim_{1-\kappa}, z\sim\mu^I}\left[\widehat{f}_{I\to z}(\emptyset)^2\right]$*

Proof. Expanding the right hand side, we see it is equal to

$$\mathop{\mathbb{E}}_{I\sim_{1-\kappa}, z\sim\mu^I}\left[\mathop{\mathbb{E}}_{x,y\sim\mu^{\bar{I}}}[f(x,z)f(y,z)]\right].$$

Note that the joint distribution of (x,z) and (y,z) is $1-\kappa$ correlated, and so the result follows. $\square$

The following lemma is [BKM22b, Lemma 2.14], restated below. To interpret it, intuitively one should think of small noise stability $\mathsf{Stab}_{1-\kappa}(f) \leqslant \xi$ as saying that the degree of f is high (roughly $\log(1/\xi)/\kappa$). With this in mind, the lemma asserts that if a function f is high degree, then a random restriction of it is also high degree, albeit with some quantitative loss in the parameters.

Lemma 2.8. *There exists an absolute constant $c > 0$ such that the following holds. Let μ_1, μ_2 be distributions over $\{0,1\}$, $\alpha \in (0,1)$ and let $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$. Then $\mathbb{E}_{I\sim_\alpha [n], z\sim\mu_1^I}\left[\mathsf{Stab}_{1-\kappa}(f_{I\to z};\mu_2^{\bar{I}})\right] \leqslant \mathsf{Stab}_{1-c(1-\alpha)\kappa}(f)$.*

2.6 Small set expansion and hypercontractivity

Our arguments use the well-known hypercontractive inequality over the q-biased cube, stated below.

Theorem 2.9. *For every $r \in \mathbb{N}$ and $q \in (0,1)$ there is $C(q,r) > 0$ such that if $f\colon (\{0,1\}^n, \mu_q^{\otimes n}) \to \mathbb{R}$ is a function of degree at most d, then $\|f\|_r \leqslant C(q,r)^d \|f\|_2$.*

We will also use the following well known consequence of the hypercontractive inequality, asserting that a Boolean function $f\colon (\{0,1\}^n, \mu_q^{\otimes n}) \to \{0,1\}$ with small average has most of its mass on high levels.

Theorem 2.10. *For every $q \in (0,1)$, there is $c_q > 0$ such that the following holds. Suppose that a function $f\colon (\{0,1\}^n, \mu_q^{\otimes n}) \to \{0,1\}$ has average is at most $\zeta > 0$; then for $d = c_q \log(1/\zeta)$ it holds that*

$$W^{\leqslant d}[f] \leqslant \|f\|_2^3 \leqslant \sqrt{\zeta}\, \mathbb{E}[f].$$

In words, since the total spectral mass of f is $\|f\|_2^2 = \mathbb{E}[f]$ (since f is Boolean), Theorem 2.10 asserts that almost of the spectral mass of f lies above level d.

2.7 Markov Chains

Finally, we need the following result from [Mos10], showing that reversible connected Markov chains have a spectral gap. For us, we will identify a reversible Markov chain T over $[m]$ with the averaging operator it defines over $L_2([m]; \mu)$, where μ is the stationary distribution of T.

Lemma 2.11. *[[Mos10, Lemma 2.9]] Suppose that T is a reversible, connected Markov chain on $[m]$, in which the probability of each transition is at least α. Then $\lambda_2(T) \leqslant 1 - \frac{\alpha^2}{2}$.*

3 Proof of Theorem 1.1

This section is devoted for the proof of Theorem 1.1.

3.1 Auxiliary Facts

In this section, we prove a few basic facts about random restrictions and Fourier coefficients that were hinted in the proof overview, and will be used throughout the proof.

The following fact asserts that if a function $f\colon \{0,1\}^n \to \mathbb{R}$ has a heavy Fourier coefficient and a bounded 2-norm (over the uniform distribution), then after random restriction, it still has a heavy Fourier coefficient with noticeable probability.

Fact 3.1. *Suppose that $f\colon \{0,1\}^n \to \mathbb{R}$ is a function with $\|f\|_2 \leqslant 1$ and $\left|\widehat{f}(S)\right| \geqslant \delta$ for some S. Then for all $I \subseteq [n]$,*

$$\Pr_{a \in \{0,1\}^I}\left[\left|\widehat{f_{I\to a}}(S \cap \overline{I})\right| \geqslant \frac{\delta}{2}\right] \geqslant \frac{\delta^2}{4}.$$

Proof. Fixing I, we have

$$\widehat{f}(S) = \mathop{\mathbb{E}}_{a}\left[\chi_{S\cap I}(a)\widehat{f_{I\to a}}(S \cap \overline{I})\right],$$

so by the triangle inequality

$$\delta \leqslant \mathop{\mathbb{E}}_{a}\left[\left|\widehat{f_{I\to a}}(S \cap \overline{I})\right|\right].$$

On the other hand,

$$\mathop{\mathbb{E}}_{a}\left[\left|\widehat{f_{I\to a}}(S \cap \overline{I})\right|^2\right] \leqslant \mathop{\mathbb{E}}_{a}\left[\|f_{I\to a}\|_2^2\right] = \|f\|_2^2 \leqslant 1.$$

Hence, we get by the Paley-Zygmund inequality that

$$\Pr_{a}\left[\left|\widehat{f_{I\to a}}(S \cap \overline{I})\right| \geqslant \frac{\delta}{2}\right] \geqslant \left(1 - \frac{1}{2}\right)^2 \frac{\mathbb{E}_a\left[\left|\widehat{f_{I\to a}}(S \cap \overline{I})\right|\right]^2}{\mathbb{E}_a\left[\left|\widehat{f_{I\to a}}(S \cap \overline{I})\right|^2\right]} \geqslant \frac{\delta^2}{2}.$$

$\square$

The following fact is similar in spirit to Fact 3.1, except that the underlying measure of the function changes after random restriction. It asserts that if a function f is correlated with a character χ_S and has bounded 2-norm under some distribution, and we perform a random restriction that changes the underlying measure of the restricted function, then with noticeable probability the restriction of f is still correlated with some character χ_T.

Fact 3.2. *Let μ_1, μ_2 be distributions over $\{0,1\}$, $\alpha \in (0,1)$ and let $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$. Suppose that $f\colon (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$ is a function with $\|f\|_2 \leqslant 1$ and $|\mathbb{E}_x[f(x)\chi_S(x)]| \geqslant \delta$ for some S. Then*

$$\Pr_{I\sim_\alpha [n], a\sim\mu_1^I}\left[\left|\mathop{\mathbb{E}}_{x\sim\mu_2^{\overline{I}}}[f_{I\to a}(x)\chi_S|_{I\to a}(x)]\right| \geqslant \frac{\delta}{2}\right] \geqslant \frac{\delta^2}{4}.$$

Proof. We have

$$\widehat{f}(S) = \mathop{\mathbb{E}}_{I\sim_\alpha [n], a\sim\mu_1^I}\left[\mathop{\mathbb{E}}_{x\sim\mu_2^{\overline{I}}}[f_{I\to a}(x)\chi_S|_{I\to a}(x)]\right],$$

so by the triangle inequality

$$\delta \leqslant \mathop{\mathbb{E}}_{I\sim_\alpha [n], a\sim\mu_1^I}\left[\left|\mathop{\mathbb{E}}_{x\sim\mu_2^{\overline{I}}}[f_{I\to a}(x)\chi_S|_{I\to a}(x)]\right|\right].$$

On the other hand,

$$\mathop{\mathbb{E}}_{I\sim_\alpha [n], a\sim\mu_1^I}\left[\left|\mathop{\mathbb{E}}_{x\sim\mu_2^{\overline{I}}}[f_{I\to a}(x)\chi_S|_{I\to a}(x)]\right|^2\right] \leqslant \mathop{\mathbb{E}}_{I\sim_\alpha [n], a\sim\mu_1^I}\left[\mathop{\mathbb{E}}_{x\sim\mu_2^{\overline{I}}}\left[|f_{I\to a}(x)\chi_S|_{I\to a}(x)|^2\right]\right] = \|f\|_2^2 \leqslant 1.$$

Hence, the result follows again by the Paley-Zygmund inequality. $\square$

The third and last fact is an auxiliary statement in probability. It asserts that if we have independent random variables X and Y and an event E that depends on them that has a significant probability, then sampling $x^1, \ldots, x^{r_1} \sim X$ and $y^1, \ldots, y^{r_2} \sim Y$ all independently, the event that E holds for all pairs (x^i, y^j) for $1 \leqslant i \leqslant r^1$ and $1 \leqslant j \leqslant r^2$ has significant probability.

Fact 3.3. *Suppose X, Y are independent random variables, and E is an event depending on X, Y such that $\Pr_{x\sim X, y\sim Y}[E(x,y)] \geqslant \delta$. Then for all r_1, r_2,*

$$\Pr_{x^1,\ldots,x^{r_1}\sim X, y^1,\ldots,y^{r_2}\sim Y}\left[\bigcap_{i=1}^{r_1}\bigcap_{j=1}^{r_2} E(x^i, y^j)\right] \geqslant \delta^{r_1 r_2}.$$

Proof. By Jensen's inequality

$$\delta^{r_2} \leqslant \mathop{\mathbb{E}}_{x\sim X, y\sim Y}\left[1_{E(x,y)}\right]^{r_2} \leqslant \mathop{\mathbb{E}}_{x\sim X}\left[\mathop{\mathbb{E}}_{y\sim Y}\left[1_{E(x,y)}\right]^{r_2}\right] = \mathop{\mathbb{E}}_{x\sim X}\left[\mathop{\mathbb{E}}_{y^1,\ldots,y^{r_2}\sim Y}\left[\prod_{j=1}^{r_2} 1_{E(x,y^j)}\right]\right].$$

By Jensen's inequality again

$$\delta^{r_1 r_2} \leqslant \mathop{\mathbb{E}}_{x\sim X}\left[\mathop{\mathbb{E}}_{y^1,\ldots,y^{r_2}\sim Y}\left[\prod_{j=1}^{r_2} 1_{E(x,y^j)}\right]\right]^{r_1} = \mathop{\mathbb{E}}_{y^1,\ldots,y^{r_2}\sim Y}\left[\mathop{\mathbb{E}}_{x\sim X}\left[\prod_{j=1}^{r_2} 1_{E(x,y^j)}\right]\right]^{r_1}$$
$$\leqslant \mathop{\mathbb{E}}_{y^1,\ldots,y^{r_2}\sim Y}\left[\mathop{\mathbb{E}}_{x\sim X}\left[\prod_{j=1}^{r_2} 1_{E(x,y^j)}\right]^{r_1}\right]$$
$$= \mathop{\mathbb{E}}_{\substack{x^1,\ldots,x^{r_1}\sim X\\ y_1,\ldots,y^{r_2}\sim Y}}\left[\prod_{i=1}^{r_1}\prod_{j=1}^{r_2} 1_{E(x^i,y^j)}\right],$$

and the proof is concluded. □

3.2 Local Linear Structure

In this section, we begin the formal proof of Theorem 1.1, and first show that with each $I \subseteq [n]$ one may associate a set of characters which are the ones that can become heavy after randomly restricting the coordinates of I. Fix f as in Theorem 1.1; throughout the proof, we will have the parameters

$$0 \ll \kappa \ll s, r^{-1} \ll \zeta \ll \varepsilon \ll \xi \ll \delta, \eta \ll \beta < \alpha < 1.$$

For a set $I \subseteq [n]$ and $z \in \{0,1\}^I$, define

$$W_{I,z} = \left\{ S \subseteq \overline{I} \;\middle|\; \left|\widehat{f_{I\to z}}(S)\right| \geqslant \delta \right\}, \qquad \widetilde{W}_{I,z} = \left\{ S \subseteq \overline{I} \;\middle|\; \left|\widehat{f_{I\to z}}(S)\right| \geqslant \frac{\delta}{2} \right\},$$

where $\widehat{g}(S) = \mathbb{E}_x\left[g(x)\chi_S(x)\right]$. Note that by the premise of Theorem 1.1, we have that choosing $I \sim_{1-\beta} [n]$ and $z \sim \mu'^I$, we have that $W_{I,z} \neq \emptyset$ with probability at least η.

We now consider $I' \sim_{1-\beta/2} [n]$ and $z' \sim \mu''^I$, where $\mu'' = \frac{1-\beta}{1-\beta/2}\mu' + \frac{\beta/2}{1-\beta/2}U$. Then note that sampling I', z' can be done by sampling $I_1 \sim_{1-\beta} [n]$, $I_2 \sim_{1/2} [n] \setminus I_1$, $z(1) \sim \mu'^{I_1}$ and $z(2) \sim U^{I_2}$ and taking $I' = I_1 \cup I_2$ and $z' = z(1) \circ z(2)$. Then by our earlier observation, $W_{I_1,z(1)} \neq \emptyset$ with probability at least η; we condition on this event and take some $S \in W_{I_1,z(1)}$, thus getting from Fact 3.2 that $\left|\widehat{f}_{I'\to z'}(S \cap \overline{I_2})\right| \geqslant \delta/2$ with probability at least $\delta^2/2$, and so we get that

$$\Pr_{\substack{I_1,I_2\\ z(1),z(2)}}\left[S \cap \overline{I_2} \in \widetilde{W}_{I_1\cup I_2, z(1)\circ z(2)} \;\middle|\; S \in W_{I_1,z(1)}\right] \geqslant \frac{\delta^2}{2}.$$

Sampling I_2' independently of I_2, and $z(2), z(3)$ assignments for I_2 and $z(2)', z(3)'$ assignments for I_2' independently, we get by Fact 3.3 that

$$\Pr_{\substack{I_1,I_2,I_2'\\ z(1),z(2),z(3)\\ z(2)',z(3)'}}\left[\begin{smallmatrix} S\cap\overline{I_2}\in\widetilde{W}_{I_1\cup I_2,z(1)\circ z(2)}\cap\widetilde{W}_{I_1\cup I_2,z(1)\circ z(3)}\\ S\cap\overline{I_2'}\in\widetilde{W}_{I_1\cup I_2',z(1)\circ z(2)'}\cap\widetilde{W}_{I_1\cup I_2',z(1)\circ z(3)'}\end{smallmatrix} \;\middle|\; S \in W_{I_1,z(1)}\right] \geqslant \frac{\delta^8}{16}. \tag{8}$$

For each I', we define the set of $S \subseteq \overline{I'}$ that occur somewhat frequently as characters when restricting the coordinates of I':

$$W_{I'} = \left\{ S \subseteq \overline{I'} \mid \Pr_{z \sim \mu''^{I'}} \left[\left| \widehat{f_{I' \to z}}(S) \right| \geqslant \frac{\delta}{2} \right] \geqslant \zeta \right\}.$$

One can show that with significant probability over the choice of $I' \sim_{1-\beta/2} [n]$, the set collection $W_{I'}$ is non-empty, but we need the following stronger statement. It asserts that the probability that $\widetilde{W}_{I_1 \cup I_2, z(1) \circ z(2)} \cap \widetilde{W}_{I_1 \cup I_2, z(1) \circ z(2)'}$ intersect in T which is rare, i.e. such that $T \notin \widetilde{W}_{I_1 \cup I_2}$, is small.

Claim 3.4. *For all I', we have that*

$$\Pr_{\substack{I_1, I_2 : I_1 \cup I_2 = I' \\ z(1), z(2), z(3)}} \left[\exists T, T \in \widetilde{W}_{I_1 \cup I_2, z(1) \circ z(2)} \cap \widetilde{W}_{I_1 \cup I_2, z(1) \circ z(3)}, T \notin W_{I'} \right] \leqslant \xi.$$

Proof. For each $T \subseteq \overline{I'}$, define $X_T = \left\{ z' \in \{0,1\}^{I'} \,\middle|\, \left| \widehat{f_{I' \to z'}}(T) \right| \geqslant \frac{\delta}{2} \right\}$. We note that $T \in W_{I'}$ if and only if $\mu''(X_T) \geqslant \zeta$. We also note that:

$$\sum_T \mu''(X_T) = \sum_T \sum_{z'} \mu''(z') 1_{\left| \widehat{f_{I' \to z'}}(T) \right| \geqslant \frac{\delta}{2}} = \sum_{z'} \mu''(z') \sum_T 1_{\left| \widehat{f_{I' \to z'}}(T) \right| \geqslant \frac{\delta}{2}} \leqslant \sum_{z'} \mu''(z') \frac{\|f_{I' \to z'}\|_2^2}{(\delta/2)^2},$$

where in the last inequality we used Parseval. The last expression is equal to $\frac{\|f\|_2^2}{(\delta/2)^2} \leqslant \frac{4}{\delta^2}$.

Next, consider the distribution over $z' = z(1) \circ z(2)$ and $z'' = z(1) \circ z(3)$ as in (8). Note that this is a product distribution, in which independently for each $i \in I'$, with probability $(1-\beta)/(1-\beta/2)$ we take $z'_i = z''_i$ according to the distribution μ', and otherwise we take z'_i, z''_i independently according to U. We define the corresponding Markov chain $p_{a \to b} = \Pr\left[z''_1 = b \mid z'_1 = a \right]$, and note that it is connected, reversible and each transition has probability at least $\beta/2$. Thus, defining the corresponding averaging operator $\mathrm{T}\colon L_2(\{0,1\}, \mu'') \to L_2(\{0,1\}, \mu'')$, by Lemma 2.11 we have that $\lambda_2(\mathrm{T}) \leqslant 1 - \Omega(\beta^2)$.

Fix $T \notin W_{I'}$, so that $\mu''(X_T) < \zeta$. By Theorem 2.10, we get that for $d = \Omega_\beta(\log(1/\zeta))$ it holds that $W_{\leqslant d}[1_{X_T}; \mu''] \leqslant \varepsilon \mu''(X_T)$, hence

$$\langle 1_{X_T}, \mathrm{T}^{I'} 1_{X_T} \rangle \leqslant W_{\leqslant d}[1_{X_T}; \mu''] + \lambda_2(\mathrm{T})^d W_{>d}[1_{X_T}; \mu''] \leqslant \varepsilon \mu''(X_T) + (1 - \Omega(\beta^2))^d \mu''(X_T) \leqslant 2\varepsilon \mu''(X_T),$$

and summing over $T \notin W_{I'}$ gives

$$\sum_{T \notin W_{I'}} \langle 1_{X_T}, \mathrm{T}^{I'} 1_{X_T} \rangle \leqslant \sum_T 2\varepsilon \mu''(X_T) \leqslant \frac{8\varepsilon}{\delta^2} \leqslant \xi.$$

On the other hand, inspecting the left hand side, it is equal to

$$\begin{aligned}\sum_{T \notin W_{I'}} \langle 1_{X_T}, \mathrm{T}^{I'} 1_{X_T} \rangle = \sum_{T \notin W_{I'}} \mathop{\mathbb{E}}_{z', z''} \left[1_{z', z'' \in X_T} \right] &= \mathop{\mathbb{E}}_{z', z''} \left[\sum_{T \notin W_{I'}} 1_{T \in \widetilde{W}_{I', z'}} 1_{T \in \widetilde{W}_{I', z''}} \right] \\ &= \mathop{\mathbb{E}}_{z', z''} \left[\sum_T 1_{T \in \widetilde{W}_{I', z'} \cap \widetilde{W}_{I', z''}} 1_{T \notin W_{I'}} \right],\end{aligned}$$

which is at least the left hand side in the claim. The proof is thus concluded. $\square$

From the above claim we deduce the following claim, which asserts that choosing I_1 and independently I_2 and I_2', the collections $W_{I_1\cap I_2}$ and $W_{I_1\cap I_2'}$ contain *compatible* sets T and T' with noticeable probability.

Definition 3.5. *Let $I_1 \subseteq [n]$, and let $I_2, I_2' \subseteq [n] \setminus I_1$. We say that two sets $T \subseteq [n] \setminus (I_1 \cup I_2)$ and $T' \subseteq [n] \setminus (I_1 \cup I_2')$ are compatible if there is $S \subseteq [n]$ such that $T = S \cap \overline{I_1 \cup I_2}$ and $T' = S \cap \overline{I_1 \cup I_2'}$.*

Claim 3.6. $\Pr_{I_1,I_2,I_2'}\left[\exists S \subseteq [n], S \cap \overline{I_2} \in W_{I_1\cup I_2} \wedge S \cap \overline{I_2'} \in W_{I_1\cup I_2'}\right] \geqslant \frac{\eta\delta^8}{64}$.

Proof. Let E be the event in (8). Combining Claim 3.4 and (8), we get that

$$\Pr_{\substack{I_1,I_2,I_2'\\ z(1),z(2),z(3)\\ z(2)',z(3)'}}\left[\exists S \in W_{I_1,z(1)} : E \wedge S \cap \overline{I_2} \in W_{I_1\cup I_2} \wedge S \cap \overline{I_2} \in W_{I_1\cup I_2'}\right] \geqslant \frac{\delta^8}{16}\Pr\left[W_{I_1,z(1)} \neq \emptyset\right] - 2\xi,$$

and as the probability that $W_{I_1,z(1)}$ is non-empty is at least $\eta/2$, we get that the left hand side of the claim is at least

$$\Pr_{\substack{I_1,I_2,I_2'\\ z(1),z(2),z(3)\\ z(2)',z(3)'}}\left[\exists S \in W_{I_1,z(1)}, E \wedge S \cap \overline{I_2} \in W_{I_1\cup I_2} \wedge S \cap \overline{I_2} \in W_{I_1\cup I_2'}\right] \geqslant \frac{\delta^8}{16}\frac{\eta}{2} - 2\xi \geqslant \frac{\eta\delta^8}{64}.$$

□

Next, we show that each $|W_{I'}|$ is not too large.

Claim 3.7. *For all I', $|W_{I'}| \leqslant \frac{4}{\zeta\delta^2}$.*

Proof. Note that

$$\mathop{\mathbb{E}}_{z'}\left[\left|\left\{ S \mid 1_{S\in\widetilde{W}_{I',z'}}\right\}\right|\right] \geqslant \mathop{\mathbb{E}}_{z'}\left[\sum_{S\in W_{I'}} 1_{S\in\widetilde{W}_{I',z'}}\right] = \sum_{S\in W_{I'}} \mathop{\mathbb{E}}_{z'}\left[1_{S\in\widetilde{W}_{I',z'}}\right] \geqslant \zeta\,|W_{I'}|\,.$$

On the other hand,

$$\mathop{\mathbb{E}}_{z'}\left[\left|\left\{ S \mid 1_{S\in\widetilde{W}_{I',z'}}\right\}\right|\right] = \mathop{\mathbb{E}}_{z'}\left[\sum_{S} 1_{S\in\widetilde{W}_{I',z'}}\right] \leqslant \mathop{\mathbb{E}}_{z'}\left[\frac{\|f_{I'\to z'}\|_2^2}{(\delta/2)^2}\right] = \frac{\|f\|_2^2}{(\delta/2)^2} \leqslant \frac{4}{\delta^2},$$

and the result follows. □

Note that the distribution of $I_1 \cup I_2$ is $\sim_{1-\beta/2} [n]$, and we next want to define a function over such sets. We define $F\colon (P([n]), \mu_{1-\beta/2}^{\otimes n}) \to P([n])$ that assigns to each $I' \subseteq [n]$ a subset of $\overline{I'}$, denoted by $F[I']$, in the following way: for each input $I' \subseteq [n]$, consider $W_{I'}$. If it is non-empty, choose a random $T \in W_{I'}$ and set $F[I'] = T$. If it is empty, choose a random $T \subseteq \overline{I'}$ and output $F[I'] = T$. For convenience, we define $G\colon (P([n]), \mu_{p/2}^{\otimes n}) \to P([n])$ by $G[A] = F[[n] \setminus A]$, and note that $G[A] \subseteq A$ always.

We consider the following direct product test over the assignment G:

1. Choose $I_1 \sim_{1-\beta} [n]$ and independently $I_2, I_2' \sim_{1/2} \overline{I_1}$. Set $A = \overline{I_1 \cup I_2}$, $A' = \overline{I_1 \cup I_2'}$.

2. Take $T = G[A], T' = G[A']$.

3. Accept if $T \cap A \cap A' = T \cap A \cap A'$.

Claim 3.8. *Over the randomness of the choice of the assignment F, we have that*

$$\mathop{\mathbb{E}}_{F}\left[\Pr\left[\textit{Direct product test succeeds}\right]\right] \geqslant \frac{\eta\zeta^2\delta^{12}}{1024}.$$

Proof. By Claim 3.6, with probability at least $\frac{\eta\delta^8}{64}$ the collections $W_{I_1\cup I_2}$ and $W_{I_1\cup I_2'}$ contain a pair of compatible sets, call them T and T'. Conditioned on that, by Claim 3.7 the probability that $F[I_1 \cup I_2] = T$ and $F[I_1 \cup I_2'] = T'$ is at least $\left(\frac{\zeta\delta^2}{4}\right)^2$, in which case the direct product test between $I_1 \cup I_2$ and $I_1 \cup I_2'$ accepts. We conclude that with probability at least $\frac{\eta\delta^8}{64} \cdot \frac{\zeta^2\delta^4}{16}$ over the randomness of I_1, I_2, I_2' and F, the direct product test between $I_1 \cup I_2$ and $I_1 \cup I_2'$ accepts, and the claim is proved. □

It follows that with probability at least $\frac{\eta\zeta^2\delta^{12}}{2048}$ over the choice of randomness over the assignment F, the direct product test above succeeds with probability at least $\frac{\eta\zeta^2\delta^{12}}{2048}$. We fix such assignment F henceforth.

3.3 Applying the Direct Product Theorem

Using Corollary 1.5, we find S such that

$$\Pr_{A\sim_{\beta/2}[n]}\left[|G[A]\Delta S| \leqslant r\right] \geqslant s.$$

Next, we argue that this global consistency does not come from the A's that were randomly assigned. Let $\mathcal{A}_k$ be the set of $A \subseteq [n]$ of size k for which $\tilde{W}_A$ was empty. For each S, we note that by Chernoff's inequality, the probability that $|G[A]\Delta S| \leqslant r$ for more than $s/2$ fraction of A of size k is at most $2^{-\Omega_{r,s}(\binom{n}{k})}$ (since the events that the various A satisfy it are independent, and the probability of each one is exponentially small in n hence much smaller than s). Thus, by the union bound over all $S \subseteq [n]$ it follows that the probability this occurs for some S is at most $2^n 2^{-\Omega_{r,s}(\binom{n}{k})} \leqslant 2^{-\Omega_{r,s}(\binom{n}{k})}$, and by the Union bound over k it follows that the probability that there is k for which there is such S is at most $2^{-\Omega_{r,s}(\binom{n}{k})}$. Thus, it follows that we could have fixed the randomness of the choice of F so that F has the above property and also passes the direct product test with probability at least $\frac{\eta\zeta^2\delta^{12}}{2048}$, and doing so we conclude that then we have

$$\Pr_{A\sim_{\beta/2}[n]}\left[|G[A]\Delta S| \leqslant r, \tilde{W}_A \neq \emptyset\right] \geqslant \frac{s}{2}.$$

Define the function $g(x) = \chi_S(x)$ and consider $f'(x) = f(x)g(x)$. For A such that $|G[A]\Delta S| \leqslant r$ and $\tilde{W}_A$ is non-empty, choosing $A' \subseteq A$ by including each element $i \in A$ in A' with probability $\frac{\kappa}{r}$, we get that $G[A] \cap A' = S \cap A'$ with probability $1 - O(\kappa)$. As $G[A] \in W_{\overline{A}}$, when we choose $z \sim \mu''^{\overline{A}}$ with probability at least ζ we have $\left|\widehat{f}_{\overline{A}\to z}(G[A])\right| \geqslant \frac{\delta}{2}$, and so $\left|\widehat{f'}_{\overline{A}\to z}(G[A]\Delta S)\right| \geqslant \frac{\delta}{2}$ (note that we have switched from f to f'). Thus, choosing $z' \sim U^{A\setminus A'}$ we get that

$$\mathop{\mathbb{E}}_{A',z'}\left[\widehat{f'}_{\overline{A}\to z, A\setminus A'\to z'}(\emptyset)^2\right] \geqslant \mathop{\mathbb{E}}_{A',z'}\left[\widehat{f'}_{\overline{A}\to z, A\setminus A'\to z'}(G[A]\Delta S \cap A')^2\right] - \Pr_{A'}\left[(G[A]\Delta S) \cap A' \neq \emptyset\right],$$

which is at least $\Omega(\delta^2) - O(\kappa) \geqslant \Omega(\delta^2)$. On the other hand, by Lemma 2.7 the left hand side is equal to $\mathsf{Stab}_{1-\kappa}(f'_{\overline{A}\to z})$. Thus, we get from Lemma 2.8 that for some absolute constant $c > 0$ we have

$$\mathsf{Stab}_{1-c(1-\beta)\kappa}(f') \geqslant \mathop{\mathbb{E}}_{A,z}\left[\mathsf{Stab}_{1-\kappa}(f'_{\overline{A}\to z})\right] \geqslant \mathop{\mathbb{E}}_{A,z}\left[1_{G[A]\in W_{\overline{A}}} 1_{|G[A]\Delta S|\leqslant r} \mathsf{Stab}_{1-\kappa}(f'_{\overline{A}\to z})\right] \geqslant \Omega(s\delta^2).$$

This means that for $d = O\left(\frac{\log(1/s\delta^2)}{(1-\beta)\kappa}\right)$, we have that $W_{\leqslant d}[f'] \geqslant \Omega(s\delta^2)$, hence f' is $\Omega(s\delta^2)$-correlated with the function $f'' = f'^{\leqslant d}$, and therefore f is $\Omega(s\delta^2)$-correlated with the function gf'', as desired.

4 Proof of Theorem 1.3

In this section, we prove Theorem 1.3. Throughout this section, we fix ν to be the distribution from the setting of Theorem 1.3.

4.1 From Linearity Testing to Large Fourier Coefficients under Random Restrictions

The following lemma demonstrates the connection between Theorems 1.1 and Theorem 1.3, and we state it in a general form that will be necessary for our argument to go through. The lemma asserts that if $f_1, \dots, f_4$ are functions such that $f_2, \dots, f_4$ have a bounded 12-norm, and f_1 has a bounded 2-norm, for which $\left|\mathbb{E}_{(x,y,z,w)\sim\nu^n}\left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right|$ is bounded away from 0, then with significant probability, after a random restriction f_1 must be correlated with a character χ_S.

To be more precise, we fix a small enough absolute constant β, and note that we may write $\nu = (1-\beta)\nu' + \beta\mu$ where μ is the uniform distribution over $\{\,(a,b,c,d) \in \{0,1\}^n \mid a+b+c+d = 0\}$. Then

Lemma 4.1. *For all $\varepsilon > 0$ and $M \geqslant 1$ there is $\delta > 0$ such that the following holds. Suppose that $f_1, \dots, f_4 \colon \{0,1\}^n \to \mathbb{R}$ are functions with $\|f_1\|_2 \leqslant M$ and $\|f_i\|_{12} \leqslant M$ for $i = 2,3,4$; further suppose that*

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right| \geqslant \varepsilon.$$

Then, $\Pr_{\substack{I\sim_{1-\beta}[n]\\ z\sim\nu'^I}}\left[\exists S \subseteq \bar{I}, \left|\widehat{f_{1\,I\to z}}(S)\right| \geqslant \delta\right] \geqslant \delta$

Proof. By assumption, ε is at most

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right| &= \left|\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\mu'^I}}\left[\mathop{\mathbb{E}}_{(x,y,z,w)\sim\mu^{[n]\setminus I}}\left[f_{1\,I\to a}(x)f_{2\,I\to b}(y)f_{3\,I\to c}(z)f_{4\,I\to c}(w)\right]\right]\right| \\ &= \left|\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\sum_{S\subseteq\bar{I}}\widehat{f_{1\,I\to a}}(S)\widehat{f_{2\,I\to b}}(S)\widehat{f_{3\,I\to c}}(S)\widehat{f_{4\,I\to d}}(S)\right]\right|.\end{aligned}$$

Let E be the event that $|f_{1\,I\to a}(S)| \leqslant \xi$ for all S. Then we get that

$$\varepsilon \leqslant \mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\sum_{S\subseteq\bar{I}}\left(\xi 1_E + \left|\widehat{f_{1\,I\to a}}(S)\right| 1_{\overline{E}}\right)\left|\widehat{f_{2\,I\to b}}(S)\widehat{f_{3\,I\to c}}(S)\widehat{f_{4\,I\to d}}(S)\right|\right].$$

By AM-GM inequality,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\sum_{S\subseteq\bar{I}}\left|\widehat{f_{2I\to b}}(S)\widehat{f_{3I\to c}}(S)\widehat{f_{4I\to d}}(S)\right|\right]\\
&\leqslant\frac{1}{3}\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\sum_{S\subseteq\bar{I}}\left|\widehat{f_{2I\to b}}(S)\right|^3+\sum_{S\subseteq\bar{I}}\left|\widehat{f_{3I\to c}}(S)\right|^3+\sum_{S\subseteq\bar{I}}\left|\widehat{f_{4I\to d}}(S)\right|^3\right]\\
&\leqslant\frac{1}{3}\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\|f_{2I\to b}\|_1\|f_{2I\to b}\|_2^2+\|f_{2I\to c}\|_1\|f_{3I\to c}\|_2^2+\|f_{2I\to d}\|_1\|f_{4I\to d}\|_2^2\right]\\
&\leqslant\frac{1}{3}\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\|f_{2I\to b}\|_2^3+\|f_{3I\to c}\|_2^3+\|f_{4I\to d}\|_2^3\right]\\
&\leqslant\frac{1}{3}\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\|f_{2I\to b}\|_3^3+\|f_{3I\to c}\|_3^3+\|f_{4I\to d}\|_3^3\right],
\end{aligned}$$

which is equal to $\frac{1}{3}\left(\|f_2\|_3^3+\|f_3\|_3^3+\|f_4\|_3^3\right)\leqslant M^3$. Similarly, we get that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[1_{\bar{E}}\sum_{S\subseteq\bar{I}}\left|\widehat{f_{1I\to a}}(S)\widehat{f_{2I\to b}}(S)\widehat{f_{3I\to c}}(S)\widehat{f_{4I\to d}}(S)\right|\right]\\
&\leqslant\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[1_{\bar{E}}\|f_{1I\to a}\|_1\|f_{2I\to b}\|_1\sum_{S\subseteq\bar{I}}\left|\widehat{f_{3I\to c}}(S)\widehat{f_{4I\to d}}(S)\right|\right]\\
&\leqslant\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}[1_{\bar{E}}\|f_{1I\to a}\|_1\|f_{2I\to b}\|_1\|f_{3I\to c}\|_2\|f_{4I\to d}\|_2],
\end{aligned}$$

where in the last inequality we used Cauchy-Schwarz and Parseval. By Hölder's inequality we may bound the last expression as

$$\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[1_{\bar{E}}^4\right]^{1/4}\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\|f_{1I\to a}\|_1^{4/3}\|f_{2I\to b}\|_1^{4/3}\|f_{3I\to c}\|_2^{4/3}\|f_{4I\to d}\|_2^{4/3}\right]^{3/4},$$

which again using Hölder's inequality is at most

$$\Pr\left[\bar{E}\right]^{1/4}\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\|f_{1I\to a}\|_1^2\right]^{1/2}\mathop{\mathbb{E}}_{\substack{I\sim_{1-\beta}[n]\\(a,b,c,d)\sim\nu'^I}}\left[\|f_{2I\to b}\|_1^4\|f_{3I\to c}\|_2^4\|f_{4I\to d}\|_2^4\right]^{1/4}.$$

Bounding $\|f_{1I\to a}\|_1\leqslant\|f_{1I\to a}\|_2$ so that the first expectation is at most $\|f_1\|_2$ and using Hölder's inequality again on the rest, we get that the above expression is at most

$$\Pr\left[\bar{E}\right]^{1/4}\|f_1\|_2\|f_2\|_{12}\|f_3\|_{12}\|f_4\|_{12}\leqslant M^4\Pr\left[\bar{E}\right]^{1/4}$$

Combining, we conclude that

$$\varepsilon \leqslant \xi M^3 + \Pr\left[\bar{E}\right]^{1/4} M^4.$$

We take $\xi = \frac{\varepsilon}{2M^3}$, and get that $\Pr\left[\bar{E}\right] \geqslant \left(\frac{\varepsilon}{2M^4}\right)^4$, and the claim is proved for $\delta = \left(\frac{\varepsilon}{2M^4}\right)^4$. $\square$

Combining Lemma 4.1 and Theorem 1.1 we get the following corollary:

Corollary 4.2. *For all $\varepsilon > 0$ and $M \geqslant 1$ there are $d \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose that $f_1, \ldots, f_4\colon \{0,1\}^n \to \mathbb{R}$ are functions with $\|f_1\|_2 \leqslant M$ and $\|f_i\|_{12} \leqslant M$ for $i = 2, 3, 4$; further suppose that*

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right| \geqslant \varepsilon.$$

Then, there is $S \subseteq [n]$ such that f_1 is δ correlated with $\chi_S g$ where $g = (f_1\chi_S)^{\leqslant d}$.

In the following section, we strengthen this assertion to a list-decoding type statement.

4.2 The List Decoding Argument

As explained before, Corollary 4.2 implies that f_1 is δ-correlated with a function of the form $\chi_S g_1$ where $g_1 = (f_1 \cdot \chi_S)^{\leqslant d}$, where $d \in \mathbb{N}, \delta > 0$ depend only on ε and M. We would like a stronger statement, saying that all of the advantage in the expectation of $f_1, \ldots, f_4$ comes from such structures in f_1, and we show such statement in this section.

Define

$$\mathcal{S}_{d,\delta}(f) = \left\{ S \mid f \text{ is } \delta\text{-correlated with } \chi_S(f\chi_S)^{\leqslant d} \right\}.$$

Contrary to the case of the uniform distribution, the size of $\mathcal{S}_{d,\delta}(f)$ can be large and may even depend on the dimension n; this is possible because for $S_1, S_2 \subseteq [n]$ that have a small symmetric difference, the functions χ_{S_1} and χ_{S_2} are correlated. Thus, we replace the bound on the size of $\mathcal{S}$ by the following notion.

Definition 4.3. *We say a collection of sets $S_1, \ldots, S_r$ are D-separated if $|S_i \Delta S_j| \geqslant D$ for all i, j.*

First, we observe that separated characters are near orthogonal, even when multiplied by a low-degree function.

Claim 4.4. *Suppose S_i, S_j are D-separated, and let g_i, g_j be functions of degree at most d with $\|g_i\|_4, \|g_j\|_4 \leqslant M$. Then*

$$\langle \chi_{S_i} g_i, \chi_{S_j} g_j \rangle \leqslant M^2 4^{2d} e^{-\frac{q(1-q)D}{2}}.$$

Proof. By definition,

$$\langle \chi_{S_i} g_i, \chi_{S_j} g_j \rangle = \mathop{\mathbb{E}}_{x\sim\mu_q^{\otimes n}}\left[\chi_{S_i\Delta S_j}(x) g_i(x) g_j(x)\right] = \mathop{\mathbb{E}}_{x\sim\mu_q^{\otimes n}}\left[\chi^{\leqslant 2d}_{S_i\Delta S_j}(x) g_i(x) g_j(x)\right],$$

where the last transition holds as each one of g_i and g_j have degree at most $2d$, hence $g_i(x)g_j(x)$ has degree at most d. Thus by Cauchy-Schwarz twice, we get that

$$\langle \chi_{S_i} g_i, \chi_{S_j} g_j \rangle \leqslant \left\|\chi^{\leqslant 2d}_{S_i\Delta S_j}\right\|_2 \|g_i g_j\|_2 \leqslant \left\|\chi^{\leqslant 2d}_{S_i\Delta S_j}\right\|_2 \|g_i\|_4 \|g_j\|_4 \leqslant M^2 \left\|\chi^{\leqslant 2d}_{S_i\Delta S_j}\right\|_2.$$

To upper bound the last 2-norm, note that by Parseval, we have that

$$\left\|\chi_{S_i\Delta S_j}^{\leqslant 2d}\right\|_2^2=\sum_{j=0}^{2d}\left\|\chi_{S_i\Delta S_j}^{=j}\right\|_2^2\leqslant 4^{2d}\sum_{j=0}^{2d}4^{-j}\left\|\chi_{S_i\Delta S_j}^{=j}\right\|_2^2\leqslant 4^{2d}\sum_{j=0}^{n}4^{-j}\left\|\chi_{S_i\Delta S_j}^{=j}\right\|_2^2=4^{2d}\left\|\mathrm{T}_{1/2}\chi_{S_i\Delta S_j}\right\|_2^2,$$

and since $\chi_{S_i\Delta S_j}$ is a product function we have that

$$\begin{aligned}\left\|\mathrm{T}_{1/2}\chi_{S_i\Delta S_j}\right\|_2^2=\left(\mathop{\mathbb{E}}_{x\sim\mu_q,y\sim_{1/2}x}[(-1)^x(-1)^y]\right)^{|S_i\Delta S_j|}&\leqslant\left(1-\Pr_{x\sim\mu_q,y\sim_{1/2}x}[x\neq y]\right)^D\\&\leqslant\left(1-\frac{q(1-q)}{2}\right)^D.\end{aligned}$$

Combining, we get that $\langle\chi_{S_i}g_i,\chi_{S_j}g_j\rangle\leqslant M^24^{2d}e^{-\frac{q(1-q)D}{2}}$. □

The following claim asserts that the collection $\mathcal{S}_{d,\delta}(f)$ cannot contain many separated sets.

Claim 4.5. *For all $\delta>0$ and $M\geqslant 1$, there is $R=\frac{2M^2}{\delta^2}$, such that for all $d\in\mathbb{N}$, there is $D(d,q,M,\delta)=\frac{100}{q^2(1-q)^2}d^2\log^2(M/\delta)$ such that the following holds. If a function $f\colon\{0,1\}^n\to\mathbb{R}$ has 2-norm at most M, then the largest collection of D-separated sets in $\mathcal{S}_{d,\delta}(f)$ consists of at most R elements.*

Proof. Assume that $S_1,\ldots,S_R\in\mathcal{S}_{d,\delta}(f)$ are D-separated for $R>\frac{2M^2}{\delta^2}$; then we take a sub-collection of size $R'=\frac{2M^2}{\delta^2}$, and to simplify notations we simply assume that $R=R'$. Denote $g_i=(f\chi_{S_i})^{\leqslant d}$.

The functions $\chi_{S_i}g_i$ are nearly orthogonal. By Claim 4.4, for all $i\neq j$ we have $\langle\chi_{S_i}g_i,\chi_{S_j}g_j\rangle\leqslant M^24^{2d}e^{-\frac{q(1-q)D}{2}}\leqslant e^{-\sqrt{D}}$, where the last inequality is by choice of D.

Concluding the proof. Note that $\langle f,\chi_{S_i}g_i\rangle=\langle f\chi_{S_i},g_i\rangle=\|g_i\|_2^2\geqslant\delta^2$. Computing, we get that

$$0\leqslant\left\|f-\sum_{i=1}^R\chi_{S_i}g_i\right\|_2^2=\|f\|_2^2+\sum_{i=1}^R\|\chi_{S_i}g_i\|_2^2-2\sum_{i=1}^R\langle f,\chi_{S_i}g_i\rangle+\sum_{i\neq j}\langle\chi_{S_i}g_i,\chi_{S_j}g_j\rangle.$$

and using our earlier observation and the fact that $\|\chi_{S_i}g_i\|_2^2=\|g_i\|_2^2$ we get that

$$R\delta^2\leqslant\|f\|_2^2+\sum_{i\neq j}\langle\chi_{S_i}g_i,\chi_{S_j}g_j\rangle\leqslant M^2+R^2e^{-\sqrt{D}}.$$

Plugging $R=\frac{2M^2}{\delta^2}$ yields that $2M^2\leqslant M^2+\frac{4M^4}{\delta^4}e^{-\sqrt{D}}$, and hence $e^{\sqrt{D}}\leqslant\frac{4M^2}{\delta^4}$, in contradiction to the choice of D. □

We would like to remark one important feature in Claim 4.5, which is that the list size R is independent of d. This allows us to perform the following iterative processes, which will help us in replacing the orthogonality property one usually has in the list decoding argument.

4.2.1 The First Iterative Process: Getting Separation

Fix M and $\varepsilon > 0$, and take d and δ for M and $\varepsilon/10$ from Corollary 4.2; we choose $0 < \delta' \ll \delta$ and choose R from Claim 4.5 for δ'^2. We take a sequence of parameters $d = D_0 \ll D_1 \ll D_2 \ll \ldots \ll D_{R+1} \ll D_{R+2}$. For each i define $\mathcal{S}_i = \mathcal{S}_{D_i,\delta'^2}(f)$. Starting with $i = 0$, we wish to find a well separated collection in $\mathcal{S}_i$ in which any two sets are far from each other. That is, we want to find a collection of sets from $\mathcal{S}_i$ in which any two sets are very far from each other, and any other $S \in \mathcal{S}_i$ is close to one of the sets in the collections.

To achieve this, we pick the largest collection of $S_1, \ldots, S_r$ in $\mathcal{S}_i$ which is D_{i+1} separated, and if there are $j \neq k$ such that $|S_j \Delta S_k| \leqslant D_{i+2}/2$, increase i by 1 and iterate. In other words, increase i by 1, and pick the largest collection of $S_1, \ldots, S_{r'}$ in $\mathcal{S}_i$ which is D_{i+1} separated, and increase i again if there are two distinct sets in this collection are $D_{i+2}/2$ close.

It is clear that if the above process terminates, say at step i, then in our collection $S_1, \ldots, S_r$ satisfies that $|S_j \Delta S_k| > D_{i+2}/2$ for all $j \neq k$, and any $S \in \mathcal{S}_i$ is D_i-close to one of the sets in the collection. Next, we prove that this process indeed terminates, and for that we denote by R_i the size of the collection $S_1, \ldots, S_r$ in iteration i. The next claim shows that whenever we iterate, the size of R_i strictly decreases:

Claim 4.6. $R_{i+1} \leqslant R_i - 1$.

Proof. Assume otherwise, and let $S_1, \ldots, S_{R_i}$ be the collection at iteration i and $T_1, \ldots, T_{R_i}$ be a part of the collection at iteration $i + 1$ (which can be picked as we assume $R_{i+1} \geqslant R_i$). We first observe that for any $j = 1, \ldots, R_i$, there is $k = 1, \ldots, R_i$ such that $|T_j \Delta S_k| \leqslant D_{i+1}$. Indeed, otherwise we could have added T_j to the list $S_1, \ldots, S_{R_i}$ at iteration i, in contradiction to its maximality. Also, we note that for a given k there is at most a single j such that $|T_j \Delta S_k| \leqslant D_{i+1}$, otherwise if we have distinct such j, j' then

$$\left|T_j \Delta T_{j'}\right| = \left|(T_j \Delta S_k) \Delta (T_{j'} \Delta S_k)\right| \leqslant |T_j \Delta S_k| + \left|T_{j'} \Delta S_k\right| \leqslant 2D_{i+1} < D_{i+2},$$

in contradiction to the fact that the T_j's are D_{i+2} separated. Thus, there is a permutation $\pi \colon \{1, \ldots, R_i\} \to \{1, \ldots, R_i\}$ such that $\left|T_j \Delta S_{\pi(j)}\right| \leqslant D_{i+1}$, and without loss of generality we assume that π is the identity.

As the process didn't stop at iteration i it means that the collection S is not $D_{i+2}/2$ separated, so there are distinct k, k' such that $|S_k \Delta S_{k'}| \leqslant D_{i+2}/2$, and we get that

$$|T_k \Delta T_{k'}| \leqslant |T_k \Delta S_k| + |T_{k'} \Delta S_{k'}| + |S_k \Delta S_{k'}| \leqslant 2D_{i+1} + \frac{D_{i+2}}{2} < D_{i+2},$$

in contradiction to the fact that the collection T is D_{i+2} separated. $\square$

Since by Claim 4.5 we have that $R_0 \leqslant R$, it follows that the process must terminate after at most R steps, and we fix i on which it stops. Thus we get that for $D' = D_i$, we look at $\mathcal{S}_{D',\delta'^2}(f)$ and find there a largest collection of sets $S_1, \ldots, S_r$, such that any $S \in \mathcal{S}_{D',\delta'^2}(f)$ is $D'' = D_{i+1}$ close to one of the S'_j, and for any $j \neq k$ we have that $|S_i \Delta S_j| > \frac{D_{i+2}}{2} = D_{\mathsf{top}}$. Our parameters satisfy $D' \ll D'' \ll D_{\mathsf{top}}$.

4.2.2 The Second Iterative Process: Avoiding Overlaps

We now run another iterative process; take $\eta > 0$; if $\left\|(\chi_{S_j} f)^{\leqslant 2D''}\right\|_2 \geqslant (1 + \eta) \left\|(\chi_{S_j} f)^{\leqslant D'}\right\|_2$ for some r, we increase D' to $2D''$, and take the new D'' to be sufficiently larger than the new D' (and still much smaller than D_{top}). Note that since we have at most R distinct j's and each one can cause at most $O_{\delta,\delta',\eta}(1)$ increases, we eventually reach D' such that $\left\|(\chi_{S_j} f)^{\leqslant 2D''}\right\|_2 \leqslant (1+\eta) \left\|(\chi_{S_j} f)^{\leqslant D'}\right\|_2$ for all j. Furthermore, since the factor by which D' increases in this process only depends on δ, δ', η, we still have that $D' \ll D'' \ll D_{\mathsf{top}}$.

4.2.3 The List Explains All of the Advantage

We are now ready to show that the list that we found, $\chi_{S_j}(\chi_{S_j} f_1)^{\leqslant D'}$ for $j = 1, \ldots, r$, explains almost all of the advantage the functions $f_1, \ldots, f_4$ have in the premise of Theorem 1.3. Towards this end, denote $f_1' = f_1 - \sum_{j=1}^{r} \chi_{S_j} g_j$, where $g_j = (\chi_{S_j} f_1)^{\leqslant D'}$.

Lemma 4.7. *We have that*

$$\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[f_1'(x) f_2(y) f_3(z) f_4(w)\right] \leqslant \frac{\varepsilon}{4}.$$

Proof. Assume towards contradiction otherwise. Before proceeding to the formal proof, we explain the idea. After showing that the 2-norm of f_1' is not much larger than M, we will be able to apply Corollary 4.2 to conclude that f_1' is correlated with a function of the form $\chi_S g$ for g of degree at most D'. We will then show that S must be close to one of the S_i's; indeed, if this is not the case then the difference $f_1 - f_1'$ is not correlated with $\chi_S g$, so it must be the case that f_1 is correlated with $\chi_S g$, and then we appeal to the maximality of the list $S_1, \ldots, S_r$ to argue that S then must be close to one of the S_i's.

But then, in a sense, the ℓ_2-mass of $(f_1\chi_S)^{\leqslant D'}$ would already "be included" in the ℓ_2 mass of $(f_1\chi_{S_i})^{\leqslant D'+L}$, where L is the distance between S_i and S. But by our iterative processes above, the ℓ_2 mass of $(f_1\chi_{S_i})^{\leqslant D'+L}$ and of $(f_1\chi_{S_i})^{\leqslant D'}$ should be very close to each other, hence we would have already removed that from f_1', therefore we get a contradiction.

We proceed to the formal argument. Clearly the 12-norm of each one of f_2, f_3, f_4 is at most M. As for the 2-norm of f_1', we have

$$\left\|f_1'\right\|_2^2 = \|f\|_2^2 - 2\sum_i \langle f, \chi_{S_i} g_i\rangle + \left\|\sum_i \chi_{S_i} g_i\right\|_2^2 = \|f\|_2^2 - \sum_i \|g_i\|_2^2 + \sum_{i\neq j}\langle \chi_{S_i} g_i, \chi_{S_j} g_j\rangle,$$

which is at most $M^2 + R^2 M^2 4^{2D'} e^{-\frac{q(1-q)}{2}D''}$, where we used Claim 4.4. By choice of D'', this is at most $2M^2$, so $\|f_1'\|_2 \leqslant 2M$. We therefore have that the functions $f_1'/2, f_2, f_3, f_4$ satisfy the conditions of Corollary 4.2, and so $f_1'/2$ is δ-correlated with a function of the function $\chi_S(f_1'\chi_S/2)^{\leqslant d}$. Thus, as

$$\langle f_1', \chi_S(f_1'\chi_S/2)^{\leqslant D'}\rangle = \left\|(f_1'\chi_S/2)^{\leqslant D'}\right\|_2^2 \geqslant \left\|(f_1'\chi_S/2)^{\leqslant d}\right\|_2^2 = \langle f_1', \chi_S(f_1'\chi_S/2)^{\leqslant d}\rangle \geqslant \delta,$$

it follows that $f_1'/2$ is δ-correlated with $\chi_S g$ for $g = (f_1'\chi_S/2)^{\leqslant D'}$.

Next, we argue that S has to be close to some S_i and that $(f_1\chi_S)^{\leqslant D'+D''}$ must have significant 2-norm.

Claim 4.8. *There is i such that $|S\Delta S_i| \leqslant D''$, and also it holds that $\left\|(f_1\chi_S)^{\leqslant D'+D''}\right\|_2 \geqslant \delta'^2$.*

Proof. We have $2\delta \leqslant \langle f_1', \chi_S g\rangle = \langle f_1, \chi_S g\rangle - \sum_i \langle \chi_{S_i} g_i, \chi_S g\rangle$. If $\langle f_1, \chi_S g\rangle \geqslant \delta$, then we get that $f_1\chi_S$ is δ-correlated with a degree d function of 2-norm at most 1, hence

$$\left\|(f_1\chi_S)^{\leqslant D'}\right\|_2^2 = \sup_{\substack{h\colon \{0,1\}^n\to\mathbb{R} \text{ degree} \leqslant D' \\ \|h\|_2\leqslant 1}} \langle f\chi_S, h\rangle^2 \geqslant \langle f\chi_S, g\rangle^2 \geqslant \delta^2,$$

so $S \in \mathcal{S}_{D',\delta^2}$. By the maximality of $S_1, \ldots, S_r$, we conclude $|S\Delta S_i| \leqslant D''$ for some i.

Else, we have that $|\langle \chi_{S_i} g_i, \chi_S g\rangle| \geqslant \frac{3\delta''}{R}$ for some i, and using Claim 4.4 we find that $|S_i\Delta S| \leqslant O(\log M + D' + \log(R/3\delta'')) \leqslant D''$.

In any case, we get that $|S\Delta S_i| \leqslant D''$ for some i. Consider the function $\chi_{S\Delta S_i}(f\chi_{S_i})^{\leqslant D'}$ which has degree at most $D' + D''$; then

$$\langle \chi_S f, \chi_{S\Delta S_i}(f\chi_{S_i})^{\leqslant D'}\rangle = \langle f, \chi_{S_i}(f\chi_{S_i})^{\leqslant D'}\rangle = \|g_i\|_2^2 \geqslant \delta'^2,$$

so the function $\chi_S f$ is δ'^2-correlated with a function of degree at most $D' + D''$ which has 2-norm 1, hence

$$\left\|(f\chi_S)^{\leqslant D'+D''}\right\|_2^2 = \sup_{\substack{h\colon \{0,1\}^n\to\mathbb{R} \text{ degree} \leqslant D'+D'' \\ \|h\|_2\leqslant 1}} \langle f\chi_S, h\rangle^2 \geqslant \delta'^4. \qquad \square$$

Assume without loss of generality that in Claim 4.8 we have $i = 1$. Then S is at least $D_{\mathsf{top}} - D''$ far from S_j for all $j \neq 1$, hence using Claim 4.4

$$|\langle f_1', \chi_S g\rangle - \langle f_1, \chi_S g\rangle + \langle \chi_{S_1} g_1, \chi_S g\rangle| = \left|\langle \sum_{j\neq 1} \chi_{S_j} g_j, \chi_S g\rangle\right| \leqslant 2^{O_{M,R,\delta}(D'')-\Omega_{M,R,\delta}(D_{\mathsf{top}})} \leqslant \frac{\delta'}{2},$$

so $|\langle f_1, \chi_S g\rangle - \langle \chi_{S_1} g_1, \chi_S g\rangle| \geqslant \frac{\delta'}{2}$. On the other hand, the next claim shows an upper bound on this difference, which is a contradiction provided η is small enough, thereby finishing the proof.

Claim 4.9. $|\langle \chi_{S_1} g_1, \chi_S g\rangle - \langle f_1, \chi_S g\rangle| \leqslant \sqrt{6\eta} M^2$

Proof. Recall that $\left\|(f_1\chi_{S_1})^{\leqslant 2D''}\right\|_2 \leqslant (1+\eta)\left\|(f_1\chi_{S_1})^{\leqslant D'}\right\|_2$, so

$$\begin{aligned}
\left\|(f_1\chi_{S_1})^{\leqslant 2D''} - (f\chi_{S_1})^{\leqslant D'}\right\|_2^2 &= \left\|(f_1\chi_{S_1})^{\leqslant 2D''}\right\|_2^2 - 2\langle (f_1\chi_{S_1})^{\leqslant 2D''}, (f_1\chi_{S_1})^{\leqslant D'}\rangle + \left\|(f_1\chi_{S_1})^{\leqslant D'}\right\|_2^2 \\
&= \left\|(f_1\chi_{S_1})^{\leqslant 2D''}\right\|_2^2 - \left\|(f_1\chi_{S_1})^{\leqslant D'}\right\|_2^2 \\
&\leqslant 3\eta \left\|(f_1\chi_{S_1})^{\leqslant D'}\right\|_2^2 \\
&\leqslant 3\eta M^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
\langle \chi_{S_1} g_1, \chi_S g\rangle = \langle g_1, \chi_{S\Delta S_1} g\rangle &= \langle (f_1\chi_{S_1})^{\leqslant D'}, \chi_{S\Delta S_1} g\rangle \\
&= \langle (f_1\chi_{S_1})^{\leqslant 2D''}, \chi_{S\Delta S_1} g\rangle + \langle (f_1\chi_{S_1})^{\leqslant D'} - (f_1\chi_{S_1})^{\leqslant 2D''}, \chi_{S\Delta S_1} g\rangle.
\end{aligned}$$

By Cauchy-Schwarz,

$$\left|\langle (f\chi_{S_1})^{\leqslant D'} - (f\chi_{S_1})^{\leqslant 2D''}, \chi_{S\Delta S_1} g\rangle\right| \leqslant \left\|(f_1\chi_{S_1})^{\leqslant 2D''} - (f_1\chi_{S_1})^{\leqslant D'}\right\|_2 \|\chi_{S\Delta S_1} g\|_2 \leqslant \sqrt{3\eta M^2}\left\|f_1'\right\|_2,$$

which is at most $\sqrt{6\eta}M^2$. Thus, $\left|\langle \chi_{S_1} g_1, \chi_S g\rangle - \langle (f_1\chi_{S_1})^{\leqslant 2D''}, \chi_{S\Delta S_1} g\rangle\right| \leqslant \sqrt{6\eta}M^2$. To finish the proof, note that since the degree of $\chi_{S\Delta S_1} g$ is less than $2D''$ we have

$$\langle (f_1\chi_{S_1})^{\leqslant 2D''}, \chi_{S\Delta S_1} g\rangle = \langle f_1\chi_{S_1}, \chi_{S\Delta S_1} g\rangle = \langle f_1, \chi_S g\rangle. \qquad \square$$

$\square$

4.3 Applying the List Decoding Argument

Armed with Lemma 4.7, we can now switch the functions f_i with functions of the form $\chi_{S_i}(f_i\chi_{S_i})^{\leqslant d_i}$ and retain that the expectation is large. More precisely, from Lemma 4.7 it follows that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[(f_1-f_1')(x)f_2(y)f_3(z)f_4(w)\right]\right|\geqslant\frac{3\varepsilon}{4},$$

so there is i such that for $h_1(x)=\chi_{S_i}(x)(f_1\chi_{S_i})^{\leqslant D'}(x)$ it holds that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[h_1(x)f_2(y)f_3(z)f_4(w)\right]\right|\geqslant\varepsilon'=\frac{3\varepsilon}{4R}.$$

Thus, starting with the assumption that $\left|\mathbb{E}_{(x,y,z,w)\sim\nu^n}\left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right|\geqslant\varepsilon$, we managed to change f_1 to a function of the form h_1 while only needing to decrease ε to ε'. Namely, we proved the following lemma:

Lemma 4.10. *For all $\varepsilon>0$ and $M\geqslant 1$ there are D and $\varepsilon'>0$ such that the following holds. Suppose that $f_1,\ldots,f_4\colon\{0,1\}^n\to\mathbb{R}$ are functions with $\|f_1\|_2\leqslant 1$ and $\|f_i\|_{12}\leqslant M$ for all i; further suppose that*

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right|\geqslant\varepsilon.$$

Then for some $S\subseteq[n]$, and $d\leqslant D$, defining the function $h_1(x)=\chi_S(f_1\chi_S)^{\leqslant d}$ we have that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[h_1(x)f_2(y)f_3(z)f_4(w)\right]\right|\geqslant\varepsilon'.$$

We iterate lemma 4.10 to change each one of the functions f_i to an h_i, and get

Lemma 4.11. *For all $\varepsilon>0$ and $M\geqslant 1$ there are $D_1,\ldots,D_4$ and $\varepsilon'>0$ such that the following holds. Suppose that $f_1,\ldots,f_4\colon\{0,1\}^n\to\mathbb{R}$ are functions with $\|f_i\|_2\leqslant 1$ and $\|f_i\|_{12}\leqslant M$ for all i for which*

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right|\geqslant\varepsilon.$$

Then for some $S_1,\ldots,S_4\subseteq[n]$, and $d_1\leqslant D_1,\ldots,d_4\leqslant D_4$, defining the function $h_i=\chi_{S_i}(f_i\chi_{S_i})^{\leqslant d_i}$ we have that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[h_1(x)h_2(y)h_3(z)h_4(w)\right]\right|\geqslant\varepsilon'.$$

Henceforth, we fix ε' and S_1,S_2,S_3,S_4 as well as h_1,h_2,h_3,h_4 as given in Lemma 4.11. Denote $T=S_1\cap S_2\cap S_3\cap S_4$. To finish this section, we show that each one of the S_i's is close to T.

Claim 4.12. *For all $D\in\mathbb{N}$ and $\varepsilon'>0$, there is $t\in\mathbb{N}$ such that if we have $g_1,\ldots,g_4$ are functions of degree at most D at 2-norm at most 1, and*

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[\chi_{S_1}(x)g_1(x)\chi_{S_2}(y)g_2(y)\chi_{S_3}(z)g_3(z)\chi_{S_4}(w)g_4(w)\right]\right|\geqslant\varepsilon',$$

then $|S_i\Delta T|\leqslant t$ for all $i=1,\ldots,4$, where $T=S_1\cap\ldots\cap S_4$.

Proof. Define $t = \max_i |S_i \Delta T|$ and

$$G(x, y, z, w) = g_1(x)g_2(y)g_3(z)g_4(w), \qquad G'(x, y, z, w) = \chi_{S_1\setminus T}(x)\chi_{S_2\setminus T}(y)\chi_{S_3\setminus T}(z)\chi_{S_4\setminus T}(w),$$

and think of them as functions from $((\{0,1\}^4)^n, \nu^{\otimes n})$ to $\mathbb{R}$. We note that as $xyzw = 1$ in the support of ν, we have that

$$\left|\langle G', G\rangle\right| = \left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^n}\left[\chi_{S_1}(x)g_1(x)\chi_{S_2}(y)g_2(y)\chi_{S_3}(z)g_3(z)\chi_{S_4}(w)g_4(w)\right]\right| \geqslant \varepsilon'.$$

Also, the degree of G is at most $4D$, so we get that

$$\left|\langle G', G\rangle\right| = \left|\langle G'^{\leqslant 4D}, G\rangle\right| \leqslant \left\|G'^{\leqslant 4D}\right\|_2 \|G\|_2 \leqslant \left\|G'^{\leqslant 4D}\right\|_2 \|g_1\|_8 \|g_2\|_8 \|g_3\|_8 \|g_4\|_8,$$

where we used Cauchy-Schwarz multiple times. By Theorem 2.9 we have that $\|g_i\|_8 \leqslant C(q)^D \|g_i\|_2 \leqslant C(q)^D$. Also by Parseval we get that $\left\|G'^{\leqslant 4D}\right\|_2 \leqslant 2^{4D}\left\|\mathrm{T}_{1/2}G'^{\leqslant 4D}\right\|_2 \leqslant 2^{4D}\left\|T_{1/2}G'\right\|_2$. Combining all, we get that

$$\varepsilon' \leqslant \left|\langle G', G\rangle\right| \leqslant C'(q, D)\left\|T_{1/2}G'\right\|_2.$$

To upper bound $\left\|T_{1/2}G'\right\|_2$, we note that G' is a product function, i.e. we may write it as $G' = \prod\limits_{i=1}^{n} G'_i$ for G'_i that depends only on the ith coordinate of (x, y, z, w), so $\left\|T_{1/2}G'\right\|_2 = \prod\limits_{i=1}^{n}\left\|T_{1/2}G'_i\right\|_2$. Each variable i it depends on, i.e. such that G'_i is not constant, is a variable that appears in at least in one of $S_1, \ldots, S_4$ but not in T, so there are at least t of these variables. We claim that there is $\lambda \in (0,1)$ depending only on q such that $\left\|T_{1/2}G'_i\right\|_2 \leqslant \lambda$ for any such i; indeed, suppose for simplicity that $G'_i(x, y, z, w) = x_i y_i z_i$, and note that

$$\left\|T_{1/2}G'_i\right\|_2^2 = \mathop{\mathbb{E}}_{\substack{(x,y,z,w),(x',y',z',w')\sim\nu \\ \text{that are } 1/2\text{-correlated}}}\left[x_i y_i z_i x'_i y'_i z'_i\right] \leqslant 1 - \Omega_q(1).$$

Thus, $\left\|T_{1/2}G'\right\|_2 \leqslant \lambda^t$, and plugging that above yields that $\varepsilon' \leqslant C'(q, D)\lambda^t$, and re-arranging yields that $t \leqslant \frac{\log(C'(q,D)/\varepsilon')}{\log(1/\lambda)}$, as desired. □

We fix T and t from Claim 4.12 and define $S'_i = S_i \Delta T$ and $h'_i = \chi_{S'_i}(f_i\chi_{S_i})^{\leqslant D_i}$; note that then each h'_i is a function of degree at most $t + D_i$, and

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[h'_1(x)h'_2(y)h'_3(z)h'_4(w)\right]\right| = \left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[h_1(x)h_2(y)h_3(z)h_4(w)\right]\right| \geqslant \varepsilon'.$$

Summarizing, in this section we proved the following lemma.

Lemma 4.13. *For all $\varepsilon > 0$, there are $t, D \in \mathbb{N}$ and $\varepsilon' > 0$ such that the following holds. Suppose $f_1, \ldots, f_4\colon \{0,1\}^n \to [-1, 1]$ are functions such that*

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right| \geqslant \varepsilon.$$

Then there are $D_1, \ldots, D_4 \in \mathbb{N}$ that are at most D, $S_1, \ldots, S_4 \subseteq [n]$ and $S'_1 \subseteq S_1, \ldots, S'_4 \subseteq S_4$ of size at most t, such that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[(\chi_{S'_1}(\chi_{S_1}f_1)^{\leqslant D_1})(x)(\chi_{S'_2}(\chi_{S_2}f_2)^{\leqslant D_2})(y)(\chi_{S'_3}(\chi_{S_3}f_3)^{\leqslant D_3})(z)(\chi_{S'_4}(\chi_{S_4}f_4)^{\leqslant D_4})(w)\right]\right| \geqslant \varepsilon'.$$

4.4 Preparation for the Invariance Principle Argument: Applying Noise

In the next and final part of the argument, we would like to apply the invariance principle on the expectation in the conclusion of Lemma 4.13 in order to relate it to an expectation over Gaussian space. There are, however, a few issues that prevent us from doing so directly. First, the functions $(\chi_{S_i} f_i)^{\leqslant D_i}$ may not be bounded (and in fact they are likely not), so if we apply the invariance principle on them directly we would need to consider a more general problem in Gaussian space, about unbounded functions, whose answer may be different than the one we're looking for. Second, the distribution over (x, y, z, w) is not connected, and therefore we cannot appeal to a black-box invariance principle. Third, the functions in the conclusion of Lemma 4.13 need not have small low-degree influence, and we would need this in order to appeal to any form of the invariance principle.

In this section, we resolve the first two issues. Namely, we show that we may switch the truncations above to applications of the noise operator T_ρ, and furthermore that we can apply the noise operator on each one of these functions and still keep the expectation in consideration substantial. We achieve these two statement using the same argument, similar to an idea from [FKLM20], asserting that if we have any integer D, then the truncation operator $f \to f^{\leqslant D}$ can be approximated, in ℓ_2, by a polynomial P, applied on the noise operator T_ρ, i.e. that $\left\|P(\mathrm{T}_\rho)f - f^{\leqslant D}\right\|_2$ is small for all f, for some fixed polynomial P. Thus, noting that each $(\chi_{S_i'}(\chi_{S_i} f_i)^{\leqslant D_i})$ has degree at most $t + D_i$, we can replace this function with $P_1(\mathrm{T}_\rho)(\chi_{S_i'}(\chi_{S_i} f_i)^{\leqslant D_i})$ for an appropriate polynomial P_1 and get a similar expectation. We can then switch the internal truncation operator with $P_2(\mathrm{T}_\rho)\chi_{S_i} f_i$ and still get a significant expectation, so expanding out the polynomials P_1 and P_2 we get that at least one of the terms that are product is significant.

We proceed to the formal description of this step, and begin with an auxiliary fact which constructs the type of polynomials P we use in our arguments.

Fact 4.14. *For all $\eta, \xi > 0$ and $s \in [\eta, 1 - \eta]$ there is a polynomial $P\colon [0, 1] \to [0, 1]$ such that: (1) for $x \leqslant s$, $P(x) \leqslant \xi$, (2) for $x \geqslant s + \eta$, $P(x) \geqslant 1 - \xi$.*

Proof. Define the function $f\colon [0, 1] \to [0, 1]$ which is 0 for $x \leqslant s$, 1 for $x \geqslant s + \eta$, and between s and $s + \eta$ we linearly interpolate so that f is continuous. By the density of polynomials, i.e. the Stone-Weirstrass theorem, there is a univariate polynomial $P\colon [0, 1] \to [0, 1]$ such that $\|P - f\|_\infty \leqslant \xi$, and the result follows. $\square$

Using Fact 4.14, we prove in the next claim that the truncation operator of degree D can be approximated by a polynomial applied on the noise operator $\mathrm{T}_{1/2}$.

Claim 4.15. *Let $D \in \mathbb{N}$ and $\xi > 0$. Then there exist a polynomial P such that for all $f\colon (\{0, 1\}^n, \mu_q^{\otimes n}) \to \mathbb{R}$ it holds that*

$$\left\|P(\mathrm{T}_{1/2})f - f^{\leqslant D}\right\|_2 \leqslant \xi \left\|f\right\|_2$$

Proof. Set $\rho = 1/2$, and note that the eigenvalues of T_ρ are ρ^j for $j = 0, 1, \ldots, n$. Note that for $j \leqslant D$ we have that $\rho^j \geqslant \rho^D$, and for $j > D + 1$ we have that $\rho^j \leqslant \rho^{D+1}$. Letting $\eta = \rho^D - \rho^{D+1} > 0$, we may use Fact 4.14 to find a polynomial P such that $P(\eta^j) \geqslant 1 - \xi$ for $j \leqslant D$ and $P(\eta^j) \leqslant \xi$ for $j > D$.

Note that the eigenvalues of $P(\mathrm{T}_\rho)$ are $P(\eta^j)$, hence we get that

$$P(\mathrm{T}_\rho)f = \sum_{j=0}^{n} P(\rho^j) f^{=j},$$

hence by Parseval

$$\begin{aligned}\left\|P(\mathrm{T}_{1/2})f - f^{\leqslant D}\right\|_2^2 &= \sum_{j=0}^{D} \left|P(\rho^j) - 1\right|^2 \left\|f^{=j}\right\|_2^2 + \sum_{j=D+1}^{n} P(\rho^j)^2 \left\|f^{=j}\right\|_2^2 \\ &\leqslant \sum_{j=0}^{D} \xi^2 \left\|f^{=j}\right\|_2^2 + \sum_{j=D+1}^{n} \xi^2 \left\|f^{=j}\right\|_2^2 \\ &= \xi^2 \left\|f\right\|_2^2,\end{aligned}$$

concluding the proof. □

We are now ready to state and prove the main result of this section.

Lemma 4.16. *For all $\varepsilon > 0$, there are $t \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose $f_1, \dots, f_4 \colon \{0,1\}^n \to [-1,1]$ are functions such that*

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right| \geqslant \varepsilon.$$

Then there are $\rho_1, \dots, \rho_4 \in (0,1/2]$, $\rho_1', \dots, \rho_4' \in (0,1/2]$, $S_1, \dots, S_4 \subseteq [n]$ and $S_1' \subseteq S_1, \dots, S_4' \subseteq S_4$ of size at most t, such that defining $g_i = \mathrm{T}_{\rho_i}(\chi_{S_i} f)$, we have

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[\mathrm{T}_{\rho_1'}(\chi_{S_1'}g_1)(x)\mathrm{T}_{\rho_2'}(\chi_{S_2'}g_2)(y)\mathrm{T}_{\rho_3'}(\chi_{S_3'}g_3)(z)\mathrm{T}_{\rho_4'}(\chi_{S_4'}g_4)(w)\right]\right| \geqslant \delta.$$

Proof. We use Lemma 4.13 and find $(S_i, S_i')_{i=1,\dots,4}$ as well as $t, D \in \mathbb{N}$ and $\varepsilon' > 0$ from there. We also take $D_1, \dots, D_4$ and denote $h_i = (\chi_{S_i} f_i)^{\leqslant D_i}$. For notational convenience, we denote $x^1 = x$, $x^2 = y$, $x^3 = z$ and $x^4 = w$

First, we show that one may introduce the noise operators on the outside, and we demonstrate how to do it for the first function. For this, we use the parameters

$$0 \ll \xi \ll D^{-1}, t^{-1}, \varepsilon' \ll \varepsilon.$$

We apply Claim 4.15 for $t+D$, and find a polynomial P_1 from Claim 4.15 that ξ-approximates the truncation operator on degree $t+D$. Thus,

$$\begin{aligned}&\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[\prod_{i=1}^{4}(\chi_{S_i'}h_i)(x^i)\right] - \mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[(P_1(\mathrm{T}_{1/2})\chi_{S_1'}h_1)(x)\prod_{i=2}^{4}(\chi_{S_i'}h_i)(x^i)\right]\right| \\ &= \left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[(P_1(\mathrm{T}_{1/2})\chi_{S_1'}h_1 - \chi_{S_1'}h_1)(x^1)\prod_{i=2}^{4}(\chi_{S_i'}h_i)(x^i)\right]\right| \\ &\leqslant \left\|P_1(\mathrm{T}_{1/2})\chi_{S_1'}h_1 - \chi_{S_1'}h_1\right\|_2 \sqrt{\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[(P_1(\mathrm{T}_{1/2})\chi_{S_1'}h_1 - \chi_{S_1'}h_1)(x^1)\prod_{i=2}^{4}(\chi_{S_i'}h_i)(x^i)^2\right]},\end{aligned} \tag{9}$$

where we used Cauchy-Schwarz. Applying Hölder's inequality and Theorem 2.9 we get that this is at most

$$\left\|P(\mathrm{T}_{1/2})\chi_{S_1'}h_1 - \chi_{S_1'}h_1\right\|_2 \prod_{i=2}^{4}\left\|\chi_{S_i'}h_i\right\|_6 \leqslant C(D,t,q)\left\|P(\mathrm{T}_{1/2})\chi_{S_1'}h_1 - \chi_{S_1'}h_1\right\|_2 \prod_{i=2}^{4}\left\|\chi_{S_i'}h_i\right\|_2.$$

Note that $\left\|\chi_{S'_i} h_i\right\|_2 = \|h_i\|_2 \leqslant \|\chi_{S_i} f_i\|_2 = \|f_i\|_2 \leqslant 1$. Also note that by Claim 4.15 we have

$$\left\|P_1(\mathrm{T}_{1/2})\chi_{S'_1} h_1 - \chi_{S'_1} h_1\right\|_2 \leqslant \xi \left\|\chi_{S'_1} h_1\right\|_2 \leqslant \xi.$$

Combining, we get that

$$(9) \leqslant \xi C(D, tq) \leqslant \frac{\varepsilon'}{10}.$$

Thus, we get that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[(P(\mathrm{T}_{1/2})\chi_{S'_1} h_1)(x^1)\prod_{i=2}^{4}(\chi_{S'_i} h_i)(x^i)\right]\right| \geqslant \frac{9\varepsilon'}{10}.$$

Expand out $P(z) = \sum_{j=0}^{d} \alpha_j z^j$, and note that $\alpha_0 = P(0) \leqslant \xi$, so we get that

$$\sum_{j=1}^{d} |\alpha_j| \left|\mathbb{E}\left[\mathrm{T}_{1/2^j}(\chi_{S'_1} h_1)(x^1)\prod_{i=2}^{4}(\chi_{S'_i} h_i)(x^i)\right]\right| \geqslant \frac{9\varepsilon}{10} - \alpha_0 \left|\mathbb{E}\left[\prod_{i=2}^{4}(\chi_{S'_i} h_i)(x^i)\right]\right|.$$

Note that using Cauchy-Schwarz and Theorem 2.9 as before we have that $\left|\mathbb{E}\left[\prod_{i=2}^{4}(\chi_{S'_i} h_i)(x^i)\right]\right| \leqslant C(D, t, q)$, so we get that

$$\sum_{j=1}^{d} |\alpha_j| \left|\mathbb{E}\left[\mathrm{T}_{1/2^j}(\chi_{S'_1} h_1)(x^1)\prod_{i=2}^{4}(\chi_{S'_i} h_i)(x^i)\right]\right| \geqslant \frac{4\varepsilon'}{5}.$$

We conclude that there is a j such that

$$\left|\mathbb{E}\left[\mathrm{T}_{1/2^j}(\chi_{S'_1} h_1)(x^1)\prod_{i=2}^{4}(\chi_{S'_i} h_i)(x^i)\right]\right| \geqslant \frac{4\varepsilon}{5\sum_{j=1}^{d} |\alpha_j|} \geqslant \varepsilon_1 = \Omega_{\varepsilon,\xi}(1),$$

and we choose $\rho'_1 = 1/2^j$.

We can now iterate this argument with ε' being ε_1 and taking parameters sufficiently small compared to it. Eventually, we find $\rho'_1, \ldots, \rho'_4 \leqslant 1/2$ such that

$$\left|\mathbb{E}\left[\prod_{i=1}^{4}\mathrm{T}_{\rho'_i}(\chi_{S'_i} h_i)(x^i)\right]\right| \geqslant \varepsilon_4,$$

where $\varepsilon_4 > 0$ is a function of the original ε.

Next, we show that we can replace h_1 with $\mathrm{T}_{\rho_1}(\chi_{S_1} f_1)$ for some $\rho_1 \leqslant 1/2$. The argument is similar to before, and below the elaborate. Take the parameters (to avoid cumbersome notations, we re-use ξ)

$$0 \ll \xi \ll D^{-1}, t^{-1}, \varepsilon' \ll \varepsilon_4,$$

and take a polynomial P_2 from Claim 4.15 that ξ-approximates the truncation to degree D operator. Then

$$\begin{aligned}
&\left|\mathbb{E}\left[\prod_{i=1}^{4} \mathrm{T}_{\rho_i'}(\chi_{S_i'} h_i)(x^i)\right] - \mathbb{E}\left[\mathrm{T}_{\rho_1'}(\chi_{S_1'} P_2(\mathrm{T}_{1/2})(\chi_{S_1} f_1))(x^1)\prod_{i=2}^{4} \mathrm{T}_{\rho_i'}(\chi_{S_i'} h_i)(x^i)\right]\right| \\
&= \left|\mathbb{E}\left[\mathrm{T}_{\rho_1'}\left(\chi_{S_1'}(\chi_{S_1} f_1) - P_2(\mathrm{T}_{1/2})(\chi_{S_1} f_1)\right)(x^1)\prod_{i=2}^{4} \mathrm{T}_{\rho_i'}(\chi_{S_i'} h_i)(x^i)\right]\right| \\
&\leqslant \left\|\mathrm{T}_{\rho_1'}\left(\chi_{S_1'}(\chi_{S_1} f_1) - P_2(\mathrm{T}_{1/2})(\chi_{S_1} f_1)\right)\right\|_2 \sqrt{\mathbb{E}\left[\prod_{i=2}^{4} \mathrm{T}_{\rho_i'}(\chi_{S_i'} h_i)(x^i)^2\right]} \\
&\leqslant \left\|\mathrm{T}_{\rho_1'}\left(\chi_{S_1'}(\chi_{S_1} f_1) - P_2(\mathrm{T}_{1/2})(\chi_{S_1} f_1)\right)\right\|_2 \prod_{i=2}^{4} \left\|\mathrm{T}_{\rho_i'}(\chi_{S_i'} h_i)\right\|_6. \qquad (10)
\end{aligned}$$

Note that by Theorem 2.9

$$\left\|\mathrm{T}_{\rho_i'}(\chi_{S_i'} h_i)\right\|_6 \leqslant \left\|\chi_{S_i'} h_i\right\|_6 \leqslant C(D,t,q)\left\|\chi_{S_i'} h_i\right\|_2 \leqslant C(D,t,q)\left\|f_i\right\|_2 \leqslant C(D,t,q).$$

From Claim 4.15, we get

$$\left\|\mathrm{T}_{\rho_1'}\left(\chi_{S_1'}(\chi_{S_1} f_1) - P_2(\mathrm{T}_{1/2})(\chi_{S_1} f_1)\right)\right\|_2 \leqslant \left\|\chi_{S_1'}(\chi_{S_1} f_1) - P_2(\mathrm{T}_{1/2})(\chi_{S_1} f_1)\right\|_2 \leqslant \xi\left\|\chi_{S_1} f_1\right\|_2 \leqslant \xi.$$

Together, we get that (10) $\leqslant C(D,t,q)\xi \leqslant \varepsilon_4/10$, so we conclude that

$$\left|\mathbb{E}\left[\mathrm{T}_{\rho_1'}(\chi_{S_1'} P_2(\mathrm{T}_{1/2})(\chi_{S_1} f_1))(x^1)\prod_{i=2}^{4} \mathrm{T}_{\rho_i'}(\chi_{S_i'} h_i)(x^i)\right]\right| \geqslant \frac{9\varepsilon_4}{10}.$$

Repeating the argument that now expands our P_2, saying that the constant term is negligible and finding the maximal contribution, we that if $\alpha_0, \ldots, \alpha_d$ are the coefficients of P_2, then there is $j \geqslant 1$ such that

$$\left|\mathbb{E}\left[\mathrm{T}_{\rho_1'}(\chi_{S_1'} \mathrm{T}_{1/2^j}(\chi_{S_1} f_1))(x^1)\prod_{i=2}^{4} \mathrm{T}_{\rho_i'}(\chi_{S_i'} h_i)(x^i)\right]\right| \geqslant \frac{\frac{9\varepsilon_4}{10} - C(D,q,t)\alpha_0}{\sum\limits_{j=1}^{d}|\alpha_j|} \geqslant \varepsilon_5 = \Omega_{\varepsilon_4,\xi}(1).$$

We choose $\rho_1 = 1/2^j$. Continuing in this way for h_2, h_3 and h_4, the lemma is proved. □

4.5 Preparation for the Invariance Principle Argument: Resilience and Regularity

The invariance principle of [MOO05] applies to functions with small influences, however it was observed by Mossel [Mos20] that appealing to an influence regularity lemma, one may apply the invariance principle to *resilient functions*. We too intend to apply the invariance principle in our setting, and towards this end we need to show the functions in our setting to be resilient (or rather, that they can be assumed to be resilient), as well as state and prove a regularity lemma for our setting. In this section, we establish these preparatory steps, and in the next section we use them in order to apply the invariance principle (we remark that the presentation and focus herein is also similar to [CFM+22]).

4.5.1 Resilient Functions

Definition 4.17. *Let μ be a probability measure over $\{0,1\}$. A function $f\colon (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$ is called (r,ε) resilient if for any $S \subseteq [n]$ of size at most r and any $s \in \{0,1\}^S$,*

$$\left| \mathbb{E}_{x\sim\mu^{\otimes n}}\left[f(x) \mid x_S = s\right] - \mathop{\mathbb{E}}_{x\sim\mu^{\otimes n}}\left[f(x)\right] \right| \leqslant \varepsilon.$$

In words, restricting any set of at most r coordinates changes the average of f by at most ε.

To relate this notion to our setting, we show that if a function f is not correlated with any χ_S, then $\mathrm{T}_{\rho'}(\chi_{S'}\mathrm{T}_\rho(\chi_S f))$ is resilient for all S, $S' \subseteq S$ of bounded size and $\rho, \rho' \in (0,1)$.

Lemma 4.18. *For all $\varepsilon > 0$, $r, t \in \mathbb{N}$ and $q \in (0,1)$ there is a $\delta > 0$ such that the following holds. Suppose $f\colon (\{0,1\}^n, \mu_q) \to \mathbb{R}$ is a function such that $|\langle f, \chi_S\rangle| \leqslant \delta$ for all $S \subseteq [n]$. Then for all S and $S' \subseteq S$ of size at most t, $\rho, \rho' \in (0,1)$, the function $g = \mathrm{T}_{\rho'}(\chi_{S'}\mathrm{T}_\rho(\chi_S f))$ is (r,ε) resilient. Moreover, $|\mathbb{E}[g]| \leqslant \varepsilon$.*

Proof. Fix R of size at most r, $z \in \{0,1\}^S$ and denote $h(x) = 1_{x_R = z}$. Then we have to show that

$$|\langle g, h\rangle - \mathbb{E}[g]\,\mathbb{E}[h]| \leqslant \varepsilon\,|\mathbb{E}[h]|, \qquad |\mathbb{E}[g]| \leqslant \varepsilon.$$

First, note that

$$|\mathbb{E}[g]| = |\mathbb{E}[\chi_{S'}\mathrm{T}_\rho(\chi_S f)]| = |\langle \chi_{S'}, \mathrm{T}_\rho(\chi_S f)\rangle| = |\langle \mathrm{T}_\rho \chi_{S'}, \chi_S f\rangle| = |\langle \chi_S \mathrm{T}_\rho \chi_{S'}, f\rangle|.$$

Note that the function $\mathrm{T}_\rho\chi_{S'}$ only depends on coordinates form S', and the set $\{\chi_T\}_{T\subseteq S'}$ is a basis for such functions, so we may write

$$\mathrm{T}_\rho\chi_{S'}(x) = \sum_{T\subseteq S'} \alpha_T \chi_T(x),$$

where the coefficients α_T satisfy that $|\alpha_T| = O_{q,r}(1)$. It follows that

$$|\mathbb{E}[g]| \leqslant |\langle \chi_S \mathrm{T}_\rho \chi_{S'}, f\rangle| \leqslant \sum_{T\subseteq S'} |\alpha_T|\,|\langle \chi_S\chi_T, f\rangle| \leqslant \sum_{T\subseteq S'} |\alpha_T|\,|\langle \chi_S\chi_T, f\rangle| = O_{q,r}(\delta) \leqslant \frac{\varepsilon}{2}$$

for sufficiently small δ. Thus, to finish the proof it suffices to bound $\langle g, h\rangle$.

Similarly to before, $\langle g, h\rangle = \langle f, \chi_S \mathrm{T}_\rho(\chi_{S'}\mathrm{T}_{\rho'} h)\rangle$. As h depends only on coordinates from R, we may write

$$h(x) = \sum_{R'\subseteq R} \alpha_{R'}\chi_{R'},$$

where $|\alpha_{R'}| = O_{q,r}(1)$ for all R', so that

$$|\langle g, h\rangle| \leqslant \sum_{R'\subseteq R} |\alpha_{R'}|\,|\langle g, h\rangle \langle f, \chi_S \mathrm{T}_\rho(\chi_{S'}\chi_{R'})\rangle|.$$

Again, for S', R' we may write

$$\mathrm{T}_\rho(\chi_{S'}\chi_{R'}) = \mathrm{T}_\rho(\chi_{S'\oplus R'}) = \sum_{L\subseteq S'\oplus R'} \beta_L \chi_L,$$

where $|\beta_L| = O_{q,r}(1)$ for all L, so

$$|\langle g, h\rangle| \leqslant \sum_{R'\subseteq R}\sum_{L\subseteq S'\oplus R'} |\alpha_{R'}|\,|\beta_L|\,|\langle f, \chi_S\chi_L\rangle| \leqslant \sum_{R'\subseteq R}\sum_{L\subseteq S'\oplus R'} |\alpha_{R'}|\,|\beta_L|\,\delta = O_{q,r}(\delta) \leqslant \frac{\varepsilon}{2}(q(1-q))^r$$

for sufficiently small δ. The proof is concluded since $|\mathbb{E}[h]| \geqslant (q(1-q))^r$. $\square$

In words, Lemma 4.18 tells us that if the functions $f_1, \dots, f_4$ in the setting of Theorem 1.3 do not satisfy the conclusion of the theorem, then the functions in the conclusion of Lemma 4.16 are resilient.

4.5.2 The Regularity Lemma

Next, we describe the regularity lemma we intend to use, and towards this end we define the notion of regular functions.

Definition 4.19. *We say a function $f\colon (\{0,1\}^n, \mu_q^{\otimes n}) \to \mathbb{R}$ is (d,τ) regular if $\max_{i\in[n]} I_i^{\leqslant d}[f] \leqslant \tau$.*

The following lemma asserts that given functions $f_1, \dots, f_\ell$ with bounded 2-norm, one can find a set T consisting of constantly many coordinates such that randomly restricting the functions $f_1, \dots, f_\ell$ on T yields that all functions are regular with high probability.

Lemma 4.20. *For all $\ell, d \in \mathbb{N}$, $\xi, \tau > 0$ and $q \in (0,1)$ there is J such the following holds. Suppose $f_1, \dots, f_\ell\colon (\{0,1\}^n, \mu_q^{\otimes n}) \to \mathbb{R}$ are functions of 2-norm at most 1; then there exists $T \subseteq [n]$ of size at most J such that*

$$\Pr_{z\sim\mu_q^T}[\exists j = 1, \dots, \ell \text{ such that } (f_j)_{T\to z} \text{ is not } (d,\tau) \text{ regular}] \leqslant \xi.$$

Proof. Let $\rho = 1/2$; starting with $T = \emptyset$, we will add elements to T, and inspect the potential function

$$p(T) = \sum_{j=1}^{\ell} \mathop{\mathbb{E}}_{z\sim\mu_q^T}[\mathsf{Stab}_\rho((f_j)_{T\to z})].$$

At each point in the argument, we will inspect the set of restrictions Z for which there is j such that $(f_j)_{T\to z}$ is not (d,τ) regular; for $z \in Z$, we denote such j by j_z and choose some i_z such that $I_{i_z}{}^{\leqslant d}[(f_{j_z})_{T\to z}] \geqslant \tau$. If $\mu_q(Z) \leqslant \xi$, we are done, and otherwise we take $T' = T \cup \bigcup_{z\in Z}\{i_z\}$, and show that $p(T') \geqslant p(T) + \Omega_{d,\xi}(1)$. We then iterate the process with T', and note that as for every T we have

$$p(T) \leqslant \sum_{j=1}^{\ell} \mathop{\mathbb{E}}_{z\sim\mu_q^T}\left[\|(f_j)_{T\to z}\|_2^2\right] = \sum_{j=1}^{\ell} \|f_j\|_2^2 \leqslant \ell,$$

the process must terminate after $O_{\ell,d,\xi}(1)$ steps and then we have $\mu_q(Z) \leqslant \xi$. As $|T'| \leqslant |T| + |Z| \leqslant |T| + ((1-q)q)^{-|T|}\frac{1}{\xi}$, it follows that the size of the final set T is also $O_{\ell,d,\xi}(1)$, so we will be done.

It remain to show that $p(T') \geqslant p(T) + \Omega_{d,\xi}(1)$. First, observe by a direct computation that for a function g and a coordinate i we have that

$$\mathop{\mathbb{E}}_{a\sim\mu_q}\left[\mathsf{Stab}_\rho(g_{\{i\}\to a})\right] = \mathsf{Stab}_\rho(g) + (1-\rho)I_i[\mathrm{T}_\rho g; \mu_q].$$

Fix $j = 1, \dots, \ell$, let E_j be the event that for z the function $(f_j)_{T\to z}$ is not (d,τ) regular and let i_z be a coordinate whose degree d influence is at least d. We get that

$$\mathop{\mathbb{E}}_{z\sim\mu_q^{T'}}[\mathsf{Stab}_\rho((f_j)_{T'\to z})] = \mathop{\mathbb{E}}_{\substack{z\sim\mu_q^T\\ z'\sim\mu_Q T\setminus T'}}\left[\mathsf{Stab}_\rho((f_j)_{\substack{T\to z\\ T'\to z'}})\right] \geqslant \mathop{\mathbb{E}}_{z\sim\mu_q^T, a\sim\mu_q}\left[\mathsf{Stab}_\rho((f_j)_{\substack{T\to z\\ \{i_z\}\to a}})\right],$$

where we think of z as fixed, and used the above observation for each coordinate in $T' \setminus T$ except for i_z, lower bounding the influence by 0. Using the observation again we get that

$$\begin{aligned}
\mathop{\mathbb{E}}_{z\sim\mu_q^{T'}}\left[\mathsf{Stab}_\rho((f_j)_{T'\to z})\right] &\geqslant \mathop{\mathbb{E}}_{z\sim\mu_q^{T}}\left[\mathsf{Stab}_\rho((f_j)_{T\to z}) + (1-\rho)I_{i_z}[\mathrm{T}_\rho (f_j)_{T\to z};\mu_q]\right]\\
&\geqslant \mathop{\mathbb{E}}_{z\sim\mu_q^{T}}\left[\mathsf{Stab}_\rho((f_j)_{T\to z}) + (1-\rho)\rho^d 1_{E_j}(z)\right]\\
&= \mathop{\mathbb{E}}_{z\sim\mu_q^{T}}\left[\mathsf{Stab}_\rho((f_j)_{T\to z})\right] + (1-\rho)\rho^d \Pr_{z\sim\mu_q^{T}}[E_j].
\end{aligned}$$

Summing over j yields that

$$p(T') \geqslant p(T) + (1-\rho)\rho^d \sum_{j=1}^{\ell} \Pr_{z\sim\mu_q^T}[E_j] \geqslant p(T) + (1-\rho)\rho^d \sum_{j=1}^{\ell} \Pr_{z\sim\mu_q^T}[E_j] \geqslant p(T) + (1-\rho)\rho^d \mu_q(Z),$$

which is at least $p(T) + \Omega_{d,\xi}(1)$, and we are done. □

In the next section, we will combine the regularity lemma and the resilience of the functions discussion in this section and show the expectation in the conclusion of Lemma 4.16 is close 0.

4.6 The Invariance Principle Argument

In this section, we show how to use the tools from the previous section in order to apply the invariance principle on the expectation in the conclusion of Lemma 4.16. Towards this end, we first present the basic set-up of the invariance principle.

Definition 4.21. *Let ν be a probability measure over $\{0,1\}^4$. The covariance matrix of ν, denoted as $\mathcal{G}[\nu]$, is a matrix in $\mathbb{R}^{4\times 4}$ whose i,j entry is equal to*

$$\mathcal{G}[\nu]_{i,j} = \sup_{\substack{f_i\colon (\{0,1\},\nu_i)\to\mathbb{R}, \mathbb{E}[f_i]=0\\ f_j\colon (\{0,1\},\nu_i)\to\mathbb{R}, \mathbb{E}[f_j]=0}} \frac{\mathbb{E}_{(x^i,x^j)\sim\nu_{i,j}}\left[f_i(x^i)f_j(x^j)\right]}{\sqrt{\mathbb{E}_{x^i\sim\nu_i}\left[f_i(x^i)^2\right]}\sqrt{\mathbb{E}_{x^j\sim\nu_j}\left[f_j(x^j)^2\right]}}.$$

For our distribution ν, since it is pairwise independent it follows that $\mathcal{G}[\nu] = I$. Given a covariance matrix, one can define the associated jointly distributed Gaussian distribution. For us, as $\mathcal{G}[\nu] = I$, the relevant Gaussian distribution will be that of $G^1,\ldots,G^4$ with co-variance matrix I, i.e. independently distributed standard Gaussian random variables. We denote this distribution by μ.

Definition 4.22. *We define the function* $\mathsf{clip}\colon \mathbb{R}\to[-1,1]$ *by* $\mathsf{clip}(z) = z$ *for* $z\in[-1,1]$, $\mathsf{clip}(z) = 1$ *if* $z > 1$ *and* $\mathsf{clip}(z) = -1$ *if* $z \leqslant -1$.

Given a function $f\colon (\{0,1\}^n,\mu_q)\to[-1,1]$, we may expand f according to the q-biased Fourier basis and write

$$f(x_1,\ldots,x_n) = \sum_{S\subseteq[n]} \widehat{f}(S;\mu_q)\chi_S^q(x).$$

Define the associated multi-linear polynomial $F(z_1,\ldots,z_n) = \sum\limits_{S\subseteq[n]} \widehat{f}(S;\mu_q)\prod\limits_{i\in S} z_i$, so that $f(x_1,\ldots,x_n) = F(\chi^q_{\{1\}}(x_1),\ldots,\chi^q_{\{n\}}(x_n))$. With this notations, and towards applying the invariance principle we think of F as a function over Gaussian space $\mathbb{R}^n$. We have the following consequence of the invariance principle of [MOO05, Mos10]

Lemma 4.23. *Let ν be a distribution over $\{0,1\}^4$ and μ be a distribution over $\mathbb{R}^4$ with the same covariance matrix as of ν.*

For all $\gamma, \varepsilon > 0$ there are $\tau > 0$ and $d \in \mathbb{N}$ such that if $f_1, \ldots, f_4 \colon (\{0,1\}^n, \mu_q^{\otimes n}) \to [-1,1]$ are function such that $\max_{i\in[n]} I_i^{\leqslant d}[g_j] \leqslant \tau$ for all $j = 1, \ldots, 4$, and $\gamma_1, \ldots, \gamma_4 \in [\gamma, 1]$. Then, looking at $g_j = \mathrm{T}_{1-\gamma_j} f_j$ and considering its associated multi-linear polynomial G_j, we have

$$\left| \mathop{\mathbb{E}}_{(x^1,\ldots,x^4)\sim\nu^{\otimes n}} \left[\prod_{j=1}^{4} \mathrm{T}_{1-\gamma_j} f_j(x^j) \right] - \mathop{\mathbb{E}}_{(z^1,\ldots,z^4)\sim\mu^{\otimes n}} \left[\prod_{j=1}^{4} \mathsf{clip}(G_j)(z^j) \right] \right| \leqslant \varepsilon.$$

Proof. The proof is the same as [CFM+22, Theorem 5.9]. □

Let $f_1, \ldots, f_4$ be as in Theorem 1.3, and take $g_i = \mathrm{T}_{\rho_i}(\chi_{S_i} f_i)$ and $h_i = \chi_{S_i'} g_i$ as in Lemma 4.16. The following lemma shows an upper bound on the expectation in Lemma 4.16 in the case that each f_i is not correlated with any function of the form χ_S. In the next section, we combine it with Lemma 4.16 to prove Theorem 1.3.

Lemma 4.24. *For all $t \in \mathbb{N}$, $q \in (0,1)$ and $\xi > 0$, there are $r \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose $f_1, \ldots, f_4 \colon (\{0,1\}^n, \mu_q^{\otimes n}) \to [-1,1]$ are functions that satisfy that $|\langle f_j, \chi_S \rangle| \leqslant \delta$ for all $j = 1, \ldots, 4$ and $S \subseteq [n]$, then*

$$\left| \mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} [h_1(x)h_2(y)h_3(z)h_4(w)] \right| \leqslant \xi,$$

where $g_j = \mathrm{T}_{\rho_j}(\chi_{S_j} h_j)$, $h_j = \mathrm{T}_{\rho_j'}(\chi_{S_j'} g_j)$ and $\left|S_j'\right| \leqslant t$, $0 < \rho_j, \rho_{j'} \leqslant 1/2$ for $j = 1, \ldots, 4$.

Proof. We use the parameters

$$0 < \delta \ll r^{-1}, \varepsilon \ll d^{-1} \ll J^{-1} \ll \tau \ll t^{-1}, \xi \leqslant 1.$$

Let $f_1, \ldots, f_4$ be functions as in the statement of Lemma 4.24. By Lemma 4.20, we may find a set of coordinates $T \subseteq [n]$ of size at most J such that, denoting by E_i the event that $(f_i)_{T\to u}$ is (r, τ) regular for $u \sim \mu_q^T$, we have that $\Pr\left[\bar{E}_1 \vee \ldots \vee \bar{E}_4\right] \leqslant \frac{\xi}{10}$. Thus, denoting $H_i = \chi_{S_i'} \mathrm{T}_{\rho_i}(f_i \chi_{S_i})$ and recalling that $h_i = \mathrm{T}_{\rho_i} H_i$, we may write

$$\begin{aligned}
\left| \mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} [h_1(x)h_2(y)h_3(z)h_4(w)] \right| &= \left| \mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}} \left[\mathrm{T}_{\rho_1'} H_1(x) \mathrm{T}_{\rho_2'} H_2(y) \mathrm{T}_{\rho_3'} H_3(z) \mathrm{T}_{\rho_4'} H_4(w) \right] \right| \\
&= \left| \mathop{\mathbb{E}}_{\substack{(x,y,z,w)\sim\nu^{\otimes n} \\ (x',y',z',w')\sim_{\vec{\rho}'}(x,y,z,w)}} \left[H_1(x')H_2(y')H_3(z')H_4(w') \right] \right|,
\end{aligned}$$

where by the notation $(x', y', z', w') \sim_{\vec{\rho}'} (x, y, z, w)$ we mean that for each $i = 1, \ldots, n$ independently, we take $x_i' = x_i$ with probability ρ_1' and otherwise re-sample according to μ_q, independently we take $y_i' = y_i$ with probability ρ_2' and otherwise re-sample according to μ_q, independently we take $z_i' = z_i$ with probability

ρ'_3 and otherwise re-sample according to μ_q, and independently we take $w'_i = w_i$ with probability ρ'_4 and otherwise re-sample according to μ_q. We re-write the last expectation as

$$\left|\mathop{\mathbb{E}}_{\substack{(a,b,\alpha,\beta)\sim\nu^T\\(a',b',\alpha',\beta')}}\left[\mathop{\mathbb{E}}_{\substack{(x,y,z,w)\sim\nu^{[n]\setminus T}\\(x',y',z',w')\sim_{\vec{\rho}'}(x,y,z,w)}}\left[(H_1)_{T\to a'}(x')(H_2)_{T\to b'}(y')(H_3)_{T\to\alpha'}(z')(H_4)_{T\to\beta'}(w')\right]\right]\right|,$$

which is the same as

$$\left|\mathop{\mathbb{E}}_{(a',b',\alpha',\beta')\sim\nu'^T}\left[\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{[n]\setminus T}}\left[\mathrm{T}_{\rho_1}(H_1)_{T\to a'}(x)\mathrm{T}_{\rho_2}(H_2)_{T\to b'}(y)\mathrm{T}_{\rho_3}(H_3)_{T\to\alpha'}(z)\mathrm{T}_{\rho_4}(H_4)_{T\to\beta'}(w)\right]\right]\right|.$$

Above, ν' is a distribution which can be explicitly described, however it is only important for us that its marginal on each coordinate is μ_q. Thus, denoting $E = E_1(a')\cap E_2(b')\cap E_3(\alpha')\cap E_4(\beta')$, we can upper bound the last quantity as

$$\mathop{\mathbb{E}}_{(a',b',\alpha',\beta')\sim\nu'^T}\left[1_E\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{[n]\setminus T}}\left[\mathrm{T}_{\rho'_1}(H_1)_{T\to a'}(x)\mathrm{T}_{\rho'_2}(H_2)_{T\to b'}(y)\mathrm{T}_{\rho'_3}(H_3)_{T\to\alpha'}(z)\mathrm{T}_{\rho'_4}(H_4)_{T\to\beta'}(w)\right]\right|\right] + \Pr\left[\bar{E}\right], \tag{11}$$

where we used the fact that $\|H_j\|_\infty \leqslant 1$ for all $j = 1,\dots,4$. We have $\Pr\left[\bar{E}\right] \leqslant \frac{\xi}{10}$. For the first term, fix a', b', α', β' such that E holds; then the functions $H'_1 = (H_1)_{T\to a'}$, $H'_2 = (H_2)_{T\to b'}$, $H'_3 = (H_3)_{T\to\alpha'}$ and $H'_4 = (H_4)_{T\to\beta'}$ are all (d,τ) regular, so by Lemma 4.23 we get

$$\begin{aligned}&\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{[n]\setminus T}}\left[\mathrm{T}_{\rho'_1}(H_1)_{T\to a'}(x)\mathrm{T}_{\rho'_2}(H_2)_{T\to b'}(y)\mathrm{T}_{\rho'_3}(H_3)_{T\to\alpha'}(z)\mathrm{T}_{\rho'_4}(H_4)_{T\to\beta'}(w)\right]\right|\\ &\leqslant\left|\mathop{\mathbb{E}}_{(z^1,z^2,z^3,z^4)\sim\mu^{n-J}}\left[\mathsf{clip}(G_1)(z^1)\mathsf{clip}(G_2)(z^2)\mathsf{clip}(G_3)(z^3)\mathsf{clip}(G_4)(z^4)\right]\right| + \frac{\xi}{10},\end{aligned} \tag{12}$$

where G_1 is the multi-linear polynomial associated with $\mathrm{T}_{\rho'_1}(H_1)_{T\to a'}$, and similarly for G_2, G_3 and G_4. Note that the covariance matrix of ν is the identity matrix, hence the covariance matrix of μ is also the identity, so the random variables $z^1,\dots,z^4$ are independent, hence

$$\left|\mathop{\mathbb{E}}_{(z^1,z^2,z^3,z^4)\sim\mu^{n-J}}\left[\mathsf{clip}(G_1)(z^1)\mathsf{clip}(G_2)(z^2)\mathsf{clip}(G_3)(z^3)\mathsf{clip}(G_4)(z^4)\right]\right| = \prod_{j=1}^{4}\left|\mathbb{E}[\mathsf{clip}(G_j)]\right|. \tag{13}$$

Fix j; without loss of generality, $j = 1$. Using Lemma 4.23 again, this time with the functions f_2, f_3, f_4 being the constant 1 function, we get that

$$\left|\mathbb{E}[\mathrm{T}_{\rho'_1}(H_1)_{T\to a'}] - \mathbb{E}[\mathsf{clip}(G_1)]\right| \leqslant \frac{\xi}{10}.$$

However, we have $\left|\mathbb{E}[\mathrm{T}_{\rho'_1}(H_1)_{T\to a'}]\right| = \left|\mathbb{E}[(H_1)_{T\to a'}]\right|$, and by the proof of Lemma 4.18 the function H_1 is (r,ε) resilient, and as $r > J$ we get that

$$\left|\mathbb{E}[(H_1)_{T\to a'}]\right| \leqslant \varepsilon + \left|\mathbb{E}[H_1]\right| \leqslant 2\varepsilon,$$

where in the last inequality we used Lemma 4.18 again. Thus, $|\mathbb{E}[\mathsf{clip}(G_1)]| \leqslant 2\varepsilon + \frac{\xi}{10} \leqslant \frac{\xi}{5}$, and plugging into (12) we get that $\left|\mathbb{E}_{(z^1,z^2,z^3,z^4)\sim\mu^{n-J}}\left[\mathsf{clip}(G_1)(z^1)\mathsf{clip}(G_2)(z^2)\mathsf{clip}(G_3)(z^3)\mathsf{clip}(G_4)(z^4)\right]\right| \leqslant \frac{\xi}{5}$, and plugging this into (13) yields that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{[n]\setminus T}}\left[\mathrm{T}_{\rho_1}(H_1)_{T\to a'}(x)\mathrm{T}_{\rho_2}(H_2)_{T\to b'}(y)\mathrm{T}_{\rho_3}(H_3)_{T\to\alpha'}(z)\mathrm{T}_{\rho_4}(H_4)_{T\to\beta'}(w)\right]\right| \leqslant \frac{3\xi}{10}.$$

Finally, plugging this into (11) and then all the way above yields that

$$\left|\mathop{\mathbb{E}}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[h_1(x)h_2(y)h_3(z)h_4(w)\right]\right| \leqslant \frac{3\xi}{10} + \frac{\xi}{10} < \xi,$$

and we are done. □

4.6.1 Concluding the Proof of Theorem 1.3

In this section, we finish up the proof of Theorem 1.3, presented below.

Proof of Theorem 1.3. We use the parameters

$$0 < \delta \ll \eta, r^{-1} \ll t^{-1}, \varepsilon' \ll \varepsilon, q, 1-q < 1.$$

Assume that $\left|\mathbb{E}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[f_1(x)f_2(y)f_3(z)f_4(w)\right]\right| \geqslant \varepsilon$; then by Lemma 4.16 we find $h_1,\ldots,h_4$ defined therein such that $\left|\mathbb{E}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[h_1(x)h_2(y)h_3(z)h_4(w)\right]\right| \geqslant \varepsilon'$, and $S'_1,\ldots,S'_4$ are of size at most t.

Assume towards contradiction that each f_i satisfies that $|\langle f_i,\chi_S\rangle| \leqslant \delta$ for all $i = 1,\ldots,4$. Then by Lemma 4.18 each h_i is (r,η)-resilient, so by Lemma 4.24 $\left|\mathbb{E}_{(x,y,z,w)\sim\nu^{\otimes n}}\left[h_1(x)h_2(y)h_3(z)h_4(w)\right]\right| < \varepsilon'$, and contradiction. □

5 Direct product testing

This section is devoted for the proof of Theorem 1.4.

Notations. Given a string $x \in \{0,1\}^n$ and a subset $S \subseteq [n]$, we use the notation $x|_S$ to denote the part of the string x restricted to the set S. Given $x, y \in \{0,1\}^n$, we use the notation $x \overset{\geqslant\alpha}{\neq} y$ to denote that the set $\{i \in [n] \mid x_i \neq y_i\}$ is of size at least α. Similarly, we use $x \overset{\leqslant\alpha}{\neq} y$ to denote that the strings x and y differ at at most α locations.

Inclusion graphs are graphs whose vertices are subsets of some finite universe, and two vertices (subsets) are connected by an edge iff one is contained in the other. Consider the bipartite inclusion graph $G(n,k,t) = G(X\cup Y, E)$ between $X = \binom{[n]}{k}$ and $Y = \binom{[n]}{t}$ for some $k < t$. The following lemma from [IKW12] will be useful for our analysis.

Lemma 5.1. *Let $G(n,k,t)$ be the inclusion graph for $k < t$. Let $Y' \subseteq Y$ be any subset of measure $\rho < 1/2$. For any constant $0 < \nu < 1$, we have that for all but at most $O_\nu\left(\frac{\log 1/\rho}{t/k}\right)$ fraction of vertices $x \in X$, we have*

$$\left|Pr_{y\in N(x)}[y \in Y'] - \rho\right| \leqslant \nu\rho.$$

Here, $N(x)$ is the neighbors of the vertex x in G.

The set up for the direct product testing. Fix constants $q, q' \in (0,1)$ such that $q' < q$ and an integer $\beta \geqslant 0$. Suppose we are given a table $F : \binom{[n]}{qn} \to \{0,1\}^{qn}$ where $F[S] \in \{0,1\}^{qn}$ can be thought of as assigning a bit to every element in S (by associating some fixed ordering on the elements of $[n]$). Consider the agreement test (Agreement-Test) parameterized by (q, q', β) given in Figure 5. If the table F is coming

- Pick a random set $A_0 \cup B_0$ of size qn where $|A_0| = q'n$
- Select a random set $B_1 \subseteq [n] \setminus A_0$ of size $(q - q')n$
- Check if $F[A_0 \cup B_0]|_{A_0} \overset{\leqslant\beta}{\neq} F[A_0 \cup B_1]|_{A_0}$

Figure 1: Agreement-Test with parameters (q, q', β).

from a global string $a \in \{0,1\}^n$, i.e., $F[S] = a|_S$, then it is clear that the test accepts with probability 1 (even when $\beta = 0$). We wish to conclude that the table F has a similar structure even if the test passes with probability at least ε for *every* constant $\varepsilon > 0$. We show this is the case when $\beta = 0$.[4] We will use the fact that the parameter β can be set to a non-zero value in the proof of the following main theorem.

Theorem 5.2 (Restatement of Theorem 1.4)**.** *For all $0 < q' < q < 1$ and $\varepsilon > 0$, there are $r \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose that $F : \binom{[n]}{qn} \to \{0,1\}^{qn}$ satisfies*

$$\Pr_{(S_1,S_2)\sim\mathcal{D}_{q,q'}} [F[S_1]|_{S_1\cap S_2} = F[S_2]|_{S_1\cap S_2}] \geqslant \varepsilon.$$

Then there exists a function $g : [n] \to \{0,1\}$ such that for at least a δ fraction of $S \in \binom{[n]}{qn}$, we have $|\{i \in S \mid F[S]_i \neq g(i)\}| \leqslant r$.

To get the local structure in Section 5.1, we follow the same proof strategy as in [IKW12], except that we change the parameters from a key definition of "*excellent*" tailored to our setting. An important distinction between the proof strategy of [IKW12] and our work is in the final step of showing consistency between different functions with local structure, we crucially use the small-set expansion property of a certain graph defined on a multi-slice of $\{0,1,2\}^n$. The is shown in Section 5.2.

Throughout the next two subsections, we use ε to denote the passing probability of the Agreement-Test.

5.1 Local structure

In this section, we prove the local structure stated in Lemma 5.6 below. We need a few definitions to state the lemma.

Consider selecting a random set of size qn as follow. First sample a subset $A \subseteq [n]$ of size $q'n$ and then select a set $B \subseteq [n] \setminus A$ of size $(q - q')n$ uniformly at random. Output (A, B). We need the following few definitions that are similar to the definitions from the work [IKW12], adapted towards analyzing the Agreement-Test.

Definition 5.3. *(consistency) Fix a set (A, B). A subset $B' \subseteq [n] \setminus A$ is said to be β-consistent with (A, B) if $F[A, B]|_A \overset{\leqslant\beta}{\neq} F[A, B']|_A$. Let $Cons_\beta(A, B)$ be the set of all the sets that are β-consistent with (A, B).*

[4] It can be extended for any non-zero β, but for simplicity we only analyze the test with $\beta = 0$.

Definition 5.4. *(goodness) A set (A, B) is called (η, β)-good if*

$$\Pr_{B' \subseteq [n] \setminus A} [B' \in Cons_\beta(A, B)] \geqslant \eta.$$

Definition 5.5. *(excellence) A set (A, B) is called (η, β, r, γ)-excellent if it is (η, β)-good and*

$$\Pr_{E, D_1, D_2} [(E, D_i) \in Cons(A, B) \text{ for i=1,2 } \& F[A, E, D_1]|_E \neq F[A, E, D_2]|_E] \leqslant \gamma.$$

Here, $|E| = r$, $|D_i| = (q - q')n - r$, and (E, D_i) is a random subset of $[n] \setminus A$ of size $(q - q')n$.

For the proof of our Theorem 1.4, we are going to think of the parameter r as a constant independent of n, i.e., $r = O_\varepsilon(1)$.

Fix any (η, β, r, γ)-excellent pair (A_0, B_0). We define a function $g_{A_0,B_0} : [n] \to \{0, 1\}$ based on the majority vote of the table F restricted to the sets in $Cons_\beta(A_0, B_0)$. More formally, for $x \in [n] \setminus A_0$, we set

$$g_{A_0,B_0}(x) := \underset{B \in Cons_\beta(A_0,B_0) | B \ni x}{\mathsf{Majority}} F[A_0, B]|_x.$$

If there is no such B that contains x then we set $g_{A_0,B_0}(x) := 0$. We also set $g_{A_0,B_0}(A_0) = F[A_0, B_0]|_{A_0}$.

Based on these definitions, we prove the following local structure, which is the main lemma from this subsection. This is called a local structure as the functions g_{A_0,B_0} enjoy strong consistency (similar to what we need for the global function in Theorem 1.4) but only *locally* with the sets in $Cons_\beta(A_0, B_0)$.

Lemma 5.6. *For every $q, q' \in (0, 1)$ such that $q' < q$, $\alpha, \beta \geqslant 0$, $\varepsilon > 0$ and $r \geqslant 1$, if (A_0, B_0) is $(\varepsilon/2, \beta, r, \gamma)$-excellent then*

$$\Pr_{B \in Cons_\beta(A_0,B_0)} [F[A_0, B]|_B \overset{>\alpha}{\neq} g_{A_0,B_0}(B)] \leqslant \nu,$$

provided that $\nu = \Omega\left(\frac{\gamma(q-q')n}{\alpha\varepsilon}\right)$ and $\nu\varepsilon \gg \Omega\left(\frac{r^2 \ln 1/\varepsilon}{((q-q')n)}\right)$. Furthermore, a random pair (A_0, B_0) is $(\varepsilon/2, \beta, r, \gamma)$-excellent with probability at least $\frac{\varepsilon}{2} - \frac{r(r+2\beta)}{q'n\gamma}$, if the test **Agreement-Test** *passes with probability at least $\varepsilon > 0$.*

In order to get an intuitive understanding the lemma statement, consider the following setting of the parameters involved in the statement. Set β, r, α to be $\mathsf{poly}(1/\varepsilon)$ such that $\frac{\alpha}{2\beta . r^2} \gg \frac{1}{\varepsilon^4}$, $\gamma = \frac{r(r+2\beta)}{q'n} \frac{4}{\varepsilon}$ and $\nu = \Theta(\varepsilon)$. With these settings, the lemma states that if the Agreement-Test passes with probability at least ε, then there is at least $\varepsilon/4$ fraction of the pairs (A_0, B_0) that are $(\varepsilon/2, \beta, r, \gamma)$-excellent. For each of these excellent pairs (A_0, B_0), the function g_{A_0,B_0} satisfies the following property:

$$\Pr_{B \in Cons_\beta(A_0,B_0)} [F[A_0, B]|_B \overset{>\mathsf{poly}(1/\varepsilon)}{\neq} g_{A_0,B_0}(B)] \leqslant O(\varepsilon).$$

In other words, if the test passes with probability $\varepsilon > 0$, then at least $\varepsilon/4$ fraction of the functions g_{A_0,B_0} defined above have the property that for at least $1 - O(\varepsilon)$ fraction of the sets in $B \in Cons_\beta(A_0, B_0)$, the function g_{A_0,B_0} agrees with the table $F[A_0, B] \in \{0, 1\}^{qn}$ on all but $\mathsf{poly}(1/\varepsilon)$ many locations.

We now being to prove the lemma. For notational convenience, in the remaining part of this subsection, we call a pair (A, B) good if it is $(\varepsilon/2, \beta)$-good. Similarly, we call a pair (A, B) excellent if it is $(\varepsilon/2, \beta, r, \gamma)$-excellent. Furthermore, we also suppress the subscript of $Cons_\beta$ and simply write it as $Cons$ as the subscript β will stay the same throughout this subsection.

We start by showing that a random pair is good and excellent with high probability.

Claim 5.7. *If* $\Pr_{A_0,B_0,B_1}[F[A_0, B_0]|_{A_0} \overset{\leqslant\beta}{\neq} F[A_0, B_1]|_{A_0}] \geqslant \varepsilon$, *then a random* (A_0, B_0) *is* $(\varepsilon/2, \beta)$*-good with probability at least* $\varepsilon/2$.

Proof. Averaging argument. □

The next claim shows that almost all the good pairs are excellent, provided $\gamma \gg \frac{r(r+2\beta)}{q'n}$.

Claim 5.8. $\Pr_{A_0,B_0}[(A_0, B_0)$ *is* $(\varepsilon/2, \beta)$*-good but not* $(\varepsilon/2, \beta, r, \gamma)$*-excellent*$] \leqslant \frac{1}{\gamma}\frac{r(r+2\beta)}{q'n}$.

Proof. Consider the following two events.

1. Event Z_1: (A_0, B_0) is good but
$$\Pr_{E,D_1,D_2}[(E, D_i) \in Cons(A_0, B_0) \text{ for i=1,2 } \& F[A_0, E, D_1]|_E \neq F[A_0, E, D_2]|_E] > \gamma.$$

2. Event Z_2: (A_0, B_0) is good, $(E, D_i) \in Cons(A_0, B_0)$ for $i = 1, 2$ and
$$F[A_0, E, D_1]|_{A_0\cup E} \overset{\geqslant 2\beta+1}{\neq} F[A_0, E, D_2]|_{A_0\cup E}.$$

We are interested in $\Pr[Z_1]$. We have $\Pr[Z_1] = \Pr[Z_1\&Z_2]/\Pr[Z_2 \mid Z_1] \leqslant \Pr[Z_2]/\Pr[Z_2 \mid Z_1]$. Note that $\Pr[Z_2 \mid Z_1] \geqslant \gamma$.

We now upper bound $\Pr[Z_2]$. The sets from the event Z_2 can be equivalently sampled as follows. First, sample a random subset A' of $[n]$ of size $q'n + r$ and then pick random sets $D_1, D_2 \subseteq [n] \setminus A'$ of size $(q - q')n - r$, conditioned on the event $F[A', D_1]|_{A'} \overset{\geqslant 2\beta+1}{\neq} F[A', D_2]|_{A'}$. Let $A'' \subseteq A'$ be the set of coordinates where the disagreement occurs. Set A_0 to be a random subset of A' of size $q'n$ and set $E = A' \setminus A_0$.

Observe that if $(E, D_i) \in Cons(A_0, B_0)$ for $i = 1, 2$ then $F[A_0, E, D_1]|_{A_0} \overset{\leqslant 2\beta}{\neq} F[A_0, E, D_2]|_{A_0}$ and hence $|A''| \leqslant r + 2\beta$, if the former event has to happen. Thus, if $(E, D_i) \in Cons(A_0, B_0)$, the expected size of $E \cap A''$ is at most $\frac{(r+2\beta)\cdot r}{q'n+r}$. Therefore, by Markov's inequality, the probability that $|E \cap A''| \geqslant 1$ is at most $\frac{r\cdot(r+2\beta)}{(q'n+r)} \leqslant \frac{r(r+2\beta)}{q'n}$. Therefore, $\Pr[Z_2] \leqslant \frac{r(r+2\beta)}{q'n}$ and the claim follows. □

The following important claim shows that the function g_{A_0,B_0} defined above for an excellent pair (A_0, B_0) enjoys strong agreement with F restricted to the sets in $Cons(A_0, B_0)$.

Claim 5.9. *For every* $q, q' \in (0, 1)$ *and* $r \geqslant 1$, *if* (A_0, B_0) *is excellent then*
$$\Pr_{B\in Cons(A_0,B_0)}[F[A_0, B]|_B \overset{>\alpha}{\neq} g_{A_0,B_0}(B)] \leqslant \nu,$$
provided that $\nu = \Omega\left(\frac{\gamma(q-q')n}{\alpha\varepsilon}\right)$ *and* $\nu\varepsilon \gg O\left(\frac{r^2 \ln 1/\varepsilon}{((q-q')n)}\right)$.

The proof of Lemma 5.6 follows from the above claim along with the Claim 5.7 and Claim 5.8.

Before we prove this claim, we need a few notations and a few claims. Fix an excellent pair (A_0, B_0). Let $Cons^x = \{B \in Cons(A_0, B_0) \mid x \in B\}$. Note that this set is used to define $g_{A_0,B_0}(x)$. For a subset $E \subseteq [n] \setminus A_0$ of size r, let B_E be the set of all $B \subseteq [n] \setminus A_0$ such that $|B| = (q - q')n$ and $E \subseteq B$. Also, let $Cons_E = Cons(A_0, B_0) \cap B_E$.

Claim 5.10. *For at least* $1 - O\left(\frac{\ln 1/\varepsilon}{(q-q')n}\right)$ *fraction of* $x \in [n] \setminus A_0$*, we have* $|Cons^x| \geqslant \frac{\varepsilon}{6}|B_x|$*. For at least* $1 - O\left(\frac{\ln 1/\varepsilon}{(q-q')n/r}\right)$ *fraction of* $E \subseteq [n] \setminus A_0$*, we have* $|Cons_E| \geqslant \frac{\varepsilon}{6}|B_E|$

Proof. Claim 3.8 and Claim 3.12 from [IKW12] (follows by invoking Lemma 5.1 for the graphs $G(n, r, (q-q')n)$ and $G(n, 1, (q - q')n)$). □

Claim 5.11. *For x as above, for at least* $1 - O\left(\frac{\ln 1/\varepsilon}{((q-q')n/r)}\right)$ *fraction of r-sets E containing x,*

$$\Pr_{B \in Cons_E}[F[A_0, B]|_x = g_{A_0,B_0}(x)] \geqslant \frac{1}{10}.$$

Proof. Consider the bipartite inclusion graph between the sets $E_x = \{X \subseteq [n] \setminus \{x\} \mid |X| = r - 1\}$ and $B_x = \{Y \subseteq [n] \setminus \{x\} \mid |Y| = (q - q')n - 1\}$. We know that the set $Cons^x = \{Y \cup \{x\} \in Cons(A_0, B_0)\}$ has density at least $\varepsilon/6$ in B_x. Since $g(x)$ is defined as the majority among $Cons^x$, for at least $\varepsilon/12$ fraction of the sets in B_x, $g(x)$ agrees with the sets on x. Let this subset be Q.

By the sampler property of the bipartite inclusion graph $G(n, r - 1, (q - q')n - 1)$ from Lemma 5.1, we have that for all but at most $O\left(\frac{\ln 1/\varepsilon}{(q-q')n/r}\right)$ fraction of the sets in E_x are such that, among the sets in B_x containing E_x, the measure of those B_x that fall into $Cons^x$ is in between $\frac{1}{3}\frac{\varepsilon}{6}$ and $\frac{5}{3}\frac{\varepsilon}{6}$. Simultaneously, the measure of those $B_x \supset E_x$ that fall into Q is between $\frac{1}{3}\frac{\varepsilon}{12}$ and $\frac{5}{3}\frac{\varepsilon}{12}$, for all but at most at most $O\left(\frac{\ln 1/\varepsilon}{(q-q')n/r}\right)$ fraction of the sets E_x. Therefore, for at least $1 - O\left(\frac{\ln 1/\varepsilon}{(q-q')n/r}\right)$ fraction of the sets E_x, we have

$$\Pr_{Y|B_x \supset E_x}[F[A_0, Y, x]|_x = g(x) \mid Y \cup \{x\} \in Cons^x] \geqslant \frac{\frac{1}{3}\frac{\varepsilon}{12}}{\frac{5}{3}\frac{\varepsilon}{6}} \geqslant \frac{1}{10}.$$

□

Claim 5.12. *For at least* $1 - O\left(\frac{r \ln 1/\varepsilon}{((q-q')n/r)}\right)$ *fraction of E, we have for all $x \in E$,*

$$\Pr_{B \in Cons_E}[F[A_0, B]|_x = g_{A_0,B_0}(x)] \geqslant \frac{1}{10}.$$

Proof. Let $\delta = O\left(\frac{\ln 1/\varepsilon}{((q-q')n/r)}\right)$. Suppose for contradiction, for at least $r\delta$ fraction of E, there exists an $x \in E$ such that

$$\Pr_{B \in Cons_E}[F[A_0, B]|_x = g_{A_0,B_0}(x)] \leqslant \frac{1}{10}.$$

This implies that if we select a random E and then a random $x \in E$, then with probability at least δ, the above inequality does not hold. Now, selecting a random E and an $x \in E$ is same as first selecting a random x and then a random $E \ni x$. Therefore, this contradicts Claim 5.11. □

We are now ready to prove Claim 5.9.

Proof of Claim 5.9. Let $\delta = O\left(\frac{r \ln 1/\varepsilon}{((q-q')n/r)}\right)$. Using Claim 5.10 and Claim 5.12, we get that for at least $1-\delta$ fraction of $E \in [n]\setminus A_0$, $|Cons_E| \geqslant \frac{\varepsilon}{6}|B_E|$ and for all $x \in E$, $\Pr_{B \in Cons_E}[F[A_0, B]|_x = g_{A_0,B_0}(x)] \geqslant \frac{1}{10}$. Towards contradiction, let us assume

$$\Pr_{B \subseteq [n]\setminus A_0}[B \in Cons(A_0, B_0) \& F[A_0, B]|_B \overset{>\alpha}{\neq} g_{A_0,B_0}(B)] > \nu\varepsilon.$$

Consider picking a random B satisfying the above event. Pick a random r-subset E of B. With probability at least $1-O(\delta)$, we have that such an E satisfies $|Cons_E| \geqslant \frac{\varepsilon}{6}|B_E|$ and for all $x \in E$, $\Pr_{B' \in Cons_E}[F[A_0, B']|_x = g_{A_0,B_0}(x)] \geqslant \frac{1}{10}$, provided $\nu\varepsilon \gg \delta$. Let E' be the set of $x \in E$ such that $F[A_0, B]|_x \neq g_{A_0,B_0}(x)$. The probability that $|E'| \neq 0$ is at least $\frac{\alpha}{(q-q')n}$.

$$\Pr_{B' \in Cons_E}[F[A_0, B]|_E \neq F[A_0, B']|_E] \mid |E'| \neq 0, B \in Cons(A_0, B_0) \text{ \& } F[A_0, B]|_B \overset{>\alpha}{\neq} g_{A_0,B_0}(B)] \geqslant \frac{1}{10}.$$

Removing the conditioning,

$$\Pr_{B \subseteq [n]\setminus A_0, E \subseteq B, B' \supseteq E}[B, B' \in Cons(A_0, B_0) \text{ \& } F[A_0, B]|_E \neq F[A_0, B']|_E] \geqslant \Omega\left(\frac{\alpha\nu\varepsilon}{(q-q')n}\right),$$

which is a contradiction to the fact that (A_0, B_0) is (r, γ) excellent, provided $\nu \geqslant \Omega\left(\frac{\gamma(q-q')n}{\alpha\varepsilon}\right)$. □

5.2 Consistency between the functions $g_{A,B}$

Now that we have a function g_{A_0,B_0} for every excellent pair (A_0, B_0), the last step is to show that these functions are similar to each other and hence there is a global function g that (almost) agrees with at least $\delta(\varepsilon)$ fraction of the entries from the table $F[.]$. In order to show this, we differ from the analysis given in [IKW12]. Towards this, we define another version of Cons, that we call the newCons, as follows.

Fix a constant $c \in (0, 1)$.[5] Consider selecting a random set of size qn as follow. First, sample a subset $A \subseteq [n]$ of size $q'n$ and then select a set $B \subseteq [n] \setminus A$ of size $(q-q')n$ uniformly at random. Select a random subset of A of size $cq'n$ and call it D. Let $E = A \setminus D$. Output (D, E, B), where $A = D \cup E$. Consider the modified agreement test (Test') given in Figure 5.2.

- Pick a random set $(D \cup E \cup B)$ of size qn where $|D| = cq'n$, $|E| = (1-c)q'n$ and $|B| = (q-q')n$.
- Select a random subset $E' \subseteq [n] \setminus (D \cup E)$ of size $(1-c)q'n$.
- Select a random set $B' \subseteq [n] \setminus (D \cup E')$ u.a.r. where $|B'| = (q-q')n$.
- Check if $F[D \cup E \cup B]|_D = F[D \cup E' \cup B']|_D$

Figure 2: Modified agreement test Test'

[5]The constant c will depend on q, q'.

Note that Test' is similar to the agreement test Agreement-Test that we wish to analyze, except that the we change the parameters from (q', q, β) to $(cq'n, q, 0)$ for $c \in (0,1)$. Another (minor) difference is that we require the sets E and E' to be disjoint in the above distribution, whereas in the Agreement-Test, the sets B_0 and B_1 are uncorrelated. As the distribution of (E', B') depends on (D, E), for notational convenience, we denote this marginal distribution by $\mathcal{D}(D, E)$.

We now relate the two tests Agreement-Test and Test′ in order to show the consistency between the functions $g_{A,B}$s. Towards this, we define the notion of consistency, goodness, and excellence tailored to Test'.

Definition 5.13. *(consistency) Fix a set (D, E, B) where $A = D \cup E$. A subset (E', B') in the support of $\mathcal{D}(D, E)$ is said to be consistent with (D, E, B) if $F[D, E, B]|_D = F[D, E', B']|_D$. Let $newCons(D, E, B)$ be the set of all the sets (E', B') that are consistent with (D, E, B).*

Definition 5.14. *(goodness) A set (D, E, B) is called good if*

$$\Pr_{(E',B')\sim\mathcal{D}(D,E)} [(E', B') \in newCons(D, E, B)] \geqslant \varepsilon^2/2.$$

Definition 5.15. *(excellence) A set (D, E, B) is called $(\tilde{r}, \tilde{\gamma})$-excellent if it is good and*

$$\Pr_{\substack{(E_1,B_1),(E_2,B_2)\sim\mathcal{D}(D,E)| \\ E:=E_1\cap E_2, |E|=\tilde{r}}} \left[\begin{smallmatrix} (E_i,B_i)\in Cons(D,E,B) \text{ for i=1,2 } \& \\ F[D\cup E_1\cup B_1]|_E \neq F[D\cup E_2\cup B_2]|_E \end{smallmatrix}\right] \leqslant \tilde{\gamma}.$$

The following claim shows that if the Agreement-Test passes with probability at least ε, then the test Test'passes with probability ε^2.

Claim 5.16. *If $\Pr_{A_0,B_0,B_1}[F[A_0, B_0]|_{A_0} = F[A_0, B_1]|_{A_0}] \geqslant \varepsilon$, then there is a constant $c \in \left[\frac{q'}{2q}, \frac{2q'}{q}\right]$ such that the test Test'passes with probability at least ε^2. Furthermore, a random triple (D, E, B), with $|D| = cq'n$, is good with probability at least $\varepsilon^2/2$.*

Proof. Consider the following distribution.

- Select $S \subseteq [n]$ of size qn u.a.r.
- Select $A, A' \subseteq S$ each of size $q'n$ u.a.r.
- Select $B \subseteq [n] \setminus A$ of size $(q - q')n$ u.a.r.
- Select $B' \subseteq [n] \setminus A'$ of size $(q - q')n$ u.a.r.

We observe the following properties of the above distribution.

1. The pairs $(A, S \setminus A)$ and (A, B) are distributed according to the test distribution Agreement-Test. The same holds for the pairs $(A', S \setminus A')$ and (A', B')
2. For a fixed S, the pairs (A, B) and (A', B') are independent.

Consider the following expectation.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{S,(A,B),(A',B')}\left[\mathbf{1}_{F[S]|_A=F[A,B]|_A} \,\&\, \mathbf{1}_{F[S]|_{A'}=F[A',B']|_{A'}}\right]\\
&= \mathop{\mathbb{E}}_{S}\left[\mathop{\mathbb{E}}_{(A,B),(A',B')}\left[\mathbf{1}_{F[S]|_A=F[A,B]|_A} \,\&\, \mathbf{1}_{F[S]|_{A'}=F[A',B']|_{A'}}\right]\right]\\
&= \mathop{\mathbb{E}}_{S}\left[\mathop{\mathbb{E}}_{(A,B)}\left[\mathbf{1}_{F[S]|_A=F[A,B]|_A}\right]^2\right] && \text{(Property 2. above)}\\
&\geqslant \left(\mathop{\mathbb{E}}_{S}\left[\mathop{\mathbb{E}}_{(A,B)}\left[\mathbf{1}_{F[S]|_A=F[A,B]|_A}\right]\right]\right)^2 && \text{(Jensen's inequality)}\\
&= \varepsilon^2
\end{aligned}$$

Note that the events $F[S]|_A = F[A,B]|_A$ and $F[S]|_{A'} = F[A',B']|_{A'}$ together imply that $F[A,B]|_{A\cap A'} = F[A',B']|_{A\cap A'}$. Therefore, we have

$$\Pr_{(A,B),(A',B')}[F[A,B]|_{A\cap A'} = F[A',B']|_{A\cap A'}] \geqslant \varepsilon^2.$$

Based on how the sets A and A' are distributed, we have with $1-\exp(-n)$ probability, the size of $A\cap A'$ lies in $\left[\frac{q'^2 n}{2q}, \frac{2q'^2 n}{q}\right]$. By an averaging argument, there exists a constant $c \in \left[\frac{q'}{2q}, \frac{2q'}{q}\right]$ such that

$$\Pr_{(A,B),(A',B')||A\cap A'|=cq'n}[F[A,B]|_{A\cap A'} = F[A',B']|_{A\cap A'}] \geqslant \varepsilon^2.$$

Now, if we let $D = A\cap A'$, $E = A\setminus D$ and $E' = A'\setminus D$, then the pairs (D,E,B) and (D,E',B') are distributed according to the distribution in Test'. Hence, the acceptance probability of Test'is at least ε^2. The claim now follows from the averaging argument similar to the one in the proof of Claim 5.7. □

Claim 5.17. *A random good set is $(\tilde{r},\tilde{\gamma})$-excellent with probability at least* $1 - \frac{\tilde{r}^2}{cq'n\tilde{\gamma}}$.

Proof. Although the distribution in Test'is slightly different from the one in Agreement-Test, the proof of this claim is along the same lines (with $\beta = 0$) as the proof of Claim 5.8. □

Similar to the previous analysis, for an$(\tilde{r},\tilde{\gamma})$-excellent triple (D_0,E_0,B_0), we define a function $g_{D_0,E_0,B_0} : [n] \to \{0,1\}$ based on the majority vote of the table F restricted to the sets in $newCons(D_0,E_0,B_0)$. More formally, for $x \in [n]\setminus D_0$, we set

$$g_{D_0,E_0,B_0}(x) := \mathop{\mathsf{Majority}}_{\substack{(E,B)\in newCons(D_0,E_0,B_0)|\\ E\cup B\ni x}} F[D_0,E,B]|_x.$$

If there is no such $E\cup B$ that contains x then we set $g_{D_0,E_0,B_0}(x) := 0$. We also set $g_{D_0,E_0,B_0}(x)(D_0) = F[D_0,E_0,B_0]|_{D_0}$.

We have the following claim analogous to Claim 5.9. The proof of this claim is analogous to the proof of Claim 5.9 and hence we omit the proof.

Claim 5.18. *For every $q, q' \in (0,1)$ and $\tilde{r} \geqslant 1$, if (D_0, E_0, B_0) is $(\tilde{r}, \tilde{\gamma})$-excellent then*

$$\Pr_{(E,B)\in newCons(D_0,E_0,B_0)} [F[D_0, E, B]|_{E,B} \overset{>\tilde{\alpha}}{\neq} g_{D_0,E_0,B_0}(E, B)] \leqslant \tilde{\nu},$$

provided that $\tilde{\nu} = \Omega\left(\frac{\tilde{\gamma}(q-cq')n}{\tilde{\alpha}\varepsilon^2}\right)$ and $\tilde{\nu}\varepsilon \gg O\left(\frac{\tilde{r}^2 \ln 1/\varepsilon}{((q-cq')n)}\right)$.

Setting of the parameters. At this point, we concretely set the parameters so as to have $\nu, \tilde{\nu} \leqslant \varepsilon^3$ and a random good set is excellent with probability at least $1-O(\varepsilon^3)$ in Claims 5.8 and 5.17. This can be achieved using the following setting of the parameters.

$$\tilde{r} = \frac{1}{\varepsilon}, \qquad \tilde{\gamma} = \frac{1}{\varepsilon^5}\frac{1}{cq'n}, \qquad \tilde{\alpha} = \frac{1}{\varepsilon^{10}}$$

With these settings, we can have $\tilde{\nu} \leqslant \varepsilon^3$ in Claim 5.18. Furthermore, we set

$$\beta = 2\tilde{\alpha}, \qquad r = \frac{1}{\varepsilon}, \qquad \gamma = \frac{16r(r+2\beta)}{\varepsilon^2 q'n} \approx \frac{1}{\varepsilon^{13}}\frac{1}{q'n}, \qquad \alpha = \frac{1}{\varepsilon^{18}}.$$

With these settings, we can have $\nu \leqslant \varepsilon^3$ in Claim 5.9.

We will fix the parameters as stated above for the rest of the section. With these setting of parameters, we now relate the functions g_{D_0,E_0,B_0} and g_{A_0,B_0} towards showing that there is a global function g that agrees with the table F on $\delta(\varepsilon)$ fraction of the sets. The following two definitions will come handy for the rest of the argument.

Definition 5.19. *For an excellent triple (D_0, E_0, B_0), let $newCons^\star(D_0, E_0, B_0) \subseteq newCons(D_0, E_0, B_0)$ be the sets (E, B) such that $F[D_0, E, B]|_{E,B} \overset{\leqslant\tilde{\alpha}}{\neq} g_{D_0,E_0,B_0}(E, B)]$. Similarly, for an excellent pair (A_0, B_0), $Cons^\star(A_0, B_0) \subseteq Cons(A_0, B_0)$ be the sets B such that $F[A_0, B]|_B \overset{\leqslant\alpha}{\neq} g_{A_0,B_0}(B)$.*

Note that if we have $\tilde{\nu} \leqslant \varepsilon^2/4$, then using Claim 5.18, we have

$$\Pr_{(E',B')\sim\mathcal{D}(D_0,E_0)} [(E', B') \in newCons^\star(D_0, E_0, B_0)] \geqslant \varepsilon^2/4. \tag{14}$$

Definition 5.20. *Fix (D_0, E_0, B_0) that is good. A pair $(E, B) \in newCons^\star(D_0, E_0, B_0)$ is called a* dense pair*, if $\Pr_{B'\subseteq[n]\setminus(D_0\cup E)} [(E, B') \in newCons^\star(D_0, E_0, B_0)] \geqslant \varepsilon^2/8$.*

The next claim shows that many sets in $newCons^\star(D_0, E_0, B_0)$ combined with D_0 give excellent sets for Agreement-Test.

Claim 5.21. *Let (D_0, E_0, B_0) be an $(\tilde{r}, \tilde{\gamma})$-excellent triple and suppose $\tilde{\nu} < \frac{\varepsilon^2}{4}$ and $\gamma \geqslant \frac{16r(r+4\tilde{\alpha})}{\varepsilon^2 q'n}$, then,*

$$\Pr_{(E,B)\in newCons^\star(D_0,E_0,B_0)} [(E, B)\ is\ dense \& (D_0 \cup E, B)\ is\ (\varepsilon^2/8, 2\tilde{\alpha}, r, \gamma) - excellent] \geqslant \varepsilon^2/16.$$

Proof. Since (D_0, E_0, B_0) is good, we have that at least $\varepsilon^2/8$ fraction of $(E, B) \in newCons^\star(D_0, E_0, B_0)$ are dense pairs. We will show that for such pair (E, B), $(D_0 \cup E, B)$ is $(\varepsilon^2/8, 2\tilde{\alpha})$-good. Fix any such (E, B). By the definition of a dense pair, we have

$$\Pr_{B'\subseteq[n]\setminus D_0\cup E}\left[\begin{array}{c} g_{D_0,E_0,B_0}(D_0\cup E)\overset{\leqslant\tilde{\alpha}}{\neq} F[D_0,E,B]|_{D_0\cup E}\ \& \\ g_{D_0,E_0,B_0}(D_0\cup E)\overset{\leqslant\tilde{\alpha}}{\neq} F[D_0,E,B']_{D_0\cup E}\end{array}\right] \geqslant \varepsilon^2/8.$$

Note that both the events in the above probability follow from the fact that $(E,B)\in newCons^\star(D_0,E_0,B_0)$ and for a random B', $(E,B')\in newCons^\star(D_0,E_0,B_0)$ with probability at least $\varepsilon^2/8$. Since g_{D_0,E_0,B_0} is a function defined on $[n]$, this implies that

$$\Pr_{B'\subseteq[n]\setminus D_0\cup E}\left[F[D_0,E,B]|_{D_0\cup E} \overset{\leqslant 2\tilde{\alpha}}{\neq} F[D_0,E,B']_{D_0\cup E}\right] \geqslant \varepsilon^2/8,$$

and hence $(D_0\cup E,B)$ is $(\varepsilon^2/8,2\tilde{\alpha})$-good. Since, using Claim 5.8, at least $1-\varepsilon^2/16$ fraction of $(\varepsilon^2/8,2\alpha)$-good pairs are $(\varepsilon^2/8,2\tilde{\alpha},r,\gamma)$-excellent, we have that at least $\varepsilon^2/16$ pairs $(E,B)\in newCons^\star(D_0,E_0,B_0)$, $(D_0\cup E,B)$ is excellent. □

The following claim shows that the sets in $Cons^\star$ and $newCons^\star$ are correlated.

Claim 5.22. *Fix any $(\tilde{r},\tilde{\gamma})$-excellent triple (D_0,E_0,B_0). For at least $\Omega(\varepsilon^4)$ fraction of the pairs $(E,B)\subseteq [n]\setminus D_0$ where $E\cap E_0=\emptyset$ and $B\subseteq[n]\setminus D_0\cup E$, we have*

1. *$(D_0\cup E,B)$ is $(\varepsilon^2/8,2\tilde{\alpha},r,\gamma)$-excellent, and*
2. *there is at least $\Omega(\varepsilon^2)$ fraction of $B'\subseteq[n]\setminus(D_0\cup E)$ such that $(E,B')\in newCons^\star(D_0,E_0,B_0)$ and $B'\in Cons^\star_{2\tilde{\alpha}}(D_0\cup E,B)$.*

Proof. Using Claim 5.21 and (14), for a random $(E,B)\sim(D_0,E_0)$, with probability at least $\Omega(\varepsilon^4)$, $(E,B)\in newCons^\star(D_0,E_0,B_0)$, $(D_0\cup E,B)$ is $(\varepsilon^2/8,2\tilde{\alpha},r,\gamma)$-excellent as well as

$$\Pr_{B'\subseteq[n]\setminus(D_0\cup E)}[(E,B')\in newCons^\star(D_0,E_0,B_0)]=\Omega(\varepsilon^2).$$

This implies $g_{D_0,E_0,B_0}(D_0\cup E\cup B)\overset{\leqslant\tilde{\alpha}}{\neq} F[D_0\cup E\cup B]$. We also have $g_{D_0,E_0,B_0}(D_0\cup E\cup B')\overset{\leqslant\tilde{\alpha}}{\neq} F[D_0\cup E\cup B']$ for at least $\Omega(\varepsilon^2)$ fraction of $B'\subseteq[n]\setminus(D_0\cup E)$. Therefore, at least $\Omega(\varepsilon^4)$ fraction of (E,B) are such that for at least $\Omega(\varepsilon^2)$ fraction of B', $(E,B')\in newCons^\star(D_0,E_0,B_0)$ as well as $F[D_0\cup E\cup B]|_{D_0\cup E}\overset{\leqslant 2\tilde{\alpha}}{\neq} F[D_0\cup E\cup B']|_{D_0\cup E}$. The latter condition implies that $B'\in Cons_{2\tilde{\alpha}}(D_0\cup E,B)$ and since $\nu\leqslant\varepsilon^3$, the claim follows. □

Using the above claim, we show that for every (D_0,E_0,B_0) that is excellent, the functions g_{D_0,E_0,B_0} and $g_{D_0\cup E,B}$ are very close to each other in hamming distance for many pairs (E,B).

Claim 5.23. *Fix any $(\tilde{r},\tilde{\gamma})$-excellent triple (D_0,E_0,B_0). Then*

$$\Pr_{(E,B)\sim\mathcal{D}(D_0,E_0)}[g_{D_0,E_0,B_0}\overset{\leqslant O(\alpha)}{\neq} g_{D_0\cup E,B}]\geqslant\Omega(\varepsilon^4).$$

Proof. Select a pair (E,B) according to the distribution $\mathcal{D}(D_0,E_0)$. Using Claim 5.22, with probability at least $\Omega(\varepsilon^4)$, we have $g_{D_0,E_0,B_0}(D_0\cup E\cup B)\overset{\leqslant\tilde{\alpha}}{\neq} F[D_0\cup E\cup B]$, $(D_0\cup E,B)$ is $(\varepsilon^2/8,2\tilde{\alpha},r,\gamma)$-excellent and $g_{D_0\cup E,B}(D_0\cup E\cup B)\overset{\leqslant 2\tilde{\alpha}+\alpha}{\neq} F[D_0\cup E\cup B]$.[6] Note that in the last condition, we used the fact that if

[6]The $2\tilde{\alpha}$ is for possible disagreements on $D_0\cup E$ and the α is for possible disagreements on B.

$(D_0 \cup E, B)$ is excellent and if $B' \in Cons^\star_{2\tilde{\alpha}}(D_0 \cup E, B)$, then $(D \cup E, B')$ is also excellent and furthermore the functions $g_{D_0 \cup E, B}$ and $g_{D_0 \cup E, B'}$ are the same as they are defined using the set $Cons^\star(D_0 \cup E, B) = Cons^\star(D_0 \cup E, B')$

Combining these events and the above claim, we get that over the randomness of (E, B), with probability at least $\Omega(\varepsilon^4)$, the following happens:

1. $(D_0 \cup E, B)$ is $(\varepsilon^2/8, 2\tilde{\alpha}, r, \gamma)$-excellent,
2. $g_{D_0,E_0,B_0}(D_0 \cup E \cup B) \overset{\leqslant 4\alpha}{\neq} g_{D_0 \cup E, B}(D_0 \cup E \cup B)$, and
3. $\Pr_{B' \subseteq [n] \setminus (D_0 \cup E)}\left[(E, B') \in newCons^\star(D_0, E_0, B_0) \;\&\; B' \in Cons^\star_{2\alpha}(D_0 \cup E, B)\right] \geqslant \Omega(\varepsilon^2)$.

Form the third point we can conclude that $\Pr_{B' \subseteq [n] \setminus (D_0 \cup E)}\left[g_{D_0,E_0,B_0}(B') \overset{\leqslant 2\alpha}{\neq} g_{D_0 \cup E, B}(B')\right] \geqslant \Omega(\varepsilon^2)$. Assuming $\alpha = \mathsf{poly}(1/\varepsilon)$, we have from the following claim, $g_{D_0,E_0,B_0}|_{[n] \setminus (D_0 \cup E)} \overset{\leqslant O(\alpha)}{\neq} g_{D_0 \cup E, B}|_{[n] \setminus (D_0 \cup E)}$. The claim follows from this and the point 2 above. □

Claim 5.24. *For any $q \in (0, 1)$, $\beta \geqslant 1$ and $\delta > e^{-q\beta/8}$, given two functions $f, g : [n] \to \{0, 1\}$ such that $\Pr_{S \subseteq [n] \mid |S| = qn}[f(S) \overset{\leqslant \beta}{\neq} g(S)] \geqslant \delta$, then $|\{i \in [n] \mid f(i) \neq g(i)\}| \leqslant 4\beta/q$.*

Proof. Let $\mathcal{E}$ be the disagreement set, i.e., $\mathcal{E} = \{i \in [n] \mid f(i) \neq g(i)\}$. Suppose towards contradiction $|\mathcal{E}| \geqslant 4\beta/q$. Now suppose we sample a subset S by including each $i \in [n]$ in S independently with probability $q/2$. By the Chernoff bound, $|S| \leqslant qn$ with probability at least $1 - \exp(-\Omega(q^2 n))$. Furthermore, the expected size of $S \cap \mathcal{E}$ is 2β. Now, again by the Chernoff bound, the probability that $|S \cap \mathcal{E}|$ is at most β is at most $e^{-q\beta/4}$. This implies $\Pr_{S \subseteq [n] \mid |S| = qn}[f(S) \overset{\leqslant \beta}{\neq} g(S)] \leqslant e^{-q\beta/8}$ which is a contradiction. □

The following claim follows easily from Claim 5.23.

Claim 5.25. *There exists a constant $d_0 \in (0, 1)$ such that for $c' = d_0 c$ we have the following. For at least $\Omega(\varepsilon^2)$ fraction of excellent pairs (A_0, B_0),*

$$\Pr_{(\tilde{A}_0, B)}\left[g_{A_0, B_0} \overset{\leqslant O(\alpha)}{\neq} g_{\tilde{A}_0, B}\right] \geqslant \Omega(\varepsilon^{12}).$$

Here, $\tilde{A}_0$ is distributed uniformly conditioned on $|\tilde{A}_0| = q'n$, $|A_0 \cap \tilde{A}_0| = c'q'n$ and $B \subseteq [n] \setminus \tilde{A}_0$ of size $(q - q')n$.

Proof. Since $\Omega(\varepsilon^2)$ fraction of the triples (D_0, E_0, B_0) are $(\varepsilon^2/8, 2\tilde{\alpha}, r, \gamma)$-excellent, we have the following from Claim 5.23.

$$\Omega(\varepsilon^6) \leqslant \mathop{\mathbb{E}}_{(D_0, E_0, B_0)} \mathop{\mathbb{E}}_{(E, B) \sim \mathcal{D}(D_0, E_0)}\left[\mathbf{1}_{g_{D_0,E_0,B_0} \overset{\leqslant O(\alpha)}{\neq} g_{D_0 \cup E, B}}\right].$$

Note that $(D_0 \cup E, B)$ being $(\varepsilon^2/8, 2\tilde{\alpha}, r, \gamma)$-excellent is implicit in the event. Squaring and applying Cauchy-Schwarz we get that

$$\Omega(\varepsilon^{12}) \leqslant \mathop{\mathbb{E}}_{(D_0,E_0,B_0)} \mathop{\mathbb{E}}_{\substack{(E,B)\sim\mathcal{D}(D_0,E_0)\\(E',B')\sim\mathcal{D}(D_0,E_0)}} \left[\mathbf{1}_{g_{D_0,E_0,B_0} \overset{\leqslant O(\alpha)}{\neq} g_{D_0\cup E,B}} \mathbf{1}_{g_{D_0,E_0,B_0} \overset{\leqslant O(\alpha)}{\neq} g_{D_0\cup E',B'}} \right]$$
$$\leqslant \mathop{\mathbb{E}}_{(D_0,E_0,B_0)} \mathop{\mathbb{E}}_{\substack{(E,B)\sim\mathcal{D}(D_0,E_0)\\(E',B')\sim\mathcal{D}(D_0,E_0)}} \left[\mathbf{1}_{g_{D_0\cup E,B} \overset{\leqslant O(\alpha)}{\neq} g_{D_0\cup E',B'}} \right]$$

Now letting $A_0 = D_0 \cup E'$, $B_0 = B'$, $\tilde{A}_0 = D_0 \cup E$ and $B = B$, we see that with probability at least $1 - \exp(-\Omega((cq')^2 n))$, we have $|E \cap E'| = O(q'^2 n)$. Therefore, there exists a constant d_0 such that when we condition the above distribution on $|A_0 \cap \tilde{A}_0| = d_0 c q' n$, the expectation at still least $\Omega(\varepsilon^{12})$. The claim now follows from an averaging argument. ☐

We are now ready to prove the final global structure and the proof of Theorem 1.4 follows from the following claim.

Claim 5.26. *For all $\varepsilon > 0$, there exists $\delta > 0$ such that if*

$$\Pr_{A_0,B_0,B_1}\left[F[A_0,B_0]|_{A_0} = F[A_0,B_1]|_{A_0}\right] \geqslant \varepsilon,$$

then there is an excellent pair $(A_0^\star, B_0^\star)$ such that

$$\Pr_{T\in\binom{[n]}{qn}}\left[g_{A_0^\star,B_0^\star}(T) \overset{\leqslant \alpha^{O(1)}}{\neq} F[T]\right] \geqslant \delta.$$

Proof. We first note that if two functions $f, g : [n] \to \{0,1\}$ differ at $O(\alpha)$ locations, then if we take a random subset $S \subseteq [n]$ of size n/α, then the probability that $f(S) = g(S)$ is at least $\Omega(1)$. Using this fact and Claim 5.25, we have the following

$$\mathop{\mathbb{E}}_{S\subseteq[n]||S|=\frac{n}{\alpha^2}}\left[\mathop{\mathbb{E}}_{(A_0,B_0),(\tilde{A}_0,B)}\left[\mathbf{1}_{g_{A_0,B_0}(S)=g_{\tilde{A}_0,B}(S)}\right]\right] \geqslant \Omega(\varepsilon^{\Theta(1)}),$$

where $(\tilde{A}_0, B)$ is distributed as in Claim 5.25. Note that the event $\mathbf{1}_{g_{A_0,B_0}(S)=g_{\tilde{A}_0,B}(S)}$ implicitly implies that the pairs $(A_0, B_0), (\tilde{A}_0, B)$ are excellent. By an averaging argument, at least $\Omega(\varepsilon^{\Theta(1)})$ fraction of sets S are such that

$$\mathop{\mathbb{E}}_{(A_0,B_0),(\tilde{A}_0,B)}\left[\mathbf{1}_{g_{A_0,B_0}(S)=g_{\tilde{A}_0,B}(S)}\right] \geqslant \Omega(\varepsilon^{\Theta(1)}).$$

Given such a set S, we define a partition of $\binom{[n]}{qn}$ based on the value $g_{A,B}(S)$. In other words, we have parts identified by strings in $\{0,1\}^{n/\alpha^2}$; (A, B) and (A', B') belong to the part $\mathcal{S}_\beta$ if $g_{A,B}(S) = g_{A',B'}(S) = \beta$. We claim that there exists a β such that $|\mathcal{S}_\beta| = c_0(\varepsilon) \cdot \binom{n}{qn}$, for some $c_0(\varepsilon) > 0$.

Consider the graph $\mathcal{G}_n$ induced on the set of vertices $\{(A, B) \mid |A| = q'n, |B| = (q - q')n\}$ as follows: A random neighbor (A', B') of (A, B) in this graph is sampled conditioned on the fact that A' is distributed uniformly conditioned on $|A'| = q'n$, $|A \cap A'| = c'q'n$ and $B' \subseteq [n] \setminus A'$ is a uniformly random set of size

$(q-q')n$. Using the small set expansion property of $\mathcal{G}_n$ from Lemma 5.32, if all the parts were of size at most $\delta \cdot \binom{n}{qn}$, then

$$\Pr_{(A_0,B_0),(\tilde{A}_0,B)}[(A_0,B_0),(\tilde{A}_0,B) \in \mathcal{S}_\beta \textit{ for some } \beta] \leqslant c''(\delta),$$

where $c''(\delta) \to 0$ as $\delta \to 0$. Therefore, Taking expectation over S yields that

$$\mathop{\mathbb{E}}_{S\subseteq[n]||S|=\frac{n}{\alpha^2}}\left[\mathop{\mathbb{E}}_{(A_0,B_0),(A_0',B_0')}\left[\mathbf{1}_{g_{A_0,B_0}(S)=g_{A_0',B_0'}(S)}\right]\right] \geqslant \varepsilon^{\Theta(1)}c_0(\varepsilon)^2.$$

This implies,

$$\mathop{\mathbb{E}}_{(A_0,B_0),(A_0',B_0')}\left[\mathop{\mathbb{E}}_{S\subseteq[n]||S|=\frac{n}{\alpha^2}}\left[\mathbf{1}_{g_{A_0,B_0}(S)=g_{A_0',B_0'}(S)}\right]\right] \geqslant \varepsilon^{\Theta(1)}c_0(\varepsilon)^2.$$

and hence,

$$\Pr_{(A_0,B_0),(A_0',B_0')}\left[g_{A_0,B_0} \overset{\leqslant\alpha^{O(1)}}{\neq} g_{A_0',B_0'}\right] \geqslant \varepsilon^{\Theta(1)}c_0(\varepsilon)^2.$$

This means that there exists an excellent pair $(A_0^\star, B_0^\star)$ such that

$$\Pr_{(A_0',B_0')}\left[g_{A_0^\star,B_0^\star} \overset{\leqslant\alpha^{O(1)}}{\neq} g_{A_0',B_0'}\right] \geqslant \varepsilon^{\Theta(1)}c_0(\varepsilon)^2.$$

Now consider selecting a random set T of size qn and select a random subset A of T of size $q'n$ and let $B = T \setminus A$. Select a random set $B' \subseteq [n] \setminus A$. Using the above inequality, with probability at least $\varepsilon^{\Theta(1)}c_0(\varepsilon)^2$, $g_{A_0^\star,B_0^\star} \overset{\leqslant\alpha^{O(1)}}{\neq} g_{A,B'}$ and with probability at least $\Omega(\varepsilon^2)$, $B \in Cons^\star(A,B')$ and hence $g_{A,B'}(T) \overset{\leqslant\alpha}{\neq} F[T]$. Combining all these events, we have

$$Pr_{T\in\binom{[n]}{qn}}\left[g_{A_0^\star,B_0^\star}(T) \overset{\leqslant\alpha^{O(1)}}{\neq} F[T]\right] \geqslant \Omega(\varepsilon^{\Theta(1)}c_0(\varepsilon)^2 \cdot \varepsilon^2),$$

and the claim follows.

□

5.3 Small-set expansion property of the graphs over a multi-slice

In this section, we show the small set expansion property of the graph that was needed in the proof of Claim 5.26. For a graph $G(V,E)$, let $\mathcal{T}(G)$ be the markov operator associated with G. Also, let $\phi_G(\mu) := \min_{S\subseteq V(G),|S|\leqslant\mu|V(G)|} \Pr_{(u,v)\in E(G)}[v \notin S \mid u \in S]$. Note that if every subset of size at most μ in G expands, then $\phi_G(\mu)$ is large. For any linear operator T, its $p \to q$ norm is defined as $\|T\|_{p\to q} := \max_{v\neq 0}\frac{\|Tv\|_q}{\|v\|_p}$.

The graph $\mathcal{G}_n$. Consider the graph $\mathcal{G}_n$ induced on the set of vertices $\{(A,B) \mid |A| = q'n, |B| = (q-q')n\}$ as follows: A random neighbor (A', B') of (A,B) in this graph is sampled conditioned on the fact that A' is distributed uniformly conditioned on $|A'| = q'n$, $|A \cap A'| = cq'n$ and $B' \subseteq [n] \setminus A'$ is a uniformly random set of size $(q-q')n$. In this section, we deduce the small-set expansion property of the graph $\mathcal{G}_n$. We can view the vertices of the above graph $\mathcal{G}_n$ as the multi-slice of $\{0,1,2\}^n$ – map the vertex (A,B) to $\mathbf{x} \in \{0,1,2\}^n$ where $x_i = 1$ if $i \in A$, $x_i = 2$ if $i \in B$ and $x_i = 0$ if $i \in [n] \setminus (A \cup B)$. Let us denote the multi-slice by $\mathcal{U}_n$.

One can view the multi-slice $\mathcal{U}_n$ as a quotient space $S_n/(S_{(1-q)n} \times S_{q'n} \times S_{(q-q')n})$, which is useful in lifting the standard representation-theoretic decomposition of functions over S_n to decompositions of functions over $\mathcal{U}_n$. In order to state the relevant lemmas from [BKLM22], we need the following few definitions.

Definition 5.27. *A function $f : S_n \to \mathbb{R}$ is called a d-junta if there exists a set of coordinate $A \subseteq [n]$ of size at most d such that $f(\pi) = g(\pi(A))$ for some function $g : [n]^A \to \mathbb{R}$.*

For two function $f, g : S_n \to \mathbb{R}$, define the inner product $\langle f, g\rangle$ as $\mathbb{E}_\pi [f(\pi)g(\pi)]$.

Definition 5.28. *For $d = 0, 1, \ldots, n$ we denote by $V_d(S_n) \subseteq \{f : S_n \to R\}$ the span of d-juntas. Also, define $V_{=d}(S_n) = V_d(S_n) \cap V_{d-1}(S_n)^\perp$.*

Therefore, we can write the space of real-valued functions as $V_{=0}(S_n) \oplus V_{=1}(S_n) \oplus \ldots \oplus V_{=n-1}(S_n)$, and thus write any $f : S_n \to \mathbb{R}$ uniquely as $f = \sum_{i=0}^{n-1} f^{=i}$ where $f^{=i} \in V_{=i}(S_n)$. Let $V_{\geqslant d}(\mathcal{U}_n)$ ($V_{\leqslant d}(\mathcal{U}_n)$) be the span of functions over $\mathcal{U}_n$ whose degree is at least (at most) d.

We say a distribution μ over $([3] \times [3])^n$ commutes with the action of S_n if the following distributions are the same for all $\pi \in S_n$ and $x \in [3]^n$: a) $\mathbf{x}'$, where $(\mathbf{x}, \mathbf{x}') \sim \mu$ conditioned on $\mathbf{x} = \pi(x)$, and b) $\pi(\mathbf{x}')$, where $(\mathbf{x}, \mathbf{x}') \sim \mu$ conditioned on $\mathbf{x} = x$. The following claim shows that the operator $\mathcal{T}$ that commutes with S_n preserves the degree of the functions.

Claim 5.29. *(Claim 3.6 from [BKLM22]) Suppose $\mathcal{T}$ is an operator that commutes with the action of S_n on functions over the multi-slice $\mathcal{U}_n$. Then for each $0 \leqslant d < n$, we have that $\mathcal{T}(V_{=d}(\mathcal{U}_n)) \subseteq V_{=d}(\mathcal{U}_n)$.*

We observe the following few properties of the multi-slice $\mathcal{U}_n$ and the operator $\mathcal{T}(\mathcal{G}_n)$.

1. In a multi-slice, every symbol appears the same number of times in every element of the multi-slice. If we let k_i be the number of times the symbol i appears, then the multi-slice is called *α-balanced* if $k_i \geqslant \alpha n$ for every i. The multi-slice $\mathcal{U}_n$ is α-balanced for $\alpha = \min\{q', (q-q'), (1-q)\}$.

2. The edge distribution of the graph $\mathcal{G}_n$ is a distribution on the multi-slices $\mathcal{U}_n \times \mathcal{U}_n$. A distribution μ on $\mathcal{U}_n \times \mathcal{U}_n$ is called α-admissible if a) the distribution is symmetric under S_n, and b) for all $(a,b) \in \{0,1,2\} \times \{0,1,2\}$, the quantity $\Pr_{(\mathbf{x},\mathbf{y})\sim\mu, i\in[n]} [x_i = a \;\&\; y_i = b]$ is either at least α or 0. It can be easily observed that the edge distribution of $\mathcal{G}_n$ is α-admissible for $\alpha = \Omega_{c,q,q'}(1)$ independent of n.

3. A distribution μ on $\mathcal{U}_n \times \mathcal{U}_n$ also called *connected* iff the bipartite graph $(V_1 \cup V_2, E)$ where a) V_i is the corresponding support of the marginal distribution of μ, and b) $(\mathbf{x}, \mathbf{y}) \in E$ iff $(\mathbf{x}, \mathbf{y})$ is in the support of μ, is connected. Here again, it is easy to observe the connectedness property of the edge distribution of the graph $\mathcal{G}_n$.

4. Finally, the operator $\mathcal{T}(\mathcal{G}_n)$ commutes with the action of S_n.

One of the important characteristic of $\Omega(1)$-admissible and connected distributions, as shown in [BKLM22], is that it can be replaced by a certain product distribution $\nu^{\otimes n}$ as far as low-degree functions are concerned. Thus, this gives a way to prove analytical results for a multi-slice by invoking the corresponding results over a product distribution.

The following lemma from [FKLM20] gives an upper bound on $\|\mathcal{P}_{\leqslant d}\|_{2\to 4}$, where $\mathcal{P}_{\leqslant d}$ is the projector operator into the space $V_{\leqslant d}(\mathcal{U}_n)$. It crucially uses the fact that $\mathcal{U}_n$ is α-balanced for some constant α independent of n.

Lemma 5.30 (Theorem 2.12 from [BKLM22]). *For all $c > 0$ and $0 < q' < q < 1$ and $d \in \mathbb{N}$ there is $N = N(c, q, q', d) > 0$ and $C = C(c, q, q', d) > 0$ such that the following holds. Let $n > N$ and let $f : \mathcal{U}_n \to \mathbb{R}$ be a function of degree at most d, then $\|f\|_4 \leqslant C\|f\|_2$.*

Finally, we need the following lemma that bounds the eigenvalues of the operator $\mathcal{T}(\mathcal{G}_n)$. This lemma uses the fact the the edge distribution is $\Omega(1)$-admissible and connected.

Lemma 5.31 (Lemma 3.11 from [BKLM22]). *For all $c > 0$ and $0 < q' < q < 1$, there exist constants $C > 0$ and $\delta > 0$ such that for all $d \in \mathbb{N}$, if $f \in V_{>d}(\mathcal{U}_n)$, we have $\|\mathcal{T}(\mathcal{G}_n)f\|_2 \leqslant C(1+\delta)^{-d}\|f\|_2$.*

We are now ready to prove the small-set expansion property of the graph $\mathcal{G}_n$.

Lemma 5.32. *For every $c > 0$ and $0 < q' < q < 1$, there exists $N = N(c, q', q)$ such that for all $n > N$ and $\mu > 0$, the graph $\mathcal{G}_n$ defined above has*

$$\phi_{\mathcal{G}_n}(\mu) \geqslant 1 - \mu',$$

where $\mu' \to 0$ as $\mu \to 0$.

Proof. Fix any set $S \subseteq V(\mathcal{G}_n)$ of density at most μ. Let $f : \mathcal{U}_n \to \{0,1\}$ be the indicator function of S. As stated at the begining of the section, we can write f as $f = \sum_{i=0}^{n-1} f^{=i}$ where $f^{=i} \in V_{=i}(\mathcal{U}_n)$. Let $f = f_1 + f_2$, where the component $f_1 = \sum_{i=0}^{d} f^{=i}$ and $f_2 = \sum_{i=d+1}^{n-1} f^{=i}$ for some d to be fixed later. By letting $\mathcal{T} := \mathcal{T}(\mathcal{G}_n)$, we have

$$\phi_{\mathcal{G}_n}(S) = 1 - \frac{\langle f, \mathcal{T} f\rangle}{\mu}.$$

By letting $\tau := C(1+\delta)^{-d}$ from Lemma 5.31, we can bound

$$\begin{aligned}
\langle f, \mathcal{T} f\rangle &= \langle f_1, \mathcal{T} f_1\rangle + \langle f_2, \mathcal{T} f_2\rangle && \text{(Using Claim 5.29 and orthogonality of spaces } V_{=i}(\mathcal{U}_n)\text{)}\\
&\leqslant \|f_1\|_2^2 + \tau\|f_2\|_2^2 && \text{(Using Claim 5.31)}\\
&\leqslant \|f_1\|_2^2 + \tau\mu.
\end{aligned}$$

Therefore,

$$\phi_{\mathcal{G}_n}(S) \geqslant 1 - \tau - \frac{\|f_1\|_2^2}{\mu}. \tag{15}$$

If we let $\mathcal{P}_{\leqslant d}$ be the projector operator into the subspace $V_{\leqslant d}(\mathcal{U}_n)$, then we have

$$\|\mathcal{P}_{\leqslant d}\|_{4/3\to 2} = \max_{g\neq 0} \frac{\|\mathcal{P}_{\leqslant d} g\|_2}{\|g\|_{4/3}} \geqslant \frac{\|\mathcal{P}_{\leqslant d} f\|_2}{\|f\|_{4/3}} = \frac{\|f_1\|_2}{\|f\|_{4/3}}.$$

Since $\|f\|_{4/3} = \mu^{3/4}$, we have $\|f_1\|_2^2 \leqslant \|\mathcal{P}_{\leqslant d}\|_{4/3\to 2}^2 \cdot \mu^{3/2}$. Using the fact that $\|\mathcal{P}_{\leqslant d}\|_{4/3\to 2}^2 = \|\mathcal{P}_{\leqslant d}\|_{2\to 4}^2$, we get $\|f_1\|_2^2 \leqslant \|\mathcal{P}_{\leqslant d}\|_{2\to 4}^2 \cdot \mu^{3/2}$. Therefore,

$$\phi_{\mathcal{G}_n}(S) \geqslant 1 - \tau - \|\mathcal{P}_{\leqslant d}\|_{2\to 4}^2 \cdot \mu^{1/2}. \tag{16}$$

Finally, using Lemma 5.30, we can bound $\|\mathcal{P}_{\leqslant d}\|_{2\to 4}^2$ as follows.For any function g, let g_1 be the component of g from $V_{\leqslant d}$ and $g_2 := g - g_1$ be orthogonal to g_1.

$$\|\mathcal{P}_{\leqslant d}\|_{2\to 4}^2 = \max_{g:=g_1+g_2\neq 0} \frac{\|\mathcal{P}_{\leqslant d} g\|_4^2}{\|g\|_2^2} = \max_{g:=g_1+g_2} \frac{\|g_1\|_4^2}{\|g_1\|_2^2 + \|g_2\|_2^2} \leqslant \max_{g_1\in V_{\leqslant d}} \frac{\|g_1\|_4^2}{\|g_1\|_2^2} \leqslant C(c, q, q', d),$$

where the last inequality uses Lemma 5.30. Plugging this into (16), we get

$$\phi_{\mathcal{G}_n}(\mu) \geqslant 1 - C(1+\delta)^{-d} - C(c, q, q', d) \cdot \mu^{1/2}.$$

If we choose d to be some growing function in $1/\mu$ such that $C(c, q, q', d) \leqslant \frac{1}{\mu^{1/4}}$, then it is easy to observe that $\phi(\mu) \to 1$ as $\mu \to 0$. $\square$

5.4 Direct product testing: from sets to a product distribution

In this section, we show that Corollary 1.5 follows from Theorem 1.4. We need the following two claims.

Claim 5.33. *Fix $q \in (0,1)$ and $N = \omega(n^2)$. We have,*

$$\frac{\binom{N-n}{qN-t}}{\binom{N}{qN}} = q^t(1-q)^{n-t}(1 \pm o(1)).$$

Proof. Expanding the left-hand side,

$$\begin{aligned}
\frac{\binom{N-n}{qN-t}}{\binom{N}{qN}} &= \frac{(N-n)!qN!(N-qN)!}{N!(qN-t)!(N-qN-n+t)!} \\
&= \frac{qN.(qN-1)\dots(qN-t+1)\cdot(N-qN).(N-qN-1)\dots(N-qN-n+t+1)}{N.(N-1)\dots(N-n+1)} \\
&= q^t(1-q)^{n-t}\frac{N(N-1/q)\dots(N-(t-1)/q)\cdot N(N-1/(1-q))\dots(N-(n-t-1)/(1-q))}{N(N-1)\dots(N-(n-1))} \\
&= q^t(1-q)^{n-t}\frac{(1-1/qN)\dots(1-(t-1)/qN)\cdot(1-1/(1-q)N)\dots(1-(n-t-1)/(1-q)N)}{(1-1/N)\dots(1-(n-1)/N)} \\
&= q^t(1-q)^{n-t}\cdot\frac{e^{-\Theta(t^2/qN)}\cdot e^{-\Theta((n-t)^2/(1-q)N)}}{e^{-\Theta(n^2/N)}} \\
&= q^t(1-q)^{n-t}(1\pm o(1)),
\end{aligned}$$

and the claim follows. $\square$

Claim 5.34. *Fix $q \in (0,1)$ and $N = \omega(n^2)$. Consider the following two distributions on $P([n])$:*

- *$\mathcal{D}_1$: Select a subset $A \subseteq [n]$ by including $i \in [n]$ to A with probability q for each i independently.*

- *$\mathcal{D}_2$: Select a random subset $S \subseteq [N]$ of size qN and output $S|_{[n]}$.*

Then, the statistical distance between $\mathcal{D}_1$ and $\mathcal{D}_2$ is at most $o_n(1)$.

Proof. We will compare the point-wise probabilities $p_1, p_2 : P([n]) \to \mathbb{R}$ assigned by the two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. Fix any set $A \subseteq [n]$ of size t. We have $p_1(A) = q^t(1-q)^{n-t}$. Now, in order to sample A from $\mathcal{D}_2$, it must be the case that $S|_{[n]} = A$ and therefore, we have

$$p_2(A) = \frac{\binom{N-n}{qN-t}}{\binom{N}{qN}},$$

which is $q^t(1-q)^{n-t}(1 \pm o(1))$ using Claim 5.33. $\square$

Given a function $G : (P[n], \mu^{\otimes n}) \to P([n])$ where $G(A)$ can be thought of as a string in $\{0,1\}^{|A|}$ by specifying a fixed arbitrary ordering on $[n]$, we define a map $\tilde{G} : \binom{[N]}{qN} \to \{0,1\}^{qN}$ as follows: For a set $S \in \binom{[N]}{qN}$, define $\tilde{G}(S)|_{S \cap [n]} = G(S \cap [n])$ and $\tilde{G}(S)|_{S \setminus [n]} = 0^{|S \setminus [n]|}$.

Corollary 5.35 (Restatement of Corollary 1.5)**.** *For all $\alpha, \varepsilon > 0$ and $0 < q < \frac{1}{2-\alpha}$, there are $r \in \mathbb{N}$ and $\delta > 0$ such that the following holds. Suppose that $G\colon (P[n], \mu_q^{\otimes n}) \to P([n])$ satisfies*

$$\Pr_{(A,A') \sim \mathcal{D}_{q,\alpha}} \left[G[A] \cap (A \cap A') = G[A'] \cap (A \cap A')\right] \geqslant \varepsilon.$$

Then there exists $S \subseteq [n]$ such that $\Pr_{A \sim_q [n]} [|G[A] \Delta (S \cap A)| \leqslant r] \geqslant \delta$.

Proof. We know that

$$\Pr_{(A,A') \sim \mathcal{D}_{q,\alpha}} [G[A] \cap (A \cap A') = G[A'] \cap (A \cap A')] \geqslant \varepsilon.$$

Instead of checking consistency on $A \cap A'$, we select a set $A'' \subseteq A \cap A'$ by independently including $i \in A \cap A'$ to A'' with probability $q'/\alpha q$. This can only increase the acceptance probability, and hence

$$\Pr_{\substack{(A,A') \sim \mathcal{D}_{q,\alpha} \\ A'' \sim_{q'/\alpha q} A \cap A'}} [G[A] \cap A'' = G[A'] \cap A''] \geqslant \varepsilon. \tag{17}$$

We denote the overall distribution on (A, A', A'') from the above probability by $\mathcal{D}$. Now consider selecting the pairs (A_0, B_0) and (A_0, B_1) according to the Agreement-Test with parameters $(q, q', 0)$ for the table $\tilde{G}$. Let $\tilde{A} = A_0 \cup B_0 \cap [n]$, $\tilde{A}' = A_0 \cup B_1 \cap [n]$ and $\tilde{A}'' = A_0 \cap [n]$. We use $\tilde{\mathcal{D}}$ to denote the distribution on $(\tilde{A}, \tilde{A}', \tilde{A}'')$. We show that the statistical distance between the distributions $\mathcal{D}$ and $\tilde{\mathcal{D}}$ is at most $o(1)$.

Claim 5.36. *The statistical distance between the distributions $\mathcal{D}$ and $\tilde{\mathcal{D}}$ is at most $o(1)$ when $\alpha q = q' + \frac{(q-q')^2}{(1-q')}$.*

Proof. For simplicity, consider the following distribution which is a refinement of the distribution $\mathcal{D}$. Fix $0 \leqslant p_1, p_2, p_3, p_4 \leqslant 1$ such that $\sum_i p_i \leqslant 1$. For each $i \in [n]$ independently, $i \in A \setminus A'$ with probability p_1 and $i \in A' \setminus A$ with probability p_2, $i \in (A \cap A') \setminus A''$ with probability p_3, and $i \in A''$ with probability p_4. Note that by choosing $p_1 = p_2 = (1-\alpha)q$, $p_4 = q'$ and $p_3 = \alpha q - q'$, we recover the given distribution $\mathcal{D}$ and hence we will fix these values of p_i throughout the claim.

Fix a triple (A, A', A''). Let p and $\tilde{p}$ be the probability masses given to (A, A', A'') by the distributions $\mathcal{D}$ and $\tilde{\mathcal{D}}$, respectively. Let $a = |A \setminus A'|$, $b = |A' \setminus A|$, $c = |(A \cap A') \setminus A''|$ and $d = |A''|$. We have

$$p = p_1^a \cdot p_2^b \cdot p_3^c \cdot p_4^d \cdot (1 - (p_1 + p_2 + p_3 + p_4))^{n-(a+b+c+d)}$$

We can compute $\tilde{p}$ as follows:

$$\begin{aligned}
\tilde{p} &= \Pr_{A_0,B_0,B_1}\left[A_0|_{[n]} = A'', B_0|_{[n]} = A \setminus A'', B_1|_{[n]} = A' \setminus A''\right] \\
&= \Pr_{A_0}\left[A_0|_{[n]} = A''\right] \cdot \Pr_{A_0,B_0}\left[B_0|_{[n]} = A \setminus A'' \mid A_0|_{[n]} = A''\right] \Pr_{A_0,B_1}\left[B_1|_{[n]} = A' \setminus A'' \mid A_0|_{[n]} = A''\right] \\
&= \frac{\binom{N-n}{q'N-d}}{\binom{N}{q'N}} \cdot \frac{\binom{N-q'N-(n-d)}{(q-q')N-(a+c)}}{\binom{N-q'N}{(q-q')N}} \cdot \frac{\binom{N-q'N-(n-d)}{(q-q')N-(b+c)}}{\binom{N-q'N}{(q-q')N}} \\
&= (1 \pm o(1)) \cdot q'^d (1-q')^{n-d} \cdot \left(\frac{q-q'}{1-q'}\right)^{a+c} \left(\frac{1-q}{1-q'}\right)^{n-(a+c+d)} \cdot \left(\frac{q-q'}{1-q'}\right)^{b+c} \left(\frac{1-q}{1-q'}\right)^{n-(b+c+d)},
\end{aligned}$$

where the last equality follows from Claim 5.33. It can be shown that $p = \tilde{p}$ with the following setting of $p_i s$

$$p_1 = p_2 = \frac{(q-q')(1-q)}{(1-q')}, p_3 = \frac{(q-q')^2}{(1-q')}, \text{ and } p_4 = q'.$$

Therefore, when $\alpha q = q' + \frac{(q-q')^2}{(1-q')}$, the above identities hold along with $p_1 = p_2 = (1-\alpha)q$, $p_4 = q'$ and $p_3 = \alpha q - q'$. This finishes the proof of this claim. $\square$

Using the above claim and Equation (17), we conclude

$$\Pr[G[\tilde{A}] \cap \tilde{A}'' = G[\tilde{A}'] \cap \tilde{A}''] \geqslant \varepsilon - o(1).$$

Since $\tilde{G}(S)|_i = 0$ for every $i > n$ and $S \ni i$, we have

$$\Pr_{(A_0,B_0),(A_0,B_1)}[\tilde{G}[A_0, B_0]|_{A_0} = \tilde{G}[A_0, B_1]|_{A_0}] \geqslant \varepsilon - o(1),$$

Therefore, using Theorem 1.4, we conclude that there exists a global function $\tilde{g} : [N] \to \{0,1\}$ such that

$$\Pr_{S \in \binom{[N]}{qN}}\left[\tilde{G}[S] \overset{\leqslant \alpha}{\neq} \tilde{g}(S)\right] \geqslant \delta,$$

where $\alpha = O_{q,q',\varepsilon}(1)$. Furthermore, based on how we came up with the global function $\tilde{g}$, we have $\tilde{g}(i) = 0$ for all $i \in (n, N]$ as $\tilde{G}(S)|_i = 0$ for all $S \in \binom{[N]}{qN}$ and $i \in S$. If we let $g : [n] \to \{0,1\}$ be the function $\tilde{g}$ restricted to the domain $[n]$, we have

$$\Pr_{A \sim \mu_q^{\otimes n}}\left[G[A] \overset{\leqslant \alpha}{\neq} g(A)\right] \geqslant \delta - o(1).$$

Here, we used Claim 5.34 that shows the statistical distance between the distribution $\mu_q^{\otimes n}$ and the distribution on $S|_{[n]}$ where S is a uniformly random set from $\binom{[N]}{qN}$ is at most $o(1)$. $\square$

References

[BCH+96] Mihir Bellare, Don Coppersmith, JOHAN Hastad, Marcos Kiwi, and Madhu Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.

[BDN17] Amey Bhangale, Irit Dinur, and Inbal Livni Navon. Cube vs. Cube Low Degree Test. In *8th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 67, pages 40:1–40:31, Dagstuhl, Germany, 2017.

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

[BKM22a] Amey Bhangale, Subhash Khot, and Dor Minzer. On Approximability of Satisfiable k-CSPs: I. In *Proceedings of the 54th annual Symposium on Theory of Computing (STOC)*, page 976–988. ACM, 2022.

[BKM22b] Amey Bhangale, Subhash Khot, and Dor Minzer. On Approximability of Satisfiable k-CSPs: II. In *(To appear in STOC 2023)*, 2022+.

[BKM22c] Amey Bhangale, Subhash Khot, and Dor Minzer. On Approximability of Satisfiable k-CSPs: IV. In *(Manuscript)*, 2022+.

[BLR90] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proceedings of the 22nd annual Symposium on Theory of Computing (STOC)*, pages 73–83, 1990.

[CFM+22] Gilad Chase, Yuval Filmus, Dor Minzer, Elchanan Mossel, and Nitin Saurabh. Approximate polymorphisms. In *Proceedings of the 54th annual Symposium on Theory of Computing (STOC)*, page 195–202. ACM, 2022.

[DDG+15] Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct sum testing. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS)*, pages 327–336, 2015.

[DFH19] Irit Dinur, Yuval Filmus, and Prahladh Harsha. Analyzing boolean functions on the biased hypercube via higher-dimensional agreement tests. In *Proceedings of the 30th annual Symposium on Discrete Algorithms (SODA)*, pages 2124–2133. SIAM, 2019.

[DG08] Irit Dinur and Elazar Goldenberg. Locally testing direct product in the low error range. In *Proceedings of the 49th annual symposium on Foundations of Computer Science (FOCS)*, pages 613–622. IEEE, 2008.

[DLN17] Irit Dinur and Inbal Livni Navon. Exponentially small soundness for the direct product z-test. In *32nd Computational Complexity Conference (CCC)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[DR06] Irit Dinur and Omer Reingold. Assignment testers: Towards a combinatorial proof of the PCP theorem. *SIAM Journal on Computing*, 36(4):975–1024, 2006.

[DS14] Irit Dinur and David Steurer. Direct product testing. In *29th Conference on Computational Complexity (CCC)*, pages 188–196. IEEE, 2014.

[FKLM20] Yuval Filmus, Guy Kindler, Noam Lifshitz, and Dor Minzer. Hypercontractivity on the symmetric group. *arXiv preprint arXiv:2009.05503*, 2020.

[GS00] Oded Goldreich and Shmuel Safra. A combinatorial consistency lemma with application to proving the pcp theorem. *SIAM Journal on Computing*, 29(4):1132–1154, 2000.

[IKW12] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query pcps. *SIAM Journal on Computing*, 41(6):1722–1768, 2012.

[KLX10] Tali Kaufman, Simon Litsyn, and Ning Xie. Breaking the ε-soundness bound of the linearity test over gf (2). *SIAM Journal on Computing*, 39(5):1988–2003, 2010.

[KS09] Swastik Kopparty and Shubhangi Saraf. Tolerant linearity testing and locally testable codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 601–614. Springer, 2009.

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

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

[Mos20] Elchanan Mossel. Gaussian bounds for noise correlation of resilient functions. *Israel Journal of Mathematics*, 235(1):111–137, 2020.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il