On Approximability of Satisfiable k-CSPs: III

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

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

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

ABSTRACT

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 towards understanding approximability of satisfiable Constraint Satisfaction Problems.

Towards proving our structural theorem, we analyze a 2-query direct product test for the table $F : \binom{[n]}{qn} \rightarrow \{0,1\}^{qn}$ 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 $g : [n] \rightarrow \{0,1\}$ 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 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 the previous results and hence our analysis involves new techniques, including the use of the small-set expansion property of graphs defined on multi-slices.

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 $f : \{0,1\}^n \rightarrow \{0,1\}$ 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 where the authors studied the test with soundness $1 - \varepsilon$, for ε close to 0.

CCS CONCEPTS

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

KEYWORDS

constraint satisfaction problems, hardness of approximation, linearity test, direct product test

ACM Reference Format:
Amey Bhangale, Subhash Khot, and Dor Minzer. 2023. On Approximability of Satisfiable k-CSPs: III. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC '23), June 20–23, 2023, Orlando, FL, USA*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3564246.3585121

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, 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 [3, 4]. 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 $f : \{0,1\}^n \rightarrow \{0,1\}$, 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 $f_{I\rightarrow z} : \{0,1\}^{n-|I|} \rightarrow \{0,1\}$ where $f_{I\rightarrow z}(x) = f(x, z|_I)$, i.e., we fix the variables from I according to z. In this work, we study the properties of f if $f_{I\rightarrow z}$ 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.

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

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

ACM ISBN 978-1-4503-9913-5/23/06.
https://doi.org/10.1145/3564246.3585121

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

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

In other 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. *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*

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

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 by the study of satisfiable CSPs. In particular, in [4] the authors proved an analytical lemma [4, 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 [5] 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 [3]), 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.

Upon trying to think of Theorem 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

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

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 [6], 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 [1, 16].

For any $q \neq 1/2$, one can recover the result for the 99% regime using the same local-correction techniques [9, 17] 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 already by itself gives some structural result for functions f satisfying (2), and to see that, we re-write (2) as

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

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. 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. *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 $s \in \{0,1\}^S$,*

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

In 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 $g_1, \ldots, g_4$ 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 2. *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 2 for real-valued functions with bounded 12-norms, as well as a list-decoding version. We refer to the full-version of the paper the proof of the above theorem.

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 [9, 10, 12, 13, 15] 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 [15]). 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 [9, 12] (in which one wants to draw the conclusion when the test passes with probability $1-\varepsilon$) and in the 1% regime [10, 13, 15] (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, and hence proving Theorem 2), 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 $g_1, \ldots, g_m : [n] \to \{0,1\}$ 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 3. *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*

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

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 $|\{i \in S \mid F[S]_i \neq g(i)\}| \leqslant r$.

We refer to the full-version of the paper for the proof of the theorem. 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 [9] 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 [7].

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 3, asserts that in this case, G must be correlated to a global subset $S \subseteq [n]$.

Corollary 1. *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 a function $G\colon (P[n], \mu_q^{\otimes n}) \to P([n])$ 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 [6] gave the 3-query lineary test under uniform distribution in the 99% regime. [1, 16] 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 [17] gives $O_p(1)$-query linearity test with soundness $1-\varepsilon$ for ε close to 0. David, Dinur, Goldenberg, Kindler and Shinkar [8] 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 [9] 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 [14] in which they show that it can be testable with constantly many queries. Dinur and Reingold [12] gave a 2-query direct product test in the 99% accepting regime. Dinur and Goldenberg [10] 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 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 [10] and obtained similar results. Dinur and Livni Navon [11] 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 [9] 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 [9], 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. 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

[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.

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. 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, 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 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 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 2: Linearity Testing Over a Biased Hypercube.* We begin with an overview of the proof of Theorem 2 and the overall idea is as follows. We know that [1] 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 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 2, i.e. such that

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

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 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

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

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

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

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

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

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 [19], and more specifically to a version from [18]. 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

$$\mathop{\mathbb{E}}_{(z^1,z^2,z^3,z^4)\sim\tilde{\nu}^{\otimes n}}\left[P_1(z^1)P_2(z^2)P_3(z^3)P_4(z^4)\right],$$

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 3: 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 $g(i) = \mathrm{Majority}_{S, S\ni i} F[S]|_i$, 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 [9, 12] 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 [15]. This framework, that we will explain next, has been used in [2, 11] 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 $g_{S,\sigma}(i) = \mathrm{Majority}_{S', S'\ni i, F[S']|_{S\cap S'}=\sigma|_{S'\cap S}} F[S']|_i$, 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 $C_{S,\sigma} \subseteq \binom{[n]}{qn}$ 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 $C_{S,\sigma}$ and $C_{S',\sigma'}$ 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 [10, 15] 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 $C_{S,\sigma}$ and $C_{\tilde{S},\sigma'}$ 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 [20] 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 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

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

[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.

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. *The degree d weight of a function $f : (\Omega, \mu) \to \mathbb{R}$ 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 3. *For a function $f : (\Omega = \Omega_1 \times \ldots \times \Omega_n, \mu = \mu_1 \times \ldots \times \mu_n) \to \mathbb{R}$ 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 4. *For a function $f : (\Omega = \Omega_1 \times \ldots \times \Omega_n, \mu = \mu_1 \times \ldots \times \mu_n) \to \mathbb{R}$, $d \in \mathbb{N}$ 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 5. *Let $q \in (0, 1)$, and denote $\sigma = \sqrt{q(1-q)}$ the standard deviation of a q-biased random coin. We define the function $\chi_i^q : \{0, 1\} \to \mathbb{R}$ as*

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

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

For the q-biased measure, one can show that for a function $f : (\{0, 1\}^n, \mu_q^{\otimes n}) \to \mathbb{R}$, 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 $f : (\{0, 1\}^n, \mu^{\otimes n}) \to \mathbb{R}$, a set of coordinates $I \subseteq [n]$ and a partial input $z \in \{0, 1\}^I$, the restricted function $f_{I\to z} : \{0, 1\}^{[n]\setminus I} \to \mathbb{R}$ 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 $f : (\{0, 1\}^n \mu^{\otimes n}) \to \mathbb{R}$ 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 [3, 4], 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 6. *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 $\mathrm{T}_{\mu,\rho} : L_2(\{0, 1\}, \mu) \to L_2(\{0, 1\}, \mu)$ the corresponding averaging operator defined as $\mathrm{T}_{\mu,\rho} f(x) = \mathbb{E}_{y\sim_{\rho,\mu} x}[f(x)]$.

For multi-variate functions $f : (\{0, 1\}^n, \mu^{\otimes n}) \to \mathbb{R}$, 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 $T_{\mu,\rho}^{\otimes n}$. When the measure μ and n are clear from context, we often omit them from the notation.

Definition 7. *Let μ be a distribution over $\{0,1\}$, let $\rho \in [0,1]$ and let $f : (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$ 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 1. *Let μ be a distribution over $\{0,1\}$, and let $f : (\{0,1\}^n, \mu^{\otimes n}) \to \mathbb{R}$ be a function. Then*

$$\mathsf{Stab}_{1-\kappa}(f) = \mathop{\mathbb{E}}_{I\sim_{1-\kappa}, z\sim\mu^I}\left[\widehat{f}_{I\to z}(\emptyset)^2\right].$$

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. □

The following lemma is [4, 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. *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 4. *For every $r \in \mathbb{N}$ and $q \in (0,1)$ there is $C(q,r) > 0$ such that if $f : (\{0,1\}^n, \mu_q^{\otimes n}) \to \mathbb{R}$ 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 with small average has most of its mass on high levels.

Theorem 5. *For every $q \in (0,1)$, there is $c_q > 0$ such that the following holds. Suppose that a function $f : (\{0,1\}^n, \mu_q^{\otimes n}) \to \{0,1\}$ 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 5 asserts that almost of the spectral mass of f lies above level d.

2.7 Markov Chains

Finally, we need the following result from [18], 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 3. *[[18, 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

This section is devoted for the proof of Theorem 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 $f : \{0,1\}^n \to \mathbb{R}$ 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 1. *Suppose that $f : \{0,1\}^n \to \mathbb{R}$ is a function with $\|f\|_2 \leqslant 1$ and $\left|\widehat{f}(S)\right| \geqslant \delta$ for some S. Then for all $I \subseteq [n]$,*

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

Proof. Fixing I, we have

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

so by the triangle inequality

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

On the other hand,

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

Hence, we get by the Paley-Zygmund inequality that

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

□

The following fact is similar in spirit to Fact 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 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*

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

Proof. We have

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

so by the triangle inequality

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

On the other hand,

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

Hence, the result follows again by the Paley-Zygmund inequality. □

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. *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, 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; 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

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

where $\widehat{g}(S) = \mathbb{E}_x[g(x)\chi_S(x)]$. Note that by the premise of Theorem 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 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 that

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

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

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

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 1. *For all I', we have that*

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

Proof. For each $T \subseteq \overline{I'}$, define the set

$$X_T = \left\{ z' \in \{0,1\}^{I'} \;\middle|\; \left| \widehat{f_{I'\to z'}}(T) \right| \geqslant \frac{\delta}{2} \right\}.$$

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 $p_{a\to b} = \Pr\left[z''_1 = b \;\middle|\; z'_1 = a \right]$, 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 3 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 5, 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

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

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. ◻

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 8. *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 2. *We have*

$$\Pr_{I_1,I_2,I'_2}\left[\exists S \subseteq [n], S\cap \overline{I_2} \in W_{I_1\cup I_2} \wedge S\cap \overline{I'_2} \in W_{I_1\cup I'_2} \right] \geqslant \frac{\eta\delta^8}{64}.$$

Proof. Let E be the event in (8). Combining Claim 1 and (8), we get that

$$\begin{aligned}&\Pr_{\substack{I_1,I_2,I'_2\\ z(1),z(2),z(3)\\ z(2)',z(3)'}}\left[\exists S \in W_{I_1,z(1)} : E \wedge S\cap \overline{I_2} \in W_{I_1\cup I_2} \wedge S\cap \overline{I_2} \in W_{I_1\cup I'_2} \right] \\ &\geqslant \frac{\delta^8}{16}\Pr\left[W_{I_1,z(1)} \neq \emptyset \right] - 2\xi,\end{aligned}$$

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

$$\begin{aligned}&\Pr_{\substack{I_1,I_2,I'_2\\ z(1),z(2),z(3)\\ z(2)',z(3)'}}\left[\exists S \in W_{I_1,z(1)}, E \wedge S\cap \overline{I_2} \in W_{I_1\cup I_2} \wedge S\cap \overline{I_2} \in W_{I_1\cup I'_2} \right] \\ &\geqslant \frac{\delta^8}{16}\frac{\eta}{2} - 2\xi \geqslant \frac{\eta\delta^8}{64}.\end{aligned}$$

◻

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

Claim 3. *For all I', $|W_{I'}| \leqslant \frac{4}{\zeta\delta^2}$.*

Proof. Note that

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

On the other hand,

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

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 a function $F\colon (P([n]), \mu_{1-\beta/2}^{\otimes n}) \to P([n])$ 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 4. *Over the randomness of the choice of the assignment F, we have that*

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

Proof. By Claim 2, 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 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, 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

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

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

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

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.

REFERENCES

[1] Mihir Bellare, Don Coppersmith, JOHAN Hastad, Marcos Kiwi, and Madhu Sudan. 1996. Linearity testing in characteristic two. *IEEE Transactions on Information Theory* 42, 6 (1996), 1781–1795. https://doi.org/10.1109/18.556674
[2] Amey Bhangale, Irit Dinur, and Inbal Livni Navon. 2017. Cube vs. Cube Low Degree Test. In *8th Innovations in Theoretical Computer Science Conference (ITCS)*, Vol. 67. Dagstuhl, Germany, 40:1–40:31. https://doi.org/10.4230/LIPIcs.ITCS.2017.40
[3] Amey Bhangale, Subhash Khot, and Dor Minzer. 2022. On Approximability of Satisfiable k-CSPs: I. In *Proceedings of the 54th annual Symposium on Theory of Computing (STOC)* (Rome, Italy). ACM, 976–988. https://doi.org/10.1145/3519935.3520028
[4] Amey Bhangale, Subhash Khot, and Dor Minzer. 2022+. On Approximability of Satisfiable k-CSPs: II. In *(To appear in STOC 2023)*.
[5] Amey Bhangale, Subhash Khot, and Dor Minzer. 2022+. On Approximability of Satisfiable k-CSPs: IV. In *(Manuscript)*.
[6] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. 1990. Self-testing/correcting with applications to numerical problems. In *Proceedings of the 22nd annual Symposium on Theory of Computing (STOC)*. 73–83. https://doi.org/10.1145/100216.100225
[7] Mark Braverman, Subhash Khot, Noam Lifshitz, and Dor Minzer. 2022. An Invariance Principle for the Multi-slice, with Applications. In *Proceedings of the 62nd annual symposium on Foundations of Computer Science (FOCS)*. IEEE, 228–236. https://doi.org/10.1109/FOCS52979.2021.00030
[8] Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. 2015. Direct sum testing. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS)*. 327–336. https://doi.org/10.1145/2688073.2688078
[9] Irit Dinur, Yuval Filmus, and Prahladh Harsha. 2019. Analyzing Boolean functions on the biased hypercube via higher-dimensional agreement tests. In *Proceedings of the 30th annual Symposium on Discrete Algorithms (SODA)*. SIAM, 2124–2133. https://doi.org/10.1137/1.9781611975482.128
[10] Irit Dinur and Elazar Goldenberg. 2008. Locally testing direct product in the low error range. In *Proceedings of the 49th annual symposium on Foundations of Computer Science (FOCS)*. IEEE, 613–622. https://doi.org/10.1109/FOCS.2008.26
[11] Irit Dinur and Inbal Livni Navon. 2017. Exponentially small soundness for the direct product z-test. In *32nd Computational Complexity Conference (CCC)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. https://doi.org/10.4230/LIPIcs.CCC.2017.29
[12] Irit Dinur and Omer Reingold. 2006. Assignment testers: Towards a combinatorial proof of the PCP theorem. *SIAM J. Comput.* 36, 4 (2006), 975–1024. https://doi.org/10.1137/S0097539705446962

[13] Irit Dinur and David Steurer. 2014. Direct product testing. In *29th Conference on Computational Complexity (CCC)*. IEEE, 188–196. https://doi.org/10.1109/CCC.2014.27

[14] Oded Goldreich and Shmuel Safra. 2000. A combinatorial consistency lemma with application to proving the PCP theorem. *SIAM J. Comput.* 29, 4 (2000), 1132–1154. https://doi.org/10.1137/S0097539797315744

[15] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. 2012. New direct-product testers and 2-query PCPs. *SIAM J. Comput.* 41, 6 (2012), 1722–1768. https://doi.org/10.1137/09077299X

[16] Tali Kaufman, Simon Litsyn, and Ning Xie. 2010. Breaking the ε-soundness bound of the linearity test over GF (2). *SIAM J. Comput.* 39, 5 (2010), 1988–2003. https://doi.org/10.1137/080715548

[17] Swastik Kopparty and Shubhangi Saraf. 2009. Tolerant linearity testing and locally testable codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*. Springer, 601–614. https://doi.org/10.1007/978-3-642-03685-9_45

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

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

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

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