Gowers Uniformity, Influence of Variables, and PCPs

Alex Samorodnitsky[*]   Luca Trevisan[†]

October 12, 2005

Abstract

Gowers [Gow98, Gow01] introduced, for $d \geq 1$, the notion of *dimension-d uniformity* $U^d(f)$ of a function $f : G \to \mathbb{C}$, where G is a finite abelian group. Roughly speaking, if a function has small Gowers uniformity of dimension d, then it "looks random" on certain structured subsets of the inputs.

We prove the following "inverse theorem." Write $G = G_1 \times \cdots \times G_n$ as a product of groups. If a bounded balanced function $f : G_1 \times \cdots G_n \to \mathbb{C}$ is such that $U^d(f) \geq \varepsilon$, then one of the coordinates of f has *influence* at least $\varepsilon/2^{O(d)}$. Other inverse theorems are known [Gow98, Gow01, GT05, Sam05], and U^3 is especially well understood, but the properties of functions f with large $U^d(f)$, $d \geq 4$, are not yet well characterized.

The *dimension-d Gowers inner product* $\langle \{f_S\} \rangle_{U^d}$ of a collection $\{f_S\}_{S \subseteq [d]}$ of functions is a related measure of pseudorandomness. The definition is such that if all the functions f_S are equal to the same fixed function f, then $\langle \{f_S\} \rangle_{U^d} = U^d(f)$.

We prove that if $f_S : G_1 \times \cdots \times G_n \to \mathbb{C}$ is a collection of bounded functions such that $|\langle \{f_S\} \rangle_{U^d}| \geq \varepsilon$ and at least one of the f_S is balanced, then there is a variable that has influence at least $\varepsilon^2/2^{O(d)}$ for at least four functions in the collection.

Finally, we relate the acceptance probability of the "hypergraph long-code test" proposed by Samorodnitsky and Trevisan to the Gowers inner product of the functions being tested and we deduce the following result: if the Unique Games Conjecture is true, then for every $q \geq 3$ there is a PCP characterization of NP where the verifier makes q queries, has almost perfect completeness, and soundness at most $2q/2^q$. For infinitely many q, the soundness is $(q+1)/2^q$. Two applications of this results are that, assuming that the unique games conjecture is true, it is hard to approximate Max kCSP within a factor $2k/2^k$ (or even $(k+1)/2^k$, for infinitely many k), and it is hard to approximate Independent Set in graphs of degree D within a factor $(\log D)^{O(1)}/D$.

1   Introduction

We return to the study of the relation between number of queries and error probability in probabilistically checkable proofs.

---

[*]Institute of Computer Science, Hebrew University. `salex@huji.ac.il`

The PCP Theorem [AS98, ALM+98] states that it is possible to encode certificates of satisfiability for SAT instances (and, more generally, for every problem in NP) in such a way that a probabilistic verifier can check the validity of the certificate with high confidence after inspecting only a constant number of bits. We write $SAT \in \mathrm{PCP}_{c,s}[r(n), q]$ if there is a verifier that uses at most $r(n)$ random bits, where n is the size of the formula, accepts encoding of valid proofs with probability at least c (the *completeness* probability of the verifier) and accepts purported encodings of proofs with probability at most s (the *soundness* error of the verifier) if the formula is unsatisfiable. The PCP Theorem states that there exists a constant k such that $SAT \in \mathrm{PCP}_{1,1/2}[O(\log n), k]$. Improvements and variants of the PCP Theorem and their applications to the study of the approximability of optimization problems are too many to summarize here, and we refer the reader to the chapter on hardness of approximation in Vazirani's book [Vaz01] and to some recent survey papers [Aro02, Fei02, Tre04].

In this paper we are interested in the following question: for a given number of queries, what is the highest confidence that we can have in the validity of the proof? That is, for a given value of q, what is the smallest value $s = s(q)$ for which $SAT \in \mathrm{PCP}_{1-\delta,s+\delta}[O(\log n), q]$ for every $\delta > 0$? We call this parameter s the *soundness* of the PCP construction. A good measure of the trade-off between the number q of queries and the soundness s is the so-called *amortized query complexity*, defined as $\bar{q} = q/(\log_2 s^{-1})$.

A simple argument shows that, unless $P = NP$, s must be at least $1/2^q$, that is, the amortized query complexity must be at least 1. A more careful argument gives a lower bound of $2/2^q$ [Tre98a] on the soundness, which was recently improved to $\Omega(\frac{q}{\log q} \cdot \frac{1}{2^q})$ by Hast [Has05]. (Hast's result can also be stated as giving a lower bound of $1 + (1 - o(1))\frac{\log q}{q}$ to the amortized query complexity of a q-query PCP.) The PCP Theorem shows that we can have $s = 1/2^{\Omega(q)}$, and the authors showed that can have $s \leq 2^{2\sqrt{q}}/2^q$ [ST00]. (That is, the amortized query complexity can be as low as $1 + O(1/\sqrt{q})$.) Our proof was simplified by Håstad and Wigderson [HW03], and the soundness was improved to $s \leq 2^{\sqrt{2q}}/2^q$ by Engebretsen and Holmerin [EH05]. As we discuss below, $2^{\Theta(\sqrt{q})}/2^q$ was a natural limit for the soundness achievable with current techniques.

In this paper, assuming Khot's Unique Games Conjecture [Kho02], we present an improvement to $s = (q + 1)/2^q$. Our analysis is based on a theorem, which is probably of independent interest, bounding the Gowers uniformity of a given function in terms of the influence of its variables.

1.1 Linearity Tests and PCP

The *linearity testing problem* captures most of the technical difficulties of the the construction of PCP constructions, and it is a good starting point.

Let us call a function $f : \{0, 1\}^n \to \{-1, 1\}$ *linear* if it is a homomorphism between the group $\{0, 1\}^n$ (where the operation is bitwise XOR, written additively) and the group $\{-1, 1\}$ (where the operation is multiplication). Equivalently, f is linear if and only if it can be written as $f(x_1, \ldots, x_n) = (-1)^{\sum_{i \in S} x_i}$ for some set $S \subseteq [n]$. We use the notation $\chi_S(x) := (-1)^{\sum_{i \in S} x_i}$.

In the linearity testing problem we are given oracle access to a boolean function $f : \{0, 1\}^n \to \{-1, 1\}$ and we would like to distinguish between the following extreme settings:

1. f is linear;

2. for every S, the agreement between f and χ_S is at most $1/2 + \varepsilon$.

By *agreement* between a function f and a function g we mean the fraction of inputs on which they are equal. We say that a test has *error-probability* at most e if in case (1) it accepts with probability 1 and in case (2) it accepts with probability at most $e + \varepsilon'$, where $\varepsilon' \to 0$ when $\varepsilon \to 0$.

Blum, Luby and Rubinfeld [BLR93] define a very simple such test, that makes only three queries into f:

BLR-Test (f)
   choose x, y uniformly at random in $\{0, 1\}^n$
   accept if and only if $f(x) \cdot f(y) = f(x + y)$

Bellare et al. [BCH$^+$96] give a tight analysis of this test, showing that if it accepts with probability at least $1/2 + \varepsilon$, then f has agreement at least $1/2 + \varepsilon$ with some linear function. According to our above definition, the BLR test has error probability at most $1/2$.

There are at least two ways in which such a result needs to be extended before it can be used in a PCP construction.

First of all, we would like to consider a case where two or more functions are given as an oracle, and the test wants to distinguish between the following cases:

1. The functions given as an oracle are all equal to the same linear function

2. No two functions have agreement more than $1/2 + \varepsilon$ (or less than $1/2 - \varepsilon$) with the same linear function

There is a natural extension of the BLR test to this setting:

3-functions-BLR-Test (f, g, h)
   choose x, y uniformly at random in $\{0, 1\}^n$
   accept if and only if $f(x) \cdot g(y) = h(x + y)$

Aumann et al. [AHRS01] show that if this test accepts with probability $1/2 + \varepsilon$, then there is a linear function χ_S such that f, g, h have all agreement at least $1/2 + \varepsilon/3$ with either χ_S or $-\chi_S$.

The second change is that, for the sake of PCP constructions, we are especially interested in linear functions χ_S with a small S. We call functions of the form $\chi_{\{i\}}(x) = (-1)^{x_i}$ *long codes*. In a *long code test* we are given several functions and, in a commonly used definition, we want to distinguish the following cases:

1. The functions given as an oracle are all equal to the same long code;

2. For every small S, no two functions have agreement more than $1/2 + \varepsilon$ (or less than $1/2 - \varepsilon$) with χ_S.

We say that such a test has error probability at most e if, whenever it accepts with probability more than $e + \varepsilon$, there are constants ε', d, depending only on ε, and a set S, $|S| \leq d$, such that at least two of the given functions have agreement at least $1/2 + \varepsilon'$ with χ_S or $-\chi_S$.

A test satisfying this definition can be obtained from the BLR test by adding noise to each query. Let μ_δ be the probability distribution over $\{0,1\}^n$ defined by picking n independent copies of a biased coin that returns 0 with probability $1-\delta$ and 1 with probability δ, and consider the following test:

δ-noisy-3-functions-BLR-Test (f, g, h)

   choose x, y uniformly at random in $\{0,1\}^n$

   sample η_1, η_2, η_3 indipendently according to μ_δ

   accept if and only if $f(x + \eta_1) \cdot g(y + \eta_2) = h(x + y + \eta_3)$

It is easy to see that in case (1), that is, when f, g, h are equal to the same long code, then the test accepts with probability at least $(1 - \delta)^3$. Håstad [Hås01] shows[1] that if the test accepts with probability at least $1/2 + \varepsilon$, then there is a set S of size at most $\mathrm{poly}(\varepsilon^{-1}, \delta^{-1})$ such that f, g, h all have agreement at least $1/2 + \varepsilon/3$ with χ_S or $-\chi_S$.

Having a q-query test for this problem with error probability at most e is enough to construct a PCP characterization of NP with query complexity q and soundness error about e, provided that Khot's *unique games conjecture* [Kho02] is true. (Further refinements are needed to derive an unconditional result in [Hås01].)

Given this logical path from the basic linearity testing problem to the task of constructing PCPs, our plan in [Tre98b, ST98, ST00] was to devise linearity tests with good trade-offs between number of queries and error probability, and then "translate" such tests into a PCP construction.

In [ST00] we devise a linearity tester whose asymptotic trade-off between number of queries and error probability is optimal. The test, for $k \geq 2$ is defined as follows:

Complete-Graph-Test (f, k)

   choose $x_1, \ldots, x_k$ uniformly at random in $\{0,1\}^n$

   accept if and only if

     for every $i \neq j$, $f(x_i) \cdot f(x_j) = f(x_i + x_j)$

Note that the test has query complexity $q = k + \binom{k}{2}$ and runs $\binom{k}{2}$ correlated copies of the BLR test. If f is linear, then it is clearly accepted with probability 1. If f has agreement at most $1/2 + \varepsilon$, then we already know that each of the $\binom{k}{2}$ tests accepts with probability at most $1/2 + \varepsilon$. In [ST00] we show that the $\binom{k}{2}$ tests behave almost mutually independently, and the probability that all accept is at most $1/2^{\binom{k}{2}} + \varepsilon' \approx 1/2^{q - \sqrt{2q}}$. We also extended the test to the noisy case, the case of several functions, and the setting (which we do not describe in this paper) which is sufficient to derive an unconditional PCP characterization of NP.[2]

One might have thought that the following test would have achieved an even better trade-off between number of queries and error probability:

---

[1]Such an analysis is implicit in Håstad's paper, and the result, as stated here, appears explicitly in [AHRS01].

[2]In this last step, we lost something in the soundness error, which became $1/2^{q - 2\sqrt{q}}$, where q is the number of queries. This was recently improved to $1/2^{q - \sqrt{2q}}$, the same bound of the basic linearity test, by Engebretsen and Holmerin [EH05].

> Complete-Hypergraph-Test (f, k)
>   choose $x_1, \ldots, x_k$ uniformly at random in $\{0,1\}^n$
>   **accept** if and only if
>   for every $S \subseteq [k]$: $|S| \geq 2$. $\prod_{j \in S} f(x_j) = f(\sum_{j \in S} x_j)$

In the hypergraph test we make $q = 2^k - 1$ queries and run $2^k - k - 1$ correlated copies of the BLR test. If the tests behaved almost mutually independently for f far from linear, then the test would have error probability $\approx (q + 1)/2^q$.

Unfortunately, the above statement fails already when $k = 3$ if we consider the function $f(x) = (-1)^{x_1 x_2 + \cdots + x_{n-1} x_n}$. Such a function is very far from linear, but the 7 conditions of the hypergraph test for $k = 3$ are not almost independent. In fact, much more generally, we prove in [ST00] that if we run a q-query hypergraph test on f, then the test accepts with probability at least $2^{-q + \Omega(\sqrt{q})}$. In Section 7 we generalize this result and prove that any linearity test that makes q queries and that accepts linear functions with probability $\geq c$ must accept f with probability at least $(1 - c) + 2^{-q + \Omega(\sqrt{q})}$.

Even though there is no more room for improvement in the basic linearity testing problem (or its multi-function version, which is only harder), there is still hope for PCP, by using a more relaxed formulation of the long code test. Several recent papers [DS05, Kho02, KR03, KKMO04, CKK$^+$05, KV05] define a long code test based not on closeness to linear functions χ_S with small S but based on the notion of degree-d influence. We will define such a notion later (Section 2). For now, it suffices to say that if we can solve a certain (relaxed) variant of the linearity test with a given query complexity and error probability, then we can also test the long code with the same query complexity and error probability. We say that the relaxed test has error probability e if:

1. if f is linear, then the test accepts with probability 1;

2. if the test accepts with probability $e + \varepsilon$, then there is a coordinate i that has influence at least $\delta(\varepsilon)$ for f.

Influence (as opposed to "degree-d influence") has a simple definition for boolean functions: the influence of i for f is the probability that f is a non-constant function after we randomly fix all coordinates of f except the i-th. If f has agreement $1/2 + \varepsilon$ with a non-constant linear function χ_S, then f has variables of influence at least 2ε, but there are functions f that have influential variables even though they are very far from all linear functions (this is why this test is a relaxation). Intriguingly, in the function $f(x) = (-1)^{x_1 x_2 + \cdots + x_{n-1} x_n}$ all variables have influence $1/2 \pm o(1)$.

As we shall see later, we prove that the complete hypergraph test has error probability only $1/\left(2^{2^k - k - 1}\right)$ with respect to the above definition. That is, if a function f is accepted by the test with probability more than $1/\left(2^{2^k - k - 1}\right) + \varepsilon$, then one of the variables of f has influence at least $\delta(\varepsilon)$. We then generalize the test to the setting of several function (in which case we show that at least two functions have an influential variable in common) and to the case of testing the long code, resulting in a conditional PCP construction assuming the unique games conjecture.

1.2 Szemerédi's Theorem and Gowers Norms

We will use a definition that has been introduced by Gowers [Gow98, Gow01] in his seminal papers on a new proof of Szemerédi's Theorem.

Szemerédi's Theorem states that any subset of the integers of positive density contains arbitrarily long arithmetic progressions. The quantitative statement is that for every δ, k there is a $n(\delta, k)$ such that if A is an arbitrary subset of the integers $\{1, \ldots, N\}$, $|A| \geq \delta N$, and $N \geq n(\delta, k)$, then A must contain an arithmetic progression of length k, that is, there are integers a, b such that $a, a + b, \ldots, a + (k-1)b$ all belong to A.

In Szemerédi's original proof, which introduced the famous Szemerédi Regularity Lemma, $n(\delta, k)$ grows *very* fast with k and δ^{-1}: it is a tower of exponentials whose height is a tower of exponentials whose height is polynomial in k and δ^{-1}.

The case $k = 3$ had been settled earlier by Roth, with a simpler analytical proof that gave a reasonable (doubly exponential) relation between $n(\delta, 3)$ and δ^{-1}. Roth's proof (like all other proofs about arithmetic sequences in dense sets) is based on an iterative argument on δ: if δ is a sufficiently large constant (say, 90%), then a random length-3 progression is contained in A with positive probability, and so A contains some length-3 progressions. Otherwise (blurring, for the moment, the distinction between progressions mod N and true progressions), one writes the fraction of length-3 progressions in A as the number one would expect if A where a random set of density δ, that is, a δ^3 fraction of all length-3 progressions, minus an error term that equals $\sum_{g \neq 0} \hat{f}_A^3(g)$, the summation of the cubes of the non-zero Fourier coefficients of f_A, the characteristic function of A. One then considers two cases: if all Fourier coefficients of f_A are small, then the error term is smaller than δ^3, and so A contains a positive fraction of all length-3 progressions in $\mathbb{Z}_N$. If f_A has a large Fourier coefficient, then one can reduce the task of finding a length-3 progression in A to the task of finding a length-3 progression in a certain subset A' of $\{1, \ldots, N'\}$ of density $\delta' > \delta + \Omega(\delta^2)$. In this reduction, N' is about $\sqrt{N}$, one does no more than $O(\delta^{-1})$ such reductions, so it's enough to start with $N = 2^{2^{O(1/\delta)}}$. We mention these technical details because they are very similar to the analysis of the basic BLR linearity test in [BCH+96], where the acceptance probability of the verifier is written as $\frac{1}{2} + \frac{1}{2} \sum_\alpha \hat{f}^3(\alpha)$, and so (i) if all Fourier coefficients of f are small, then $\sum_\alpha \hat{f}^3(\alpha)$ is small, and the test accepts with probability close to $1/2$, while (ii) if one Fourier coefficient is large, then f is close to a linear function (this part is trivial).

A way to look at both proofs is to think of a function f as being "pseudorandom" if all its Fourier coefficients are small, and of a set $A \subseteq [N]$ to be pseudorandom if all the non-zero Fourier coefficients of its characteristic function are small. Then one proves that, for a pseudorandom function, the values $f(x), f(y), f(x + y)$ are nearly independent random bits, and so, in particular, $f(x)f(y) = f(x+y)$ happens with probability approximately $1/2$. For a psedorandom set of density δ, the events $x \in A$, $(x + r) \in A$, $(x + 2r) \in A$ also behave nearly independently for random x, r, and they happen simultaneously with probability approximately δ^3.

To prove Szemerédi's Theorem for $k \geq 4$ one might try to show that a pseudorandom set, as defined above, has approximately the expected number of length-k progression. This, unfortunately, does not seem to be true, and sets whose characteristic function is defined in terms of a degree-2 polynomial are basic counterexamples even for $k = 4$. (See [Gow01].) Recall that something similar happens in the hypergraph test, where a function defined in terms of a degree-2 polynomial is very far from linear (and so all its Fourier coefficients are small), but the tests performed in the

6

hypergraph test do not behave independently.

Gowers [Gow98, Gow01] resolves this problem by introducing a more refined measure of pseudo-randomness. For $d \geq 1$ and $f : G \rightarrow \mathbb{R}$, where G is a group, he defines the quantity

$$U^d(f) := \mathop{\mathbb{E}}_{x, x_1, \ldots, x_d} \left[\prod_{S \subseteq [d]} f \left(x + \sum_{i \in S} x_i \right) \right]$$

and there is a similar definition for $f : G \rightarrow \mathbb{C}$, in which all evaluations of $f()$ when $|S|$ is odd are conjugated. (See definitions in Sections 2 and 6.)

Gowers goes on to prove that if $A \subseteq \mathbb{Z}_N$ is a subset of density δ, $f : Z_N \rightarrow [-1, 1]$ is a normalized version of the characteristic function, and $U^d(f)$ is sufficiently small (as a function of d and δ, but not of N) then A must contain arithmetic progressions of length $d + 1$. The *hard* part of Gowers's work is then to show that if $U^d(f)$ is large, then finding progressions in A reduces to finding progressions in a set $A' \subseteq \mathbb{Z}_{N'}$ of density strictly larger than δ.

Towards this goal, Gowers proves certain structural properties of functions $f : \mathbb{Z}_N \rightarrow \mathbb{C}$, N prime, having non-trivially large U^d value. Such functions are shown to have a certain "local correlation" with degree $(d-1)$-polynomials. Green and Tao [GT05] study functions $f : G \rightarrow \mathbb{C}$ with large U^3, and, provided that the order of G is not divided by and 2 and 3, prove a certain "global correlation" between such functions and degree-2 polynomials.[3] Samorodnitsky [Sam05] proves such a result for functions $f : \{0, 1\}^n \rightarrow \mathbb{R}$.

Not much is known about functions $f : G \rightarrow \mathbb{C}$ having large U^d when $d \geq 4$ and G is a general group.

1.3   Our Results

We prove that if $f : G_1 \times \cdots \times G_n \rightarrow \mathbb{C}$ is a bounded balanced function and $U^d(f) \geq \varepsilon$, then there is a variable that has influence at least $\varepsilon / 2^{O(d)}$ for f. Above, we only defined influence in the boolean case, but a more general definition applies to functions mapping from an arbitrary product space into the complex numbers.

Green and Tao introduce a notion related to the U^d measures of pseudorandomness. For 2^d functions $\{f_S\}_{S \subseteq [d]}$, $f_S : G \rightarrow \mathbb{R}$, their *Gowers inner product* is defined as

$$\langle \{f_S\} \rangle_{U^d} := \mathop{\mathbb{E}}_{x, x_1, \ldots, x_d} \left[\prod_{S \subseteq [d]} f_S \left(x + \sum_{i \in S} x_i \right) \right]$$

In the case of complex-valued functions, the terms with odd $|S|$ are conjugated. Note that if all the functions f_S are identical to the same function f then $\langle \{f_S\} \rangle_{U^d} = U^d(f)$.

Our second result is that if $f : G_1 \times \cdots \times G_n \rightarrow \mathbb{C}$ are bounded functions such that at least one of them is balanced, and $\langle \{f_S\} \rangle_{U^d} \geq \varepsilon$, then there is a variable that has influence at least $\varepsilon^2 / 2^{O(d)}$ for at least *four* of the functions in the collection.

---

[3] The "globality" of this result deteriorates with the order of the group, and a result as stated is false for, say, $\mathbb{Z}_N$, N prime.

Finally, we come back to the analysis of the hypergraph test. Håstad and Wigderson [HW03] significantly simplified the analysis of the graph test of [ST00] by using an averaging argument that reduces the analysis of the graph test to the analysis of the 3-functions BLR test.

We apply a similar averaging argument and reduce the analysis of the hypergraph test to the task of bounding expressions of the form

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_d} \prod_{S\subseteq[d]} f_S\left(\sum_i x_i\right) \tag{1}$$

where $d \leq k$ and the functions $f_S : \{0,1\}^n \to \{-1,1\}$ are derived in a certain way from the function being tested. The expression looks very similar to the expression for the Gowers inner product, and in fact it is easy to see that if the Expression in (1) is large, then the Gowers inner product of a related set of functions is also large. From our results, it follows that at least four of these new functions must share an influential variable, from which it follows that two of f_S must share an influential variable and so the function being tested has an influential variable. (See Section 7.)

This analysis easily extends to the case of testing multiple functions and to the noisy case. We present the analysis of the noisy, multi-function case in Section 8.

This leads, under the unique games conjecture, to a PCP characterization of NP with q queries and error probability $(q+1)/2^q$, if q is of the form $2^k - 1$. As corollary, we obtain $(q+1)/2^q$ hardness of approximation for $qCSP$ and $(\text{poly}\log D)/D$ hardness of approximation for independent set in graphs of maximum degree D. In terms of amortized query complexity, we show that $1 + (1+o(1))\frac{\log q}{q}$ is achievable assuming the unique games conjecture, and Hast's algorithm [Has05] implies that this is tight up to the lower order term.

1.4 Organization of the paper

We develop the basic theory of influence of variables and Gowers uniformity for the case $G = \mathbb{Z}_2$ in Section 2. We prove our connection between influence of variables and Gowers uniformity in the case $G = \mathbb{Z}_2^n$ in Sections 3, 4, 5. Section 6 shows how to generalize our results to the case of general abelian groups; only the result of Section 3 requires a new proof, while the other results follow by making syntactic changes in the proof for the boolean case. An application to generalized linearity testing is presented in Section 7, together with lower bounds establishing the tightness of our analysis. An application to PCP is presented in Section 8.

The paper ends up being quite long because, for the sake of readability, we first prove some results in important special cases before proceeding to the more general statements. A reader interested only in the PCP application may skip Section 6. A reader interested only in our results on Gowers uniformity may skip Sections 7 and 8.

2 Preliminaries

In this section we develop the theory of Gowers uniformity and Fourier transforms for functions $f : \mathbb{Z}_2^n \to \mathbb{R}$. Generalized definitions for the setting of functions $f : G_1 \times \cdots \times G_n \to \mathbb{C}$ will be given later in Section 6.

For a positive integer n, we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. We use the convention that $[0]$ is the empty set.

For two reals $a < b$ we use $[a, b]$ to denote the interval $\{x \in \mathbb{R} : a \leq x \leq b\}$.

2.1   Fourier Analysis

For a subset $S \subseteq [n]$, define the function $\chi_S : \{0,1\}^n \to \mathbb{R}$ as

$$\chi_S(x_1, \ldots, x_n) = (-1)^{\sum_{i \in S} x_i}$$

We use the convention that an empty sum equals zero, so $\chi_\emptyset(x) = 1$ for every x.

For two functions $f, g : \{0,1\}^n \to \mathbb{R}$, define their inner product as

$$\langle f, g \rangle := \mathop{\mathbb{E}}_{x \in \{0,1\}^n} f(x)g(x)$$

Then it is easy to see that the functions χ_S are *orthonormal* with respect to this inner product, that is, for every S,

$$\langle \chi_S, \chi_S \rangle = 1$$

and, for every $S \neq T$,

$$\langle \chi_S, \chi_T \rangle = 0$$

This implies that the functions χ_S are linearly independent and, since there are 2^n such functions and the space of functions $f : \{0,1\}^n \to \mathbb{R}$ has dimension 2^n, it follows that the χ_S are a *basis* for such space.

Every function $f : \{0,1\}^n \to \mathbb{R}$ can therefore be written as

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S(x)$$

where the coefficients $\hat{f}(S)$ in the linear combination satisfy

$$\hat{f}(S) = \langle f, \chi_S \rangle$$

and are called the *Fourier coefficients* of f. The function $\hat{f}()$ mapping sets S into coefficients $\hat{f}(S)$ is the *Fourier transform* of f. We will make use of the following equality, called *Parseval's identity* (or *Plancherel's identity*).

$$\sum_{S} \hat{f}^2(S) = \mathop{\mathbb{E}}_{x} f^2(x) \tag{2}$$

In particular, if $f : \{0,1\}^n \to [-1, 1]$, then $\sum_S \hat{f}^2(S) \leq 1$, and so $|\hat{f}(S)| \leq 1$ for every S.

2.2 Influence of Variables

If $f : \{0,1\}^n \to \{-1,1\}$ is a boolean function, then the influence of i for f is defined as

$$\text{Inf}_i(f) := \Pr_x[f(x) \neq f(x + e_i)] \tag{3}$$

where $e_i \in \{0,1\}^n$ is the vector that has a 1 in the i-th position and zeroes everywhere else. That is, $\text{Inf}_i(f)$ is the probability that, if we pick a random x, the value of f at x is different from the value of f at the point that we get by flipping the i-th bit of f.

It is easy that see that $\text{Inf}_i(f)$ satisfies the following identities.

$$\text{Inf}_i(f) = \frac{1}{4} \mathbb{E}_x(f(x) - f(x + e_i))^2 = \sum_{S : i \in S} \hat{f}^2(S) \tag{4}$$

For a general real-valued function $f : \{0,1\}^n \to \mathbb{R}$, we will define the influence of i for f as

$$\text{Inf}_i(f) := \frac{1}{4} \mathbb{E}_x(f(x) - f(x + e_i))^2 = \sum_{S : i \in S} \hat{f}^2(S)$$

Note that Equation 3 is not valid if f is a general real-valued function.

We will make use of the following simple fact, that is valid for every function $f : \{0,1\}^n \to \mathbb{R}$: if $S \neq \emptyset$ and $i \in S$, then

$$\text{Inf}_i(f) \geq \hat{f}^2(S)$$

and, in particular, for every $f : \{0,1\}^n \to \mathbb{R}$,

$$\max_i \text{Inf}_i(f) \geq \max_{S \neq \emptyset} \hat{f}^2(S) \tag{5}$$

For boolean functions $f : \{0,1\}^n \to \{-1,1\}$, the stronger inequality $\max_i \text{Inf}_i(f) \geq \max_{S \neq \emptyset} |\hat{f}(S)|$ holds, but we will not use it.

For a function $f : \{0,1\}^n \to \mathbb{R}$, a threshold $d \leq n$, and an index $i \in [n]$, we define the *degree-d influence of i* for f as

$$\text{Inf}_i^{\leq d}(f) := \sum_{S : i \in S, |S| \leq d} \hat{f}^2(S) \ .$$

We will make use of the following simple fact: if f is a bounded function, then not too many variables can have large low-degree influence. Specifically:

$$\sum_{i=1}^n \text{Inf}_i^{\leq d}(f) = \sum_{i=1}^n \sum_{S : i \in S, |S| \leq d} \hat{f}^2(S) = \sum_{S : i \in S, |S| \leq d} |S| \hat{f}^2(S) \leq d \sum_S \hat{f}^2(S)$$

In particular, if $f : \{0,1\}^n \to [-1,1]$, then $\sum_{i=1}^n \text{Inf}_i^{\leq t}(f) \leq d$, and so at most d/ε variables can have degree-d influence larger than ε.

2.3 Cross-Influence

For the application to PCP, the following definition will be useful. Let $\mathcal{F} = f_1, \ldots, f_k$ be a collection of (not necessarily distinct) functions $f_j : \{0, 1\}^n \to \mathbb{R}$. Then the *cross-influence* of i for $\mathcal{F}$ is defined as

$$\mathrm{XInf}_i(f_1 \ldots, f_k) := \max_{j \neq h} \min\{\mathrm{Inf}_i(f_j), \mathrm{Inf}_i(f_h)\} \tag{6}$$

In other words, $\mathrm{XInf}_i(f_1, \ldots, f_k) \geq \varepsilon$ if and only if there are at least two functions f_j, f_h, with $j \neq h$, such that $\mathrm{Inf}_i(f_j) \geq \varepsilon$ and $\mathrm{Inf}_i(f_h) \geq \varepsilon$. Conversely, $\mathrm{XInf}_i(f_1, \ldots, f_k) \leq \varepsilon$ if there is at most one function f_j such that $\mathrm{Inf}_i(f_j) > \varepsilon$.

More generally, for a threshold t, we define the *t-cross influence* of i for $\mathcal{F}$ as

$$t\mathrm{XInf}_i(f_1, \ldots, f_k) := \max_{j_1, \ldots, j_t \in [k], \text{ all distinct}} \min\{\mathrm{Inf}_i(f_{j_1}), \ldots, \mathrm{Inf}_i(f_{j_t})\} \tag{7}$$

That is, $t\mathrm{XInf}_i(\mathcal{F})$ is the smallest ε such that there are at least t functions $f_{j_1}, \ldots, f_{j_t}$ in $\mathcal{F}$ such that coordinate i has influence at least ε for all of them. Note that $\mathrm{XInf}(\mathcal{F}) = 2\mathrm{XInf}(\mathcal{F})$.

If all the functions f_j are equal to the same function f, then $k\mathrm{XInf}_i(f_1, \ldots, f_k) = \mathrm{XInf}_i(f_1, \ldots, f_k) = \mathrm{Inf}_i(f)$.

Similarly, we define the *degree-d cross-influence* of $\mathcal{F}$ as

$$\mathrm{XInf}_i^{\leq d}(f_1 \ldots, f_k) := \max_{j \neq h} \min\{\mathrm{Inf}_i^{\leq d}(f_j), \mathrm{Inf}_i^{\leq d}(f_h)\} \tag{8}$$

2.4 Gowers Uniformity

Definition 1 (Gowers Uniformity) *Let $f : \{0, 1\}^n \to \mathbb{R}$ be a function, and $d \geq 1$ be an integer. The* dimension-d uniformity *of f is defined as*

$$U^d(f) := \mathop{\mathbb{E}}_{x, x_1, \ldots, x_d} \prod_{S \subseteq [d]} f\left(x + \sum_{i \in S} x_i\right)$$

Remark 1 *Here we use a terminology and notation that is a hybrid between the one of Gowers [Gow98, Gow01] and the one of Green and Tao [GT04, GT05]. What we call* dimension-d *uniformity is called* degree-$(d-1)$ *uniformity by Gowers, and no notation is introduced for it. Gowers also introduces the notation $||f||_d$, which equals, in our notation, $\left(U^d(f)\right)^{1/2^d}$. Gowers proves that $|| \cdot ||_d$ is a norm, and he does not give it a name. Green and Tao use the notation $||f||_{U^d}$ for $\left(U^d(f)\right)^{1/2^d}$, and call it the* (dimension-d) Gowers norm.

Here are expressions for the first few values of d:

$$
\begin{aligned}
U^1(f) &= \operatorname*{\mathbb{E}}_{x,y} f(x)f(x+y) = \left(\operatorname*{\mathbb{E}}_{x} f(x) \right)^2 \\
U^2(f) &= \operatorname*{\mathbb{E}}_{x,y,z} f(x)f(x+y)f(x+z)f(x+y+z) \\
&= \operatorname*{\mathbb{E}}_{y} \left(\operatorname*{\mathbb{E}}_{x} f(x)f(x+y) \right)^2 \\
U^3(f) &= \operatorname*{\mathbb{E}}_{x,y,z,w} f(x)f(x+y)f(x+z)f(x+y+z)f(x+w)f(x+y+w)f(x+z+w)f(x+y+z+w) \\
&= \operatorname*{\mathbb{E}}_{y,z} \left(\operatorname*{\mathbb{E}}_{x} f(x)f(x+y)f(x+z)f(x+y+z) \right)^2
\end{aligned}
$$

The above examples suggest the use of the following notation.

For a function $f : \{0,1\}^n \to \mathbb{R}$ and elements $x_1, \ldots, x_d \in \{0,1\}^n$, define

$$
f_{x_1,\ldots,x_d}(x) := \prod_{S \subseteq [d]} f\left(x + \sum_{i \in S} x_i \right)
$$

Then we have

$$
U^d(f) = \operatorname*{\mathbb{E}}_{x,x_1,\ldots,x_d} f_{x_1,\ldots,x_d}(x) = \operatorname*{\mathbb{E}}_{x_1,\ldots,x_{d-1}} \left(\operatorname*{\mathbb{E}}_{x} f_{x_1,\ldots,x_{d-1}}(x) \right)^2
$$

Define the *dimension-d Gowers inner product* of a collection $\{f_S\}_{S \subseteq [d]}$ of (not necessarily different) functions $f_S : \{0,1\}^n \to \mathbb{R}$, as

$$
\langle | \{f_S\} \rangle_{U^d} := \operatorname*{\mathbb{E}}_{x,x_1,\ldots,x_d} \left[\prod_{S \subseteq [d]} f_S\left(x + \sum_{i \in S} x_i \right) \right] \tag{9}
$$

Note that, in particular, if all the functions f_S are equal to the same function f then $\langle | \{f_S\} \rangle_{U^d} = U^d(f)$.

2.5 Unique Games

A unique game [Kho02] is a constraint satisfaction problem such that every constraint is of the form $y = f_{x,y}(x)$, where x, y are variables ranging over a finite set Σ, which we call the *alphabet*, specified as part of the input, and $f_{x,y} : \Sigma \to \Sigma$ is a permutation. Given a unique game, we are interested in finding the assignment of values to the variables that satisfies the largest number of constraints.

More formally, a unique game is a tuple $(V, E, \Sigma, \{f_{x,y}\}_{(x,y) \in E})$ where V is a set of variables, E is a set of pairs of variables (corresponding to constraints), and, for every $(x,y) \in E$, the function $f_{x,y} : \Sigma \to \Sigma$ is a permutation. Note that (V, E) is a graph, which we call the *constraint graph* of the unique game. We want to find an assignment $A : V \to \Sigma$ that maximizes the number of satisfied constraints, that is, the number of pairs $(x,y) \in E$ such that $A(y) = f_{x,y}(A(x))$. The *value*

of an assignment is the fraction of constraints satisfied by the assignment; the value of a unique game is the value of an optimum assignment.

For example, the following is a unique game with $V = \{v_1, v_2, v_3, v_4\}$ and $\Sigma = \{a, b, c\}$:

$$v_3 = \begin{pmatrix} a\ b\ c \\ c\ b\ a \end{pmatrix}(v_1)$$

$$v_3 = \begin{pmatrix} a\ b\ c \\ a\ c\ b \end{pmatrix}(v_2)$$

$$v_1 = v_2$$

$$v_4 = \begin{pmatrix} a\ b\ c \\ b\ c\ a \end{pmatrix}(v_2)$$

Where we use the notation $\begin{pmatrix} a & b & c \\ f(a) & f(b) & f(c) \end{pmatrix}$ to represent a function $f : \{a, b, c\} \to \{a, b, c\}$. The reader can verify that the value of the above unique game is $3/4$.

The *unique games conjecture* is that for every $\gamma > 0$ there is a $\sigma = \sigma(\gamma)$ such that it is NP-hard to distinguish unique games of value $\geq 1 - \gamma$ from unique games of value $\leq \gamma$, even when restricted to instances where $|\Sigma| \leq \sigma$ and where the constraint graph is bipartite.

For our application, we will need a variant of unique games, that we call d-ary unique game. In a d-ary unique game, a constraint is specified by a d-tuple $v_1 \ldots, v_d$ of variables and a d-tuple of permutations $f_1, \ldots, f_d : \Sigma \to \Sigma$. An assignment $A : V \to \Sigma$ *strongly satisfies* the constraint if $f_1(A(v_1)), \ldots, f_d(A(v_d))$ are all equal; an assignment *weakly satifies* the constraint it if $f_1(A(v_1)), \ldots, f_d(A(v_d))$ are not all different.

The following result is a rephrasing of a result by Khot and Regev [KR03].

Theorem 2 *If the unique games conjecture is true, then for every d and every γ there is a $\sigma = \sigma(d, \gamma)$ such that, given a d-ary unique game with alphabet size σ, it is NP-hard to distinguish the case in which there is an assignment that strongly satisfies at least a $1 - \gamma$ fraction of constraints from the case where every assignment weakly satisfies at most a γ fraction of constraints.*

We define the *strong value* of an assignment to a d-ary unique game as the fraction of constraints that are strongly satisfied by the assignment. The strong value of a unique game is the largest strong value among all assignments. The *weak value* of an assignment and of a unique game are similarly defined. Note that the weak value is always at least as large as the strong value.

3 Influence of Product of Functions

In this section we prove a bound on the influence of a function of the form $f(x) := f_1(x) \cdot f_2(x) \cdots f_k(x)$ in terms of the influence of the functions f_j. Such a bound will be very useful in the proofs of our main results.

In the boolean case, the bound is just a simple union bound.

Lemma 3 *Let $f_1, \ldots, f_k : \{0,1\}^n \to \{-1,1\}$ be boolean functions, and define $f(x) = f_1(x) \cdot f_2(x) \cdots f_k(x)$.*

Then, for every $i \in [n]$

$$\mathrm{Inf}_i(f) \leq \sum_j \mathrm{Inf}_i(f_j) \tag{10}$$

PROOF: Using the formula for the influence of boolean functions, we see that

$$
\begin{aligned}
\mathrm{Inf}_i(f) &= \Pr_x[f(x) \neq f(x + e_i)] \\
&= \Pr_x[f_1(x) \cdots f_k(x) \neq f_1(x + e_i) \cdots f_k(x + e_i)] \\
&\leq \sum_{j=1}^k \Pr_x[f_j(x) \neq f_j(x + e_i)] \\
&= \sum_j \mathrm{Inf}_i(f_j)
\end{aligned}
$$

□

For general real-valued functions, we cannot hope to achieve the nice bound of Equation 10. Suppose, for example, that $n = 1$ and that all functions f_j are defined as follows: $f_j(0) = 1 - \varepsilon$, $f_j(1) = 1$.

Then, we have $I_1(f_j) = \frac{1}{4}\varepsilon^2$. When we define $f(x) := \prod_j f_j(x)$, we get $f(0) = (1-\varepsilon)^k$ and $f(1) = 1$, and so $I_i(f) = \frac{1}{4}(1 - (1 - \varepsilon)^k)^2$, which is about $\frac{1}{4}k^2\varepsilon^2$ for small ε, or about k times the sum of the influences of the functions f_j. The following Lemma achieves such a tight bound.

Lemma 4 *Let $f_i : \{0,1\}^n \to [-1,1]$ be k functions, and define*

$$f(x) := f_1(x) \cdots f_k(x)$$

Then, for every $i \in [n]$, $\mathrm{Inf}_i(f) \leq k \cdot \sum_j \mathrm{Inf}_i(f_j)$

PROOF: We begin by proving the following claim:

$$\forall a_1, \ldots, a_k, b_1, \ldots, b_k \in [-1,1]. \quad \left| \prod_i a_i - \prod_i b_i \right| \leq \sum_i |a_i - b_i| \tag{11}$$

We prove Inequality 11 by first expressing the right-hand side as a telescoping sum of a sequence of "hybrids," and then by using the triangle inequality and the fact that all a_i and b_i have absolute value at most 1.

$$
\begin{aligned}
&|a_1 \cdot a_2 \cdots a_k - b_1 \cdot b_2 \cdots b_k| \\
=\ & |(a_1 \cdot a_2 \cdots a_k) - (b_1 \cdot a_2 \cdots a_k) + (b_1 \cdot a_2 \cdots a_k) - (b_1 \cdot b_2 \cdot a_3 \cdots a_k) + (b_1 \cdot b_2 \cdot a_3 \cdots a_k) - \\
& \cdots + (b_1 \cdots b_{k-1} \cdot a_k) - (b_1 \cdots b_k)| \\
=\ & |(a_2 \cdots a_k) \cdot (a_1 - b_1) + (b_1 \cdot a_3 \cdots a_k) \cdot (a_2 - b_2) + \cdots + (b_1 \cdots b_{k-1}) \cdot (a_k - b_k)| \\
\leq\ & |a_2 \cdots a_k| \cdot |a_1 - b_1| + |b_1 \cdot a_3 \cdots a_k| \cdot |a_2 - b_2| + \cdots + |b_1 \cdots b_{k-1}| \cdot |a_k - b_k| \\
\leq\ & |a_1 - b_1| + |a_2 - b_2| + \cdots + |a_k - b_k|
\end{aligned}
$$

If we square both sides and apply Cauchy-Schwartz, we get

$$\left(\prod_i a_i - \prod_i b_i \right)^2 \leq \left(\sum_i |a_i - b_i| \right)^2$$
$$\leq k \cdot \sum_i (a_i - b_i)^2$$

To summarize our progress so far, we have proved the following claim:

$$\forall a_1, \ldots, a_k, b_1, \ldots, b_k \in [-1, 1]. \quad \left(\prod_i a_i - \prod_i b_i \right)^2 \leq k \cdot \sum_i (a_i - b_i)^2 \tag{12}$$

We are now ready to prove the Lemma. For every x, using Inequality 12, we have

$$(f(x) - f(x + e_i))^2 = \left(\prod_i f_i(x) - \prod_i f_i(x + e_i) \right)^2 \leq k \cdot \sum_j (f_j(x) - f_j(x + e_i))^2$$

and of course the same inequality remains valid if we take the average over x, so we have

$$\text{Inf}_i(f) = \frac{1}{4} \mathbb{E}_x (f(x) - f(x + e_i))^2 \leq k \frac{1}{4} \mathbb{E}_x \sum_j (f_j(x) - f_j(x + e_i))^2 \leq k \sum_j \text{Inf}_i(f_j)$$

$\square$

4   Low Influence Implies Small Gowers Uniformity

We are going to show that, for balanced functions $f : \{0,1\}^n \to \{-1,1\}$, if $U^d(f)$ is large, then one of the variables of f has high influence.

Lemma 5 *Let $f : \{0,1\}^n \to \{-1,1\}$ be a function and $d \geq 1$ be an integer. Then*

$$U^d(f) \leq U^1(f) + (2^{d-1} - 1) \max_i \text{Inf}_i(f)$$

PROOF:The case $d = 1$ is trivial. Let $d \geq 2$, and define $I := \max_i \text{Inf}_i(f)$. We will prove

$$U^d(f) \leq U^{d-1}(f) + 2^{d-2}I \tag{13}$$

which immediately implies the statement of the lemma.
We write

$$U^d(f) = \mathbb{E}_{x_1,\ldots,x_{d-2}} \left[\mathbb{E}_{x,y,z} f_{x_1,\ldots,x_{d-2}}(x) f_{x_1,\ldots,x_{d-2}}(x+y) f_{x_1,\ldots,x_{d-2}}(x+z) f_{x_1,\ldots,x_{d-2}}(x+y+z) \right]$$

$$= \mathbb{E}_{x_1,\ldots,x_{d-2}} \sum_\alpha \hat{f}^4_{x_1,\ldots,x_{d-2}}(\alpha)$$

15

$$= \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \sum_{\alpha \neq \emptyset} \hat{f}^4_{x_1,\ldots,x_{d-2}}(\alpha) + \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \hat{f}^4_{x_1,\ldots,x_{d-2}}(\emptyset)$$

We separately bound the two terms in the last expression.

For every $a_1,\ldots,a_{d-2} \in \{0,1\}^n$, using Lemma 3, we get that, for every $i \in \alpha$, $\mathrm{Inf}_i(f_{a_1,\ldots,a_{d-2}}) \leq 2^{d-2}\mathrm{Inf}_i(f) \leq 2^{d-2}I$ and so we have that, for every $a_1,\ldots,a_{d-2} \in \{0,1\}^n$,

$$\sum_{\alpha \neq \emptyset} \hat{f}^4_{a_1,\ldots,a_{d-2}}(\alpha) \leq \max_{\alpha \neq \emptyset} \hat{f}^2_{a_1,\ldots,a_{d-2}}(\alpha) \leq \max_i \mathrm{Inf}_i(f_{a_1,\ldots,a_{d-2}}) \leq 2^{d-2}I$$

and so

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \sum_{\alpha \neq \emptyset} \hat{f}^4_{x_1,\ldots,x_{d-2}}(\alpha) \leq 2^{d-2}I$$

Regarding the other term,

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \hat{f}^4_{x_1,\ldots,x_{d-2}}(\emptyset) \leq \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \hat{f}^2_{x_1,\ldots,x_{d-2}}(\emptyset) = \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \left(\mathop{\mathbb{E}}_x f_{x_1,\ldots,x_{d-2}}(x) \right)^2 = U^{d-1}(f)$$

$\square$

The same argument also applies to general bounded real-valued functions.

Lemma 6 *Let $f : \{0,1\}^n \to [-1,1]$ be a function and $d \geq 1$ be an integer. Then*

$$U^d(f) \leq U^1(f) + 4^d \max_i \mathrm{Inf}_i(f)$$

PROOF: Define $I := \max_i \mathrm{Inf}_i(f)$. It suffices to prove that, for $d \geq 2$,

$$U^d(f) \leq U^{d-1}(f) + 2^{2d-4} \max_i \mathrm{Inf}_i(f) \tag{14}$$

We repeat the proof of Lemma 5 verbatim, except that we use Lemma 4 instead of Lemma 3 to get an upper bound for $\mathrm{Inf}_i(f_{a_1,\ldots,a_{d-2}})$. Because of the worse bound in Lemma 4, we only get the bound

$$\mathrm{Inf}_i(f_{a_1,\ldots,a_{d-2}}) \leq 2^{2d-4}\mathrm{Inf}_i(f) \leq 2^{2d-4}I$$

and the rest of the proof proceeds with no change, except for the term 2^{2d-4} instead of 2^{d-2}. $\square$

We remark that our bound for the boolean case is nearly tight.

Lemma 7 *For every fixed $d \geq 2$ and large n, there is a function $f : \{0,1\}^n \to \{-1,1\}$ such that $U^1(f) = o_n(1)$, $U^d(f) = 1$ and $\max_i \mathrm{Inf}_i(f) \leq \frac{1}{2^{d-2}} + o_n(1)$.*

PROOF: Consider the function

$$f(x_1,\ldots,x_n) := (-1)^{x_1 x_2 \cdots x_{d-1} + x_d \cdots x_{2d-2} + \cdots}$$

$\square$

5 Low Cross-Influence Implies Small Gowers Inner Product

The main result of this section is that if a collection of functions has small cross-influence, then it has small Gowers inner product, provided that at least one of the functions is balanced.

Lemma 8 *For $d \geq 2$, let $\{f_S\}_{S \subseteq [d]}$ be a collection of functions $f_S : \{0,1\}^n \to [-1,1]$ such that*

- *$f_{[d]}$ is balanced, that is, $\mathbb{E}_x f_{[d]}(x) = 0$;*

- *$4\mathrm{XInf}_i(\{f_S\}) \leq \varepsilon$ for every i.*

Then

$$\langle \{f_S\} \rangle_{U^d} \leq \sqrt{\varepsilon} \cdot 2^{O(d)}$$

Before proving Lemma 8, we establish a variant of a result of Aumann et al. [AHRS01] which will be useful in the inductive step of the proof of Lemma 8.

Lemma 9 *For every four bounded functions $f_1, f_2, f_3, f_4 : \{0,1\}^n \to [-1,1]$,*

$$\left| \sum_\alpha \hat{f}_1(\alpha) \hat{f}_2(\alpha) \hat{f}_3(\alpha) \hat{f}_4(\alpha) \right| \leq 4 \max_\alpha \min\{ |\hat{f}_1(\alpha)|, |\hat{f}_2(\alpha)|, |\hat{f}_3(\alpha)|, |\hat{f}_4(\alpha)| \}$$

PROOF: Let

$$\varepsilon := \max_\alpha \min\{ |\hat{f}_1(\alpha)|, |\hat{f}_2(\alpha)|, |\hat{f}_3(\alpha)|, |\hat{f}_4(\alpha)| \}$$

For $i = 1, \ldots, 4$, let S_i be the family of all sets α such that $|\hat{f}_i(\alpha)| \leq \varepsilon$. By definition, the union of the families S_i contains all subsets $\alpha \subseteq [n]$. We can then write

$$
\begin{aligned}
& \left| \sum_\alpha \hat{f}_1(\alpha) \hat{f}_2(\alpha) \hat{f}_3(\alpha) \hat{f}_4(\alpha) \right| \\
\leq\ & \sum_\alpha |\hat{f}_1(\alpha)| \cdot |\hat{f}_2(\alpha)| \cdot |\hat{f}_3(\alpha)| \cdot |\hat{f}_4(\alpha)| \\
\leq\ & \sum_{i=1}^{4} \sum_{\alpha \in S_i} |\hat{f}_1(\alpha)| \cdot |\hat{f}_2(\alpha)| \cdot |\hat{f}_3(\alpha)| \cdot |\hat{f}_4(\alpha)| \\
<\ & \varepsilon \left(\sum_{\alpha \in S_1} |\hat{f}_2(\alpha)| \cdot |\hat{f}_3(\alpha)| \cdot |\hat{f}_4(\alpha)| + \sum_{\alpha \in S_2} |\hat{f}_1(\alpha)| \cdot |\hat{f}_3(\alpha)| \cdot |\hat{f}_4(\alpha)| \right. \\
& \left. + \sum_{\alpha \in S_3} |\hat{f}_1(\alpha)| \cdot |\hat{f}_2(\alpha)| \cdot |\hat{f}_4(\alpha)| + \sum_{\alpha \in S_4} |\hat{f}_1(\alpha)| \cdot |\hat{f}_2(\alpha)| \cdot |\hat{f}_3(\alpha)| \right) \\
\leq\ & 4\varepsilon
\end{aligned}
$$

The last inequality follows from the fact that for every three functions $f, g, h : \{0, 1\}^n \to [-1, 1]$ we have

$$\sum_\alpha |\hat{f}(\alpha)||\hat{g}(\alpha)||\hat{h}(\alpha)| \leq \sum_\alpha |\hat{f}(\alpha)||\hat{g}(\alpha)| \leq \sqrt{\sum_\alpha \hat{f}^2(\alpha)} \sqrt{\sum_\alpha \hat{g}^2(\alpha)} \leq 1$$

□

We proceed with the proof of our main result of this section.

PROOF:[Of Lemma 8] We want to prove

$$\langle\{f_S\}\rangle_{U^d} \leq \tau(\varepsilon, d) \tag{15}$$

for a function $\tau(\varepsilon, d) = \sqrt{\varepsilon} \cdot 2^{O(d)}$ that we specify later.

We proceed by induction on d.

The case $d = 1$

For $d = 1$, we have two functions $f_\emptyset, f_{\{1\}}$ such that $\mathbb{E}_x f_{\{1\}}(x) = 0$ and we want an upper bound to $\langle\{f_S\}\rangle_{U^1}$. We see that

$$\langle f_\emptyset, f_{\{1\}}\rangle_{U^1} = \left(\mathbb{E}_x f_\emptyset(x)\right) \cdot \left(\mathbb{E}_y f_{\{1\}}(y)\right) = 0$$

We have proved the base case of Equation 15 with $\tau(\varepsilon, 1) = 0$.

The inductive step

Suppose now that, for $d \geq 1$, the lemma is true up to dimension d, and we want to prove it for dimension $d + 1$.

We have 2^{d+1} functions $\{f_S\}_{S \subseteq [d+1]}$ and we want to upper bound the Gowers inner product

$$\langle\{f_S\}\rangle_{U^{d+1}} = \mathbb{E}_{x, x_1, \ldots, x_{d+1}} \prod_{S \subseteq [d+1]} f_S \left(x + \sum_{i \in S} x_i\right) \tag{16}$$

For every $x_1, \ldots, x_{d-1}$, define the four functions[4]

$$A_{x_1, \ldots, x_{d-1}}(x) := \prod_{S \subseteq [d-1]} f_S \left(x + \sum_{i \in S} x_i\right)$$

$$B_{x_1, \ldots, x_{d-1}}(x) := \prod_{S \subseteq [d-1]} f_{S \cup \{d\}} \left(x + \sum_{i \in S} x_i\right)$$

---

[4]The case $d = 1$ is somewhat degenerate: $[d - 1]$ is the empty set, and $x_1, \ldots, x_{d-1}$ is an empty sequence. So we simply have $A := f_\emptyset$, $B := f_{\{1\}}$, $C := f_{\{2\}}$ and $D := f_{\{1,2\}}$, with no subscripts.

$$C_{x_1,\ldots,x_{d-1}}(x) := \prod_{S \subseteq [d-1]} f_{S \cup \{d+1\}}\left(x + \sum_{i \in S} x_i\right)$$

$$D_{x_1,\ldots,x_{d-1}}(x) := \prod_{S \subseteq [d-1]} f_{S \cup \{d,d+1\}}\left(x + \sum_{i \in S} x_i\right)$$

with this notation, we can rewrite the expression (16) as

$$\langle \{f_S\} \rangle_{U^{d+1}} = \mathop{\mathbb{E}}_{x,x_1,\ldots,x_{d+1}} A_{x_1,\ldots,x_{d-1}}(x) B_{x_1,\ldots,x_{d-1}}(x+x_d) C_{x_1,\ldots,x_{d-1}}(x+x_{d+1}) D_{x_1,\ldots,x_{d-1}}(x+x_d+x_{d+1})$$

and using the Fourier expansion and simplifying,

$$
\begin{aligned}
\langle \{f_S\} \rangle_{U^{d+1}} &= \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-1}} \sum_{\alpha} \hat{A}_{x_1,\ldots,x_{d-1}}(\alpha) \hat{B}_{x_1,\ldots,x_{d-1}}(\alpha) \hat{C}_{x_1,\ldots,x_{d-1}}(\alpha) \hat{D}_{x_1,\ldots,x_{d-1}}(\alpha) \\
&= \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-1}} \sum_{\alpha \neq \emptyset} \hat{A}_{x_1,\ldots,x_{d-1}}(\alpha) \hat{B}_{x_1,\ldots,x_{d-1}}(\alpha) \hat{C}_{x_1,\ldots,x_{d-1}}(\alpha) \hat{D}_{x_1,\ldots,x_{d-1}}(\alpha) \\
&\quad + \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-1}} \hat{A}_{x_1,\ldots,x_{d-1}}(\emptyset) \hat{B}_{x_1,\ldots,x_{d-1}}(\emptyset) \hat{C}_{x_1,\ldots,x_{d-1}}(\emptyset) \hat{D}_{x_1,\ldots,x_{d-1}}(\emptyset)
\end{aligned}
$$

We bound the two terms separately.

For the first term, we have that, for every fixed $\mathbf{x} = (x_1, \ldots, x_{d-1})$,

$$
\begin{aligned}
&\sum_{\alpha \neq \emptyset} \hat{A}_{\mathbf{x}}(\alpha) \hat{B}_{\mathbf{x}}(\alpha) \hat{C}_{\mathbf{x}}(\alpha) \hat{D}_{\mathbf{x}}(\alpha) \\
\leq\ & 4 \max_{\alpha \neq \emptyset} \min\{|\hat{A}_{\mathbf{x}}(\alpha)|, |\hat{B}_{\mathbf{x}}(\alpha)|, |\hat{C}_{\mathbf{x}}(\alpha)|, |\hat{D}_{\mathbf{x}}(\alpha)|\} \\
\leq\ & 4 \sqrt{\max_i \min\{\mathrm{Inf}_i(A_{\mathbf{x}}), \mathrm{Inf}_i(B_{\mathbf{x}}), \mathrm{Inf}_i(C_{\mathbf{x}}), \mathrm{Inf}_i(D_{\mathbf{x}})\}}
\end{aligned}
$$

And we observe that

$$\min\{\mathrm{Inf}_i(A_{\mathbf{x}}), \mathrm{Inf}_i(B_{\mathbf{x}}), \mathrm{Inf}_i(C_{\mathbf{x}}), \mathrm{Inf}_i(D_{\mathbf{x}})\} \leq 2^{2d-2} \varepsilon \tag{17}$$

by using Lemma 4. To verify this claim,[5] let δ be the minimum in the above expression. Then $\mathrm{Inf}_i(A_{\mathbf{x}}) \geq \delta$; recall $A_{\mathbf{x}}$ is defined as a product of 2^{d-1} functions of the form $f_S(x + \sum_S x_i)$, $S \subseteq [d-1]$, where we think of the x_i as constants, and so for at least one $S \subseteq [d-1]$ we have $\mathrm{Inf}_i(f_S) \geq \delta/2^{2d-2}$. We argue similarly for $B_{\mathbf{x}}$, $C_{\mathbf{x}}$ and $D_{\mathbf{x}}$, and we find three more functions for

---

[5]Again, the $d = 1$ case is degenerate but easy to check: Equation 17 reduces to

$$\min\{\mathrm{Inf}_i(f_\emptyset), \mathrm{Inf}_i(f_{\{1\}}), \mathrm{Inf}_i(f_{\{2\}}), \mathrm{Inf}_i(f_{\{1,2\}})\} \leq \varepsilon$$

which is precisely our assumption that the 4-cross influence of the functions is at most ε.

which coordinate i has influence at most $\delta/2^{2d-2}$, and these four functions are distinct. Therefore, $\varepsilon \geq 4\mathrm{XInf}_i(\{f_S\}) \geq \delta/2^{2d-2}$.

So we have that, for every fixed $\mathbf{x} = (x_1, \ldots, x_{d-1})$

$$\sum_{\alpha \neq \emptyset} \hat{A}_{\mathbf{x}}(\alpha)\hat{B}_{\mathbf{x}}(\alpha)\hat{C}_{\mathbf{x}}(\alpha)\hat{D}_{\mathbf{x}}(\alpha) \leq 4\sqrt{2^{2d-2}\varepsilon}$$

and so, in particular

$$\mathbb{E}_{\mathbf{x}} \sum_{\alpha \neq \emptyset} \hat{A}_{\mathbf{x}}(\alpha)\hat{B}_{\mathbf{x}}(\alpha)\hat{C}_{\mathbf{x}}(\alpha)\hat{D}_{\mathbf{x}}(\alpha) \leq 4\sqrt{2^{2d-2}\varepsilon}$$

The second term can be written as

$$\mathbb{E}_{\mathbf{x}} \hat{A}_{\mathbf{x}}(\emptyset)\hat{B}_{\mathbf{x}}(\emptyset)\hat{C}_{\mathbf{x}}(\emptyset)\hat{D}_{\mathbf{x}}(\emptyset)$$

$$= \mathbb{E}_{x,\mathbf{x},y,z,w} A_{\mathbf{x}}(x)B_{\mathbf{x}}(x+y)C_{\mathbf{x}}(x+z)D_{\mathbf{x}}(x+w) \tag{18}$$

$$\leq \mathbb{E}_{x,\mathbf{x},w} A_{\mathbf{x}}(x)B_{\mathbf{x}}(x+a)C_{\mathbf{x}}(x+b)D_{\mathbf{x}}(x+w) \tag{19}$$

where a, b are values for y, z that maximize the expectation in Equation 18.

If we expand the Expression in (19), we get

$$\mathbb{E}_{x,x_1,\ldots,x_{d-1},w} \prod_{S \subseteq [d-1]} \begin{array}{l} f_S\left(x + \sum_{i \in S} x_i\right) f_{S \cup \{d\}}\left(x + a + \sum_{i \in S} x_i\right) \\ f_{S \cup \{d+1\}}\left(x + b + \sum_{i \in S} x_i\right) f_{S \cup \{d,d+1\}}\left(x + w + \sum_{i \in S} x_i\right) \end{array} \tag{20}$$

which we are going to re-write as the d-dimensional Gowers inner product of a new set of functions, so that we can invoke the inductive hypothesis. For every $S \subseteq [d-1]$, define

$$g_S(x) := f_S(x) \cdot f_{S \cup \{d\}}(x+a) \cdot f_{S \cup \{d+1\}}(x+b)$$

and define

$$g_{S \cup \{d\}}(x) := f_{S \cup \{d,d+1\}}(x)$$

Expression (20) becomes

$$\mathbb{E}_{x,x_1,\ldots,x_{d-1},w} \prod_{S \subseteq [d-1]} g_S\left(x + \sum_{\in S} x_i\right) \cdot g_{S \cup \{d\}}\left(x + \sum_{i \in S} x_i + w\right)$$

$$= \mathbb{E}_{x,x_1,\ldots,x_{d-1},x_d} \prod_{S \subseteq [d]} g_S\left(x + \sum_{\in S} x_i\right) = \langle\{g_S\}\rangle_{U^d}$$

after the change of variable $w \to x_d$.

By definition, the function $g_{[d]} = f_{[d+1]}$ is balanced, and, by construction and by Lemma 4, the 4-cross-influence of the functions g_S is at most 9ε. So we have $U^d(\{g_S\}) \leq \tau(9\varepsilon, d)$.

We have thus solved the case of dimension $d + 1$, with

$$\tau(\varepsilon, d + 1) = 2^{d+1} \cdot \sqrt{\varepsilon} + \tau(9\varepsilon, d)$$

Together with the base case

$$\tau(\varepsilon, 1) = 0$$

the recursion gives $\tau(\varepsilon, d) = 2^{O(d)} \cdot \sqrt{\varepsilon}$. $\square$

6 Generalizing to Complex-Valued Functions on Arbitrary Groups

In this section we generalize our result to the setting of functions $f : G_1 \times \cdots \times G_n \to \mathbb{C}$ where each G_i is a finite abelian group.

We fix a group $G = G_1 \times \cdots \times G_n$ for the rest of this section.

We write group operations in G and in the groups G_i additively. We will recover the results of the previous sections when each G_i is $\mathbb{Z}_2$. We denote an element of G as a tuple $\mathbf{g} = (g_1, \ldots, g_n)$ where g_i is an element of G_i. The zero element of G is the tuple $\mathbf{0} = (0, \ldots, 0)$.

If $z = a + bi$ is a complex number, we define its conjugate $\overline{z} := a - bi$ and its absolute value $|z| := \sqrt{a^2 + b^2} = \sqrt{z \cdot \overline{z}}$.

6.1 Complex Random Variables

If X is a random variable that takes on finitely many complex values, and $\mu(x)$ is the probability that X takes value x, then the average of X is the complex number

$$\mathbb{E}[X] := \sum_x \mu(x) \cdot x$$

and the variance of X is the *real* number

$$\mathbf{Var}[X] := \mathbb{E}[|X - \mathbb{E}[X]|^2] .$$

6.2 Fourier Analysis

A function $\chi : G_0 \to \mathbb{C}$ is a *character* of a group G_0 if $\chi(0) = 1$ and $\chi(a + b) = \chi(a)\chi(b)$ for every $a, b \in G_0$. We also have $\chi(-a) = \overline{\chi(a)}$.

It is well known that a finite abelian group G_0 has precisely $|G_0|$ characters, and that there is an isomorphism between G_0 and the set of characters that associates to each group element g precisely one character χ_g so that χ_0 is the constant 1 function and $\chi_g \chi_h = \chi_{g+h}$.

For each group G_i, let $\{\chi_g^i\}_{g \in G_i}$ be the characters of G_i, indexed according to the above isomorphism.

For each $\mathbf{g} = (g_1, \ldots, g_n)$ define the function $\chi_{g_1, \ldots, g_n} : G \to \mathbb{C}$

$$\chi_{g_1,\ldots,g_n}(x_1,\ldots,x_n) := \prod_{i\in[n]} \chi^i_{g_i}(x_i) \ .$$

Then the functions $\chi_{\mathbf{g}}$ are the characters of the group G.

Define the following inner product among functions $f, h : G \to \mathbb{C}$:

$$\langle f, h \rangle := \mathbb{E}_x f(x)\overline{h(x)}$$

Then one can verify that the functions $\chi_{\mathbf{g}}$ are orthonormal with respect to the inner product, and form a basis for the space of functions $f : G \to \mathbb{C}$. Every such function can be written as

$$f(x) = \sum_{\mathbf{g}\in G} \hat{f}(\mathbf{g})\chi_{\mathbf{g}}(x)$$

where $\hat{f}(\mathbf{g}) = \langle f, \chi_{\mathbf{g}} \rangle$.

We again have Placherel's identity

$$\mathbb{E}_x |f(x)|^2 = \sum_{\mathbf{g}} |\hat{f}(\mathbf{g})|^2$$

and we observe that

$$\hat{f}(\mathbf{0}) = \mathbb{E}_x f(x)$$

6.3 Influence of Variables

The following definitions could be given in a much more general setting, but the following will suffice for the purpose of this paper.

Let $\Sigma_1, \ldots, \Sigma_n$ be finite and $f : \Sigma_1 \times \cdots \times \Sigma_n \to \mathbb{C}$ be a function, then the *influence of i for f* is defined as

$$\mathrm{Inf}_i(f) := \mathbb{E}_{x_1,\ldots,x_{i-1},x_{x+1},\ldots,x_n} \left[\mathbf{Var}_{x_i}[f(x_1,\ldots,x_n)] \right]$$

where the variables x_i are mutually independent, and each x_i is uniformly distributed over Σ_i. The reader should verify that if each $\Sigma_i = \{0, 1\}$ and if f takes only real values, then we recover the definition we gave earlier. Note that the influence is always a non-negative real number.

Returning to our setting of functions $f : G_1 \times \cdots \times G_n \to \mathbb{C}$, we have the following fact:

$$\mathrm{Inf}_i(f) = \sum_{(g_1,\ldots,g_n):g_i\neq 0} |\hat{f}(g_1,\ldots,g_n)|^2$$

and, in particular,

$$\max_i \text{Inf}_i(f) \geq \max_{\mathbf{g} \neq \mathbf{0}} |\hat{f}(\mathbf{g})|^2$$

6.4   Influence of Products of Functions

The following result is the only real difficulty in generalizing our results from previous sections. (The rest just follows the same proofs with some changes in notation.)

Lemma 10 *Let $f, g : \Sigma_1 \times \cdots \times \Sigma_n \to \mathbb{C}$ be functions such that $|f(x)| \leq 1$ and $|g(x)| \leq 1$ for every x.*
Then, for every $i \in [n]$,
$$\text{Inf}_i(fg) \leq 3 \cdot (\text{Inf}_i(f) + \text{Inf}_i(g))$$

To prove the lemma, it is enough to prove the following bound

Lemma 11 *Let $f, g : \Sigma \to \mathbb{C}$ be functions such that $|f(x)| \leq 1$ and $|g(x)| \leq 1$ for every x.*
Then,
$$\mathbf{Var}_x[f(x)g(x)] \leq 3 \cdot (\mathbf{Var}_x[f(x)] + \mathbf{Var}_x[g(x)])$$

PROOF: Recall that $\mathbf{Var}_x[f(x)g(x)] := \mathbb{E}_x |f(x)g(x) - \mathbb{E}_y[f(y)g(y)]|^2$. For every x,

$$
\begin{aligned}
\left| f(x)g(x) - \mathbb{E}_y[f(y)g(y)] \right| &= \left| f(x)g(x) - \mathbb{E}_y[f(y)]g(x) + \mathbb{E}_y[f(y)]g(x) - \mathbb{E}_y[f(y)g(y)] \right| \\
&= \left| g(x) \cdot \left(f(x) - \mathbb{E}_y[f(y)] \right) + \mathbb{E}_y [f(y) \cdot (g(x) - g(y))] \right| \\
&\leq |g(x)| \cdot \left| f(x) - \mathbb{E}_y[f(y)] \right| + \mathbb{E}_y [|f(y)||g(x) - g(y)|] \\
&\leq \left| f(x) - \mathbb{E}_y[f(y)] \right| + \mathbb{E}_y [|f(y)| \, |g(x) - g(y)|]
\end{aligned}
$$

We can now compute the variance of fg:

$$
\begin{aligned}
\mathbf{Var}_x[f(x)g(x)] &:= \mathbb{E}_x \left| f(x)g(x) - \mathbb{E}_y[f(y)g(y)] \right|^2 \\
&\leq \mathbb{E}_x \left| \left| f(x) - \mathbb{E}_y[f(y)] \right| + \mathbb{E}_y [|f(y)| \, |g(x) - g(y)|] \right|^2 \\
&\leq 2 \mathbb{E}_x \left[\left| f(x) - \mathbb{E}_y[f(y)] \right|^2 + \left| \mathbb{E}_y [|f(y)| \, |g(x) - g(y)|] \right|^2 \right] \\
&= 2 \mathbb{E}_x \left| f(x) - \mathbb{E}_y[f(y)] \right|^2 + 2 \mathbb{E}_x \mathbb{E}_y \left[|f(y)|^2 |g(x) - g(y)|^2 \right] \\
&\leq 2 \mathbf{Var}_x[f(x)] + 4 \mathbf{Var}_x[g(x)]
\end{aligned}
$$

23

Where the last step follows from the fact that

$$\mathbb{E}_{x,y} |g(x) - g(y)|^2 = \mathbb{E}_{x,y} \left| g(x) - \mathbb{E}_z[g(z)] + \mathbb{E}_z[g(z)] - g(y) \right|^2 \le 2 \mathbb{E}_x \left| g(x) - \mathbb{E}_z[g(z)] \right|^2 + 2 \mathbb{E}_y \left| \mathbb{E}_z[g(z)] - g(y) \right|^2$$

Similarly, we could prove

$$\mathbf{Var}_x[f(x)g(x)] \le 4 \mathbf{Var}_x[f(x)] + 2 \mathbf{Var}_x[g(x)]$$

And the average of the two bonds gives us the desired result. $\square$

A simple induction shows that

$$\mathrm{Inf}_i(f_1 \cdots f_k) \le 3 \cdot k^{\log_2 3} \cdot \sum_j \mathrm{Inf}_i(f_j)$$

6.5 Gowers Uniformity

The definition of U^d for complex-valued functions is as follows: for $d \ge 1$ and $f : G \to \mathbb{C}$,

$$U^d(f) := \mathbb{E}_{x,x_1,\ldots,x_d} \left(\prod_{S \subseteq [d], |S| \text{ even}} f\left(x + \sum_{i \in S} x_i \right) \right) \cdot \left(\prod_{S \subseteq [d], |S| \text{ odd}} \overline{f\left(x + \sum_{i \in S} x_i \right)} \right)$$

The inductive definition of U^d is perhaps simpler. For a function $f : G \to \mathbb{C}$ and elements $x_1, \ldots, x_d \in G$, define $f_{x_1,\ldots,x_d} : G \to \mathbb{C}$ inductively as follows:

$$f_{x_1,\ldots,x_d}(x) := f_{x_1,\ldots,x_{d-1}}(x) \overline{f_{x_1,\ldots,x_{d-1}}(x + x_d)}$$

and then define

$$U^d(f) := \mathbb{E}_{x,x_1,\ldots,x_d} f_{x_1,\ldots,x_d}(x)$$

It is possible to show that $U^d(f)$ is always a non-negative real number, because we have

$$
\begin{aligned}
U^d(f) &= \mathbb{E}_{x,x_1,\ldots,x_d} f_{x_1,\ldots,x_{d-1}}(x) \overline{f_{x_1,\ldots,x_{d-1}}(x + x_d)} \\
&= \mathbb{E}_{x_1,\ldots,x_{d-1}} \left(\mathbb{E}_x f_{x_1,\ldots,x_{d-1}}(x) \right) \left(\overline{\mathbb{E}_x f_{x_1,\ldots,x_{d-1}}(x)} \right) \\
&= \mathbb{E}_{x_1,\ldots,x_{d-1}} \left| \mathbb{E}_x f_{x_1,\ldots,x_{d-1}}(x) \right|^2
\end{aligned}
$$

Explicit formulas for the case $d = 1, 2$ are:

24

$$U^1(f) = \mathop{\mathbb{E}}_{x,y} f(x)\overline{f(x+y)} = (\mathop{\mathbb{E}}_x f(x)) \cdot \overline{\mathop{\mathbb{E}}_x f(x)} = |\mathop{\mathbb{E}}_x f(x)|^2$$

and

$$U^2(f) = \mathop{\mathbb{E}}_{x,y,z} f(x)\overline{f(x+y)f(x+z)}f(x+y+z) = \sum_{\mathbf{g}} \hat{f}(\mathbf{g})\overline{\hat{f}(\mathbf{g})\hat{f}(\mathbf{g})}\hat{f}(\mathbf{g}) = \sum_{\mathbf{g}} |\hat{f}(\mathbf{g})|^4$$

6.6   Gowers Uniformity and Influence

Theorem 12 *Let $f : G_1 \times \cdots \times G_n \to \mathbb{C}$ be a function such that $|f(x)| \le 1$ for every x. Then, for every $d \ge 1$,*

$$U^d(f) \le U^1(f) + 2^{O(d)} \max_i \mathrm{Inf}_i(f)$$

PROOF: The case $d = 1$ is trivial. For $d \ge 2$, it suffices to prove

$$U^d(f) \le U^{d-1}(f) + 2^{cd} \max_i \mathrm{Inf}_i(f)$$

for an absolute constant c. Let $I := \max_i \mathrm{Inf}_i(f)$.

We follow the proof of [Lemma 5](#) and write

$$U^d(f) = \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \left[\mathop{\mathbb{E}}_{x,y,z} f_{x_1,\ldots,x_{d-2}}(x)\overline{f_{x_1,\ldots,x_{d-2}}(x+y)f_{x_1,\ldots,x_{d-2}}(x+z)}f_{x_1,\ldots,x_{d-2}}(x+y+z) \right]$$

$$= \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \sum_{\mathbf{g}} |\hat{f}_{x_1,\ldots,x_{d-2}}(\mathbf{g})|^4$$

$$= \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \sum_{\mathbf{g}\neq\mathbf{0}} |\hat{f}_{x_1,\ldots,x_{d-2}}(\mathbf{g})|^4 + \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} |\hat{f}_{x_1,\ldots,x_{d-2}}(\mathbf{0})|^4$$

We separately bound the two terms in the last expression.

For every $a_1,\ldots,a_{d-2} \in G$, and every $\mathbf{g} = (g_1,\ldots,g_n) \neq \mathbf{0}$, using [Lemma 10](#), we get that, for every $i : g_i \neq 0$, $\mathrm{Inf}_i(f_{a_1,\ldots,a_{d-2}}) \le 2^{cd}\mathrm{Inf}_i(f) \le 2^{cd}I$ for an absolute constant c, and so we have that, for every $a_1,\ldots,a_{d-2} \in G$,

$$\sum_{\mathbf{g}\neq\mathbf{0}} |\hat{f}_{a_1,\ldots,a_{d-2}}(\mathbf{g})|^4 \le \max_{\mathbf{g}} |\hat{f}_{a_1,\ldots,a_{d-2}}(\mathbf{g})|^2 \le \max_i \mathrm{Inf}_i(f_{a_1,\ldots,a_{d-2}}) \le 2^{cd}I$$

and so

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \sum_{\mathbf{g}\neq\mathbf{0}} |\hat{f}_{x_1,\ldots,x_{d-2}}(\mathbf{g})|^4 \le 2^{cd}I$$

Regarding the other term,

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} |\hat{f}_{x_1,\ldots,x_{d-2}}(\mathbf{0})|^4 \le \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} |\hat{f}_{x_1,\ldots,x_{d-2}}(\mathbf{0})|^2 = \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-2}} \left| \mathop{\mathbb{E}}_x f_{x_1,\ldots,x_{d-2}}(x) \right|^2 = U^{d-1}(f)$$

$\square$

6.7 Gowers Inner Product

Let $\{f_S\}_{S\subseteq[d]}$ be a collection of functions $f_S : G \to \mathbb{C}$. Then their Gowers inner product is the complex number

$$\langle\{f_S\}\rangle_{U^d} := \mathop{\mathbb{E}}_{x,x_1,\ldots,x_d}\left[\prod_{S\subseteq[d]:|S|\text{ even}} f_S\left(x+\sum_{i\in S}x_i\right)\cdot \prod_{S\subseteq[d]:|S|\text{ odd}} \overline{f_S\left(x+\sum_{i\in S}x_i\right)}\right] \qquad (21)$$

6.8 Gowers Inner Product and Cross-Influence

We generalize Lemma 8 to the case of products of arbitrary groups.

Lemma 13 *For $d \geq 1$, let $\{f_S\}_{S\subseteq[d]}$ be a collection of functions $f_S : G_1 \times \cdots \times G_n \to \mathbb{C}$ such that*

- *$|f_S(x)| \leq 1$ for every S and every x;*

- *$f_{[d]}$ is balanced, that is, $\mathbb{E}_x f_{[d]}(x) = 0$;*

- *$4\mathrm{XInf}_i(\{f_S\}) \leq \varepsilon$ for every i.*

Then

$$|\langle\{f_S\}\rangle_{U^d}| \leq \sqrt{\varepsilon} \cdot 2^{O(d)}$$

We give an outline of the changes needed to adapt the proof of Lemma 8.

First, we need the following bound, whose proof is identical to the proof of Lemma 9.

Lemma 14 *Let $f_1, f_2, f_3, f_4 : G \to \mathbb{C}$ be functions such that $|f_i(x)| \leq 1$ for every i and every x. Then*

$$\sum_{\mathbf{g}} |\hat{f}_1(\mathbf{g})| \cdot |\hat{f}_2(\mathbf{g})| \cdot |\hat{f}_3(\mathbf{g})| \cdot |\hat{f}_4(\mathbf{g})| \leq 4\max_{\mathbf{g}} \min\{|\hat{f}_1(\mathbf{g})|,|\hat{f}_2(\mathbf{g})|,|\hat{f}_3(\mathbf{g})|,|\hat{f}_4(\mathbf{g})|\}$$

As in the proof Lemma 8, we proceed by induction on d and prove that, under the hypothesis of the lemma, $\langle\{f_S\}\rangle_{U^d}| \leq \tau(\varepsilon, d)$ for a function τ that satisfies $\tau(\varepsilon, d) = \sqrt{\varepsilon}2^{O(d)}$. For $d = 1$, We have $\langle f_\emptyset, f_{\{1\}}\rangle_{U^1} = (\mathbb{E}_x f_\emptyset(x)) \cdot \left(\mathbb{E}_y \overline{f_{\{1\}}(y)}\right) = 0$, and so we have the base case with $\tau(\varepsilon, 1) = 0$.

For the inductive step, we consider the $(d + 1)$-dimensional Gower inner product

$$\langle\{f_S\}\rangle_{U^{d+1}} = \mathop{\mathbb{E}}_{x,x_1,\ldots,x_{d+1}}\left[\prod_{S\subseteq[d+1]:|S|\text{ even}} f_S\left(x+\sum_{i\in S}x_i\right)\cdot \prod_{S\subseteq[d+1]:|S|\text{ odd}} \overline{f_S\left(x+\sum_{i\in S}x_i\right)}\right] \qquad (22)$$

It is easier to rewrite it as

$$\langle\{f_S\}\rangle_{U^{d+1}} = \mathop{\mathbb{E}}_{x,x_1,\ldots,x_{d+1}}\left[\prod_{S\subseteq[d+1]:|S|} F_S\left(x+\sum_{i\in S}x_i\right)\right] \qquad (23)$$

26

where we define $F_S := f_S$ if $|S|$ is even and $F_S := \overline{f_S}$ if $|S|$ is odd.

For every $x_1, \ldots, x_{d-1}$, define the four functions

$$A_{x_1,\ldots,x_{d-1}}(x) := \prod_{S \subseteq [d-1]} F_S\left(x + \sum_{i \in S} x_i\right)$$

$$B_{x_1,\ldots,x_{d-1}}(x) := \overline{\prod_{S \subseteq [d-1]} F_{S \cup \{d\}}\left(x + \sum_{i \in S} x_i\right)}$$

$$C_{x_1,\ldots,x_{d-1}}(x) := \overline{\prod_{S \subseteq [d-1]} F_{S \cup \{d+1\}}\left(x + \sum_{i \in S} x_i\right)}$$

$$D_{x_1,\ldots,x_{d-1}}(x) := \prod_{S \subseteq [d-1]} F_{S \cup \{d,d+1\}}\left(x + \sum_{i \in S} x_i\right)$$

with this notation, we can rewrite the expression (23) as

$$\langle \{f_S\}\rangle_{U^{d+1}} = \mathop{\mathbb{E}}_{x,x_1,\ldots,x_{d+1}} A_{x_1,\ldots,x_{d-1}}(x)\overline{B_{x_1,\ldots,x_{d-1}}(x+x_d)C_{x_1,\ldots,x_{d-1}}(x+x_{d+1})}D_{x_1,\ldots,x_{d-1}}(x+x_d+x_{d+1})$$

and using the Fourier expansion and simplifying,

$$
\begin{aligned}
|\langle \{f_S\}\rangle_{U^{d+1}}| &= \left|\mathop{\mathbb{E}}_{\mathbf{x}=(x_1,\ldots,x_{d-1})} \sum_{\mathbf{g}} \hat{A}_{\mathbf{x}}(\mathbf{g})\overline{\hat{B}_{\mathbf{x}}(\mathbf{g})\hat{C}_{\mathbf{x}}(\mathbf{g})}\hat{D}_{\mathbf{x}}(\mathbf{g})\right| \\[2mm]
&\leq \left|\mathop{\mathbb{E}}_{\mathbf{x}=(x_1,\ldots,x_{d-1})} \sum_{\mathbf{g}\neq(0,\ldots,0)} \hat{A}_{\mathbf{x}}(\alpha)\overline{\hat{B}_{\mathbf{x}}(\mathbf{g})\hat{C}_{\mathbf{x}}(\mathbf{g})}\hat{D}_{\mathbf{x}}(\mathbf{g})\right| \\[2mm]
&\quad + \left|\mathop{\mathbb{E}}_{\mathbf{x}=(x_1,\ldots,x_{d-1})} \hat{A}_{\mathbf{x}}(0,\ldots,0)\overline{\hat{B}_{\mathbf{x}}(0,\ldots,0)\hat{C}_{\mathbf{x}}(0,\ldots,0)}\hat{D}_{\mathbf{x}}(0,\ldots,0)\right|
\end{aligned}
$$

We bound the two terms separately.

For the first term, we have that, for every fixed $\mathbf{x} = (x_1, \ldots, x_{d-1})$,

$$
\begin{aligned}
&\left|\sum_{\mathbf{g}\neq(0,\ldots,0)} \hat{A}_{\mathbf{x}}(\mathbf{g})\overline{\hat{B}_{\mathbf{x}}(\mathbf{g})\hat{C}_{\mathbf{x}}(\mathbf{g})}\hat{D}_{\mathbf{x}}(\mathbf{g})\right| \\[2mm]
\leq\ & \sum_{\mathbf{g}\neq(0,\ldots,0)} |\hat{A}_{\mathbf{x}}(\mathbf{g})| \cdot |\overline{\hat{B}_{\mathbf{x}}(\mathbf{g})}| \cdot |\overline{\hat{C}_{\mathbf{x}}(\mathbf{g})}| \cdot |\hat{D}_{\mathbf{x}}(\mathbf{g})| \\[2mm]
=\ & \sum_{\mathbf{g}\neq(0,\ldots,0)} |\hat{A}_{\mathbf{x}}(\mathbf{g})| \cdot |\hat{B}_{\mathbf{x}}(\mathbf{g})| \cdot |\hat{C}_{\mathbf{x}}(\mathbf{g})| \cdot |\hat{D}_{\mathbf{x}}(\mathbf{g})| \\[2mm]
\leq\ & 4\sqrt{\max_i \min\{\mathrm{Inf}_i(A_{\mathbf{x}}), \mathrm{Inf}_i(B_{\mathbf{x}}), \mathrm{Inf}_i(C_{\mathbf{x}}), \mathrm{Inf}_i(D_{\mathbf{x}})\}}
\end{aligned}
$$

27

And we observe that, for every $\mathbf{x} = (x_1, \ldots, x_{d-1})$ and every i,

$$\min\{\mathrm{Inf}_i(A_\mathbf{x}), \mathrm{Inf}_i(B_\mathbf{x}), \mathrm{Inf}_i(C_\mathbf{x}), \mathrm{Inf}_i(D_\mathbf{x})\} \leq 2^{cd}\varepsilon$$

for some absolute constant c.

The second term can be written as

$$\left| \underset{\mathbf{x}}{\mathbb{E}} \, \hat{A}_\mathbf{x}(0, \ldots, 0) \overline{\hat{B}_\mathbf{x}(0, \ldots, 0) \hat{C}_\mathbf{x}(0, \ldots, 0)} \hat{D}_\mathbf{x}(0, \ldots, 0) \right|$$

$$= \left| \underset{x,\mathbf{x},y,z,w}{\mathbb{E}} A_\mathbf{x}(x) \overline{B_\mathbf{x}(x+y) C_\mathbf{x}(x+z)} D_\mathbf{x}(x+w) \right| \tag{24}$$

$$\leq \underset{y,z}{\mathbb{E}} \left| \underset{x,\mathbf{x},w}{\mathbb{E}} A_\mathbf{x}(x) \overline{B_\mathbf{x}(x+y) C_\mathbf{x}(x+z)} D_\mathbf{x}(x+w) \right|$$

$$\leq \left| \underset{x,\mathbf{x},w}{\mathbb{E}} A_\mathbf{x}(x) \overline{B_\mathbf{x}(x+a) C_\mathbf{x}(x+b)} D_\mathbf{x}(x+w) \right| \tag{25}$$

where a, b are values for y, z that maximize the expectation in Equation 24.

If we expand the Expression in (25), we get

$$\mathbb{E}_{x,x_1,\ldots,x_{d-1},w} \prod_{S \subseteq [d-1]} \begin{array}{l} F_S \left(x + \sum_{i \in S} x_i \right) F_{S \cup \{d\}} \left(x + a + \sum_{i \in S} x_i \right) \\ F_{S \cup \{d+1\}} \left(x + b + \sum_{i \in S} x_i \right) F_{S \cup \{d,d+1\}} \left(x + w + \sum_{i \in S} x_i \right) \end{array} \tag{26}$$

which we are going to re-write as the d-dimensional Gowers inner product of a new set of functions, so that we can invoke the inductive hypothesis. For every $S \subseteq [d-1]$, if S is even, define

$$g_S(x) := F_S(x) \cdot F_{S \cup \{d\}}(x+a) \cdot F_{S \cup \{d+1\}}(x+b)$$

and define

$$g_{S \cup \{d\}}(x) := \overline{F_{S \cup \{d,d+1\}}(x)}$$

If S is odd, define

$$g_S(x) := \overline{F_S(x) \cdot F_{S \cup \{d\}}(x+a) \cdot F_{S \cup \{d+1\}}(x+b)}$$

and

$$g_{S \cup \{d\}}(x) := F_{S \cup \{d,d+1\}}(x)$$

Expression (26) becomes $\langle \{g_S\} \rangle_{U^d}$ after the change of variable $w \to x_d$.

By definition, the function $g_{[d]}$ is either $f_{[d+1]}$ of $\overline{f_{[d+1]}}$ and, in either case, $g_{[d]}$ is balanced. By construction and by Lemma 10, the cross-influence of the functions g_S is at most $c'\varepsilon$ for an absolute constant c'.

This proves the inductive step with $\tau(\varepsilon, d+1) = 4 \cdot \sqrt{2^{cd}\varepsilon} + \tau(c'\varepsilon, d-1)$ where c, c' are absolute constants. Together with $\tau(\varepsilon, 1) = 0$ we have $\tau(\varepsilon, d) = 2^{O(d)}\sqrt{\varepsilon}$ as desired.

7 A Tight Analysis of Linearity Testing

Consider the following promise problem. Given a function $f : \{0,1\}^n \to \{-1,1\}$ and a small $\varepsilon > 0$, we want to distinguish the two cases

1. f is linear;

2. $U^d(f) \leq \varepsilon$.

We refer to such a test as a "relaxed linearity test of degree $(d-1)$." As usual, we say that a test has completeness c and soundness s if the test acceptes with probability $\geq c$ in case (1) and with probability $\leq s + \varepsilon'$ in case (2), where $\varepsilon' \to 0$ when $\varepsilon \to 0$. If a test makes q queries and has soundness s, then its *amortized query complexity* is $\bar{q} = q/\log s^{-1}$.

For $d = 2$, this problem is the linearity testing problem. For $d = 3$, the only functions such that $U^3(f) \geq \varepsilon$ are functions that are correlated with degree-2 polynomials [Sam05], and so the test is required to distinguish linear functions from functions that are far from being quadratic. For $d \geq 4$, it is conjectured that the only functions with $U^d(f) \geq \varepsilon$ are those that are correlated with a degree-$(d-1)$ polynomial, and, if so, such a test distinguishes linear functions from functions that are far from low-degree polynomials. By our results, such a test also distinguishes linear functions from functions where all variables have low influence.

We give a tight analysis of the error probability of such tests for a given number of queries.

7.1 The Linear Gowers Inner Product

For the sake of our analysis of the Hypergraph Test, it is convenient to study expressions of the following form. Let $\{f_S\}_{S \subseteq [d]}$ be a collection of 2^d functions $f_S : \{0,1\}^n \to [-1,1]$, $d \geq 1$, and define their *linear* Gowers Inner Product as

$$\langle \{f_S\} \rangle_{LU^d} := \mathop{\mathbb{E}}_{x_1,\dots,x_d} \prod_{S \subseteq [d]} f\left(\sum_{i \in S} x_i\right)$$

As usual, an empty sum is zero. For example:

$$
\begin{aligned}
\langle f_\emptyset, f_{\{1\}} \rangle_{LU^1} &:= \mathop{\mathbb{E}}_{x} f_\emptyset(\mathbf{0}) f_{\{1\}}(x) \\
\langle f_\emptyset, f_{\{1\}}, f_{\{2\}}, f_{\{1,2\}} \rangle_{LU^2} &:= \mathop{\mathbb{E}}_{x,y} f_\emptyset(\mathbf{0}) f_{\{1\}}(x) f_{\{2\}}(y) f_{\{1,2\}}(x+y)
\end{aligned}
$$

Where $\mathbf{0} = (0,\dots,0)$ is the all-zero vector of $\{0,1\}^n$. We call it a *linear* inner product because, for functions $f_S : \{0,1\}^n \to \mathbb{R}$, the Gowers inner product is defined by picking at random an *affine* subspace of $\{0,1\}^n$ of dimension d, and then taking the product of the functions on all points of the subspace. In the above expression, we do something similar but on a *linear* subspace.

We prove that if the linear Gowers inner product of a collection of functions is large, then the regular Gowers inner product of a related collection of functions must also be large.

Lemma 15 *Let $f_S : \{0,1\}^n \to [-1,1]$ be functions, $S \subseteq [d]$, and define the collection $\{g_T\}_{T \subseteq [d]}$ as $g_T = f_{T \cup \{d\}}$. Then*

$$|\langle \{f_S\}\rangle_{LU^d}| \leq \sqrt{\langle \{g_S\}\rangle_{U^d}}$$

PROOF:

$$
\begin{aligned}
|\langle \{f_S\}\rangle_{LU^d}| \quad &:= \quad \left| \mathop{\mathbb{E}}_{x_1,\ldots,x_d} \prod_{S \subseteq [d]} f_S\left(\sum_{i \in S} x_i\right) \right| \\[2ex]
&\leq \quad \sqrt{ \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-1}} \left(\prod_{S \subseteq [d], d \notin S} f_S\left(\sum_{i \in S} x_i\right) \right)^2 \times } \\
&\qquad\quad \overline{ \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-1}} \left(\mathop{\mathbb{E}}_{x_d} \prod_{S \subseteq [d], d \in S} f_S\left(\sum_{i \in S} x_i\right) \right)^2 } \\[2ex]
&\leq \quad \sqrt{ \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-1}} \left(\mathop{\mathbb{E}}_{x_d} \prod_{S \subseteq [d], d \in S} f_S\left(\sum_{i \in S} x_i\right) \right)^2 } \\[2ex]
&= \quad \sqrt{ \mathop{\mathbb{E}}_{x_1,\ldots,x_{d-1}} \left(\mathop{\mathbb{E}}_{x} \prod_{S \subseteq [d], d \in S} f_S\left(x + \sum_{i \in S - \{d\}} x_i\right) \right) \cdot \left(\mathop{\mathbb{E}}_{y} \prod_{S \subseteq [d], d \in S} f_S\left(y + \sum_{i \in S - \{d\}} x_i\right) \right) } \\[2ex]
&= \quad \sqrt{ \mathop{\mathbb{E}}_{x,y,x_1,\ldots,x_{d-1}} \prod_{T \subseteq [d-1]} f_{T \cup \{d\}}\left(x + \sum_{i \in T} x_i\right) f_{T \cup \{d\}}\left(y + \sum_{i \in T} x_i\right) } \\[2ex]
&\leq \quad \sqrt{\langle \{g_T\}\rangle_{U^d}}
\end{aligned}
$$

after the change of variable $y \to x + x_d$. $\square$

Lemma 16 *Let $f_S : \{0,1\}^n \to [-1,1]$ be functions, $S \subseteq [d]$, such that $\langle \{f_S\}\rangle_{LU^d} \geq \varepsilon$ and $\mathbb{E}_x f_{[d]}(x) = 0$.*

Then there is a variable i such that $\mathrm{XInf}_i(\{f_S\}) \geq \varepsilon^4 / 2^{O(n)}$.

PROOF: Define $g_T := f_{T \cup \{d\}}$. Then we have

1. $\mathbb{E}_x g_{[d]}(x) = \mathbb{E}_x f_{[d]}(x) = 0$,

2. $\langle \{g_T\}\rangle_{U^d} \geq (\langle \{f_S\}\rangle_{LU^d})^2 \geq \varepsilon^2$.

From Lemma 8 we derive that there is a variable i such that $4\mathrm{XInf}_i(\{g_T\}) \geq \varepsilon^4 / 2^{O(d)}$. Each function f_S occurs at most twice in the collection $\{g_T\}$, and so it must be $\mathrm{XInf}_i(\{f_S\}) \geq 4\mathrm{XInf}_i(\{g_T\}) \geq \varepsilon^4 / 2^{O(d)}$. $\square$

7.2 Positive Results on Relaxed Linearity Testing

Given a hypergraph $H = ([k], E)$, we can define a relaxed linearity test associated with H by

H-Test
  choose $x, x^1, \ldots, x^k$ uniformly at random in $\{0, 1\}^n$
  accept if and only if
    $\forall e \in E: \ \prod_{i \in e} f\left(x^i\right) = f\left(\sum_{i \in e} x^i\right)$

Then we have the following result.

Theorem 17 *Let $d \geq 2$ and let $H = ([k], E)$ be an hypergraph such that each edge of H contains at most d vertices.*

Them the H-Test is a degree-$(d-1)$ relaxed linearity test of completeness 1 and soundness at most $1/2^{|E|}$.

We remark that this result was first proved in [Sam05], using a different approach.

For the proof of Theorem 17 and of results in the next section, it will be convenient to use the following "Vazirani XOR Lemma," whose proof is immediate.

Lemma 18 *Let $X_1, \ldots, X_m$ be random variables taking values in $\{-1, 1\}$. Then*

$$\mathbf{Pr}[X_1 = 1 \wedge X_2 = 1 \wedge \cdots \wedge X_m = 1] = \frac{1}{2^m} \sum_{S \subseteq [m]} \prod_{i \in S} X_i$$

Using Lemma 18, we see that the probability that the H-Test accepts a function f is equal to

$$\frac{1}{2^{|E|}} \sum_{E' \subseteq E} \prod_{e \in E} \left(\prod_{i \in e} f(x^i) \right) \cdot f\left(\sum_{i \in e} x^i \right)$$

$$= \frac{1}{2^{|E|}} + \frac{1}{2^{|E|}} \sum_{E' \subseteq E, E' \neq \emptyset} \prod_{e \in E} \left(\prod_{i \in e} f(x^i) \right) \cdot f\left(\sum_{i \in e} x^i \right)$$

We will also need two results from [Gow01, GT04]. The first one is [Gow01, Lemma 3.8], and it states that for every collection $\{f_S\}$ of functions:

$$\left| \langle \{f_S\} \rangle_{U^d} \right| \leq \prod_S \left(U^d \left(f_S \right) \right)^{1/2^d}$$

The other is [GT04, 5.7] and states that for every f and $d \geq 2$,

$$U^{d-1}(f) \leq \sqrt{U^d(f)}$$

We now proceed with the proof of Theorem 17.

PROOF:[Of Theorem 17] It's clear that a linear function is accepted with probability 1.

If the H-Test accepts with f probability more than $1/2^{|E|} + \varepsilon$, then there is a non-empty $E' \subseteq E$ such that

$$\prod_{e \in E'} \left(\prod_{i \in e} f(x^i) \right) \cdot f \left(\sum_{i \in e} x^i \right) \geq \varepsilon \tag{27}$$

Let d' be the size of the largest edge in E', and, without loss of generality, assume that the edge $(1, \ldots, d')$ is in E'. Fix the variables $x^{d'+1}, \ldots, x^k$ to values that maximize (27). Then, (27) becomes

$$\prod_{S \subseteq [d']} f_S \left(\sum_{i \in S} x_i \right) \geq \varepsilon$$

where $f_\emptyset$ is the constant function equal to the product of the terms of (27) that depend only on $x^{d'+1}, \ldots, x^k$, $f_{\{1\}}(x^1)$ is the product of the terms of (27) that depend only on x^1 and $x^{d'+1}, \ldots, x^k$, and so on. In particular, $f_{[d']}(x^1 + \cdots + x^{d'}) = f(x^1 + \cdots + x^{d'})$.

We thus have

$$\langle \{f_S\} \rangle_{LU^{d'}} \geq \varepsilon$$

By Lemma 15, there are functions $\{g_S\}$ such that $g_{[d']} = f$ and $\langle \{g_S\} \rangle_{U^{d'}} \geq \varepsilon^2$.

Since all the functions involved are boolean, and thus their uniformity norms are at most 1, we conclude

$$\varepsilon^2 \leq \left| \langle \{g_S\} \rangle_{U^{d'}} \right| \leq \min_S |U^{d'}(g_S)|^{1/2^{d'}} \leq |U^{d'}(f)|^{1/2^{d'}}$$

And, finally, $U^d(f) \geq \left(U^{d'}(f) \right)^{2^{d-d'}} \geq \varepsilon^{2^{d+1}}$. $\square$

In particular, taking H to be the complete hypergraph on k vertices with at most d vertices per edge, leads to a hypergraph linearity test with $\sum_{i=1}^d \binom{k}{i}$ queries, and soundness $\frac{1}{2^{\sum_{i=2}^d \binom{k}{i}}}$. The amortized query complexity of this test, for $k \gg d$, is $\bar{q} \leq 1 + O\left(\frac{1}{q^{(d-1)/d}} \right)$.

Next, we show that no linearity test can do better.

We do it in two steps. The first is to show that the amortized query complexity of any hypergraph test cannot be better than $1 + \Omega\left(\frac{1}{q^{(d-1)/d}} \right)$ for q queries.

The second step builds on the first, and shows that *any* non-adaptive linearity test with perfect completeness cannot do better under this promise, namely it will have amortized query complexity of at least $1 + \Omega\left(\frac{1}{q^{(d-1)/d}} \right)$ for q queries.

7.3 Lower Bound for the H-Test

We prove a lower bound for the H-Test by describing an explicit function $f : \{0,1\}^n \to \{-1,1\}$, which has small d-th uniformity norm, and for any hypergraph H the acceptance probability of the H-Test on f is large.

Proposition 19 *Fix $d \geq 2$ and fix an hypergraph $H = ([k], E)$ such that all edges have at most d vertices. For n a multiple of d, let $f : \{0, 1\}^n \to \{-1, 1\}$ be defined as*

$$f(x_1, \ldots, x_n) := (-1)^{x_1 x_2 \cdots x_d + x_{d+1} \cdots x_{2d} + \cdots}$$

Then $U^d(f) \leq 2^{-\Omega(n)}$, and the H-Test accepts f with probability at least

$$\max \left\{ 1/2^{|E|}, 2^{-\sum_{i=2}^{d} \binom{k}{i}} \right\}$$

The lower bound on amortized query complexity of hypergraph tests follows immediately, if we recall that the number of queries is $|E| + k$, and consider the two cases: $|E| \leq \sum_{i=2}^{d} \binom{k}{i}$, or $|E| > \sum_{i=2}^{d} \binom{k}{i}$.

PROOF: We first show $U^d(f)$ to be small.

Some additional notation: Let $g_i(x) = x_{i \cdot d + 1} x_{i \cdot d + 2} \cdots x_{(i+1) \cdot d}$, and let $g(x) = x_1 x_2 \cdots x_d + x_{d+1} \cdots x_{2d} + \cdots = \sum_{i=0}^{n/d-1} g_i$. Let $f_i = (-1)^{g_i}$.

Then $f = \prod_{i=0}^{n/d-1} f_i$ and $U^d(f) = \prod_{i=0}^{n/d-1} U^d(f_i) = \left(U^d(f_0)\right)^{n/d}$. Thus it remains to show that $U^d(f_0)$ is bounded away from 1.

$$U^d(f_0) = U^d \left((-1)^{g_0}\right) = \mathop{\mathbb{E}}_{x, x^1, \ldots, x^d} (-1)^{\sum_{S \subseteq [d]} g_0 \left(x + \sum_{i \in S} x^i\right)}$$

We may assume that the variables x^i live in $\{0, 1\}^d$, and then g_0 is just the AND function. Therefore $\sum_{S \subseteq [d]} g_0 \left(x + \sum_{i \in S} x^i\right)$ counts the number of times the complement $\mathbf{1} - x$ of x is representable as a linear combination of $x^1 \ldots x^d$. This number is odd $(= 1)$ iff $x^1 \ldots x^d$ are linearly independent and is even otherwise. Thus

$$U^d(f_0) = 1 - 2 \mathbf{Pr}[x^1 \ldots x^d \text{ are linearly independent}]$$

and this is easily seen (and well-known) to be a positive constant bounded away from 1.

The proof that the acceptance probability of f is high closely follows the proof of Proposition 15 in [ST00]. That proposition is a special case $d = 2$ of Proposition 19. We will repeat parts of the proof since many definitions need to be modified, and for completeness, but will omit proofs of intermediary steps if they are similar.

First, the probability of f to be accepted is

$$\frac{1}{2^{|E|}} \sum_{\mathcal{S} \subseteq E} \mathop{\mathbb{E}}_{x^1, \ldots, x^k} \left[\prod_{T \in \mathcal{S}} \prod_{i \in T} f(x^i) \cdot f \left(\sum_{i \in T} x^i \right) \right] \tag{28}$$

In order to simplify this expression, we need to introduce some notation. Let $\mathcal{F} = \{\{1\}, \ldots, \{k\}\} \cup E := \{F_1 \ldots F_q\}$, be a family of all the vertices and the edges of H, viewed as subsets of $\{1, \ldots, k\}$. Let $\mathbf{A}$ be a $k \times q$ zero-one matrix whose q columns are given by $F_1 \ldots F_q$, which we view as $0, 1$ vectors of length k (in particular, the first k columns of $\mathbf{A}$ form the $k \times k$ identity matrix). Let u_T, for $T \in E$ be a zero-one vector of length q which is 1 if F_i is either T or a singleton, corresponding to a vertex than T passes through; and 0 otherwise.

For $\mathcal{R} \subseteq 2^{[k]}$, let $\mathbb{E}(f, \mathcal{R}) := \mathbb{E}_{x^1,\ldots,x^k} \left[\prod_{R \in \mathcal{R}} f \left(\sum_{l \in R} x^l \right) \right].$

Let $U = \text{Span } (u_T)_{T \in E}$ be a t-dimensional subspace of $\mathbb{Z}_2^q$ then, for a boolean f, (28) is

$$\frac{1}{2^{|E|}} \sum_{\mathcal{S} \subseteq E} \mathop{\mathbb{E}}_{x^1,\ldots,x^k} \left[\prod_{i \; : \; \bigoplus_{T \in \mathcal{S}} u_T(i)=1} f \left(\sum_{l \in F_i} x^l \right) \right] =$$

$$\frac{1}{2^t} \sum_{u=(u(1),\ldots u(q)) \in U} \mathop{\mathbb{E}}_{x^1,\ldots,x^k} \left[\prod_{i \; : \; u(i)=1} f \left(\sum_{l \in F_i} x^l \right) \right] =$$

$$\frac{1}{2^t} \sum_{u=(u(1),\ldots u(q)) \in U} \mathbb{E}(f, \{F_i : u(i) = 1\}). \tag{29}$$

We will show many of the terms $\mathbb{E}(f, \{F_i : u(i) = 1\})$ are 1.

Definition 20 *A family $\mathcal{R} \subseteq 2^{[k]}$ is an "even cover", iff every subset $T \subseteq [k]$, $|T| \leq d$ is covered an even number of times by the sets $R \in \mathcal{R}$.*

The proofs of the following three lemmas are easily adaptable from the proofs of the corresponding statements (Lemmas 17,18,19) in [ST00].

Lemma 21 *Let $\mathcal{R} \subseteq 2^{[k]}$. If $\mathcal{R}$ is an even cover, than $\mathbb{E}(f, \mathcal{R}) = 1$. For any $\mathcal{R}$, $\mathbb{E}(f_n, \mathcal{R}) \geq 0$.*

PROOF: Omitted. $\square$

Lemma 22 *The number of vectors $u \in U$, such that the family $\mathcal{R} = \{F_i \; : \; u(i) = 1\}$ is an even cover, is at least $\max \left\{ 1, 2^{t - \sum_{i=2}^d \binom{k}{i}} \right\}$.*

PROOF: Omitted. $\square$

Lemma 23
$$t \leq |E|.$$

PROOF: Omitted. $\square$

Proposition 19 now follows from lemma 21, lemma 22, lemma 23 and (29). $\square$

7.4 Lower Bounds for Arbitrary Test

We now pass to our more general result, that holds for any test, including tests with completeness smaller than 1. Recall that a test making q-queries is called *non-adaptive* if it makes the q queries simultaneously, instead of using the answer to the first query to decide how to make the second query, and so on.

Theorem 24 *let $\mathcal{T}$ be a non-adaptive test that makes q queries and that is a degree $(d-1)$ relaxed linearity test with completeness c and soundness s for functions $f : \{0,1\}^n \to \{-1,1\}$.*

Then

$$1 - c + s \geq 2^{-q + \Omega\left(\frac{1}{q^{(d-1)/d}}\right)} .$$

Equivalently, the amortized query complexity of $\mathcal{T}$ is at least $1 + \Omega\left(\frac{1}{q^{(d-1)/d}}\right)$.

PROOF: We will show that $(1 - c) + s$ is at least $\varepsilon = \varepsilon_H - 2^{-\Omega(n)}$, where ε_H is the best error achievable by a hypergraph test after q queries. The theorem will then follow from Proposition 19.

By von Neumann's minimax theorem (also known as Yao's pronciple) it is enough to construct two distributions P and Q on boolean functions, P supported on linear functions and Q on functions with small U^d norm, such that such that for any subset $X = \{x^1...x^q\}$ of the boolean cube the distributions P' and Q' induced by P and Q on $\{0,1\}^q$ by evaluating a function $f \sim P$ (correspondingly $f \sim Q$) on X are at most $1 - \varepsilon'$ apart.

For a function $f : \{0,1\}^n \to \mathbb{R}$, and an $n \times n$ matrix A over $G\mathbb{F}_2$ let f_A be given by $f_A(x) = f(Ax)$. Now, fix a non-zero linear function ℓ, and let $f = f(x_1, \ldots, x_n) := (-1)^{x_1 x_2 \cdots x_d + x_{d+1} \cdots x_{2d} + \cdots}$. The distribution P is taken to be uniform over the functions $\{\ell_A : A \text{ is invertible}\}$ and the distribution Q is taken to be uniform over the functions $\{f_A : A \text{ is invertible}\}$.

Observe that P is supported on linear functions. In fact, P is uniform over all non-zero linear functions. On the other hand f has a small U^d norm, by Proposition 19. The following lemma shows the same for the functions f_A.

Lemma 25 *For a function $f : \{0,1\}^n \to \mathbb{R}$, and an invertible $n \times n$ matrix A over $G\mathbb{F}_2$ holds*

$$U^d(f) = U^d(f_A)$$

PROOF:

$$U^d(f_A) = \underset{x,x^1,\ldots,x^d}{\mathbb{E}} \prod_{S \subseteq [d]} f_A\left(x + \sum_{i \in S} x^i\right) = \underset{x,x^1,\ldots,x^d}{\mathbb{E}} \prod_{S \subseteq [d]} f\left(Ax + \sum_{i \in S} Ax^i\right) =$$

$$\underset{x,x^1,\ldots,x^d}{\mathbb{E}} \prod_{S \subseteq [d]} f\left(x + \sum_{i \in S} x_i\right) = U^d(f)$$

The third equality follows from the fact that if a $(d+1)$-tuple $(x, x^1, ..., x^d)$ is distributed uniformly in $\{0,1\}^{n(d+1)}$, then so is the $(d+1)$-tuple $(Ax, Ax^1, ..., Ax^d)$. $\square$

Since the distributions P and Q are invariant under invertible linear transformations, the induced distributions P' and Q' are determined by the linear structure of X, namely linear dependencies between $x^1 \ldots x^q$. Let the linear rank of the vectors $x^1, \ldots, x^q$ over $G\mathbb{F}_2$ be k. We will assume $x^1, \ldots, x^k$ are linearly independent, and $x^{k+i} = \sum_{j=1}^{k} a_{ij} x^j$ for $i = 1, \ldots, q - k$.

In the remainder of the proof we assume $n \gg k$ and call an event negligible if its probability is exponentially small in n.

The distributions P' and Q' are almost precisely modelled by the following experiments: Choose $y^1, \ldots, y^k$ independently at random, set $y^{k+i} = \sum_{j=1}^{k} a_{ij} y^j$, for $i = 1, \ldots, q - k$, and return $\ell(y^1), \ldots, \ell(y^q)$ or $f(y^1), \ldots, f(y^q)$ correspondingly. The only caveat comes from the negligible event that $y^1, \ldots, y^k$ are linearly dependent.

For $n \gg k$, the distribution P' is, up to a negligible probability, given by choosing the first k bits uniformly at random, and setting the other $q - k$ bits according to specified linear dependencies.

Consider the distribution Q'. By proposition 19 the probability that for $i = 1, \ldots, q - k$ holds $f(y^{k+i}) = \sum_{j=1}^{k} a_{ij} f(y^j)$ is at least ε_H. Observe that in this case the q-tuple we deal with has no linear contradictions and thus belongs to the support of P'. Call such q-tuple "linear". We have just proved that, up to a negligible factor,

$$\mathbf{Pr}\{z \; : \; z \text{ is "linear"}\} \geq \varepsilon_H$$

We have

$$\|P' - Q'\| = \frac{1}{2} \cdot \sum_z \left| P'(z) - Q'(z) \right| = \frac{1}{2} \cdot \sum_z \left[\max\{P'(z), Q'(z)\} - \min\{P'(z), Q'(z)\} \right] \leq$$

$$1 - \frac{1}{2} \cdot \sum_z \min\{P'(z), Q'(z)\} \leq 1 - \frac{1}{2} \cdot \sum_{\text{"linear"} \; z} \min\{P'(z), Q'(z)\}$$

To complete the proof of the theorem we observe that, up to negligible factors, $P'(z) \geq 2^{-k}$ for all "linear" z, and

$$Q'(z) = Q'(z_1, \ldots, z_q) \leq \mathbf{Pr}\{f(y_1) = z_1, \ldots, f(y_k) = z_k\} = \prod_{i=1}^{k} \mathbf{Pr}\{f(y_i) = z_i\} = 2^{-k}$$

Therefore

$$\sum_{\text{"linear"} \; z} \min\{P'(z), Q'(z)\} \geq \sum_{\text{"linear"} \; z} P'(z) - 2^{-\Omega(n)} \geq \varepsilon_H - 2^{-\Omega(n)}$$

$\square$

8 The PCP Construction

8.1 The Long-Code Test

We say that a function $g : \{0, 1\}^n \to \{-1, 1\}$ is a *codeword of the Long Code* (or, simply, is a long code) if there exists a coordinate i such that $g(x_1, \ldots, x_n) = (-1)^{x_i}$, that is, if $g = \chi_{\{i\}}$. Note that

if g is a long code then there is a coordinate that has degree-1 influence 1 for g. This is the extreme case of large low-degree influence for a bounded function.

Given a collection of K balanced functions $g_1, \ldots, g_K : \{0,1\}^n \to \{-1,1\}$, we are interested in designing a test that distinguishes the following two cases:

- The functions g_j are all equal to the same long code, that is, for some $i \in [n]$ and for all $j \in [K]$, $g_j(x) = (-1)^{x_i}$;

- The degree-d cross-influence of the collection $g_1, \ldots, g_K$ is less than than δ, for some small δ and large d.

More formally, we have the following definition.

Definition 26 (K-Function Long Code Test) *A test that is given oracle access to K functions $g_1, \ldots, g_K : \{0,1\}^n \to \{-1,1\}$ is said to have soundness s and completeness c if the following conditions hold.*

- *If the functions g_j are equal to the same long code, then the test accepts with probability $\geq c$;*

- *For every $\varepsilon > 0$ there is a $\tau = \tau(\varepsilon) > 0$ and $d = d(\varepsilon)$ such that if the test accepts with probability $\geq s + \varepsilon$, then there is a variable of degree-d cross-influence at least τ for the functions g_j.*

Let $H = ([t], E)$ be a hypergraph on t vertices.

For $0 < \gamma < 1/2$, define the distribution μ_γ over $\{0,1\}^n$ so that $\mu_\gamma(x) = \gamma^{w(x)}(1-\gamma)^{n-w(x)}$, where $w(x)$ is the number of ones in x.

The γ-noisy H-test is a $(t + |E|)$-function long code test defined as follows:

γ-noisy-H-Test $(\{g^a\}_{a \in [t] \cup E})$
   choose $x, x_1, \ldots, x_t$ uniformly at random in $\{0,1\}^n$
   for every $i \in [t]$, sample η^i from μ_γ
   for every $e \in E$, sample η^e from μ_γ
   accept if and only if
     $\forall e \in E.\ \prod_{i \in e} g^{(i)}\left(\eta^i + x_i\right) = g^{(e)}\left(\eta^{(e)} + \sum_{i \in e} x_i\right)$

Remark 2 *In the definition of K-function long-code test, we index functions by integers $1, \ldots, K$, while in the definition of the Hypergraph Test we let the given functions be indexed by elements of $[t] \cup E$. We hope the reader is not too bothered by this abuse of notation.*

8.2 Analysis of the Hypergraph Test

In this section we prove the following theorem.

Theorem 27 *For every hypergraph $H = ([t], E)$, and every $\gamma > 0$, the γ-noisy H-test is a $(t+|E|)$-function long code test with completeness $1 - (t+1)\gamma|E|$ and soundness $1/2^{|E|}$.*

The completeness part is clear.

For the soundness part, as in Section 7.2, we can write the probability that the test accepts a given set of oracle functions as

$$\sum_{E' \subseteq E} \frac{1}{2^{|E|}} \underset{x_1,\dots,x_t,\{\eta^i,\eta^e\}}{\mathbb{E}} \prod_{e \in E} \prod_{i \in e} g^{(i)}\left(\eta^i + x_i\right) \cdot g^{(e)}\left(\eta^{(e)} + \sum_{i \in e} x_i\right)$$

$$= \frac{1}{2^{|E|}} + \sum_{E' \subseteq E,\ E' \neq \emptyset} \frac{1}{2^{|E|}} \underset{x_1,\dots,x_t,\{\eta^i,\eta^e\}}{\mathbb{E}} \prod_{e \in E} \prod_{i \in e} g^{(i)}\left(\eta^i + x_i\right) \cdot g^{(e)}\left(\eta^{(e)} + \sum_{i \in e} x_i\right)$$

and so if thest accepts with probability at least $2^{-|E|} + \varepsilon$ there is a subset of tests $E' \subseteq E$ such that

$$\underset{x_1,\dots,x_t,\{\eta^i,\eta^e\}}{\mathbb{E}} \prod_{e \in E'} \prod_{i \in e} g^{(i)}\left(\eta^i + x_i\right) \cdot g^{(e)}\left(\eta^{(e)} + \sum_{i \in e} x_i\right) > \varepsilon$$

It remains to prove the following lemma (E in the lemma plays the role of E' above).

Lemma 28 *Let $g_j : \{0,1\}^n \to \{-1,1\}$ be functions and $H = ([t], E)$ be a hypergraph such that*

$$\underset{x_1,\dots,x_t,\{\eta^i,\eta^e\}}{\mathbb{E}} \prod_{e \in E} \prod_{i \in e} g^{(i)}\left(\eta^i + x_i\right) \cdot g^{(e)}\left(\eta^{(e)} + \sum_{i \in e} x_i\right) > \varepsilon \tag{30}$$

where the η's are sampled according to μ_γ. Then there is a variable i that has degree-$d(\varepsilon, \gamma)$ cross-influence at least $\delta(\varepsilon, \gamma) > 0$ for the functions $\{g^{(a)}\}_{a \in [t] \cup E}$.

PROOF: Since the $g^{(a)}$ map to $\{-1,1\}$, we can write Equation 30 equivalently as

$$\underset{x_1,\dots,x_t,\{\eta^i,\eta^e\}}{\mathbb{E}} \left(\prod_{i \in \text{Odd}} g^{(i)}\left(\eta^{(i)} + x_i\right) \right) \cdot \left(\prod_{e \in E} g^{(e)}\left(\eta^{(e)} \sum_{i \in e} x_i\right) \right) > \varepsilon \tag{31}$$

Where Odd denotes the set of vertices of odd degree in the hypergraph $([t], E)$.

Now, define $G(x) = \mathbb{E}_\eta\, g(\eta + x)$, where η is sampled from μ_γ.

Then (30) becomes

$$\underset{x_1,\dots,x_t}{\mathbb{E}} \left(\prod_{i \in \text{Odd}} G^{(i)}\left(x_i\right) \right) \cdot \left(\prod_{e \in E} G^{(e)}\left(\sum_{i \in e} x_i\right) \right) > \varepsilon \tag{32}$$

Let k be the maximum size of an hyperedge in E, and assume, without loss of generality that H has an hyperedge $e = (1, 2, 3, \dots, k)$. Fix the variables $x_{k+1}, \dots, x_t$ in (32) in such a way that the average over $x_1, x_2, \dots, x_k$ is still at least ε. In particular, (32) becomes

$$\underset{x_1,x_1,\dots,x_k}{\mathbb{E}} f_\emptyset(\mathbf{0}) f_{\{1\}}\left(x_1\right) f_{\{2\}}\left(x_2\right) \cdots G^{(e)}\left(x_1 + x_2 + \dots + x_k\right) > \varepsilon \tag{33}$$

38

Where $f_\emptyset(x)$ is the constant function equal product of all the terms in (32) that depend exclusively on the fixed variables $x_{(k+1)}, \ldots, x_{(t)}$, $f_{\{1\}}(x_1)$ is the product of all the terms that depend only on x_1 and on $x_{k+1}, \ldots, x_t$, and so on.

Note that each f_S is a product of shifts of functions $G^{(a)}$. Furthermore, the index sets of a's for distinct f_S are disjoint.

If we call $f_{[k]} := G^{(e)}$, then Equation 33 says that

$$\langle\{f_S\}\rangle_{LU^k} > \varepsilon$$

Choose $\delta = \delta(k, \varepsilon) = \varepsilon^4/2^{O(k)}$ to be small enough so that Lemma 16 implies that there are two functions f_S, f_T, with $S \neq T$, and a variable i, such that $\text{Inf}_i(f_S)$ and $\text{Inf}_i(f_T)$ are both at least δ.

By previous observations on the relation between the f_S and the $G^{(a)}$ and by Lemma 4, we have that there is a δ' and two functions G^a, G^b such that $\text{Inf}_i(G^a)$ and $\text{Inf}_i(G^b)$ are both at least δ'.

Consider now the Fourier transform of a function $G(x) := \mathbb{E}_\eta g(\eta + x)$. It is easy to see that the Fourier coefficients of G satisfy $\hat{G}(\alpha) = (1 - 2\gamma)^{|\alpha|}\hat{g}(\alpha)$. Therefore, for every degree bound d, we have

$$\text{Inf}_i(G) = \sum_{\alpha:i\in\alpha} \hat{G}^2(\alpha) = \sum_{\alpha:i\in\alpha} (1 - 2\gamma)^{|\alpha|}\hat{g}^2(\alpha) \leq (1 - 2\gamma)^d + \sum_{\alpha:i\in\alpha,|\alpha|\leq d} g^2(\alpha) \leq (1 - 2\gamma)^d + \text{Inf}^{\leq d}(g)$$

This means that we can get $\text{Inf}_i^{\leq d}\left(g^{(a)}\right), \text{Inf}_i^{\leq d}\left(g^{(b)}\right) \geq \delta'/2$ if we choose $d = O\left(\gamma^{-1}\log(\delta'^{-1})\right)$. $\square$

8.3 Composition and PCP Construction

The following theorem follows from Theorem 2 and Theorem 27 using standard techniques.

We need a couple of definitions. If $f : \{0,1\}^n \to \{-1,1\}$ is a boolean function, then we define its *folding* as the boolean function $\overline{f}$ defined as follows: $\overline{f}(0, x_2, \ldots, x_n) := f(0, x_2, \ldots, x_n)$ and $\overline{f}(1, x_2, \ldots, x_n) := -f(0, 1 - x_2, \ldots, 1 - x_n)$.

The definition satisfies the following useful properties: (i) if f is a long code, then $f = \overline{f}$, and, (ii) for every f, $\mathbb{E}_x \overline{f}(x) = 0$, that is, $\hat{\overline{f}}(\emptyset) = 0$.

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function, and $\pi : [n] \to [n]$ be a permutation in S_n. Then we let $f \circ \pi : \{0,1\}^n \to \mathbb{R}$ be the function such that $f \circ \pi(x_1, \ldots, x_n) := f(x_{\pi(1)}, \ldots, x_{\pi(n)})$. Here the interesting properties are that, for every coordinate i and degree bound t,

$$\text{Inf}_{\pi(i)}^{\leq t}(f) = \text{Inf}_i^{\leq t}(f \circ \pi)$$

and that if f is the long code of i, g is the long code of j, and $\pi(i) = \pi'(j)$, then $f \circ \pi$ and $g \circ \pi'$ are the same function.

Theorem 29 (Main) *Suppose that the Unique Games Conjecture is true. Then, for every $\delta > 0$, and for every $q \geq 3$,*

$$\text{NP} = \text{PCP}_{1-\delta, 2q/2^q+\delta}[O(\log n), q]$$

and, if we can write $q = 2^k - 1$, then

$$\text{NP} = \text{PCP}_{1-\delta,(q+1)/2^q+\delta}[O(\log n), q]$$

PROOF: Let $H = ([t], E)$ be an hypergraph such that $t + |E| = q$ and $t \leq 1 + \log_2 q$. (Or $t = \log_2(q+1)$ if it is possible to write $q = 2^t - 1$.)

Fix a small constant δ, and consider the δ-noisy H-test. From Theorem 27 we have that there are constants $\tau(\delta)$ and $d(\delta)$ such that if $\{g^a\}_{a \in [t] \cup E}$ are functions accepted by the test with probability at least $1/2^{|E|}$ then the degree-$d(\delta)$ cross-influence of the functions is at least $\tau(\delta)$.

Let γ be a constant smaller than $\delta \tau^2 / d^2$.

From Theorem 2 we know that, if the Unique Games conjecture is true, there is a reduction from SAT to q-ary unique games with the property that a satisfiable instance φ of SAT is mapped into a unique game U_φ of strong value at least $1 - \gamma$ and an unsatisfiable instance φ of SAT is mapped into a unique game U_φ of weak value at most γ.

A PCP for a formula φ is a long code for each of the variables of the unique game U_φ. Let n be size of the alphabet of U_φ.

We consider the following verifier:

- Pick at random a constraint of U_φ, say that it involves the variables $v_1, \ldots, v_q$ and the permutations $\pi_1, \ldots, \pi_q$. Let $f_1, \ldots, f_q$ be the functions written in the proof being checked and which, supposedly, are the long codes of assignments for the variables $v_1, \ldots, v_q$.

- Run the δ-noisy H-test using the functions $\overline{f}_1 \circ \pi_1, \ldots, \overline{f}_q \circ \pi_q$.

Completeness of the test If φ is satisfiable, then consider the proof where for each variable v we write the function f that is the long code of $A(v)$, with $A()$ being an assignment that strongly satisfies a $1 - \gamma$ fraction of constraints. Then the test accepts with probability at least $(1-\gamma) \cdot (1 - q\delta) \geq 1 - (q+1)\delta$.

Soundness of the test Suppose that the test accepts with probability at least $1/2^{|E|} + 2\delta$. Consider the following randomized assignment: for every variable v, consider the function f written in the proof corresponding to v; consider the set of coordinates i such that $\text{Inf}_i^{\leq d}(\overline{f}) \geq \tau$; if the set is empty, give v a random value, otherwise, give v a randomly chosen value from the set. Note that if the set is non-empty then it has size at most d/τ.

Call a constraint *good* if the H-test accepts with probability at least $1/2^{|E|} + \delta$ when that constraint is chosen by test. Then at least a δ fraction of constraints is good.

Consider a good constraint, and say that it involves the variables $v_1, \ldots, v_q$ and the permutations $\pi_1, \ldots, \pi_q$. Let $f_1, \ldots, f_q$ be the functions written in the proof corresponding to the variables $v_1, \ldots, v_q$.

Since the constraint is good, there is a variable i that has degree-d cross-influence at least τ for the functions $\overline{f}_j \circ \pi_j$. That is, there are two functions f_a, f_b such that $\text{Inf}_i^{\leq d}(\overline{f}_a \circ \pi_a) \geq \tau$ and $\text{Inf}_i^{\leq d}(\overline{f}_b \circ \pi_b) \geq \tau$. Then the randomized assignment described above assigns with probability at

least τ/d the value $\pi_a^{-1}(i)$ to v_a, and with probability at least τ/d the value $\pi_b^{-1}(i)$ to v_b. When this happens, the constraint is weakly satisfied.

Overall, the randomized assignments weakly satisfies on average at least a $\delta\tau^2/d^2 > \gamma$ fraction of constraints, which proves that φ was satisfiable.

Conclusion Wrapping up, our PCP verifier has completeness at least $1 - q\delta$ and soundness at most $1/2^{|E|} + 2\delta$. Since δ was chosen arbitrarily at the beginning, the theorem follows. $\square$

8.4 Inapproximability Results

It is an immediate consequence of Theorem 29 that, assuming the unique games conjecture, Max kCSP cannot be approximated within a factor larger than $(k+1)/2^k$ if k is of the form $2^t - 1$. It follows from Theorem 29 and the reductions in [Tre01] that, assuming the unique games conjecture, the Maximum Independent Set problem in graphs of maximum degree D cannot be approximated within a factor larger $(\log D)^c/D$, for sufficiently large D, where c is an absolute constant.

Acknowledgements

We are grateful to Elchanan Mossel for his help with the definitions in Section 6 and with the proof of Lemma 10.

References

[AHRS01] Yonatan Aumann, Johan Håstad, Michael O. Rabin, and Madhu Sudan. Linear-consistency testing. *Journal of Computer and System Sciences*, 4(62):589–607, 2001. 3, 4, 17

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. Preliminary version in *Proc. of FOCS'92*. 2

[Aro02] Sanjeev Arora. How NP got a new definition: a survey of probabilistically checkable proofs. In *Proceedings of the International Congress of Mathematicians*, pages 637–648, 2002. Volume 3. 2

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. Preliminary version in *Proc. of FOCS'92*. 2

[BCH+96] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, and M. Sudan. Linearity testing over characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996. 3, 6

[BLR93] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993. Preliminary version in *Proc. of STOC'90*. 3

[CKK+05]   Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D.Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *Proceedings of the 20th IEEE Conference on Computational Complexity*, 2005. 5

[DS05]   Irit Dinur and Shmuel Safra. On the hardness of approximating minimum vertex-cover. *Annals of Mathematics*, 162(1):439–486, 2005. 5

[EH05]   Lars Engebretsen and Jonas Holmerin. More efficient queries in PCPs for NP and improved approximation hardness of maximum CSP. In *Proceedings of the 22th Symposium on Theoretical Aspects of Computer Science*, pages 194–205, 2005. 2, 4

[Fei02]   Uriel Feige. Approximation thresholds for combinatorial optimization problems. In *Proceedings of the International Congress of Mathematicians*, pages 649–658, 2002. Volume 3. 2

[Gow98]   Timothy Gowers. A new proof of Szeméredi's theorem for progressions of length four. *Geometric and Functional Analysis*, 8(3):529–551, 1998. 1, 6, 7, 11

[Gow01]   Timothy Gowers. A new proof of Szeméredi's theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001. 1, 6, 7, 11, 31

[GT04]   Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. To appear in Annals of Mathematics. math.NT/0404188, 2004. 11, 31

[GT05]   Ben Green and Terence Tao. An inverse theorem for the Gowers U^3 norm. math.NT/0503014, 2005. 1, 7, 11

[Hås01]   Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001. 4

[Has05]   Gustav Hast. Approximating Max kCSP - outperforming a random assignment with almost a linear factor. In *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming*, pages 956–968, 2005. 2, 8

[HW03]   Johan Håstad and Avi Wigderson. Simple analysis of graph tests for linearity and PCP. *Random Structures and Algorithms*, 22(2):139–160, 2003. 2, 8

[Kho02]   Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 767–775, 2002. 2, 4, 5, 12

[KKMO04]   Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 146–154, 2004. 5

[KR03]   Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within $2 - \varepsilon$. In *Proceedings of the 18th IEEE Conference on Computational Complexity*, 2003. 5, 13

[KV05]   Subhash Khot and Nisheeth Vishnoi. The unique games conjecture, integrality gap for cut problems and the embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, 2005. 5

[Sam05]   Alex Samorodnitsky. Hypergraph linearity and quadraticity tests for boolean functions. Manuscript, 2005. 1, 7, 29, 31

[ST98]   Madhu Sudan and Luca Trevisan. Probabilistically checkable proofs with low amortized query complexity. In *Proceedings of the 39th IEEE Symposium on Foundations of Computer Science*, pages 18–27, 1998. 4

[ST00]   Alex Samorodnitsky and Luca Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pages 191–199, 2000. 2, 4, 5, 8, 33, 34

[Tre98a]   Luca Trevisan. Parallel approximation algorithms by positive linear programming. *Algorithmica*, 21(1):72–88, 1998. Preliminary version in *Proc. of ESA'96*. 2

[Tre98b]   Luca Trevisan. Recycling queries in PCPs and in linearity tests. In *Proceedings of the 30th ACM Symposium on Theory of Computing*, pages 299–308, 1998. 4

[Tre01]   Luca Trevisan. Non-approximability results for optimization problems on bounded degree instances. In *Proceedings of the 33rd ACM Symposium on Theory of Computing*, pages 453–461, 2001. 41

[Tre04]   Luca Trevisan. Inapproximability of combinatorial optimization problems. Technical Report TR04-065, Electronic Colloquium on Computational Complexity, 2004. 2

[Vaz01]   Vijay Vazirani. *Approximation Algorithms*. Springer, 2001. 2