Low-degree tests at large distances

Alex Samorodnitsky*

April 16, 2006

Abstract

We define tests of boolean functions which distinguish between linear (or quadratic) polynomials, and functions which are very far, in an appropriate sense, from these polynomials. The tests have optimal or nearly optimal trade-offs between soundness and the number of queries.

In particular, we show that functions with small Gowers uniformity norms behave "randomly" with respect to hypergraph linearity tests.

A central step in our analysis of quadraticity tests is the proof of an inverse theorem for the third Gowers uniformity norm of boolean functions.

The last result has also a coding theory application. It is possible to estimate efficiently the distance from the second-order Reed-Muller code on inputs lying far beyond its list-decoding radius.

*Institute of Computer Science, Hebrew University. `salex@huji.ac.il` This paper is based upon work supported by the Israel Science Foundation under grant 039-716 and by the German-Israeli Foundation under grant I-2052.

1 Introduction

This paper returns to the general question of the relation between number of queries and the probability of error in low-degree tests.

The specific questions we deal with originate within a wider framework of Probabilistically Checkable Proofs (PCPs). The PCP theorem [2, 3] states that it is possible to encode certificates of satisfiability for SAT instances in such a way that a probabilistic verifier using logarithmic number of random bits can check the validity of the certificate with high probability of success, after looking only at a constant number of bits in the encoding. We consider here only PCPs with almost perfect *completeness*, which means that valid certificates are nearly always [1] accepted. Given this, and fixing the number of queries q, we are interested in the best possible *soundness* of the PCP, namely the probability s of accepting an encoding of a false proof.

It is easy to see that, unless $P = NP$, the lower bound $s \geq 1/2^q$ must hold. Stronger lower bounds were given in [15, 26]. The best known lower bound [7] is $s \geq \Omega\left(\frac{q}{2^q}\right)$. From the other direction, the PCP theorem shows that we can achieve $s \leq \frac{1}{2^{O(q)}}$, and it was shown in [8], following [23], that $s \leq \frac{2\sqrt{2q}}{2^q}$. In [24], assuming the Unique Games Conjecture [18], the upper bound was improved to $s \leq q/2^{q-1}$, which is of course (conditionally) best possible, up to constants.

Let us say a few words on the structure of a PCP protocol. In the common paradigm [3] the verifier of a PCP is split into two entities, the *inner* and the *outer* verifiers. Roughly speaking, the outer verifier chooses the (randomized) portion of the proof to be checked by the inner verifier. The inner verifier views the binary string it is given as a boolean function, and looks for a certain combinatorial pattern. If the pattern is not there the proof is rejected. If the inner verifier finds the appropriate property with non-negligible probability over its inputs, the outer verifier can then use this information to validate the PCP statement.

Due to the gap structure inherent in the PCP construction, the decision of the inner verifier is usually dichotomic. This is to say it must accept if the property is satisfied and reject only if the function is very far, in the appropriate sense, from having the property.

In this framework, an often considered property of a boolean function is that of being represented by a low-degree polynomial over a finite field. Here we deal only with the field of two elements and this representation is particularly simple:

Definition 1.1: A boolean function $f : \{0,1\}^n \to \{-1,1\}$ has a degree-d representation if $f(x) = (-1)^{P(x)}$, where $P(x) = P(x_1...x_n)$ is an n-variate polynomial of degree d over $\mathbb{F}_2$. ∎

In our version of the Low-Degree testing problem we are given an oracle access to a boolean function $f : \{0,1\}^n \to \{-1,1\}$ and we want to determine whether

1. The function f can be represented by a degree-d polynomial
2. It is $\frac{1}{2} - \epsilon$ far from any function with such representation.

[1] See, e.g., [27] for a precise definition

The distance between two functions is a fraction of points in which they disagree.

Low-degree tests we consider have perfect completeness, namely in case (1) they always accept. We now define the soundness of a test.

Definition 1.2: A low-degree test has soundness s if for any function f that is $\frac{1}{2} - \epsilon$ far from degree-d polynomials, the test accepts f with probability at most $s + \phi(\epsilon)$ where $\phi(x) \rightarrow_{x \rightarrow 0} 0$. ∎

Designing a low-degree test with a good trade-off between the number of queries and the soundness is a step towards a PCP construction. There are several ways in which such a result needs to be augmented to lead to a full PCP construction. We refer to the discussion in [24]. It seems, however, that in most cases in which this extension process succeeded, the obtained PCP inherited the relevant parameters (number of queries, soundness) of the low-degree test.

Degree-1 (linear) tests with asymptotically optimal asymptotic trade-off between the number of queries and the soundness where given in [23]. In the same paper these tests were extended to PCP constructions with similar parameters.

A natural way to improve the PCP parameters further is to consider additional combinatorial tests.

In this paper we study degree-2 tests and relaxed versions of the degree-1 test. Here is a brief overview of our main results.

- We define and analyze a degree-1 test with relaxed rejection criteria whose trade-off between the number of queries and the soundness is asymptotically optimal and is much better than that achievable by the standard linearity tests. A different (and easier) analysis of this test was given in [24]. In that paper we were also able to extend the test to a conditional PCP construction (assuming the Unique Games Conjecture [18]) with an optimal number of queries vs. soundness trade-off.
- We define and analyze a degree-2 test with a very good trade-off between the number of queries q and the soundness s. (We conjecture this trade-off to be asymptotically optimal.) A technical ingredient of this result has a natural interpretation in the framework of error-correcting codes. We give a tight analysis of the acceptance probability of a natural local test of [1] for the second-order Reed-Muller code at distances near the covering radius of this code. As a consequence, it turns out to be possible to estimate efficiently the distance from this code on inputs lying far beyond its list-decoding radius.

Our analysis of these tests is based on several technical assertions which could be of independent interest, and which we describe next.

- We give a tight analysis of the Abelian Homomorphism testing problem for some families of groups, including powers of $\mathbb{Z}_p$. The central technical claim, which we state here for the special case of $p = 2$, is that if a function $\phi: \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2^n$ satisfies $Pr\left(\phi(x+y) = \phi(x) + \phi(y)\right)$ with probability bounded away from zero, then there is a matrix $D \in M_{n,n}\left(\mathbb{Z}_2\right)$ such that a linear transformation $\psi : x \mapsto Dx$ coincides with ϕ on a non-negligible fraction of the inputs.

- We introduce and study the notion of a *generalized average* of a function f over $\{0,1\}^n$. A generalized average is a non-linear functional on the space of real (or complex) valued functions on the boolean cube. It is associated with a binary matrix M and it measures the average over a certain family of subsets of $\{0,1\}^n$, defined by M, of products of f over each subset. Generalized averages arize naturally in the analysis of low-degree tests. An important special case is when this family consists of all the affine subsets of $\{0,1\}^n$ of a fixed dimension d. The generalized average in this case turns out to measure (a power of) a norm of the function f. These norms are the *Gowers uniformity norms* [11] and they measure, in a certain sense, a proximity of the function to a polynomial of degree d.
  - We show that a function with a large third uniformity norm is somewhat close to an n-variate quadratic polynomial over $\mathbb{F}_2$. Similar results for finite Abelian groups of cardinality indivisible by 6 have been independently proved in [14].
  - We show that functions on which the *hypergraph linearity tests* defined in [23] fail with non-negligible probability have large uniformity norms.
  - We observe that functions with small uniformity norms are *pseudorandom* in the sense of [11], and briefly discuss pseudorandom properties of such functions in our context.

In the next sections we give a more detailed description of the background and of the results in this paper. The proofs are given in the Appendices.

Organization

We describe relaxed linearity tests in Section 2. Degree-2 tests and properties of the Reed-Muller code of order 2 are described in Section 3. Abelian homomorphism testing is discussed in Section 4. Section 5 gives more details on the technical tools used, in particular their connections with recent work in additive number theory. A notion of pseudorandomness of boolean functions which comes from additive number theory is introduced and briefly discussed.

2 Degree-1 Tests

This is the simplest and the most useful case in practice. A boolean function f has a degree-1 representation if $f(x) = (-1)^{\langle a,x\rangle + b}$, where a is a fixed vector in $\{0,1\}^n$, and b is a fixed constant in $\{0,1\}$. Hence in this case the tester has to decide whether the function is linear [2] or is far from every linear function.

A simple linearity test with three queries was defined in [6]. [3]

Choose uniformly at random $x, y \in \{0,1\}^n$ If $f(x)f(y)f(x+y) = 1$ then accept else reject

[2] or rather *affine*. In practice the function is usually tested for linearity ($b = 0$). The two testing problems are essentially equivalent, and we occasionally will, with some abuse of meaning, refer to both as *linearity testing problems*.

[3] We observe that to transform this test to an affinity (degree-1) test, it suffices to replace 1 with $f(0)$ in the definition of the test.

It is shown in [4] that if this test accepts f with probability $\frac{1}{2}+\delta$ then f is $\frac{1}{2}-2\delta$ close to a linear function. Therefore, according to our definition, this test has soundness $s=\frac{1}{2}$.

Independent repetition of $q/3$ basic tests leads to a test with q queries and soundness $s=\frac{1}{2^{2q/3}}$. To improve the trade-off between q and s, more complex tests have to be considered. It turns out that it is possible to associate such a test with any given graph. Fix a graph $G=(V,E)$ on t vertices, The following test is a *dependent* combination of the basic tests of [6].

Choose uniformly at random $x_1,\dots,x_t \in \{0,1\}^n$ If $\prod_{i\in e} f(x_i)\cdot f(\sum_{i\in e} x_i) = f(0)$ for all $e\in E$ then accept else reject

These *graph tests* were defined in [27]. A graph test associated with a graph $G=(V,E)$ runs $|E|$ correlated copies of the basic linearity test. In [23] it was shown that for functions which are far from degree-1 polynomials (this is to say, have small Fourier coefficients), these copies of the basic test behave essentially independently. More precisely, the soundness of this test is $s=1/2^{|E|}$. Of course, the total number of queries is $q=|V|+|E|$. In particular, choosing G to be the complete graph on t vertices, we obtain an affinity test with $q=\binom{t}{2}+t$ and $s=\frac{1}{2^{\binom{t}{2}}}$. This means that $s\approx\frac{2^{\sqrt{2q}}}{2^q}$.

A natural generalization of graph tests to *hypergraph tests* was given in [23]. Let $H=(V,E)$ be a hypergraph on t vertices and consider the following test:

Choose uniformly at random $x_1,\dots,x_t \in \{0,1\}^n$ If $\prod_{i\in e} f(x_i)\cdot f(\sum_{i\in e} x_i) = f^{|e|+1}(0)$ for all $e\in E$ then accept else reject

A hypergraph test runs $|E|$ copies of the basic linearity test, where $|E|$ is now the number of hyper-edges. Unfortunately, it is not true that, for functions far from degree-1 polynomials, these copies behave independently. Consider a function $f(x)=(-1)^{x_1x_2+\ldots+x_{n-1}x_n}$. This function is maximally far from all degree-1 polynomials (it is a *bent function*), but any hypergraph test with $q=|V|+|E|$ queries accepts this function with probability at least $\frac{2^{\Omega(\sqrt{q})}}{2^q}$ [23]. More generally, we show in [24] that this is true for any non-adaptive linearity test that always accepts linear functions.

Our results

The starting point of this work was the realization that the function f we have described is a quadratic polynomial and that it is accepted by a hypergraph test with non-negligible probability, because, roughly speaking, the basic ingredient of this test takes a discrete derivative of the tested function and compares it to zero. The order of the derivative is essentially given by the cardinality of the hyperedges. We will say more about this in Section 6 and in the full version of the paper. The natural question then is whether quadratic polynomials, and, more generally, low-degree polynomials, are the only obstructions to better performance by hypergraph linearity tests.

We give a partial (affirmative) answer to this question for general hypergraphs. We are able to answer this question completely for hypergraphs of maximal edge-size 3 and for quadratic polynomials. The answer is again positive. We conjecture the answer to be positive for general hypergraphs and low-degree polynomials.

We prove two claims. These are the main technical results of this paper.

The first claim is valid for any hypergraph. First, we define Gowers uniformity norms.

Definition 2.1: Let $f:\ \{0,1\}^n \to \mathbb{R}$ be a function, and $d \geq 1$ be an integer. The d-th Gowers uniformity norm (for the group $\mathbb{Z}_2^n$) is given by

$$\|f\|_{U_d} = \left[\mathbb{E}_{x,y_1,\ldots,y_d} \prod_{S\subseteq[d]} f\left(x + \sum_{i\in S} y_i\right)\right]^{1/2^d}$$

Here $x, y_1, ..., y_d$ are chosen uniformly and independently at random from $\{0,1\}^n$. ∎

Theorem 2.2: *Let $H = (V,E)$ be a hypergraph with maximal edge-size d. Then the probability that the linearity test associated with H accepts a boolean function f is bounded by*

$$\frac{1}{2^{|E|}} + \|f\|_{U_d}$$

Another (and easier) proof of this theorem and its generalization to several functions is given in [24].

The second claim is that a boolean function with a large third unformity norm is somewhat close to a quadratic polynomial.

Theorem 2.3: *Let $f:\ \{0,1\}^n \to \{-1,1\}$ be a function such that $\|f\|_{U_3} \geq \epsilon$. Then there exists a quadratic polynomial g such that the distance between f and g is at most $\frac{1}{2} - \epsilon'$. Here one can choose $\epsilon' \geq \Omega\left(exp\left\{-\left(\frac{1}{\epsilon^C}\right)\right\}\right)$ for an absolute constant C.*

Consider the following relaxed degree-1 testing problem. Given an oracle access to a boolean function $f : \{0,1\}^n \to \{-1,1\}$ and an integer $d \geq 2$ we want to determine whether

1. The function f can be represented by a degree-1 polynomial.
2. $\|f\|_{U_d} \leq \epsilon$.

Once again we want tests with perfect completeness. The soundness of the test is defined as in Definition 1.2.

Remark 2.4: Let us point out, that this test is indeed a relaxation of the standard' degree-1 test. It is known [11] that uniformity norms $\|f\|_{U_d}$ of f are monotone increasing in d. It is easy to see that the second uniformity norm is the same as the l_4 norm of the Fourier transform of f: $\|f\|_{U_2} = \left(\sum_{\alpha\in\{0,1\}^n} \hat{f}^4(\alpha)\right)^{\frac{1}{4}} \geq \max_{\alpha\in\{0,1\}^n} |\hat{f}(\alpha)|$. This means that the functions the test has to reject are at least $\frac{1}{2} - \frac{\epsilon}{2}$ far from degree-1 polynomials. ∎

It is a direct consequence of Theorem 2.2 that hypergraph tests solve the relaxed testing problem with the "right" soundness.

Theorem 2.5: *Let $d \geq 2$ and let $H = (V, E)$ be a hypergraph with maximal edge-size d. Then the hypergraph linearity test associated with H solves the relaxed degree-1 testing problem with perfect completeness and soundness $1/2^{|E|}$.*

Choosing H to be a complete d-uniform hypergraph on $t \approx q^{1/d}$ vertices leads to a test with q queries and soundness $s \leq \frac{2^{\Omega\left(q^{1/d}\right)}}{2^q}$. This trade-off is shown to be asymptotically optimal in [24].

It remains to observe that Theorem 2.5 together with Theorem 2.3 imply that the complete 3-uniform hypergraph test distinguishes between linear functions and functions which are far from quadratic polynomials with optimal soundness of $s \leq \frac{2^{\Omega\left(q^{1/3}\right)}}{2^q}$.

3 Second-Order Reed-Muller Codes

A binary error-correcting code [19] of length N and normalized distance δ is a subset of $\{0,1\}^N$ in which any two distinct elements disagree on at least δ-fraction of the domain (the coordinates). This allows for error-correction: a corrupted codeword (element of the code) with less than $\delta/2$-fraction of the errors can, in principle, be recovered by going to the unique nearest element of the code. We call $\delta/2$ the unique-decoding radius of the code.

Finding the nearest codeword can be computationally hard. Here we are interested in efficient error-correction.

An important example of a code of length $N = 2^n$ is the subset of $\{0,1\}^N$ whose elements are evaluations of n-variate degree d polynomials over $\mathbb{F}_2$. This is the *Reed-Muller (RM) code* of order d. Efficient error-correcting algorithms for RM codes were given in [21].

One can go beyond unique decoding. It is an easy consequence of the Johnson bound for constant-weight codes [19] that there is $\lambda > \delta/2$ such that there could be only a few (polynomially many in N) codewords at distance λ from a corrupted codeword. We call maximal λ with this property the list-decoding radius of the code. For many codes there are efficient list-decoding algorithms [25] that, for any λ smaller than list-decoding radius, recover all the codewords within distance λ from the corrupted codeword. To the best of our knowledge there are no such algorithms for binary RM codes of order larger than 1.

Another useful property of a code is local testability [10]. A code is locally testable if there exists an efficient randomized algorithm (test) which, given an access to a putative codeword

$f \in \{0,1\}^N$, examines a finite number of coordinates of f and decides whether f is a codeword. We want the test always to accept valid codewords, and to minimize the probability s of accepting an invalid codeword, given the number of queries q. The questions we discuss in this paper fall naturally into the framework of local testability of Reed-Muller codes. In fact, we deal with a special case (a promise problem) in which the putative codeword is promised either to lie in the code or to be $(\frac{1}{2} - \epsilon)$ far from the code. We remark that, in the general case, the probability the test accepts an invalid codeword will necessarily depend also on its distance from the code.

Example 3.1: A good example for the notions we have discussed is the first order Reed-Muller code, also known as the Hadamard code. The distance of this code is $1/2$, and therefore its unique-decoding radius is $1/4$. However, it is efficiently list-decodable for any distance $\lambda < 1/2$ [9].

The Hadamard code is also locally testable. In fact, the basic linearity test of [6] is a good 3-query local test for this code. [4] studies the dependence of the probability this test accepts an invalid codeword on its distance from the code. For distances close to $1/2$ the analysis it tight, and the probability of acceptance is shown to be upper bounded by 1 minus the distance. ∎

Local testability of Reed-Muller codes of any fixed order d was proved in [1]. The basic test in [1] (presented here with a small twist to adopt it to our setting) chooses independently at random $d+2$ vectors $x, y_1, ..., y_{d+1}$ in $\{0,1\}^n$, and computes the product of the tested function over the d-dimensional affine subspace of $\{0,1\}^n$ given by $x + Span\,(y_1, ..., y_{d+1})$. If the product is 1 the test accepts. Otherwise it rejects. This is a natural generalization of the linearity test of [6]. While that test can be interpeted as taking a random second directional derivative and checking whether it vanishes, the test of [1] amounts to checking whether a random derivative of order $d+1$ vanishes. [1] studies the dependence of the probability this test accepts an invalid codeword f on its distance from the code. (We observe that this probability is precisely $\frac{1+\|f\|_{U_d}^{2^d}}{2}$, cf. Definition 2.1). In particular it is shown that, for distances larger than 2^{-d}, the probability of acceptance is upper bounded by $1 - \Omega\left(d^{-1}2^{-d}\right)$. Thus, for $d = 2$, the probability of acceptance is upper bounded by some constant smaller than 1.

Our results

We study the probability of error of the test of [1] for the second-order Reed-Muller code and for distances close to $1/2$. We provide a tight analysis for this case, showing this probability to be essentially upper bounded by 1 minus the distance. Specifically, by Theorem 2.3, if this probability is larger that $1/2 + \epsilon$ then there is a quadratic polynomial whose distance from the tested function is at most $1/2 - \epsilon'$.

Our result has a following coding interpretation. Although the list-decoding raduis of the second-order Reed-Muller code is $1/4$ [19], it is possible to determine whether the distance of a given function $f \in \{0,1\}^N$ from the code is strictly smaller than the covering radius of the code, which is $1/2 - o(1)$. [4] More precisely, we have the following proposition.

[4] This is also, with overwhelming probability, the typical distance of an element of $\{0,1\}^N$ from the code.

Proposition 3.2: *There is a positive constant C such that, given a function $f : \{0,1\}^n \to \{-1,1\}$, and a parameter $\delta > 0$, it is possible to determine, with probability arbitrarily close to 1, and in time linear in $\frac{1}{\delta}$, which of the two following (mutually non-exclusive) options holds:*

- *The distance of f from the quadratic polynomials is at least $\frac{1}{2} - \Omega\left(\delta^{1/2}\right)$.*
- *The distance of f from the quadratic polynomials is at most $\frac{1}{2} - \exp\left\{\left(-\frac{1}{\delta}\right)^C\right\}$.*

Combining theorems 2.5 and 2.3 leads to our main result in this section, an analysis of hypergraph degree-2 (quadraticity) tests.

Given a 3-uniform hypergraph $H = ([t], E)$ on t vertices, the test is defined as follows.

Choose uniformly at random $x_1, \ldots, x_t \in \{0,1\}^n$ If for all $e = \{i,j,k\} \in E$ holds $\mathbb{E}_{x_i,x_j,x_k} f(x_i)f(x_j)f(x_k)f(x_i+x_j)f(x_i+x_k)f(x_j+x_k)f(x_i+x_j+x_k) = f(0)$ then accept else reject

Theorem 3.3: *Let $H = (V,E)$ be a 3-uniform hypergraph. Then the hypergraph quadraticity test solves the degree-2 testing problem with perfect completeness and soundness $1/2^{|E|}$.*

Choosing H to be a complete 3-uniform hypergraph on $t \approx q^{1/3}$ vertices leads to a test with q queries and soundness $s \le \frac{2^{\Omega\left(q^{2/3}\right)}}{2^q}$.

Discussion

Analyzing acceptance probability of a low-degree test at distances larger than the unique-decoding radius seems to require a different set of techniques. It general, to prove that a code is locally testable, one needs to upper bound acceptance probability by a function of the distance. This is achieved by showing that if acceptance probability of the test on an element $f \in \{0,1\}^N$ is higher than a certain threshold, there is a codeword g not far from f. In most cases the test itself is used to efficiently "decode" f, viewed as a corrupted codeword, to the unique nearest codeword g. This approach is harder to implement when there are several possible codewords to choose from, and symmetry breaking in required. The only example we are aware of is the Hadamard code. In this case one is assisted by the fact that the elements of the code are pairwise orthogonal (as vectors over the reals). In particular, for any $\epsilon > 0$, there could be only a constant number of codewords at distance smaller than $1/2 - \epsilon$ from f. This no longer holds for degree-2 polynomials. For instance, the list-decoding radius here is $1/4$. Our main tools in this case are harmonic analysis and additive number theory. In fact, a significant part of our proof follows the approach of Gowers [11] in his proof of Szemeredi's theorem for arithmetic progressions of length 4.

4 Abelian homomorphism testing

Let G and H be two finite Abelian groups. In the Abelian Homomorphism testing problem we are given an oracle access to a transformation $\phi : G \to H$ and we have to decide whether ϕ is a homomorphism or is at least δ-far from any homomorphism between G and H. This problem is a generalization of the linearity testing problem, in which case $G = H = \mathbb{Z}_2$. It was first studied in [6], where the following natural generalization of the basic linearity test was suggested: choose $x, y \in G$ at random and check whether $\phi(x + y) = \phi(x) + \phi(y)$. The analysis of this test leads to the following question.

Let $\phi : G \to H$ such that the group law for ϕ holds with positive probability.

$$Pr_{x,y \in G}\left(\phi(x) + \phi(y) = \phi(x + y)\right) \geq \epsilon.$$

Let ρ be the maximal ϵ' such that there exists a homomorphism ψ from G to H such that $Pr_x\left(\phi(x) = \psi(x)\right) \geq \epsilon'$. The question is whether ρ can be lower bounded in terms of a function of ϵ that is independent of $|G|$. In [6] this is shown to be true if $\epsilon > 7/9$. This lower bound on ϵ is also necessary [5].

If both G and H are powers of $\mathbb{Z}_2$, the lower bound on ϵ was relaxed to $\epsilon > 83/128$ [4].

Our results

We show the following theorem to be a simple consequence of two results [11, 22] in additive number theory.

Theorem 4.1: *Let p be a prime number, and let $\epsilon > 0$. Let G be a p-group of order r and let H be a power of $\mathbb{Z}_p$. Let $\phi : G \to H$ such that*

$$Pr_{x,y \in G}\left(\phi(x) + \phi(y) = \phi(x + y)\right) \geq \epsilon.$$

Then there exists a homomorphism $\psi : G \to H$ such that

$$Pr_{x \in G}\left\{\phi(x) = \psi(x)\right\} \geq c \cdot r^{-c'} \cdot \epsilon^{c''},$$

where c, c', c'' are absolute constants (independent of the groups G, H).

In particular, if both G and H are powers of $\mathbb{Z}_2$, ρ can be lower bounded by a function of ϵ, for any $\epsilon > 0$. In testing terms, this means that the acceptance probability of the basic test of [6] goes to zero as the distance from the code (the set of all homomorphisms) goes to one.

5 Tools

In this section we discuss the technical tools used in this paper. We believe these tools, and their connection to recent results in additive number theory, might be of independent interest.

5.1 Generalized averages

Let $H = (V, E)$ be a hypergraph on t vertices. Given a boolean function $f : \{0,1\}^n \to \{-1,1\}$, the acceptance probability of the linearity test associated with H on f is easily seen (cf. Appendix 7) to be an average of expressions of the following type. Let $\mathcal{S} = \{e_1, \ldots, e_T\}$ be a family of edges of H, this is to say subsets of $\{1, \ldots, t\}$. We define the *average* of f on $\mathcal{S}$ in the following way:

$$\mathbb{E}_{\mathcal{S}}(f) := \mathbb{E}_{y_1,\ldots,y_t \in \{0,1\}^n} \prod_{j=1}^{T} f\left(\sum_{i \in e_j} y_i\right). \tag{1}$$

The operator $\mathbb{E}_{\mathcal{S}}$ is naturally associated with a binary matrix A whose columns are characteristic vectors of e_j. We will also denote this operator by $\mathbb{E}_A$.

Example 5.1: Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. Then $\mathbb{E}_A(f) = \mathbb{E}_{x,y} f(x)f(y)f(x+y)$ is the basic linearity test of [6]. ∎

For $A = [1]$, the average of f over A is, of course, simply the expectation $\mathbb{E}f$. The notion of generalized average is naturally extended to real or complex valued functions on $\{0,1\}^n$.

The analysis of the probability of acceptance of a hypergraph test entails studying generalized averages of functions. In particular, we would like to upper bound such averages by expressions which are convenient to deal with.

With this in mind, we define a useful family of binary matrices.

Definition 5.2: For an integer $k \geq 1$ let A_k be a $(k+1) \times 2^k$ matrix of the following form: the last row of A_k is an all-1 vector. Removing this last row gives a $k \times 2^k$ matrix whose columns are all binary vectors of length k (in an arbitrary order). ∎

Observe that $E_{A_k}(f)$ is precisely $\|f\|_{U_k}^{2^k}$.

We prove several properties of generalized averages in Appendix 7, leading to the following main claim. This is essentially a restatement of theorem 2.2.

Theorem 5.3: *Assume that all the columns in A are distinct and have at most k ones. Then for any Boolean function f*

$$\left|\mathbb{E}_A(f)\right| \leq \left(\mathbb{E}_{A_k}(f)\right)^{\frac{1}{2^k}} = \|f\|_{U_k}$$

5.2 Gowers norms and pseudorandomness

In the previous subsection we have seen how Gowers uniformity norms $\|\cdot\|_{U_k}$ appear naturally in the analysis of linearity tests. These norms were originally defined in [11] and were instrumental in the new proof of Szemeredi's theorem on arithmetic progressions given in that paper. We refer to [11], [14] for a more detailed discussion. Here let us briefly mention that, intuitively, the k-th uniformity norm of a function is high if this function has a non-negligible correlation with a polynomial of degree $k-1$. This is to say, this function has a non-trivial combinatorial structure. On the other hand, if a function has small uniformity norms, we would like to deduce that it is 'pseudorandom', in an appropriate sense. In particular, it is shown in [11] that if a characteristic function of a subset of the integers has small uniformity norms, then the number of arithmetic progressions it contains is similar to that contained by a random subset of the same size.

This notion of pseudorandomness naturally generalizes (and strengthens) the standard notion of a boolean function (or a set) being pseudorandom if its non-zero Fourier coefficients are small. In fact, the maximal size of a Fourier coefficient is controlled by the second uniformity norm. Since uniformity norms are monotone increasing, if a function has a small k-th uniformity norm, $k \geq 2$, it is also pseudorandom in the usual sense. This is, of course, intuitively clear, since a function far from degree-$(k-1)$ polynomials is, in particular, far from linear polynomials.

In our context, a function f with a small k-th uniformity norm, is pseudorandom in the following sense. Consider a linearity test associated with a hypergraph $H = (V, E)$ with maximal edge-size k. Theorem 2.2 implies that the $|E|$ copies of the basic linearity test that H runs on f behave essentially independently.

5.3 Quadratic Fourier Analysis

We would now like to give a more specific meaning to the intuitive notion that a function with a high k-th uniformity norm should have a non-trivial combinatorial structure, presumably a non-trivial correlation with a polynomial of degree $k-1$.

Unfortunately, at this point, we can only do it for $k = 3$. By Theorem 2.3, if $\|f\|_{U_3} \geq \epsilon$ then there exists a quadratic polynomial g such that the distance between f and g is at most $1/2 - \epsilon'$, for ϵ' depending on ϵ only.

We conjecture a similar statement to be true for any fixed k. A step in this direction was made in [24], where a function with a high k-th uniformity norm is shown to have variables with large influence.

Similar results for $k = 3$, but replacing $\mathbb{Z}_2^n$ by finite Abelian groups of cardinality indivisible by 6, have been independently proved by Green and Tao [14]. The dependence of ϵ' on ϵ in both cases is super-exponential. In [14] this dependence is improved in the following way: it is shown, specializing here to $\mathbb{Z}_5$ for clarity, that one can find a subspace V of $\mathbb{Z}_5^n$ of a fixed co-dimension and a family of quadratic polynomials g_y indexed by cosets of V, such that typically f is $1/2-\epsilon''$ close to g_y on $y+V$, where the dependence of ϵ'' on ϵ is polynomial. This extension turns out to be useful in obtaining good bounds on arithmetic progressions of length 4 in subsets of $\mathbb{Z}_5^n$ (and in general finite Abelian groups). In this context, Green and Tao introduce the notion of *quadratic Fourier analysis* [12]. According to this point of view, the subject of classical Fourier analysis is to represent a function as a combination of several linear functions (elements of the Fourier basis) it has non-negligible correlation with (i.e., corresponding Fourier coefficients are large), and of a 'random' remainder (a function with small Fourier coefficients). In quadratic Fourier analysis, a function is approximated by a combination of quadratic polynomials. This approach has proven to be quite effective in additive number theory [13, 14] and in ergodic theory [16, 28], in situations in which classical Fourier analysis fails.

Theorems 2.3 and 3.3 can be viewed as an application of quadratic Fourier analysis on $\mathbb{Z}_2^n$ to boolean functions. We suggest that this tool might have other applications as well. (Among other things, it should be possible to extend Theorem 2.3 to obtain results similar to those of [14], but we haven't checked the details.)

6 Appendix A: A Proof of Theorem 2.3

We start with a short discussion on discrete directional derivatives of functions on $\{0,1\}^n$.

Let $f : \{0,1\}^n \to \{-1,1\}$ be a boolean function, and y be a vector in $\{0,1\}^n$. We define the "derivative of f in direction y" by

$$f_y(x) = f(x)f(x+y)$$

The transformation $f \mapsto f_y$ is a linear operator. This operator decreases the degree of the polynomial representation of a function: if f is representable by an n-variate polynomial of degree d, then f_y is representable by a polynomial of degree $d-1$.

We define recursively $f_{y_1,y_2} = (f_{y_1})_{y_2}$. It is easy to see that $f_{y_1,y_2} = f_{y_2,y_1}$, and in fact $f_{y_1,y_2}(x) = f_{y_2,y_1}(x) = f(x)f(x+y_1)f(x+y_2)f(x+y_1+y_2)$. Similarly, the k-th order directional derivative $f_{y_1,\ldots,y_k}$ of f with respect to $y_1,\ldots,y_k$ at a point x is given by

$$f_{y_1,\ldots,y_k}(x) = \prod_{S\subseteq[k]} f\left(x + \sum_{i\in S} y_i\right)$$

If a function f is a polynomial of degree d, then the derivative $f_{y_1,\ldots,y_k}$ is a polynomial of degree $d-k$, for all choices of linearly independent $y_1,\ldots,y_k$ [1, 19]. In particular, the $(k+1)$-th derivative of a degree-k polynomial vanishes (in our terms, it is identically 1).

Observe that, in light of the definition above, the claim of Theorem 2.3 can be interpreted as follows: if a random third derivative of a function vanishes with probability greater than 1/2 then the function is somewhat close to a quadratic polynomial.

The proof of the theorem involves several technical lemmas. The main tools are Fourier analysis on $\mathbb{Z}_2^n$ ([17]) and additive number theory.

In the following the Greek letters $\epsilon, \epsilon', \delta, \delta'$ will denote absolute positive constants (independent of n) whose value may fluctuate.

Lemma 6.1: *For a function* $f:\ \{0,1\}^n \to \mathbb{R}$

$$\|f\|_{U_3}^8 = \mathbb{E}_y \sum_\alpha \hat{f_y}^4(\alpha)$$

Proof: We start with proving

$$\|f\|_{U_2}^4 = \sum_\alpha \hat{f}^4(\alpha)$$

Indeed,

$$\|f\|_{U_2}^4 = \mathbb{E}_{x,y,z} f(x)f(x+y)f(x+z)f(x+y+z) = \mathbb{E}_x\left(f(x)\cdot \mathbb{E}_{y,z} f(x+y)f(x+z)f(x+y+z)\right)$$

Introducing new variables $u = x+y$, $v = x+z$, this equals to

$$\mathbb{E}_x\left(f(x)\cdot \mathbb{E}_{u,w} f(u)f(w)f(x+u+w)\right) = \mathbb{E}_x\left(f(x)\cdot \mathbb{E}_u (f*f)(x+u)\right) =$$

$$\mathbb{E}_x f(x)(f*f*f)(x) = \langle f, f*f*f\rangle = \left\langle \hat{f}, \hat{f}^3 \right\rangle = \sum_\alpha \hat{f}^4(\alpha)$$

Now,

$$\|f\|_{U_3}^8 = \mathbb{E}_{x,y,z,w} f(x)f(x+y)f(x+z)f(x+w)f(x+y+z)f(x+y+w)f(x+z+w)f(x+y+z+w) =$$

$$\mathbb{E}_y \mathbb{E}_{x,z,w} f_y(x) f_y(x+z) f_y(x+w) f_y(x+z+w) = \mathbb{E}_y \sum_\alpha \hat{f_y}^4(\alpha)$$

∎

Corollary 6.2: *Assuming f is boolean and $\|f\|_{U_3} \geq \epsilon$, there exist constants δ, δ' and a choice function $\phi : \{0,1\}^n \to \{0,1\}^n$ such that*

$$Pr_y\left(|\hat{f}_y(\phi(y))| \geq \delta \right) \geq \delta'.$$

Proof: The derivatives f_y are also boolean functions, and therefore

$$\mathbb{E}_y \sum_\alpha \hat{f_y}^4(\alpha) \leq \mathbb{E}_y \max_\alpha \hat{f_y}^2(\alpha) \cdot \sum_\alpha \hat{f_y}^2(\alpha) = \mathbb{E}_y \max_\alpha \hat{f_y}^2(\alpha)$$

∎

Let A be an $n \times n$ matrix over $\mathbb{F}_2$. If $g = (-1)^{\langle Ax, x\rangle + a}$ is a quadratic polynomial[5], then $f_y(x) = (-1)^{\left\langle (A+A^t)y, x\right\rangle + a} = (-1)^{\langle By, x\rangle + a}$. Here $B = A + A^t$ is a symmetric matrix with a zero diagonal. So for a quadratic polynomial the choice function $\phi(y) = By$ is linear, and of a special form. We will therefore look for similar properties of the choice function in our case.

It is sufficient to find a choice function that coincides with an appropriate linear function with positive probability. This will follow from an observation that if derivatives of two boolean functions are close on average then so are the functions themselves (up to a linear shift).

Lemma 6.3: *For boolean functions f, g:*

$$\mathbb{E}_x \left(\langle f_x, g_x\rangle\right)^2 = \sum_\alpha \widehat{fg}^4(\alpha).$$

Proof:

$$\mathbb{E}_x \left(\langle f_x, g_x\rangle\right)^2 = \mathbb{E}_x \mathbb{E}_{y_1,y_2} f_x(y_1) g_x(y_1) f_x(y_2) g_x(y_2) =$$

$$\mathbb{E}_x \mathbb{E}_{y_1,y_2} f(y_1) f(y_1+x) g(y_1) g(y_1+x) f(y_2) f(y_2+x) g(y_2) g(y_2+x) =$$

$$\mathbb{E}_x \left(\mathbb{E}_y (fg)(y)(fg)(y+x)\right)^2 = \mathbb{E}_x \left((fg)*(fg)\right)^2(x) = \sum_\alpha \widehat{fg}^4(\alpha).$$

∎

[5] Observe that, working with the field of 2 elements, we can incorporate the linear term of a quadratic form in the exponent into the quadratic term, by modifying the diagonal of the matrix appropriately.

Corollary 6.4: *Let B be a symmetric matrix with a zero diagonal such that*

$$\mathbb{E}_y \hat{f_y}^2(By) \geq \epsilon.$$

Then there exists a quadratic polynomial g such that

$$\|f - g\| \leq \frac{1}{2} - \epsilon'.$$

Proof: Let A be a matrix such that $A + A^t = B$. Consider a quadratic polynomial $h(x) = (-1)^{\langle x, Ax \rangle}$. We have

$$\mathbb{E}_x \left(\langle f_x, h_x \rangle \right)^2 = \mathbb{E}_x \hat{f_x}^2(Bx) \geq \epsilon'.$$

By lemma 6.3 there is a vector α such that $|\widehat{fh}(\alpha)| \geq \epsilon'$. This implies that there is a choice of $a \in \{0,1\}$ such that for a quadratic polynomial $g(x) = (-1)^{\langle x, Ax \rangle + \langle x, \alpha \rangle + a}$ holds

$$\|f - g\| \leq \frac{1}{2} - \epsilon'.$$

∎

We start by finding a weakly linear choice function. This is made possible by the following observation.

Lemma 6.5:

$$\mathbb{E}_{x,y} \sum_{\alpha, \beta} \hat{f_x}^2(\alpha) \hat{f_y}^2(\beta) \widehat{f_{x+y}}^2(\alpha + \beta) = \mathbb{E}_y \sum_{\alpha} \hat{f_y}^6(\alpha).$$

Proof: We start with an observation that for a boolean function f and for any x, s in $\{0,1\}^n$ holds $(f_x * f_x)(s) = (f_s * f_s)(x)$. Indeed, expanding

$$(f_x * f_x)(s) = \mathbb{E}_y f_s(y) f_s(x+y) = \mathbb{E}_y f(y) f(y+s) f(x+y) f(x+y+s).$$

Define a function $F : \{0,1\}^n \rightarrow \{-1,1\}$ by taking $F(y) = f(y) f(y + (x+s))$. Then the last expression is $(F * F)(x) = (F * F)(s)$. Expanding $(f_s * f_s)(x)$ we get the same result.

Now,

$$\mathbb{E}_{x,y} \sum_{\alpha, \beta} \hat{f_x}^2(\alpha) \hat{f_y}^2(\beta) \widehat{f_{x+y}}^2(\alpha + \beta) =$$

$$\mathbb{E}_{x,y} \sum_{\alpha, \beta} \mathbb{E}_{u,u'} f_x(u) f_x(u') w_\alpha(u+u') \mathbb{E}_{v,v'} f_y(v) f_y(v') w_\beta(v+v') \mathbb{E}_{z,z'} f_{x+y}(z) f_{x+y}(z') w_{\alpha+\beta}(z+z') =$$

$$\mathbb{E}_{x,y} \mathbb{E}_s \mathbb{E}_{u,v,z} f_x(u) f_x(u+s) f_y(v) f_y(v+s) f_{x+y}(z) f_{x+y}(z+s) =$$

$$\mathbb{E}_s \mathbb{E}_{x,y} \left(f_x * f_x \right)(s) \left(f_y * f_y \right)(s) \left(f_{x+y} * f_{x+y} \right)(s) =$$

$$\mathbb{E}_s \mathbb{E}_{x,y} \left(f_s * f_s \right)(x) \left(f_s * f_s \right)(y) \left(f_s * f_s \right)(x+y) =$$

$$\mathbb{E}_s \sum_{\alpha} \hat{f_s}^6(\alpha).$$

∎

Corollary 6.6:

$$\|f\|_{U_3} \geq \epsilon \quad \Longrightarrow \quad \mathbb{E}_{x,y} \sum_{\alpha,\beta} \hat{f_x}^2(\alpha) \hat{f_y}^2(\beta) \widehat{f_{x+y}}^2(\alpha+\beta) \geq \epsilon'.$$

Proof: By lemmas 6.1 and 6.5 and Holder's inequality. ∎

Define a product distribution on functions $\phi : \{0,1\}^n \to \{0,1\}^n$ by taking $Pr(\phi(x) = \alpha) = \hat{f_x}^2(\alpha)$. The choices for distinct values of x are independent. Let $\delta = \frac{\epsilon'}{6}$. Define a random variable L on this probability space, by taking

$$L(\phi) = Pr_{x,y} \left\{ \phi(x) + \phi(y) = \phi(x+y); \;\; \hat{f_x}^2(\phi(x)) \geq \delta; \;\; \ldots \;\; \widehat{f_{x+y}}^2(\phi(x+y)) \geq \delta \right\}.$$

Lemma 6.7:

$$\mathbb{E}_\phi L(\phi) \geq \frac{\epsilon'}{2}.$$

Proof:

$$\mathbb{E}_\phi L(\phi) = \mathbb{E}_{x,y} Pr_\phi \left\{ \phi(x) + \phi(y) = \phi(x+y); \;\; \hat{f_x}^2(\phi(x)) \geq \delta; \;\; \ldots \;\; \widehat{f_{x+y}}^2(\phi(x+y)) \geq \delta \right\} =$$

$$\mathbb{E}_{x,y} \sum_{\alpha,\beta \,:\, \hat{f_x}^2(\alpha) \geq \delta; \;\ldots\; \widehat{f_{x+y}}^2(\alpha+\beta) \geq \delta} \hat{f_x}^2(\alpha) \hat{f_y}^2(\beta) \widehat{f_{x+y}}^2(\alpha+\beta) \geq$$

$$\mathbb{E}_{x,y} \sum_{\alpha,\beta} \hat{f_x}^2(\alpha) \hat{f_y}^2(\beta) \widehat{f_{x+y}}^2(\alpha+\beta) - 3\delta \geq \epsilon' - 3\delta \geq \frac{\epsilon'}{2}.$$

∎

Take ϕ for which $L(\phi) \geq \frac{\epsilon'}{2}$. This is the choice function we choose. Our goal is to find an appropriate linear transformation $B : \{0,1\}^n \to \{0,1\}^n$ such that ϕ and B coincide on a positive fraction of the domain in which $\hat{f_x}^2(\phi(x)) \geq \delta$.

We will do this in several steps. In the first step we will find an affine transformation $x \to Dx + z$ such that $\mathbb{E}_x \hat{f_x}^2(Dx+z) \geq \epsilon'$. Then we will gradually modify this transformation to obtain a symmetric linear transformation B with a zero diagonal such that $\mathbb{E}_x \hat{f_x}^2(Bx) \geq \epsilon'$. By lemma 6.4 this will conclude the proof of the theorem.

The first step is the hardest. We will follow an approach of Gowers from his proof of Szemeredi's theorem for arithmetic progressions of length four [11]. Note that a structural theorem of Freiman for sets with small sumsets in $\mathbb{Z}$ is replaced by a theorem of Ruzsa for such sets in $\mathbb{Z}_2^n$.

Let $A = \left\{ x \; : \; \hat{f_x}^2(\phi(x)) \geq \delta \right\}$. Then, by the choice of ϕ, the cardinality m of A is a positive fraction of 2^n, and there are $\Omega\left(m^2\right)$ triples $(x, y, x+y)$ in A^3 satisfying $\phi(x) + \phi(y) = \phi(x+y)$.

Now define a subset $\mathcal{A}$ of $\{0,1\}^{2n}$ as

$$\mathcal{A} = \{(x, \phi(x)) \; : \; x \in A\}.$$

$\mathcal{A}$ is the graph of ϕ on A. We have $|\mathcal{A}| = |A| = m$, and there are $\Omega\left(m^2\right)$ triples $(a, b, a+b)$ in $\mathcal{A}^3$.

Theorem 6.8: *(Gowers [11]) For any subset $\mathcal{A}$ of an abelian group satisfying above, there a subset $\mathcal{A}'$ of $\mathcal{A}$ containing a constant fraction of the elements, such that*

$$|\mathcal{A}' + \mathcal{A}'| \leq c \cdot |\mathcal{A}'|,$$

for an absolute constant c.

Theorem 6.9: *(Ruzsa [22]) Let G be an abelian group. Assume that the order of the elements in G is bounded, and let r be the maximal order of an element. Let $\mathcal{A}' \subseteq G$ with the property above. Then*

$$\frac{|\mathcal{A}'|}{|<\mathcal{A}'>|} \geq c' \cdot r^{-c''},$$

for some absolute constants c', c''.

We assume the projection of $<\mathcal{A}'>$ on the first n coordinates to be of full rank. (Otherwise add a finite number of vectors to $<\mathcal{A}'>$ to ensure this.) Therefore, there are n vectors $v_1...v_n$ in $\{0,1\}^n$ such that the vectors $u_i = (e_i, v_i)$ are in $<\mathcal{A}'>$. Let $U =< u_1...u_n >$. Clearly $U = \{(x, Dx) \; : \; x \in \{0,1\}^n\}$, where the matrix D is defined by $De_i = v_i$, $i = 1...n$. U is a subspace of $<\mathcal{A}'>$ of a finite co-dimension. Therefore there exists a vector $c \in \{0,1\}^{2n}$ such that a constant fraction of the vectors in $\mathcal{A}$ sit in $U + c$. This is the same as to say that there is a vector $z \in \{0,1\}^n$ such that for $\Omega(2^n)$ points $x \in \mathcal{A}$ holds $\phi(x) = Dx + z$. Alternatively:

$$\mathbb{E}_x \hat{f_x}^2 (Dx + z) \geq \epsilon'.$$

We can choose $z = 0$.

Lemma 6.10: *Define a function $F \; : \; \{0,1\}^n \to \mathbb{R}$ by $F(z) = \sum_y \hat{f_y}^2 (Dy + z)$. Then*

$$\hat{F}(x) = \hat{f_x}^2 (D^t x).$$

Proof:

$$\hat{F}(x) = \mathbb{E}_z F(z) w_x(z) = \mathbb{E}_z w_x(z) \sum_y \hat{f_y}^2 (Dy + z) =$$

$$\mathbb{E}_z w_x(z) \sum_y \mathbb{E}_{u,u'} f_y(u) f_y(u') w_{Dy+z}(u + u') = \mathbb{E}_{y,u} f_y(u) f_y(u + x) w_{Dy}(x) =$$

$$\mathbb{E}_y (f_y * f_y)(x) w_{Dy}(x) = \mathbb{E}_y (f_x * f_x)(y) w_{D^t x}(y) = \hat{f_x}^2 (D^t x).$$

∎

Since the transform of F is nonnegative, F attains its maximum in 0. Therefore

$$\mathbb{E}_x \hat{f_x}^2 (Dx) \geq \mathbb{E}_x \hat{f_x}^2 (Dx + z) \geq \epsilon'.$$

We want to replace D by a symmetric matrix. The following fact is useful.

Lemma 6.11:

$$\hat{f}_y(x) = 0$$

for any x and y with $\langle x, y\rangle = 1$.

Proof:

$$\hat{f}_y(x) = \mathbb{E}_z f_y(z) w_x(z) = \mathbb{E}_z f(y)f(z)f(y+z)w_x(z) = \mathbb{E}_z f(y)f(z)f(y+z)w_x(y+z) =$$

$$w_x(y)\mathbb{E}_z f(y)f(z)f(y+z)w_x(z) = w_x(y)\hat{f}_y(x) = -\hat{f}_y(x)$$

∎

Therefore, for $g(x) = (-1)^{\langle x, Dx\rangle}$ holds

$$\mathbb{E}_x g(x)\hat{f_x}^2(Dx) = \mathbb{E}_x \hat{f_x}^2(Dx) \geq \epsilon'.$$

On the other hand

$$\mathbb{E}_x g(x)\hat{f_x}^2(Dx) = \sum_z \hat{g}(z)\mathbb{E}_x \hat{f_x}^2(D^t x + z).$$

Since the numbers $\lambda_z = \mathbb{E}_x \hat{f_x}^2(D^t x + z)$ are nonnegative and sum to one, we deduce by Jensen's inequality that

$$\sum_z \hat{g}^2(z)\mathbb{E}_x \hat{f_x}^2(D^t x + z) \geq (\epsilon')^2.$$

However

$$\sum_z \hat{g}^2(z)\mathbb{E}_x \hat{f_x}^2(D^t x + z) = \mathbb{E}_x (g * g)(x)\hat{f_x}^2(Dx) = \mathbb{E}_x \delta_{Dx, D^t x} \cdot g(x)\hat{f_x}^2(Dx).$$

Let S be a matrix defined in the following way: Set $U = \left\{x\ :\ Dx = D^t x\right\}$. Then U is a subspace of $\mathbb{Z}_2^n$. Let S be defined on U by taking $S(x) = D(x)$ on U. Then for any $x, y \in U$ holds $\langle x, Sy\rangle = \langle Sx, y\rangle$. Now the definition of S could be extended to the whole space keeping this property. Therefore S is a symmetric matrix such that

$$\mathbb{E}_x \hat{f_x}^2(Sx) \geq \epsilon.$$

It remains to deal with the diagonal of S. Let f be the vector on the diagonal of S. Since $\langle x, Sx\rangle = \langle x, f\rangle$, we have

$$\frac{1}{2^n}\sum_{x \perp f} \hat{f_x}^2(Sx) \geq \epsilon.$$

Define a matrix B by taking $Bx = Sx$ if $x \perp f$, and extending B appropriately to the whole space. Namely for $w \not\perp f$ take $Bw = z$, so that $\langle z, x\rangle = \langle w, Bx\rangle = \langle w, Sx\rangle$ for all $x \perp f$, and $\langle w, z\rangle = 0$. Then B is symmetric with zero diagonal, and

$$\mathbb{E}_x \hat{f_x}^2(Bx) \geq \epsilon'.$$

This concludes the proof of theorem 2.3, but for the dependence of ϵ' on ϵ. Tracing this dependence through the proof, it is possible to see that we can choose $\epsilon' \geq \Omega\left(\exp\left\{-\left(\frac{1}{\epsilon^C}\right)\right\}\right)$ for an absolute constant C. ∎

7 Appendix B: A Proof of Theorem 2.2

We will prove Theorem 5.3.

This will imply Theorem 2.2 as follows: Let $f : \{0,1\}^n \to \{-1,1\}$ be a boolean function. Let $H = (V, E)$ be a hypergraph with maximal edge-size d. For a subset $\mathcal{S} \subseteq E$ of edges of H, let $\sigma(S) = \sum_{e \in \mathcal{S}} (|e| + 1)$. The probability that the linearity test associated with H accepts f is given by

$$\frac{1}{2^{|E|}} \sum_{\mathcal{S} \subseteq E} f^{\sigma(\mathcal{S})}(0) \cdot \mathbb{E}_{x_1,\ldots,x_t} \left[\prod_{e \in \mathcal{S}} \prod_{i \in e} f(x_i) \cdot f\left(\sum_{i \in e} x_i \right) \right]$$

The summand corresponding to $S = \emptyset$ is 1. By Theorem 5.3 all the other summands are at most $\|f\|_{U_d}$ in absolute value. Theorem 2.2 follows.

We start with some facts on generalized averages (1). Recall that each such average is naturally associated with a $t \times T$ binary matrix A. The first observation is that some matrices (families of sets) define the same average operator.

Lemma 7.1: *Multiplying A on the left by a non-singular $t \times t$ matrix B does not change the value of the average, namely for any function f holds $\mathbb{E}_A(f) = \mathbb{E}_{BA}(f)$,*

Proof: $\mathbb{E}_A$ and $\mathbb{E}_{BA}$ are the same up to order of summation. ∎

Corollary 7.2: *We may (and will) assume that the rows of A are linearly independent, since if $rank(A) = r < t$ we can choose a non-singular $t \times t$ matrix B, so that in BA the last $t - r$ rows are zeroes, and consequently can be removed without changing the value of $\mathbb{E}_A$.*

Consider an equivalence relation on $t \times T$ binary matrices of full rank, defined by left multiplication by a non-singular $t \times t$ matrix. It is easy to see that two matrices are equivalent iff their rows span the same t-dimensional space over $\mathbb{Z}_2^T$ (or they represent the same rank-t *binary matroid* on $\{1...T\}$ [20]).

The following definition and lemmas are natural (and well-known) in the setting of matroids (see [20]).

Definition 7.3: A *hyperplane* of A is a maximal subset of $\{1, \ldots, T\}$ such that the columns of A indexed by this subset are *not* of full rank. ∎

Lemma 7.4: *A vector v is a minimal non-zero vector in the row space of A iff the complement of its support is a hyperplane of A.*

Lemma 7.5: *The row space of A is spanned by its minimal non-zero vectors.*

The key part of the proof of theorem 5.3 is the following technical proposition:

Proposition 7.6: *Let A be a full rank $t \times T$ binary matrix. Let v be a minimal vector in the row-space of A. Let A' be a $(t+1) \times 2|v|$ matrix obtained from A by the following procedure:*

1. *Delete all the columns not in the support of v, obtaining a $t \times |v|$ matrix B.*
2. *Set*

$$A' = \left[\begin{array}{c|c} B & B \\ 1 \ldots 1 & 0 \ldots 0 \end{array} \right]$$

Then for all boolean functions f,

$$\mathbb{E}_A(f) \leq \sqrt{\mathbb{E}_{A'}(f)}.$$

Example 7.7: Let

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix},$$

and take $v = \{1, 3\}$. Then

$$A' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}.$$

$\blacksquare$

Proof: Let H be the complement of v, and by lemma 7.4 a hyperplane of A. Let H_0 be a maximal independent subset (a basis) of H, of size $t-1$. We assume that $H = \{1, \ldots, |H|\}$ and $H_0 = \{1, \ldots, t-1\}$. Multiply A on the left by a non-singular $t \times t$ matrix B so that the first $t-1$ columns of A are the first $t-1$ unit vectors. Since H_0 is a basis of H, the columns of BA indexed by H will have a zero in their last coordinate, while the columns in $v = H^c$ will have one (since H is a hyperplane). Namely

$$BA = \left[\begin{array}{c|c} 10\ldots0 & \\ 01\ldots0 & \\ \vdots & \text{N} \\ 00\ldots1 & \\ \hline 00\ldots0 & 0\ldots01\ldots1 \end{array} \right].$$

We have, for any f:

$$\mathbb{E}_A(f) = \mathbb{E}_{BA}(f) = \mathbb{E}_{y_1,\ldots,y_t} f(y_1) \cdot \ldots \cdot f(y_{t-1}) \cdot \prod_{i=t}^{T} f\left(\prod_{j \in A_i} y_j\right).$$

We upper bound the right hand side in the following way, applying the Cauchy-Schwarz inequality:

$$\mathbb{E}_M(f) = \mathbb{E}_{y_1,\ldots,y_{t-1}} F(y_1, \ldots, y_{t-1}) \cdot G(y_1, \ldots, y_{t-1}) \leq$$

$$\sqrt{\mathbb{E}_{y_1,\ldots,y_{t-1}} F^2(y_1, \ldots, y_{t-1})} \cdot \sqrt{\mathbb{E}_{y_1,\ldots,y_{t-1}} G^2(y_1, \ldots, y_{t-1})},$$

where $F(y_1, \ldots, y_{t-1}) = \prod_{j=1}^{t-1} f(y_j)$, and $G(y_1, \ldots, y_{t-1}) = \mathbb{E}_{y_t} \prod_{i=t}^{T} f\left(\prod_{j \in A_i} y_j\right)$.

Observe that

$$\mathbb{E}_{y_1,\dots,y_{t-1}} F^2\left(y_1,\dots,y_{t-1}\right) = \mathbb{E}_{y_1,\dots,y_{t-1}} f^2(y_1)\cdot\ldots\cdot f^2(y_{t-1}) = \mathbb{E}1 = 1.$$

Therefore $\mathbb{E}_M(f) \le \sqrt{\mathbb{E}G^2}$. It is easily seen that $\mathbb{E}G^2$ presents an average of f on a $(t+1)\times 2(T-t+1)$ matrix $A^{(1)}$, where

$$A^{(1)} = \left[\begin{array}{c|c} \text{N} & \text{N} \\ \hline 0\dots00\dots0 & 0\dots01\dots1 \\ \hline 0\dots01\dots1 & 0\dots00\dots0 \end{array}\right].$$

We now transform the matrix $A^{(1)}$ to A' in three steps. Let u and v be the last two rows of the matrix. First, replace u with $u+v$ obtaining a new matrix $A^{(2)}$.

The second step uses booleanity of f. Note that for any matrix A and any boolean function f, deleting a pair of identical columns of A does not change the average of f on A. We delete all the columns of $A^{(2)}$ in which the last two coordinates are zero, obtaining a $(t+1)\times 2|v|$ matrix $A^{(3)}$. The third step is to multiply $A^{(3)}$ by a $(t+1)\times(t+1)$ matrix $S_1 = \begin{bmatrix} B^{-1} & 0 \\ 0 & 1 \end{bmatrix}$, obtaining A'. ∎

Now we are ready to prove Theorem 5.3. We will prove the theorem for $k=3$. The proof for larger values of k is similar.

Let A be a matrix with at most 3 ones in each column. Assume the rows of A to be independent. Let u be the first row. The first step is to replace u with a minimal non-zero vector v with a smaller support. If u is already a minimal vector let $v=u$. Otherwise there is a vector w in a row-space of A whose support is strictly smaller than that of u. One of the vectors w or $u+w$ is not spanned by the rest of the rows of A and we set u_1 to be this vector. If u_1 is minimal set $v=u_1$. Otherwise continue with u_1 instead of u. Clearly this process stops after a finite number of steps and does not change the row space of A.

Now we apply the transformation of proposition 7.6 to the new matrix A, choosing v as the appropriate minimal vector. Consider the submatrix B of the new matrix A'. The first row of B, and therefore of A' as well, is a 1-vector. Moving it to be the last row, we obtain

$$A' = \left[\begin{array}{c|c} B' & B' \\ 1\dots1 & 0\dots0 \\ 1\dots1 & 1\dots1 \end{array}\right]$$

The matrix B' has at most 2 ones in each column. Note that all the columns in B' are distinct, since so were the columns of A. There are two cases to distinguish.

B' has only one column. Then, removing dependent rows, we get to a 2×2 matrix $A_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. By Proposition 7.6, for any boolean function f holds $\left|\mathbb{E}_A(f)\right| \le \left(\mathbb{E}_{A_1}(f)\right)^{1/2} = \|f\|_{U_1} \le \|f\|_{U_3}$. In the last inequality we use monotonicity of uniformity norms.

B' has more than one column. If there are dependencies between rows of A' we remove them, keeping only a spanning set of rows, starting with the last two. In particular A' has no all-1 rows except the last.

We now repeat the procedure starting from A'. Consider the first row u of A'. $u = (u' \mid u')$ is a symmetric vector. If it is minimal set $v = u$. If not, there is a minimal vector w of smaller support such that replacing u with w does not affect the row space of A'. The vector w is in this row space, therefore it is either symmetric or antisymmetric. However if it is antisymmetric then u has to be an all-1 row, which we have excluded. Therefore we can replace u by a symmetric minimal vector v. Now apply the proposition with A' and v, and obtain a new matrix (after simplification)

$$A'' = \left[\begin{array}{c|c|c|c} B'' & B'' & B'' & B'' \\ 1\dots1 & 0\dots0 & 1\dots1 & 0\dots0 \\ 1\dots1 & 1\dots1 & 0\dots0 & 0\dots0 \\ 1\dots1 & 1\dots1 & 1\dots1 & 1\dots1 \end{array}\right]$$

The matrix B'' has at most one 1 in each column. All the columns in B' are distinct.

Once again, there are two cases. If there is only one column in B'', after simplification we get a 3×4 matrix

$$A_2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

such that for any boolean function f holds $\left|\mathbb{E}_A(f)\right| \leq (\mathbb{E}_{A_2}(f))^{1/4} = \|f\|_{U_2} \leq \|f\|_{U_3}$.

If B'' has more than one column we iterate once again. It is not hard to see that the new matrix B''' will necessarily have only one column. After simplifying, we will get to a 4×8 matrix

$$A_3 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

such that for any boolean function f holds $\left|\mathbb{E}_A(f)\right| \leq (\mathbb{E}_{A_3}(f))^{1/8} = \|f\|_{U_3}$. The theorem is proved. ∎

8 Appendix C: Other proofs

8.1 Proof of Proposition 3.2

First we estimate $\|f\|_{U_3}$ within additive precision of $O(\delta)$. This can be done by choosing at random $\Omega\left(\frac{1}{\delta}\right)$ quadruples of vectors $x, y_1, y_2, y_3 \in \{0,1\}^n$ and averaging $f_{y_1,y_2,y_3}(x)$ over the choices. Let us call this average ν. It is easy to see that for a sufficiently large number of sampled quadruples, $\left|\nu - \|f\|_{U_3}\right| \leq \delta/2$ with high probability.

Assuming this is true, there are two possibilities. First, $\nu \geq \delta$. In this case, $\|f\|_{U_3} \geq \delta/2$. Theorem 2.3 now implies that the second option of the proposition holds.

The second option is $\nu < \delta$. In this case $\|f\|_{U_3} < 3\delta/2$. Now we follow an argument from [14]. Let g be a quadratic polynomial. Recalling the interpretation of $\|\cdot\|_{U_3}^8$ as the average of the third order derivative of a function, it is easy to see that $\|fg\|_{U_3} = \|f\|_{U_3} \leq 3\delta/2$. Observe that the first uniformity "norm" of a function is the square of its expectation. By the monotonicity of uniformity norms,

$$\langle f, g\rangle = \widehat{fg}(0) \leq \|f\|_{U_1}^{1/2} \leq \|f\|_{U_3}^{1/2} \leq O\left(\delta^{1/2}\right)$$

Since both f and g are boolean functions, this implies that the distance between f and g is at least $1/2 - \Omega\left(\delta^{1/2}\right)$, and the first option of the proposition holds.

8.2 Proof of Theorem 3.3

The completeness of the test follows from the fact that it checks whether third order derivatives of the function vanish.

The fact that the soundness of the test is $1/2^{|E|}$ is an immediate consequence of Theorems 5.3 with $k = 3$ together with Theorem 2.3. Indeed, let a boolean function f be $1/2 - \epsilon$ far from quadratic polynomials. Similarly to the proof of Theorem 2.2, the acceptance probability of the test on a function f is upper bounded by $1/2^{|E|} + \|f\|_{U_3}$. By Theorem 2.3, this is at most $1/2^{|E|} + \epsilon'$, with $\epsilon' \to 0$ with ϵ.

8.3 Proof of Theorem 4.1

Combining Theorems 6.8 and 6.9 similarly to the proof of Theorem 2.3, we obtain the following claim.

Theorem 8.1: *Let p be a prime number, and let $\epsilon > 0$. Let G be a p-group of order r and let H be a power of $\mathbb{Z}_p$. Let $\phi : G \to H$ such that*

$$Pr_{x,y\in G}\left(\phi(x) + \phi(y) = \phi(x+y)\right) \geq \epsilon.$$

Then there exists a homomorphism $\psi : G \to H$ and an element $h \in H$ such that

$$Pr_{x\in G}\left(\phi(x) = \psi(x) + h\right) \geq c \cdot r^{-c'} \cdot \epsilon^{c''},$$

where c, c', c'' are absolute constants (independent of the groups G, H).

The following lemma concludes the proof of Theorem 4.1.

Lemma 8.2: *Let G be a p-group of order r, and let H be a power of $\mathbb{Z}_p$. Let $\phi : G \to H$ be such that there exists a homomorphism $\psi : G \to H$ and an element $h \in H$ such that*

$$Pr_{x\in G}\left(\phi(x) = \psi(x) + h\right) \geq \delta.$$

Then there exists a homomorphism $\psi' : G \to H$ *such that*

$$Pr_{x\in G}\left(\phi(x) = \psi'(x)\right) \geq \frac{c}{r}\cdot\delta,$$

for an absolute constant c.

Proof: Let $E = \{x \in G : \phi(x) = \psi(x) + h\}$. Let $G = \prod_{i=1}^{m} \mathbb{Z}_{p^{k_i}}$. There exists an absolute constant c, a coordinate $1 \leq i \leq m$, and a generating element $g \in \mathbb{Z}_{p^{k_i}}$ such that for at least $\frac{c}{r}$-fraction of the elements of E holds $x_i = g$. Call this set E'. Let $e_1 ... e_m$ be the standard basis of G. Consider a homomorphism $\psi' : G \to H$ defined as follows: $\psi'(e_j) = \psi(e_j)$ for $j \neq i$ and $\psi'(g \cdot e_i) = \psi(e_i) + h$. Then ψ' agrees with ϕ on E'. ∎

9 Acknowledgements

Part of this work was done while visiting Johan Håstad at KTH. I am very grateful to Johan for his hospitality and very helpful discussions. I would also like to thank Michael Ben-Or, Nati Linial, Luca Trevisan, and Benjamin Weiss for valuable conversations.

References

[1] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, D. Ron, *Testing low-degree polynomials over* $GF(2)$, RANDOM-APPROX 2003, pp. 188-199.

[2] 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.

[3] S. Arora and S. Safra *Probabilistic checking of proofs: A new characterization of NP*, Journal of the ACM, 45(1):70–122, 1998.

[4] M. Bellare, D. Coppersmith, J. Hastad, M. Kiwi, M. Sudan *Linearity testing in characteristic* 2, IEEE Trans. Inform. Theory, vol. IT-42, 6, 1996, 1782-1795.

[5] M. Ben-Or, D. Coppersmith, Personal communication to the authors of [6], 1989.

[6] M. Blum, M. Luby, R. Rubinfeld, *Self-testing/correcting with applications to numerical problems*, J. Comp. Sys. Sci., 47, 3, 1993.

[7] M. Charikar, K. Makarychev, and Y. Makarychev, personal communication to the authors of [24].

[8] L. Engebretsen and J. Holmerin, *Towards optimal lower bounds for clique and chromatic number*, TCS, 299(1-3), pp. 537-584, 2003.

[9] O. Goldreich and L. Levin, *A generic hard-core predicate for any one-way function*, STOC 1989, pp. 25-30.

[10] O. Goldreich and M. Sudan, *Locally testable codes and PCPs of almost-linear length*, FOCS 2002, pp. 13-22.

[11] W. T. Gowers, *A new proof of Szemeredi's theorem*, GAFA Vol. 11(2001), pp. 465-588.

[12] B. Green, *Montreal notes on quadratic Fourier analysis*, preprint, Mathematics ArXiv CA/0604089.

[13] B. Green and T. Tao. *The primes contain arbitrarily long arithmetic progressions*, Annals of Mathematics, to appear.

[14] B. Green, T. Tao, *An inverse theorem for the Gowers U^3 norm*, preprint, Mathematics ArXiv NT/0503014.

[15] G. Hast, *Approximating Max kCSP - outperforming a random assignment by almost a linear factor*, ICALP 2005, to appear.

[16] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Annals of Mathematics, 161(1):397–488, 2005.

[17] J. Kahn, G. Kalai, and N. Linial, *The influence of variables on boolean functions*, FOCS 1988, pp. 68-80.

[18] S. Khot, *On the power of unique 2-prover 1-round games.* STOC 2002, pp. 767-775.

[19] J. MacWilliams and N. J. A. Sloane, **The Theory of Error Correcting Codes**, Amsterdam, North-Holland, 1977.

[20] J. G. Oxley, **Matroid Theory**, New York, Oxford University Press, 1992.

[21] I. S. Reed, *A class of multiple error correcting codes and the decoding scheme*, IEEE IT, vol. 4, 1954, pp. 38-49.

[22] I. Z. Ruzsa, *An analog of Freiman's theorem in groups*, Asterisque 258, 199, pp. 323-326.

[23] A. Samorodnitsky, L.Trevisan, *A PCP Characterization of NP with Optimal Amortized Query Complexity*, STOC 2000, pp. 191-199.

[24] A. Samorodnitsky, L.Trevisan, *Gowers Uniformity, Influence of Variables, and PCPs*, STOC 2006, to appear.

[25] M. Sudan, *List decoding: algorithms and applications*, IFIP TCS 2000, pp. 25-41.

[26] L. Trevisan, *Parallel approximation algorithms by positive linear programming*, Algorithmica 21(1):72–88, 1998.

[27] L. Trevisan, *Recycling queries in PCPs and in linearity tests*, STOC 1998, pp. 299-308.

[28] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, Journal of AMS, to appear.