The Gowers Norm in the Testing of Boolean Functions

by

Victor Yen-Wen Chen

B.S., The University of Texas at Austin (2004)

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2009

Author .....
Department of Mathematics
May 8, 2009

Certified by .....
Madhu Sudan
Fujitsu Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Michel Goemans
Chairman, Applied Mathematics Committee

Accepted by .....
David Jerison
Chairman, Department Committee on Graduate Students

The Gowers Norm in the Testing of Boolean Functions

by

Victor Yen-Wen Chen

Submitted to the Department of Mathematics
on May 8, 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

A property tester is a fast, randomized algorithm that reads only a few entries of the input, and based on the values of these entries, it distinguishes whether the input has a certain property or is "different" from any input having this property. Furthermore, we say that a property tester has completeness c and soundness s if it accepts all inputs having the property with probability at least c and accepts "different" inputs with probability at most $s + o(1)$.

In this thesis we present two property testers for boolean functions on the boolean cube $\{0,1\}^n$. We summarize our contribution as follows.

- We present a new dictatorship test that determines whether the function is a dictator (of the form $f(x) = x_i$ for some coordinate i), or a function that is an "anti-dictator." Our test is "adaptive," makes q queries, has completeness 1, and soundness $O(q^3) \cdot 2^{-q}$. Previously, a dictatorship test that has soundness $(q+1) \cdot 2^{-q}$ is achieved by Samorodnitsky and Trevisan, but their test has completeness strictly less than 1. Furthermore, the previously best known dictatorship test from the PCP literature with completeness 1 has soundness $2^{O(\sqrt{q})-q}$. Our contribution lies in achieving perfect completeness and low soundness simultaneously.

- We consider properties of functions that are invariant under linear transformations of the boolean cube. Previous works, such as linearity testing and low-degree testing, have focused on linear properties. The one exception is a test due to Green for "triangle freeness": a function f satisfies this property if $f(x), f(y), f(x+y)$ do not all equal 1, for any pair $x, y \in \{0,1\}^n$. We extend this test to a more systematic study and consider non-linear properties that are described by a single forbidden pattern. Specifically, let M denote an r by k matrix over $\{0,1\}$. We say that a function f is M-free if there are no $\vec{x} = (x_1, \ldots, x_k)$, where $x_1, \ldots, x_k \in \{0,1\}^n$ such that $f(x_1), \ldots, f(x_k) = 1$ and $M\vec{x} = \vec{0}$. If M can be represented by an underlying graph, we can analyze a test that determines whether a function is M-free or "far" from one. Our test makes k queries, has completeness 1, and soundness bounded away from 1. The

technique from our work leads to alternate proofs that some previously studied linear properties are testable, albeit with worse parameters.

Our results, though quite different in terms of context, are connected by similar techniques. Our analysis of the algorithms relies on the machinery of the Gowers uniformity norm, a recent and powerful tool in additive combinatorics.

Thesis Supervisor: Madhu Sudan
Title: Fujitsu Professor of Electrical Engineering and Computer Science

Acknowledgments

There are many mentors and colleagues whom I would like to thank for helping me through the course of my graduate studies. First I am grateful to Madhu Sudan, my thesis supervisor, for the advice and encouragement he gave over the last five years. I thank him for listening to my ideas, offering his constructive feedback, and sharing his perspective on research.

Next I would like to thank Alex Samorodnitsky, who has attentively listened to my various proof attempts for the material in Chapter 3. I am indebted to him for taking a strong interest in my work and sharing his deep intuition in combinatorics with me. I also thank Mike Sipser and Peter Shor for serving on my thesis committee. In addition, being a teaching assistant for Mike's theory of computing class has helped me develop better teaching technique. I also thank Peter for being my internal advisor in the department, especially for the encouraging conversations we had each semester, even when I had no result during the early years. I also thank David Zuckerman, my undergraduate advisor at UT Austin, for introducing me to complexity theory and the discussions we had.

There are also many graduate students whom I would like to thank, in particular Kevin Matulef, Joungkeun Lim, Elena Grigorescu, Brendan Juba, Swastik Kopparty, and Mayank Varia. Kevin was my student mentor when I first started at MIT. His help was especially useful to an incoming student, and in particular, an encouraging discussion we had eventually led to my work presented in Chapter 4. Joungkeun was also very helpful during my first year, and since then has been a good friend and a source of good advice. I am also grateful to Elena for her optimism, her perspective on graduate school, and the collaboration we have on error-correcting codes. Lastly, I also thank Brendan for the numerous conversations that we had, Swastik for his enthusiasm for original ideas, and Mayank for being my officemate and vetting some of the slides for the defense talk.

Finally, without the support and love of my parents, this thesis would not be possible.

Bibliography

The result from Chapter 4 is based on the paper "Testing Linear-Invariant Non-Linear Properties" [10] which appeared in STACS 2009 and is joint with Arnab Bhattacharyya, Madhu Sudan, and Ning Xie, and I thank them for their contribution.

Contents

Chapter 1

Introduction

The subject of this thesis is concerned with the testing of boolean functions on the boolean cube $\{0,1\}^n$. We present two property testing algorithms. The first is a new dictatorship test, a useful gadget in the construction of probabilistic checkable proofs (PCPs). The second is concerned with testing whether a function has a specific linear-invariant pattern. While these two results come from two seemingly different contexts, we analyze the behavior of these algorithms by applying similar tools from additive combinatorics. Specifically, we utilize the Gowers uniformity norm of a function [15] to analyze the acceptance probability of our algorithms.

We first describe the framework and motivation of property testing. Traditionally, algorithms that run in polynomial time in the length of the input are considered practical, and linear-time algorithms are the paradigm of efficiency. However, while the computational power of computers has increased tremendously over the decades, the growth of dataset, especially those arising from the internet, has accelerated even more so. For such massive datasets, reading the input in its entirety is not computationally feasible.

The field of property testing, initiated by Blum, Luby, and Rubinfeld [11] and formally defined by Rubinfeld and Sudan [35], is concerned with "super-efficient" algorithms that perform in time sublinear in the length of the input. Instead of processing the input as a whole, such an algorithm examines the input at a few select entries, and based on the values of these entries, the algorithm tests whether the

input satisfies a certain property or "looks different" from any input that satisfies this property.

In this thesis, we focus our attention on boolean functions of the form $f : \{0,1\}^n \to \{\mathbf{T}, \mathbf{F}\}$, where $\mathbf{T}$ represents the boolean TRUE and $\mathbf{F}$ represents the boolean FALSE. The input is simply the truth table of f, the evaluation of f at every point in $\{0,1\}^n$ enumerated in some lexicographical ordering. A testing algorithm is said to have oracle access to f if the algorithm can query the entry x for every $x \in \{0,1\}^n$ and obtain the value $f(x)$. A testing algorithm, modeling the decision problems in complexity theory, accepts with high confidence if the input satisfies some specified property and rejects with high confidence if the input "looks different" from those satisfying the property.

We measure the efficiency of a testing algorithm by the number of queries it makes. Furthermore, we only consider *local* testing algorithms – those making a fixed number of queries, independent of n, into the input. We make the following formal definition:

Definition 1.0.1. Let $\mathcal{YES}_n, \mathcal{NO}_n$ be two disjoint subsets of the set of boolean functions $\{f : \{0,1\}^n \to \{\mathbf{T}, \mathbf{F}\}\}$, and let $\mathcal{YES} = \cup_{n>0} \mathcal{YES}_n$ and $\mathcal{NO} = \cup_{n>0} \mathcal{NO}_n$. Let q be an integer, and $0 < s < c < 1$. We say that $T = \{T_n\}_n$ is a *property tester* for $(\mathcal{YES}, \mathcal{NO})$ (and we say $(\mathcal{YES}, \mathcal{NO})$ is *testable*) with q *queries*, *completeness* c, and *soundness* s if for every n, T_n is a probabilistic algorithm that

- makes q oracle calls to a function f,
- accepts with probability at least c if $f \in \mathcal{YES}_n$, and
- accepts with probability at most $s + o(1)$ if $f \in \mathcal{NO}_n$.

Remark. We say that a tester T is *adaptive* if the queries it selects may depend on the values of the previous queries.

Semantically, $\mathcal{YES}$ denotes the set of boolean functions that satisfies a specified property, and $\mathcal{NO}$ denotes the set of functions that looks very "different" from the set $\mathcal{YES}$. Specifically, let $\delta_n(f,g)$ denote $2^{-n} |\{x : f(x) \neq g(x)\}|$. we say that a function f is ϵ-*far* from $\mathcal{YES}_n$ if $\delta_n(f,g) \geq \epsilon$ for every $g \in \mathcal{YES}_n$. One can see that if

$(\mathcal{YES}_n, \mathcal{NO}_n)$ is testable and f is in $\mathcal{NO}_n$, then f is ϵ-far from $\mathcal{YES}_n$ for any constant $\epsilon > 0$. To see this, suppose $\delta_n(f, g) = o(1)$. With q independent of n, any q-tester with high probability will either accept both f and g or reject both. This cannot happen if $f \in \mathcal{YES}_n$ and $g \in \mathcal{NO}_n$, since a tester must behave differently on these two functions.

We remark that in the property testing literature, the set of $\mathcal{NO}$ instances is often taken to be maximal, where for every n, $\mathcal{NO}_n$ is parametrized by a constant $\epsilon > 0$ and consists of all functions that are ϵ-far from $\mathcal{YES}_n$. This is actually the strongest requirement since by definition, provided $\mathcal{YES}'_n \subseteq \mathcal{YES}_n$ and $\mathcal{NO}'_n \subseteq \mathcal{NO}_n$, a tester for the property $(\mathcal{YES}_n, \mathcal{NO}_n)$ is also trivially a tester for the property $(\mathcal{YES}'_n, \mathcal{NO}'_n)$. In this thesis we shall work with the less stringent definition where $\mathcal{NO}_n$ is simply some meaningful subset of all functions that are far from $\mathcal{YES}_n$. In addition, when clear from the context, we may drop the subscript parameter n.

1.1 The results

We now describe the motivation and background for our results.

1.1.1 Dictatorship testing

We say a function f is *linear* if $f(x) = \sum_{i \in S} x_i$ for some subset $S \subseteq [n]$. A *dictator* function is simply a linear function where $|S| = 1$, i.e., $f(x) = x_i$ for some $i \in [n]$. In other words, a dictator function depends on exactly one variable, and we define the set of $\mathcal{YES}$ instances to be simply $\mathcal{YES}_n = \{x_i : i \in [n]\}$. Following [13, 27, 39], the set of $\mathcal{NO}$ instances for dictatorship testing consists of functions with "low-degree influences" $o(1)$ for each variable. We defer a formal definition to Chapter 3 while suppressing the technical details in our exposition.

The problems of both linearity and dictatorship testing have been intensely studied in the past both for their combinatorial interest and connection to complexity theory. In complexity theory, a dictator function is often called a *long code*, and it is first used in [7] for the constructions of probabilistic checkable proofs (PCPs), see e.g., [5, 4] for

background on the PCP Theorems.

Since then, dictatorship testing has become a valuable tool in the construction of PCPs. A PCP system is typically designed as the composition of two verifier, an outer verifier and an inner verifier. A PCP system expects the proof to be written in such a way so that the outer verifier, typically based on the verifier obtained from Raz's Parallel Repetition Theorem [34], selects some tables of the proof according to some distribution and then passes the control to the inner verifier. The inner verifier, with oracle access to these tables, makes queries into these tables and ensures that the tables are the encoding of some error-correcting codes and satisfy some joint constraint. The long code encoding is usually employed in these proof constructions, and the inner verifier simply tests whether a collection of tables (functions) are long codes satisfying some constraints. Following this paradigm, constructing a PCP with certain parameters reduces to the problem of designing a dictatorship test with similar parameters.

One question of interest is the tradeoff between the soundness and query complexity of a tester. If a tester queries the functions at every single value, then trivially the verifier can determine all the functions. One would like to construct a dictatorship test that has the lowest possible soundness while making as few queries as possible. One way to measure this tradeoff between the soundness s and the number of queries q is *amortized query complexity*, defined as $\frac{q}{\log s^{-1}}$. This investigation, initiated in [44], has since spurred a long sequence of works [41, 38, 24, 14]. All the testers from these works run many iterations of a single dictatorship test by reusing queries from previous iterations. The techniques used are Fourier analytic, and the best amortized query complexity from this sequence of works has the form $1 + O\left(\frac{1}{\sqrt{q}}\right)$.

The next breakthrough occurs when Samorodnitsky [37] introduces the notion of a *relaxed* linearity test along with new ideas from additive combinatorics. The recent paradigm in additive combinatorics is to find the right framework of structure and pseudorandomness and analyze combinatorial objects by dividing them into structured and pseudorandom components, see e.g. [43] for a survey. One success is the notion of Gowers norm [15], which has been fruitful in attacking many problems in

additive combinatorics and computer science. In [37], the problem of linearity testing is relaxed; instead of designating the set of $\mathcal{NO}$ instances to be functions that are far from being linear, the author defines the set to be the set of functions with small low degree Gowers norm. By doing so, an optimal tradeoff between soundness and query complexity is obtained for the problem of relaxed linearity testing. (Here the tradeoff is stronger than the tradeoff for the standard problem of linearity testing.)

Building on the analysis of the relaxed linearity test in [37], Samorodnitsky and Trevisan [39] construct a dictatorship test with amortized query complexity $1 + O\left(\frac{\log q}{q}\right)$. Furthermore, the test is used as the inner verifier in a conditional PCP construction (based on unique games [26]) with the same parameters. However, their dictatorship test suffers from an inherent loss of perfect completeness. Ideally one would like testers with one-sided errors. One, for aesthetic reasons, testers should always accept valid inputs. Two, for some hardness of approximation applications, in particular coloring problems (see e.g. [23] or [12]), it is important to construct PCP systems with one-sided errors.

We prove the following theorem in this thesis:

Theorem 1.1.1. *For every $q \geq 3$, there exists an (adaptive) dictatorship test that makes q queries, has completeness 1, and soundness $\frac{O(q^3)}{2^q}$; in particular it has amortized query complexity $1 + O\left(\frac{\log q}{q}\right)$.*

Our work relies on techniques developed in [24, 39, 23, 21]. Our tester is adaptive in the sense that it makes its queries in two stages. It first makes roughly $\log q$ nonadaptive queries into the function. Based on the values of these queries, the tester then selects the rest of the query points nonadaptively.

1.1.2 Linear-invariant non-linear properties

We are interested in broad classes of properties that are testable. Specifically, we examine properties that remain invariant under linear transformations of the domain. Intuitively, it is reasonable that the symmetries of a property can lead to testability, since this suggests that the value of a function at any one point of the domain is no

more important than its value at any other point. Formally, we say that a property $\mathcal{YES}$ is *linear-invariant* if for every $f \in \mathcal{YES}$ and linear map $L : \{0,1\}^n \to \{0,1\}^n$, we have that $f \circ L \in \mathcal{YES}$. Specific examples of linear-invariant properties that were previously studied include that of linearity and low-degree polynomial [11, 6, 3]. While the tests in the above mentioned works potentially use all features of the property being tested, Kaufman and Sudan [25] show that testability can be attributed primarily to the linear-invariance of the property. However, their setting only considers *linear* properties, i.e., $\mathcal{YES}$ itself is a vector space over $\{0,1\}$ and this feature plays a key role in their result.

In this thesis we consider the following question: does linear-invariance lead to testability even when the property is non-linear? The one previous work in the literature that gives examples of non-linear linear-invariant properties is Green [17], where a test for the property of being triangle-free is described. A function $f : \{0,1\}^n \to \{0,1\}$ is said to be *triangle-free* if for every $x, y \in \{0,1\}^n$ it is the case that at least one of $f(x), f(y), f(x+y)$ does not equal 1. The property of being triangle-free is easily seen to be linear-invariant and yet not linear. Green [17] shows that the natural test for this property does indeed work correctly, though the analysis is quite different from that of typical algebraic tests. In particular, Green develops an algebraic regularity lemma to analyze this test. We remark that the above example is not the principal objective in Green's work, which is mainly focused on resolving a conjecture raised by Bergelson, Host, and Kra [8] regarding the number of 3-term arithmetic progressions of the same common difference.

Motivated by Green's triangle–freeness example, we seek to understand broad classes of properties that are linear-invariant and non-linear. Now consider the following definition.

Definition 1.1.1. Let $f : \{0,1\}^n \to \{0,1\}$ and M be an r by k matrix over $\{0,1\}$. We say that f is *M-free* if there are no $x_1, \ldots, x_k \in \{0,1\}^n$ such that

- $f(x_1), \ldots, f(x_k) = 1$, and
- $M\vec{x} = \vec{0}$, where $\vec{x} = (x_1, \ldots, x_k)$.

One can easily see that this definition coincides with triangle-freeness when $r = 1$, $k = 3$, and $M = (1, 1, 1)$. Formally, we define the *property of M-free* to be $(\mathcal{YES}_M, \mathcal{NO}_\epsilon)$, where $f \in \mathcal{YES}_M$ if f is M-free and $f \in \mathcal{NO}_\epsilon$ if f is ϵ-far from $\mathcal{YES}_M$. Formulated in our language of property testing, Green [17] in the same paper conjectures the following:

Conjecture 1.1.2 (implicit in Green [17])**.** *Let $r \geq 1$, $k \geq 3$ be integers and M be any r by k matrix over $\{0, 1\}$. Then there exists a testing algorithm for the property of M–free that makes k queries, has completeness 1, and soundness bounded away from 1.*

In fact, Green shows that this conjecture holds when $r = 1$, for which the same analysis for his triangle-free test carries over. In this thesis, we make partial progress toward this conjecture. Specifically, we prove the conjecture in the special case when the matrix M can be represented by an underlying graph. We view this as a natural generalization of identifying a triangle with a linear equation with 3 variables.

Definition 1.1.2. Let M be an r by k matrix over $\{0, 1\}$. We say that the matrix M is *graphic* if there exists a graph on k edges, each edge associated with an integer from $\{1, \ldots, k\}$, such that any linear combination of the rows of M corresponds to a cycle on the graph. More specifically, for each cycle in the graph, its indicator vector (on the edge set) lies in the span of the rows of M.

We prove the following in this thesis:

Theorem 1.1.3. *Let $r \geq 1, k \geq 3$ be integers, $\epsilon > 0$, and M an r by k graphic matrix M. Then there exist a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ and a tester for the property $(\mathcal{YES}_M, \mathcal{NO}_\epsilon)$ that makes k queries, has completeness 1 and soundness $1 - \tau(\epsilon)$.*

The bound we obtain for τ is quite weak. Let $W(t)$ denote a tower of twos with height $\lceil t \rceil$. Our proof only guarantees that $\tau(\epsilon) \geq W\left(\text{poly}\left(1/\epsilon\right)\right)^{-1}$, a rather fast vanishing function. We do not know if the tower bound is inherent in τ for any property that we consider. However, we remark that in the literature, the tower behavior

persists in all the testing algorithms that are based on some notion of regularity lemmas.

Syntactically, Theorem 1.1.3 seems to indicate that infinitely many properties are testable. However, this may not be true semantically as two different graphs may actually lead to the same set of $\mathcal{YES}$ instances. Nevertheless, in our conference paper [10] where Theorem 1.1.3 appears, the paper shows that an infinite sequence of graphs exists such that the corresponding $\mathcal{YES}$ instances are distinct, and thus verifying that Theorem 1.1.3 does indeed show that infinitely many properties are testable.

Furthermore, based on the techniques from this work, we develop alternate proofs for linearity testing [11, 6] and affine subspace testing [33]. However, because of our reliance on Green's regularity lemma, the soundness parameters that we can guarantee are significantly worse than what were previously known in the literature. More details will be provided in Chapter 4.

Parallel works: After we completed our work, we learned from Asaf Shapira that independently of us, Conjecture 1.1.2 has been fully resolved in his paper [40]. Shapira's work is built on the technique developed by Král', Serra, and Vena in [28]. where they also establish Green's conjecture in the special case when the matrix M is graphic. However, their proof is different from ours. The three authors demonstrate a reduction from functions to graphs so that the number of M-patterns in a function is equal (up to scaling) to the number of copies of a certain subgraph in a graph. Then they apply known "removal lemmas" for graphs in the literature to derive a "removal lemma" for functions. In this manner, Král', Serra, and Vena show that Theorem 1.1.3 holds as well. By extending this method and utilizing known removal lemmas for hypergraphs, Shapira [40] and Král', Serra, and Vena in a followup work [29] both independently resolve Green's conjecture for any arbitrary matrix M.

1.2 Organization

In Chapter 2, we describe the background on Fourier analysis and the Gowers norm that we need to obtain our results, and we also give an overview of our proofs. The dictatorship test is presented and analyzed in Chapter 3, and we discuss linear-invariant and non-linear properties in Chapter 4. Finally in Chapter 5, we describe some open problems and possible new lines of research from this thesis.

Chapter 2

The Gowers norm and technique overview

In this chapter, we provide the necessary background needed for our results and give an overview of how we use the Gowers norm in our proofs. First, let us fix some notation. We let $[n]$ denote the set $\{1, 2, \ldots, n\}$. For a vector $v \in \{0,1\}^n$, we write $|v| = \sum_{i \in [n]} v_i$ to be the Hamming weight of v. We let $\wedge$ denote the boolean AND, where $a \wedge b = 1$ iff $a = b = 1$. For vectors $v, w \in \{0,1\}^n$, we write $v \wedge w$ to denote the vector obtained by applying AND to v and w component-wise. We abuse notation and sometimes interpret a vector $v \in \{0,1\}^n$ as a subset $v \subseteq [n]$, where $i \in v$ iff $v_i = 1$.

2.1 Fourier analysis

Throughout the rest of the thesis, we shall examine the Fourier transform of a function.

Definition 2.1.1 (Fourier transform)**.** For a real-valued function $f : \{0,1\}^n \to \mathbb{R}$, we define its Fourier transform $\widehat{f} : \{0,1\}^n \to \mathbb{R}$ to be

$$\widehat{f}(\alpha) = \mathop{\mathbb{E}}_{x \in \{0,1\}^n} f(x)\chi_\alpha(x),$$

where $\chi_\alpha(x) = (-1)^{\sum_{i\in[n]} \alpha_i x_i}$. We say $\widehat{f}(\alpha)$ is the Fourier coefficient of f at α, and the *characters* of $\{0,1\}^n$ are the functions $\{\chi_\alpha\}_{\alpha\in\{0,1\}^n}$.

It is easy to see that for $\alpha, \beta \in \{0,1\}^n$, $\mathbb{E}\,\chi_\alpha \cdot \chi_\beta$ is 1 if $\alpha = \beta$ and 0 otherwise. Since there are 2^n characters, the characters form an orthonormal basis for functions on $\{0,1\}^n$, and we have the Fourier inversion formula

$$f(x) = \sum_{\alpha\in\{0,1\}^n} \widehat{f}(\alpha)\chi_\alpha(x)$$

and Parseval's Identity

$$\sum_{\alpha\in\{0,1\}^n} \widehat{f}(\alpha)^2 = \mathop{\mathbb{E}}_{x}\left[f(x)^2\right].$$

2.2 Gowers norm

In [15], Gowers uses analytic techniques to give a new proof of Szemerédi's Theorem [42] and in particular, initiates the study of a new measure of functions. Subsequently this measure is termed the *Gowers uniformity norm* and has been intensively studied and applied in additive combinatorics.

Definition 2.2.1. Let $f : \{0,1\}^n \to \mathbb{R}$. For every $d \in \mathbb{Z}^+$, we define the *d-th dimension Gowers uniformity norm (the U_d norm) of f* to be

$$\|f\|_{U_d} = \left(\mathop{\mathbb{E}}_{x,\ x_1,\ldots,x_d\in\{0,1\}^n} \left[\prod_{S\subseteq[d]} f\left(x + \sum_{i\in S} x_i\right)\right]\right)^{\frac{1}{2^d}}.$$

The definition can be easily extended to complex-valued functions, but since we shall only work with real-valued functions, the above definition suffices for us. In particular, when f is a boolean function, one can interpret $\|f\|_{U_d}^{2^d}$ as simply the expected number of "affine parallelepipeds" of dimension d in f.

As shown by Gowers [15], for every $d \geq 2$, the expression $\|\cdot\|_{U_d}$ is indeed a norm for functions $f : \{0,1\}^n \to \mathbb{R}$. (For $d = 1$, $\|f\|_{U_1} = |\mathbb{E}\, f|$ and is a semi-norm as $\|f\|_{U_1} = 0$

does not necessarily imply that $f = 0$.) In addition, the norms are monotonically increasing, i.e.,

$$\|\cdot\|_{U_1} \leq \|\cdot\|_{U_2} \leq \ldots \leq \|\cdot\|_{U_d} \leq \ldots$$

Furthermore, for every positive integer d, if f has positive $d+1$-th Gowers norm, then f correlates with some *degree d phase function*, i.e., $(-1)^g$ for some polynomial g of degree d.

Fact 2.2.1 ([15, 19])**.** *Let $d \in \mathbb{Z}$, $\epsilon > 0$. Let $P : \{0,1\}^n \to \{0,1\}$ be a polynomial of degree at most d and $f : \{0,1\}^n \to \mathbb{R}$. Suppose $\left|\mathbb{E}_x f(x)(-1)^{P(x)}\right| \geq \epsilon$. Then $\|f\|_{U_{d+1}} \geq \epsilon$.*

It is conjectured that the inverse also holds – if a bounded function f has positive U_{d+1} Gowers norm, then f correlates with some degree d phase function. Such a formulation is known as the *Gowers Inverse Conjecture*. For $d = 1$, the inverse holds and has a short proof. Since we need this, we record the easy proof below by first noting that the U_2 norm is precisely the ℓ_4 norm of a function's Fourier transform.

Proposition 2.2.2 ([15])**.** *Let $f : \{0,1\}^n \to \mathbb{R}$. Then*

$$\|f\|_{U_2}^4 = \sum_{\alpha \in \{0,1\}^n} \widehat{f}(\alpha)^4 .$$

Proof. The equality follows by a straightforward Fourier expansion.

$$\begin{aligned}
\|f\|_{U_2}^4 &= \mathop{\mathbb{E}}_{x,y,z} f(x)f(x+y)f(x+z)f(x+y+z) \\
&= \mathop{\mathbb{E}}_{x,y} f(x)f(x+y) \sum_{\alpha,\beta} \widehat{f}(\alpha)\widehat{f}(\beta)\chi_\beta(y) \mathop{\mathbb{E}}_{z} \chi_{\alpha+\beta}(x+z) \\
&= \sum_{\alpha} \widehat{f}(\alpha)^2 \mathop{\mathbb{E}}_{x,y} f(x)f(x+y)\chi_\alpha(y) \\
&= \sum_{\alpha} \widehat{f}(\alpha)^2 \sum_{\beta,\gamma} \widehat{f}(\beta)\widehat{f}(\gamma) \mathop{\mathbb{E}}_{x} \chi_{\beta+\gamma}(x) \mathop{\mathbb{E}}_{y} \chi_{\alpha+\gamma}(y) \\
&= \sum_{\alpha} \widehat{f}(\alpha)^4.
\end{aligned}$$

□

Note that the magnitude of the Fourier coefficient $\left|\widehat{f}(\alpha)\right|$ measures the correlation between f and the linear phase function χ_α. Thus, Proposition 2.2.3 implies that the U_2 norm of a *bounded* function measures its correlation with the set of linear phase functions.

Proposition 2.2.3 (Inverse for U_2)**.** *Let $f : \{0,1\}^n \to [-1,1]$. Then there exists some $\alpha \in \{0,1\}^n$ such that $\|f\|_{U_2} \leq \left|\widehat{f}(\alpha)\right|^{1/2}$.*

Proof. By Proposition 2.2.2, it follows that

$$\begin{aligned} \|f\|^4 &= \sum_\alpha \widehat{f}(\alpha)^4 \\ &\leq \max_\alpha \widehat{f}(\alpha)^2 \cdot \sum_\alpha \widehat{f}(\alpha)^2. \end{aligned}$$

Since the magnitude of f is bounded by 1, By Parseval's Identity, $\sum_\alpha \widehat{f}(\alpha)^2 = \mathbb{E}_x\left[f(x)^2\right] \leq 1$, and thus the proposition follows.

□

For $d = 2$, showing the fact that if $\|f\|_{U_3}$ is positive, then the function f correlates with a quadratic becomes significantly more involved. An inverse theorem for U_3 norm is proved by Samorodnitsky [37] for finite fields of even characteristic and by Green and Tao [19] for odd characteristic. For $d = 3$, the conjecture is shown to be false for finite fields with low characteristic by Lovett, Meshulam, and Samorodnitsky [30] and Green and Tao [18], though some version of the inverse conjecture has been shown recently. We provide more detail on this conjecture in Section 5.3 where we describe some open problems.

For us, we simply need to understand how to apply the Gowers norm. While the expression may look cumbersome at first glance, the Gowers norm may be used to control some other expressions, which may seem harder to analyze. For instance, to count the number of $(k+1)$-term progressions of the form $(x, x+y, \ldots, x+k\cdot$

y) in a subset, one may be interested in approximating expressions of the form $\mathbb{E}_{x,y}[f_0(x)f_1(x+y)\cdots f_k(x+k\cdot y)]$, which as shown by Gowers, can be bounded above by the minimum (over $i \in \{0,1,\ldots,k\}$) of $\|f_i\|_{U_k}$, where $f_0,\ldots,f_k$ are some bounded functions over some appropriate domain. Thus, in a rough sense, questions regarding progressions are then reduced to questions regarding the Gowers norms, which are more amenable to analytic techniques.

The proof showing that the expectation $\mathbb{E}_{x,y}[f_0(x)f_1(x+y)\cdots f_k(x+k\cdot y)]$ is bounded above by the minimum Gowers norm of all the functions f_i is not difficult; it proceeds by repeated applications of the Cauchy-Schwarz inequality and substitution of variables. Collectively, statements saying that certain expectations can be bounded above by the Gowers norm are coined *von-Neumann type theorems* in the additive combinatorics literature. The most general form of the statements appears in [20] where Green and Tao show that a system of linear equations of low complexity can be analyzed by a low dimension Gowers norm.

Definition 2.2.2. Suppose $L_1,\ldots,L_k : \{0,1\}^{n\times m} \to \{0,1\}$ are binary linear equations over m variables. We say that the system of linear equations $(L_1,\ldots,L_k)$ has *complexity* d if for each $i \in [k]$, we can cover the set of linear equations $\{L_j\}_{j\in[k]\setminus\{i\}}$ by $d+1$ classes so that L_i does not lie in the span of any of these classes.

Remark. As explain in [20], the definition of complexity remains unchanged if one replace the word "cover" by "partition."

The following statement is implicit in [20]. The version we state requires the functions f_i to be over $\{0,1\}^n$ and possibly distinct; however as explained by Gowers and Wolf [16], both conditions can be easily satisfied.

Proposition 2.2.4 (von Neumann-type theorem)**.** *Let $f_1,\ldots,f_k : \{0,1\}^n \to [-1,1]$ and suppose the system of linear equations $(L_1,\ldots,L_k)$ over m variables has complexity d. Then*

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_m\in\{0,1\}^n}\left[\prod_{i\in[k]} f_i\left(L_i\left(x_1,\ldots,x_m\right)\right)\right] \le \min_{i\in[k]}\|f_i\|_{U_{d+1}}.$$

We give some examples to help the readers become familiarized with Definition 2.2.2.

Example 2.2.1. Let $x_1, \ldots, x_m$ be a collection of m variables. Consider the following $m+1$ linear equations: for $i \in [m]$, $L_i(x_1, \ldots, x_m) = x_i$, and $L_{m+1}(x_1, \ldots, x_m) = \sum_{i=1}^{m} x_i$. It is easy to check that the system of linear equations $(L_1, \ldots, L_{m+1})$ has complexity exactly 1.

Example 2.2.2. Let x, y, z be variables, and the linear equations are the following: $L_1(x, y, z) = x$, $L_2(x, y, z) = y$, $L_3(x, y, z) = z$, $L_4(x, y, z) = x + y$, $L_5(x, y, z) = x + z$, $L_6(x, y, z) = y + z$. It can also be checked that these six equations have complexity exactly 1.

Example 2.2.3. Let x, y, z be variables, and the linear equations are the following: $L_1(x, y, z) = x$, $L_2(x, y, z) = y$, $L_3(x, y, z) = z$, $L_4(x, y, z) = x + y$, $L_5(x, y, z) = x + z$, $L_6(x, y, z) = y + z$, and $L_7(x, y, z) = x + y + z$. This system of linear equations has complexity exactly 2.

Example 2.2.4. In general, let $x_1, \ldots, x_d$ be variables, and the linear equations are the $2^d - 1$ nontrivial linear combinations of these d variables. This system has complexity $d - 1$.

We shall come across the following variant of Gowers norm in Chapter 3:

Definition 2.2.3. Let $\{f_S\}_{S \subseteq [d]}$ be a collection of functions where $f_S : \{0,1\}^n \to [-1, 1]$. We define the *d-th dimension Gowers linear inner product* of $\{f_S\}$ to be

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

By the preceding example, $\langle \{f_S\} \rangle_{LU_d} \leq \min_{S \neq 0 \subseteq [d]} \|f_S\|_{U_d}$. We won't need this explicitly as we rely simply on Lemma 3.1.3 from [39], where a form of the von-Neumann type theorem was used.

2.3 Technique overview

Having described the Gowers norm, we provide some intuition behind the design and analysis of our two testers. In general, to design a testing algorithm, ideally one would like the algorithm to have perfect completeness ($c = 1$) and its soundness s to be as small as possible. To achieve this, whenever $f \in \mathcal{NO}$, many sequences of q queries into f must cause the tester to reject. So to design a tester with q queries for a given property $(\mathcal{YES}, \mathcal{NO})$, we look for a distribution $\mathcal{D}$ on $\{0,1\}^{n\times q}$, and a predicate $\psi : \{0,1\}^q \to \{0,1\}$, such that

- if $f \in \mathcal{YES}$, then $\Pr_{(v_1,\ldots,v_q)\in\mathcal{D}}[\psi(f(v_1),\ldots f(v_q)) = 1] = 1$, and
- if $f \in \mathcal{NO}$, then $\Pr_{(v_1,\ldots,v_q)\in\mathcal{D}}[\psi(f(v_1),\ldots f(v_q)) = 1] \leq s + o(1)$.

Once we have identified a distribution $\mathcal{D}$ along with a predicate ψ, it is easy to design a tester:

> TEST T: with oracle access to f,
>
> 1. Pick $v = (v_1, \ldots, v_q)$ according to the distribution $\mathcal{D}$.
> 2. Query f at the points $v_1, \ldots, v_q$.
> 3. Accept iff $\psi(f(v_1),\ldots f(v_q)) = 1$.

In other words, the design of a tester boils down to finding a pair $(\mathcal{D}, \psi)$. For our dictatorship test in Chapter 3, specifying a pair $(\mathcal{D}, \psi)$ requires some effort. On the other hand, for our M-freeness test in Chapter 4, the matrix M already specifies an inherent pair $(\mathcal{D}, \psi)$.

The analysis of a tester then reduces to estimating p, the acceptance probability of tester T. The completeness is typically easy to analyze. Elements in $\mathcal{YES}$ have an explicit description, and so showing that $\psi(f(v_1),\ldots,f(v_q)) = 1$ for each v in the support of $\mathcal{D}$ becomes an explicit computation. The soundness is harder and requires a more delicate argument.

Our proof strategy for analyzing the soundness is as follows. First consider the case when f is random. Specifically, let $0 \leq \delta < 1$ and choose a function f uniformly at random among all boolean functions with density δ. It is easy to compute p when f is random with density δ, as this can be done using standard probabilistic estimates. However, we want to estimate p when $f \in \mathcal{NO}$. To do this we need a notion of "pseudorandomness" for functions. We will also define an appropriate "extension" $\Phi(f)$ of f so that this probability,

$$\Pr_{v \in \mathcal{D}} \left[\psi(\Phi(f)(v_1), \ldots, \Phi(f)(v_q)) | \Phi(f) \text{ is pseudorandom}\right],$$

is easy to analyze and can help us in estimating p.

For the notion of pseudorandomness, we say that a function f is *d-th pseudorandom* if $\|f - \mathbb{E} f\|_{U_{d+1}} = o(1)$, and a function is *pseudorandom* if it is d-th pseudorandom for every $d \geq 1$. By Fact 2.2.1, a pseudorandom function has $o(1)$ correlation with any low-degree polynomial phase functions. The specification of a distribution $\mathcal{D}$, a predicate ψ, and an extension Φ depends on the underlying property. We shall instantiate these concepts clearly when we present the analysis of our tests.

Chapter 3

Dictatorship testing

We prove Theorem 1.1.1 in this chapter. To do so, we shall clearly specify the property $(\mathcal{YES}, \mathcal{NO})$ for dictatorship testing. We will actually prove a stronger multifunction version instead – the test has oracle access to multiple functions as opposed to just one. Within this chapter only, we assume all boolean functions are of the form $f : \{0,1\}^n \to \{\text{-}1, 1\}$, following the standard notational change in the PCP literature, and we identify the boolean TRUE with -1 and FALSE with 1.

3.1 Preliminaries

Definition 3.1.1. (dictatorship) For $i \in [n]$, the *i-th dictator* is the function $f(x) = (-1)^{x_i}$.

In the PCP literature, the i-th dictator is also known as the long code encoding of i, $\langle (-1)^{x_i} \rangle_{x \in \{0,1\}^n}$, which is simply the evaluation of the i-th dictator function at all points.

The set of $\mathcal{YES}$ instances consists of dictator functions. As discussed in Chapter 1, defining $\mathcal{NO}$ to be functions that are far from $\mathcal{YES}$ is the strongest possible requirement for a tester's soundness. In this strictest sense, a tester must reject all functions that are far from being a dictator. However, Håstad [22] notices that one can relax the soundness requirement for dictatorship testing and still construct PCPs. Such a dictatorship test, however, must still reject functions that are "egalitarian,"

such as linear phase function χ_S with $|S|$ large and $(-1)^{\mathrm{Maj}}$, where $\mathrm{Maj}(x)$ is 1 if $|x| \geq n/2$ and 0 otherwise. Nonetheless, Håstad's test is not required to reject with high probability the function $(-1)^{x_i+x_j}$ which is $\frac{1}{2}$-far from every dictator function.

Håstad [22]: A dictatorship test must reject functions that are $\left(\frac{1}{2} - o(1)\right)$-far from every "*junta*," functions that depend only on $O(1)$ variables.

Clearly a dictator function is a junta, and so trivially a function that is $\left(\frac{1}{2} - o(1)\right)$-far from every junta is $\left(\frac{1}{2} - o(1)\right)$-far from every dictator as well. Since the introduction of Khot's unique label cover [26], many papers, such as [13, 27, 39], relax the soundness requirement further. Consider the following example, $f(x) = (-1)^{x_1+\mathrm{Maj}(x_2,\ldots,x_n)}$. By computing the Fourier transform of Maj, it can be shown that f is $\left(\frac{1}{2} - o(1)\right)$-far from every junta. On the other hand, f is "dominated" by the variable x_1. In fact, the variable x_1 has positive "low-degree influence" in f. The tests in [13, 27, 39] simply require that functions with no positive "low-degree influence" be rejected.

Unique game based construction [13, 27, 39]: A dictatorship test must reject functions that have no variable with positive "low-degree influence."

Such a soundness condition is sufficient for (conditional) PCPs, and we define the notion of influence in the next section.

3.1.1 Influence of variables

For a boolean function $f : \{0,1\}^n \to \{\text{-}1, 1\}$, the *influence* of the i-variable, $\mathrm{I}_i(f)$, is defined to be $\Pr_{x\in\{0,1\}^n}[f(x) \neq f(x+e_i)]$, where e_i is a vector in $\{0,1\}^n$ with 1 on the i-th coordinate and 0 everywhere else. This corresponds to our intuitive notion of influence of the i-th variable: how likely the outcome of f changes when the i-th variable on a random input is flipped. For us, it is actually more convenient to work with the Fourier analytic definition of $\mathrm{I}_i(f)$ defined below.

Definition 3.1.2. Let $f : \{0,1\}^n \to \mathbb{R}$. For $i \in [n]$, we define the *influence of the i-th variable* of f to be

$$\mathrm{I}_i(f) = \sum_{\alpha\in\{0,1\}^n:\ \alpha_i=1} \widehat{f}(\alpha)^2.$$

It is easy to verify that the two definitions coincide when f is a boolean function. Though we do not need this fact, we include a short proof for completeness to help the readers become familiarized with the definition.

Proposition 3.1.1. *Let $f : \{0,1\}^n \to \{-1,1\}$. For $i \in [n]$,*

$$\Pr_{x\in\{0,1\}^n}[f(x) \neq f(x+e_i)] = \sum_{\alpha\in\{0,1\}^n:\alpha_i=1} \widehat{f}(\alpha)^2.$$

Proof.

$$\begin{aligned}
\Pr_{x\in\{0,1\}^n}[f(x) \neq f(x+e_i)] &= \mathop{\mathbb{E}}_{x\in\{0,1\}^n}\left[\frac{1-f(x)f(x+e_i)}{2}\right] \\
&= \frac{1}{2}\left[1-\sum_{\alpha}\widehat{f}(\alpha)^2\chi_\alpha(e_i)\right] \\
&= \frac{1}{2}\left[1+\sum_{\alpha:\alpha_i=1}\widehat{f}(\alpha)^2-\sum_{\alpha:\alpha_i=0}\widehat{f}(\alpha)^2\right] \\
&= \sum_{\alpha:\alpha_i=1}\widehat{f}(\alpha)^2,
\end{aligned}$$

where the last equality follows since $\sum_\alpha \widehat{f}(\alpha)^2 = \mathbb{E}_x f(x)^2$ by Parseval's Identity and is equal to 1.

□

We now define the notion of *low-degree influence.*

Definition 3.1.3. Let w be an integer between 0 and n. We define the *w-th degree influence of the i-th variable* of a function $f : \{0,1\}^n \to \mathbb{R}$ to be

$$\mathrm{I}_i^{\leq w}(f) = \sum_{\alpha\in\{0,1\}^n:\ \alpha_i=1,\ |\alpha|\leq w} \widehat{f}(\alpha)^2.$$

Clearly for every w, $\mathrm{I}_i^{\leq w}(f) \leq \mathrm{I}_i(f)$ by definition. While the definition of influence of a variable is very combinatorial (Proposition 3.1.1 can be generalized to non-boolean functions), the definition of low-degree influence is less motivated. We give a few examples.

Example 3.1.1. Let $f(x) = (-1)^{x_i}$ be the i-dictator. Then for every positive integer w, $\mathrm{I}_i^{\leq w}(f) = 1$ but for every $j \neq i$, $\mathrm{I}_j(f) = 0$.

Example 3.1.2. Let $f(x) = \chi_S(x)$ with $|S| = \omega(1)$. For each $i \in S$, $\mathrm{I}_i(f) = 1$ but $\mathrm{I}_i^{\leq w}(f) = 0$ if w is fixed.

Example 3.1.3. Let $f(x) = (-1)^{\mathrm{Maj}(x_1,\ldots,x_n)}$ and assume n is odd. By Proposition 3.1.1, $\mathrm{I}_i(f)$ is the probability that $\sum_{j\neq i} x_j = \frac{n-1}{2}$ for a random $x \in \{0,1\}^n$. By Stirling's approximation, $\mathrm{I}_i(f) = O\left(\frac{1}{\sqrt{n}}\right)$.

Example 3.1.4. Let $f(x) = (-1)^{x_1+\mathrm{Maj}(x_2,\ldots,x_n)}$. Then $\mathrm{I}_1^{\leq w}(f)$ is the ℓ_2 norm of the Fourier transform of the Majority function on $n-1$ variables, up to weight w, which is known to be positive.

As seen in these preceding examples, a linear phase function χ_S with $|S|$ large has many variables of influence 1 but no variable with positive low-degree influence. Furthermore, one can indeed verify that for a fixed integer w, a bounded function has only a finite number of variables with positive w-th degree influence. This easy fact is the reason why a dictatorship test based on low-degree influence is sufficient for PCPs.

We conclude this section by stating two lemmas from [39] that we shall need.

Lemma 3.1.2 ([39])**.** *Let $f_1, \ldots, f_k : \{0,1\}^n \to [-1,1]$ be a collection of k bounded real-valued functions, and define $f(x) = \prod_{i=1}^k f_i(x)$ to be the product of these k functions. Then for each $i \in [n]$,*

$$\mathrm{I}_i(f) \leq k \cdot \sum_{j=1}^{k} \mathrm{I}_i(f_j).$$

When $\{f_i\}$ are boolean functions, it is easy to see that $\mathrm{I}_i(f) \leq \sum_{j=1}^k \mathrm{I}_i(f_j)$ by the union bound. The next lemma implies that if a function has positive U_d norm for some d, then it has some variable with positive influence.

Furthermore it is shown in [39] that if a collection of functions has large (linear) Gowers inner product, then two functions must share an influential variable.

Lemma 3.1.3 ([39]). *Let $\{f_S\}_{S\subseteq[d]}$ be a collection of bounded functions of the form $f_S : \{0,1\}^n \to [-1,1]$. Suppose $\left\langle \{f_S\}_{S\subseteq[d]} \right\rangle_{LU_d} \geq \epsilon$ and $\mathbb{E}\, f_{[d]} = 0$. Then there exists some variable $i \in [n]$, some subsets $S \neq T \subseteq [d]$ such that the influences of the i-th variable in both f_S and f_T are at least $\frac{\epsilon^4}{2^{O(d)}}$.*

3.1.2 Problem statement

Now we can define a t-function dictatorship test formally.

Definition 3.1.4. We say that a test $T = T^{f_1,\ldots,f_t}$ is a *t-function dictatorship test* with completeness c and soundness s if T is given oracle access to a family of t functions $f_1, \ldots, f_t : \{0,1\}^n \to \{\text{-}1, 1\}$, such that

- if there exists some variable $i \in [n]$ such that for all $a \in [t]$, $f_a(x) = (-1)^{x_i}$, then T accepts with probability at least c, and
- for every $\epsilon > 0$, there exist a positive constant $\tau > 0$ and a fixed positive integer w such that if T accepts with probability at least $s + \epsilon$, then there exist two functions f_a, f_b where $a, b \in [t], a \neq b$ and some variable $i \in [n]$ such that $\mathrm{I}_i^{\leq w}(f_a), \mathrm{I}_i^{\leq w}(f_b) \geq \tau$.

Remark. Following our previous terminology, $(f_1, \ldots, f_t) \in \mathcal{YES}$ if all functions are the same dictator function, and $(f_1, \ldots, f_t) \in \mathcal{NO}$ if for every $a \neq b \in [t]$, $i \in [n]$, $w \in \mathbb{Z}^+$, at least one of $\mathrm{I}_i^{\leq w}(f_a)$ and $\mathrm{I}_i^{\leq w}(f_b)$ is $o(1)$.

A q-function dictatorship test making q queries, with soundness $\frac{q+1}{2^q}$ was proved in [39], but the test suffers from imperfect completeness. We obtain a $(q - O(\log q))$-dictatorship test that makes q queries, has completeness 1, soundness $\frac{O(q^3)}{2^q}$, and in particular has amortized query complexity $1 + O\left(\frac{\log q}{q}\right)$, the same as the test in [39]. By a simple change of variable, we can more precisely state the following:

Theorem 3.1.4 (Theorem 1.1.1 restated). *For infinitely many t, there exists an adaptive t-function dictatorship test that makes $t+\log(t+1)$ queries, has completeness 1, and soundness $\frac{(t+1)^2}{2^t}$.*

Our test is adaptive and selects queries in two passes. During the first pass, it picks an arbitrary subset of $\log(t+1)$ functions out of the t functions. For each function selected, our test picks a random entry y and queries the function at entry y. Then based on the values of these $\log(t+1)$ queries, during the second pass, the test selects t positions nonadaptively, one from each function, then queries all t positions at once. The adaptivity is necessary in our analysis, and it is unclear if one can prove an analogous result with only one pass.

3.1.3 Folding

As introduced by Bellare, Goldreich, and Sudan [7], we shall assume that the functions are "folded" as we only access half of the entries of a function. We require our dictatorship test to make queries in a special manner. Suppose the test wants to query f at the point $x \in \{0,1\}^n$. If $x_1 = 1$, then the test queries $f(x)$ as usual. If $x_1 = 0$, then the test queries f at the point $\vec{1} + x = (1, 1+x_2, \ldots, 1+x_n)$ and negates the value it receives. It is instructive to note that folding ensures $f(\vec{1}+x) = -f(x)$ and $\mathbb{E}\, f = 0$.

3.2 Basic Test

For ease of exposition, we first consider the following simplistic scenario. Suppose we have oracle access to just one boolean function. Furthermore we ignore the tradeoff between soundness and query complexity. We simply want a dictatorship test that has completeness 1 and soundness $\frac{1}{2}$. There are many such tests in the literature; however, we need a suitable one to build a query efficient one later on. Our basic test below is a close variant of the one proposed by Guruswami, Lewin, Sudan, and Trevisan [21].

BASIC TEST T: with oracle access to f,

1. Pick x_i, x_j, y, z uniformly at random from $\{0,1\}^n$.

2. Query $f(y)$.

3. Let $v = \frac{1-f(y)}{2}$. Accept iff

$$f(x_i)f(x_j) = f(x_i + x_j + (v\vec{1} + y) \wedge z).$$

Proof heuristic: We now specify the distribution $\mathcal{D}$, predicate ψ, extension Φ, and the notion of pseudorandomness (from our discussion in Section 2.3) that we use to analyze T. The distribution $\mathcal{D}$ and the predicate ψ should be clear from the test T itself. $\mathcal{D}$ is a distribution on 3-tuples of the form $(x_i, x_j, x_i + x_j + (v\vec{1} + y) \wedge z)$, where the four vectors x_i, x_j, y, z are generated uniformly at random from $\{0,1\}^n$. $\psi : \{\text{-}1, 1\}^3 \to \{\text{Accept, Reject}\}$ accepts iff the number of -1 in its input is even. A function f is considered pseudorandom if its U_2 norm is $o(1)$. The extension Φ is the noise operator T_δ with noise parameter δ set to $\frac{1}{4}$. In other words, $\Phi(f)(x) = \mathbb{E}_y f(x+y)$, where each bit of y is independently chosen to be 1 with probability $\frac{1}{4}$ and 0 with probability $\frac{3}{4}$.

Our proof can be summarized as follows. We want to estimate p, the acceptance probability of T. Note that when f is a random function (for each x, $f(x)$ is a uniformly random ± 1 bit), then p is $\frac{1}{2} \pm o(1)$. Proposition 3.2.2 essentially implies that p can be bounded above by $\frac{1}{2} + \|\Phi(f)\|_{U_2}^4$. If $p \geq \frac{1}{2} + \epsilon$, then $\|\Phi(f)\|_{U_2}$ must be positive. By Fact 2.2.1, $\Phi(f)$ correlates with a linear phase function χ_α. Since $\Phi(f)$ is balanced, α must be nonempty and contain some coordinate i. Thus $\mathrm{I}_i(\Phi(f)) > 0$. By examining the Fourier transform of $\Phi(f)$, it follows that $\mathrm{I}_i^{\leq w}(f) > 0$ as well. We now proceed with a formal proof.

Lemma 3.2.1. *The test T is a dictatorship test with completeness* 1.

Proof. Suppose f is the ℓ-th dictator, i.e., $f(x) = (-1)^{x_\ell}$. First note that

$$v + y_\ell = \frac{1 - (-1)^{y_\ell}}{2} + y_\ell,$$

which evaluates to 0. Thus by linearity of f

$$\begin{aligned} f(x_i + x_j + (v\vec{1} + y) \wedge z) &= f(x_i)f(x_j)f((v\vec{1} + y) \wedge z) \\ &= f(x_i)f(x_j)(-1)^{(v+y_\ell)\wedge z_\ell} \\ &= f(x_i)f(x_j) \end{aligned}$$

and the test always accepts. □

To analyze the soundness of the test T, we need to derive a Fourier analytic expression for the acceptance probability of T.

Proposition 3.2.2. *Let p be the acceptance probability of T. Then*

$$p = \frac{1}{2} + \frac{1}{2} \sum_{\alpha \in \{0,1\}^n} \widehat{f}(\alpha)^3 \, 2^{-|\alpha|} \left(1 + \sum_{\beta \subseteq \alpha} \widehat{f}(\beta)\right).$$

For sanity check, let us interpret the expression for p. Suppose $f = \chi_\alpha$ for some $\alpha \neq \vec{0} \in \{0,1\}^n$, i.e., $\widehat{f}(\alpha) = 1$ and all other Fourier coefficients of f are 0. Then clearly $p = \frac{1}{2} + 2^{-|\alpha|}$, which equals 1 whenever f is a dictator function as we have just shown. If $|\alpha|$ is large, then T accepts with probability close to $\frac{1}{2}$. We shall first analyze the soundness and then derive this analytic expression for p.

Lemma 3.2.3. *The test T is a dictatorship test with soundness $\frac{1}{2}$.*

Proof. Suppose the test T passes with probability at least $\frac{1}{2} + \epsilon$, for some $\epsilon > 0$. By applying Proposition 3.2.2, Cauchy-Schwarz, and Parseval's Identity, respectively, we obtain

$$\begin{aligned} \epsilon &\leq \frac{1}{2} \sum_{\alpha \in \{0,1\}^n} \widehat{f}(\alpha)^3 \, 2^{-|\alpha|} \left(1 + \sum_{\beta \subseteq \alpha} \widehat{f}(\beta)\right) \\ &\leq \frac{1}{2} \sum_{\alpha \in \{0,1\}^n} \widehat{f}(\alpha)^3 \, 2^{-|\alpha|} \left(1 + \left(\sum_{\beta \subseteq \alpha} \widehat{f}(\beta)^2\right)^{\frac{1}{2}} \cdot 2^{\frac{|\alpha|}{2}}\right) \\ &\leq \sum_{\alpha \in \{0,1\}^n} \widehat{f}(\alpha)^3 \, 2^{-\frac{|\alpha|}{2}}. \end{aligned}$$

Pick the least positive integer w such that $2^{-\frac{w}{2}} \leq \frac{\epsilon}{2}$. Then by Parseval's again,

$$\begin{aligned} \frac{\epsilon}{2} &\leq \sum_{\alpha \in \{0,1\}^n : |\alpha| \leq w} \widehat{f}(\alpha)^3 \\ &\leq \max_{\alpha \in \{0,1\}^n : |\alpha| \leq w} \left| \widehat{f}(\alpha) \right|. \end{aligned}$$

So there exists some $\beta \in \{0,1\}^n, |\beta| \leq w$ such that $\frac{\epsilon}{2} \leq \left| \hat{f}(\beta) \right|$. With f being folded, $\beta \neq \vec{0}$. Thus, there exists an $i \in [n]$ such that $\beta_i = 1$ and

$$\frac{\epsilon^2}{4} \leq \widehat{f}(\beta)^2 \leq \sum_{\alpha \in \{0,1\}^n : \alpha_i = 1, |\alpha| \leq w} \widehat{f}(\alpha)^2.$$

□

Now we give the straightforward Fourier analytic calculation for p.

Proof of Proposition 3.2.2. As usual, we first arithmetize p. We write

$$\begin{aligned} p &= \mathop{\mathbb{E}}_{x_i, x_j, y, z} \left(\frac{1 + f(y)}{2} \right) \left(\frac{1 + \mathrm{Acc}(x_i, x_j, y, z)}{2} \right) + \\ &\quad \mathop{\mathbb{E}}_{x_i, x_j, y, z} \left(\frac{1 - f(y)}{2} \right) \left(\frac{1 + \mathrm{Acc}(x_i, x_j, \vec{1} + y, z)}{2} \right), \end{aligned}$$

where

$$\mathrm{Acc}(x_i, x_j, y, z) = f(x_i) f(x_j) f(x_i + x_j + (y \wedge z)).$$

Since f is folded, $f(\vec{1} + y) = -f(y)$. As y and $\vec{1} + y$ are both identically distributed in $\{0,1\}^n$, we have

$$p = 2 \mathop{\mathbb{E}}_{x_i, x_j, y, z} \left(\frac{1 + f(y)}{2} \right) \left(\frac{1 + \mathrm{Acc}(x_i, x_j, y, z)}{2} \right).$$

Since $\mathbb{E} f = 0$, we can further simplify the above expression to be

$$p = \frac{1}{2} + \frac{1}{2} \mathop{\mathbb{E}}_{x_i, x_j, y, z} \left[(1 + f(y)) \, \mathrm{Acc}(x_i, x_j, y, z) \right].$$

It suffices to expand out the terms

$$\mathop{\mathbb{E}}_{x_i,x_j,y,z}[\mathrm{Acc}(x_i,x_j,y,z)]$$

and

$$\mathop{\mathbb{E}}_{x_i,x_j,y,z}[f(y)\,\mathrm{Acc}(x_i,x_j,y,z)].$$

For the first term, it is not hard to show that

$$\mathop{\mathbb{E}}_{x_i,x_j,y,z}[\mathrm{Acc}(x_i,x_j,y,z)] = \sum_{\alpha\in\{0,1\}^n} \widehat{f}(\alpha)^3 \;\; 2^{-|\alpha|},$$

by applying the Fourier inversion formula on f and averaging over x_i and x_j and then averaging over y and z over the AND operator.

Now we compute the second term. Applying the Fourier inversion formula to the last three occurrences of f and averaging over x_i and x_j, we obtain

$$\mathop{\mathbb{E}}_{x_i,x_j,y,z}[f(y)\,\mathrm{Acc}(x_i,x_j,y,z)] = \sum_{\alpha\in\{0,1\}^n} \widehat{f}(\alpha)^3 \mathop{\mathbb{E}}_{y,z}\left[f(y)\chi_\alpha(y\wedge z)\right].$$

It suffices to expand out $\mathbb{E}_{y,z}\left[f(y)\chi_\alpha(y\wedge z)\right]$. By grouping the z's according to their intersection with different possible subsets β of α, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{y,z}\left[f(y)\chi_\alpha(y\wedge z)\right] \\
&= \sum_{\beta\subseteq\alpha} \Pr_{z\in\{0,1\}^n}[z\cap\alpha=\beta] \;\; \mathop{\mathbb{E}}_{y}\left[f(y)\prod_{i:\,\alpha_i=1}(-1)^{y_i\wedge z_i}\right] \\
&= \sum_{\beta\subseteq\alpha} 2^{-|\alpha|} \;\; \mathop{\mathbb{E}}_{y}\left[f(y)\prod_{i:\,\beta_i=1}(-1)^{y_i}\right] \\
&= 2^{-|\alpha|}\sum_{\beta\subseteq\alpha}\widehat{f}(\beta).
\end{aligned}$$

Putting everything together, it is easy to see that we have the Fourier analytic expression for p as stated in the lemma. □

3.3 Hypergraph Dictatorship Test

We prove Theorem 1.1.1 in this section. The basis of our hypergraph dictatorship test will be very similar to the test in the previous section. We remark that we did not choose to present the exact same basic test for hopefully a clearer exposition.

We now address the tradeoff between query complexity and soundness. If we simply repeat the basic test a number of iterations independently, the error is reduced, but the query complexity increases. In other words, the amortized query complexity does not change if we simply run the basic test for many independent iterations. Following Trevisan [44], all the dictatorship tests that save query complexity do so by reusing queries made in previous iterations of the basic test. To illustrate this idea, suppose test T queries f at the points $x_1 + h_1$, $x_2 + h_2$, $x_1 + x_2 + h_{1,2}$ to make a decision. For the second iteration, we let T query f at the points $x_3 + h_3$ and $x_1 + x_3 + h_{1,3}$ and reuse the value $f(x_1 + h_1)$ queried during the first run of T. T then uses the three values to make a second decision. In total T makes five queries to run two iterations.

We may think of the first run of T as parametrized by the points x_1 and x_2 and the second run of T by x_1 and x_3. In general, we may have k points $x_1, \ldots, x_k$ and a graph on $[k]$ vertices, such that each edge e of the graph corresponds to an iteration of T parametrized by the points $\{x_i\}_{i \in e}$. We shall use a complete hypergraph on k vertices to save on query complexity, and we will argue that the soundness of the algorithm decreases exponentially with respect to the number of iterations.

Formally, consider a hypergraph $H = ([k], E)$. Let $\{f_a\}_{a \in [k] \cup E}$ be a collection of boolean functions of the form $f_a : \{0,1\}^n \to \{\text{-}1, 1\}$. We assume all the functions are folded, and so in particular, $\mathbb{E}\, f_a = 0$. Consider the following test:

HYPERGRAPH H-TEST: with oracle access to $\{f_a\}_{a \in [k] \cup E}$,

1. Pick $x_1, \ldots, x_k, y_1, \ldots, y_k$, and $\{z_a\}_{a \in [k] \cup E}$ independently and uniformly at random from $\{0,1\}^n$.

2. For each $i \in [k]$, query $f_i(y_i)$.

3. Let $v_i = \frac{1-f_i(y_i)}{2}$.

   Accept iff for every $e \in E$,

$$\prod_{i \in e} \left[f_i(x_i + (v_i \vec{1} + y_i) \wedge z_i) \right] = f_e \left(\sum_{i \in e} x_i + \left(\Sigma_{i \in e} (v_i \vec{1} + y_i) \right) \wedge z_e \right).$$

Design heuristic: We make a few remarks regarding the design of H-Test. The hypergraph test in [39] accepts iff for every $e \in E$, $\prod_{i \in e} f_i(x_i + \eta_i)$ equals $f_e(\sum_{i \in e} x_i + \eta_e)$, where the bits in each vector η_a are chosen independently to be 1 with some small constant, say 0.01. The noise vectors η_a rule out the possibility that linear functions with large support can be accepted. To obtain a test with perfect completeness, we use ideas from [21, 33, 23] to simulate the effect of the noise perturbation.

Note that for y, z chosen uniformly at random from $\{0,1\}^n$, the vector $y \wedge z$ is a $\frac{1}{4}$-noisy vector. As observed by Parnas, Ron, and Samorodnitsky [33], the test $f(y \wedge z) = f(y) \wedge f(z)$ distinguishes between dictators and linear functions with large support. One can also combine linearity and dictatorship testing into a single test of the form $f(x_1 + x_2 + y \wedge z)(f(y) \wedge f(z)) = f(x_1)f(x_2)$ as Hstad and Khot demonstrated [23]. However, iterating this test is too costly for us. In fact, Hstad and Khot also consider an adaptive variant that reads $k^2 + 2k$ bits to obtain a soundness of 2^{-k^2}, the same parameters as in [38], while achieving perfect completeness as well. Without adaptivity, the test in [23] reads $k^2 + 4k$ bits. While both the nonadaptive and adaptive tests in [23] have the same amortized query complexity, extending the nonadaptive test by Hstad and Khot to the hypergraph setting does not work for us. So to achieve the same amortized query complexity as the hypergraph test in [39], we also exploit adaptivity in our test.

Proof heuristic: We now specify the distribution $\mathcal{D}$, predicate ψ, the extension Φ, and the notion of pseudorandomness that we use to analyze H. The distribution $\mathcal{D}$ and the predicate ψ should be clear from the test H itself. $\mathcal{D}$ is a distribution

on $|E| + k$ points, generated from the uniformly random vectors $\{x_i, y_i\}_{i\in[k]}$ and $\{z_a\}_{z\in[k]\cup E}$. The predicate ψ is the acceptance predicate of the test. A function f is considered pseudorandom if its U_k norm is $o(1)$. For a function $f : \{0,1\}^n \to \{\text{-}1, 1\}$, define the function $\Phi(f) : \{0,1\}^{2n} \to [-1, 1]$ on $2n$ variables to be $\Phi(f)(x; y) = \mathbb{E}_{z\in\{0,1\}^n} f(x + y \wedge z)$. The function $\Phi(f)$ can be thought of as the "extension" of f.

For simplicity, suppose all the functions f_a are the same. Our proof can then be summarized as follows. We want to estimate p, the acceptance probability of H-Test. In the proof of Lemma 3.3.3, we first show that p can be bounded above by $2^{k-|E|} + \|\Phi(f)\|_{U_d}^d$, for some $d \le k$. If $p \ge 2^{k-|E|} + \epsilon$, then $\|\Phi(f)\|_{U_d}$ must be positive. Since $\Phi(f)$ is balanced, by Lemma 3.1.3, there exists some variable i so that $\mathrm{I}_i(\Phi(f))$ is positive. Proposition 3.3.4 then relates the Fourier transform of $\Phi(f)$ with f's Fourier transform, and it follows that $\mathrm{I}_i^{\le w}(f) > 0$ for some fixed w. Taking H to be the complete hypergraph then finishes the argument. We now proceed with a formal proof.

Theorem 3.3.1 (Theorem 1.1.1 restated)**.** *For infinitely many t, there exists an adaptive t-function dictatorship test with $t + \log(t + 1)$ queries, completeness 1, and soundness $\frac{(t+1)^2}{2^t}$.*

Proof. Take a complete hypergraph on k vertices, where $k = \log(t+1)$. The statement follows by applying Lemmas 3.3.2 and 3.3.3. □

Lemma 3.3.2. *The H-Test is a $(k+|E|)$-function dictatorship test that makes $|E|+2k$ queries and has completeness 1.*

Proof. The test makes k queries $f_i(y_i)$ in the first pass, and based on the answers to these k queries, the test then makes one query into each function f_a, for each $a \in [k] \cup E$. So the total number of queries is $|E| + 2k$.

Now suppose all the functions are the ℓ-th dictator for some $\ell \in [n]$, i.e., for all $a \in [k] \cup E$, $f_a = f$, where $f(x) = (-1)^{x_\ell}$. Note that for each $i \in [k]$,

$$v_i + y_i(\ell) = \frac{1 - (-1)^{y_i(\ell)}}{2} + y_i(\ell),$$

which evaluates to 0. Thus for each $e \in E$,

$$\begin{aligned}\prod_{i\in e} f_i(x_i + (v_i\vec{1} + y_i) \wedge z_i) &= f\left(\sum_{i\in e} x_i\right) \cdot \prod_{i\in e} f((v_i\vec{1} + y_i) \wedge z_i) \\ &= f\left(\sum_{i\in e} x_i\right) \cdot \prod_{i\in e} (-1)^{(v_i + y_i(\ell)) \wedge z_i(\ell)} \\ &= f\left(\sum_{i\in e} x_i\right),\end{aligned}$$

and similarly,

$$f_e\left(\sum_{i\in e} x_i + \left(\Sigma_{i\in e}(v_i\vec{1} + y_i)\right) \wedge z_e\right) = f\left(\sum_{i\in e} x_i\right).$$

Hence the test always accepts. □

Lemma 3.3.3. *The H-Test has soundness* $2^{k-|E|}$.

Before proving Lemma 3.3.3 we first prove a proposition relating the Fourier transform of a function perturbed by noise to the function's Fourier transform itself.

Proposition 3.3.4. *Let* $f : \{0,1\}^n \to \{\text{-}1, 1\}$. *Define* $g : \{0,1\}^{2n} \to [-1, 1]$ *such that*

$$g(x; y) = \mathop{\mathbb{E}}_{z\in\{0,1\}^n} f(c' + x + (c + y) \wedge z),$$

where c, c' *are some fixed vectors in* $\{0,1\}^n$. *Then*

$$\widehat{g}(\alpha; \beta)^2 = \widehat{f}(\alpha)^2\, 1_{\{\beta\subseteq\alpha\}} 4^{-|\alpha|}.$$

Proof. This is a straightforward Fourier analytic calculation. By definition,

$$\widehat{g}(\alpha; \beta)^2 = \left(\mathop{\mathbb{E}}_{x,y,z\in\{0,1\}^n} f(c' + x + (c + y) \wedge z)\chi_\alpha(x)\chi_\beta(y)\right)^2.$$

By averaging over x it is easy to see that

$$\widehat{g}(\alpha; \beta)^2 = \widehat{f}(\alpha)^2 \left(\mathop{\mathbb{E}}_{y,z\in\{0,1\}^n} \chi_\alpha((c + y) \wedge z)\chi_\beta(y)\right)^2.$$

Since the bits of y are chosen independently and uniformly at random, if $\beta \backslash \alpha$ is nonempty, the above expression is zero. So we can write

$$\widehat{g}(\alpha;\beta)^2 = \widehat{f}(\alpha)^2\, 1_{\{\beta\subseteq\alpha\}} \left(\prod_{i\in\alpha\backslash\beta} \mathop{\mathbb{E}}_{y_i,z_i} (-1)^{(c_i+y_i)\wedge z_i} \cdot \prod_{i\in\beta} \mathop{\mathbb{E}}_{y_i,z_i} (-1)^{(c_i+y_i)\wedge z_i + y_i} \right)^2 .$$

It is easy to see that the term $\mathbb{E}_{y_i,z_i}(-1)^{(c_i+y_i)\wedge z_i}$ evaluates to $\frac{1}{2}$ and the term $\mathbb{E}_{y_i,z_i}(-1)^{(c_i+y_i)\wedge z_i+y_i}$ evaluates to $(-1)^{c_i}\frac{1}{2}$. Thus

$$\widehat{g}(\alpha;\beta)^2 = \widehat{f}(\alpha)^2 \;\; 1_{\{\beta\subseteq\alpha\}}\, 4^{-|\alpha|}$$

as claimed. □

Now we prove Lemma 3.3.3.

Proof of Lemma 3.3.3. Let p be the acceptance probability of H-test. Suppose that $2^{k-|E|} + \epsilon \leq p$. We want to show that there are two functions f_a and f_b such that for some $i \in [n]$, some fixed positive integer w, some constant $\epsilon' > 0$, it is the case that $\mathrm{I}_i^{\leq w}(f_a), \mathrm{I}_i^{\leq w}(f_b) \geq \epsilon'$. As usual we first arithmetize p. We write

$$p = \sum_{v\in\{0,1\}^k} \mathop{\mathbb{E}}_{\{x_i\},\{y_i\},\{z_a\}} \prod_{i\in[k]} \frac{1+(-1)^{v_i} f_i(y_i)}{2} \prod_{e\in E} \frac{1+\mathrm{Acc}(\{x_i,y_i,v_i,z_i\}_{i\in e}, z_e)}{2},$$

where

$$\begin{aligned} \mathrm{Acc}(\{x_i,y_i,v_i,z_i\}_{i\in e}, z_e) \;\; &= \;\; \prod_{i\in e} \Big[f_i(x_i + (v_i\vec{1} + y_i) \wedge z_i) \Big] \\ &\quad \cdot f_e \left(\sum_{i\in e} x_i + \Big(\Sigma_{i\in e}(v_i\vec{1} + y_i) \Big) \wedge z_e \right). \end{aligned}$$

For each $i \in [k]$, f_i is folded, so $(-1)^{v_i} f_i(y_i) = f_i(v_i\vec{1} + y_i)$. Since the vectors $\{y_i\}_{i\in[k]}$ are uniformly and independently chosen from $\{0,1\}^n$, for a fixed $v \in \{0,1\}^k$, the vectors $\{v_i\vec{1} + y_i\}_{i\in[k]}$ are also uniformly and independently chosen from $\{0,1\}^n$.

So we can simplify the expression for p and write

$$p = \mathop{\mathbb{E}}_{\{x_i\},\{y_i\},\{z_a\}} \left[\prod_{i\in[k]} (1 + f_i(y_i)) \prod_{e\in E} \frac{1 + (\mathrm{Acc}\{x_i, y_i, \vec{0}, z_i\}_{i\in e}, z_e)}{2} \right].$$

Instead of writing $\mathrm{Acc}(\{x_i, y_i, \vec{0}, z_i\}_{i\in e}, z_e)$, for convenience we shall write $\mathrm{Acc}(e)$ to be a notational shorthand. Observe that since $1 + f_i(y_i)$ is either 0 or 2, we may write

$$p \le 2^k \mathop{\mathbb{E}}_{\{x_i\},\{y_i\},\{z_a\}} \left[\prod_{e\in E} \frac{1 + \mathrm{Acc}(e)}{2} \right].$$

Note that the product of sums $\prod_{e\in E} \frac{1+\mathrm{Acc}(e)}{2}$ expands into a sum of products of the form

$$2^{-|E|} \left(1 + \sum_{\emptyset \neq E' \subseteq E} \prod_{e\in E'} \mathrm{Acc}(e) \right),$$

so we have

$$\frac{\epsilon}{2^k} \le \mathop{\mathbb{E}}_{\{x_i\},\{y_i\},\{z_a\}} \left[2^{-|E|} \sum_{\emptyset \neq E' \subseteq E} \prod_{e\in E'} \mathrm{Acc}(e) \right].$$

By averaging, there must exist some nonempty subset $E' \subseteq E$ such that

$$\frac{\epsilon}{2^k} \le \mathop{\mathbb{E}}_{\{x_i\},\{y_i\},\{z_a\}} \left[\prod_{e\in E'} \mathrm{Acc}(e) \right].$$

Let Odd consists of the vertices in $[k]$ with odd degree in E'. Expanding out the definition of $\mathrm{Acc}(e)$, we can conclude

$$\frac{\epsilon}{2^k} \le \mathop{\mathbb{E}}_{\{x_i\},\{y_i\},\{z_a\}} \left[\prod_{i\in \mathrm{Odd}} f_i(x_i + y_i \wedge z_i) \cdot \prod_{e\in E'} f_e \left(\sum_{i\in e} x_i + \left(\sum_{i\in e} y_i \right) \wedge z_e \right) \right].$$

We now define a family of functions that represent the "noisy versions" of f_a. For $a \in [k] \cup E$, define $g'_a : \{0,1\}^{2n} \to [-1, 1]$ to be

$$g'_a(x; y) = \mathop{\mathbb{E}}_{z\in\{0,1\}^n} f_a(x + y \wedge z).$$

Thus we have

$$\frac{\epsilon}{2^k} \leq \mathop{\mathbb{E}}_{\{x_i\},\{y_i\}} \left[\prod_{i \in \mathrm{Odd}} g'_i(x_i; y_i) \cdot \prod_{e \in E'} g'_e \left(\sum_{i \in e} x_i; \sum_{i \in e} y_i \right) \right].$$

Following the approach of [24, 39], we are going to reduce the analysis of the iterated test to one hyperedge. Let d be the maximum size of an edge in E', and without loss of generality, let $(1, \ldots, d)$ be a maximal edge in E'. Now, fix the values of $x_{d+1}, \ldots, x_k$ and $y_{d+1}, \ldots, y_k$ so that the following inequality holds:

$$\frac{\epsilon}{2^k} \leq \mathop{\mathbb{E}}_{x_1,y_1,\ldots,x_d,y_d} \left[\prod_{i \in \mathrm{Odd}} g'_i(x_i; y_i) \cdot \prod_{e \in E'} g'_e \left(\sum_{i \in e} x_i; \sum_{i \in e} y_i \right) \right]. \tag{3.1}$$

We group the edges in E' based on their intersection with $(1, \ldots, d)$. We rewrite Inequality 3.1 as

$$\frac{\epsilon}{2^k} \leq \mathop{\mathbb{E}}_{(x_1,y_1),\ldots,(x_d,y_d) \in \{0,1\}^{2n}} \left[\prod_{S \subseteq [d]} \prod_{a \in \mathrm{Odd} \cup E' : a \cap [d] = S} g_a \left(\sum_{i \in S} x_i; \sum_{i \in S} y_i \right) \right], \tag{3.2}$$

where for each $a \in [k] \cup E$, $g_a(x; y) = g'_a(c'_a + x; c_a + y)$, with $c'_a = \sum_{i \in a \setminus [d]} x_i$ and $c_a = \sum_{i \in a \setminus [d]} y_i$ fixed vectors in $\{0,1\}^n$.

By grouping the edges based on their intersection with $[d]$, we can rewrite Inequality 3.2 as

$$\begin{aligned} \frac{\epsilon}{2^k} &\leq \mathop{\mathbb{E}}_{(x_1,y_1),\ldots,(x_d,y_d) \in \{0,1\}^{2n}} \left[\prod_{S \subseteq [d]} G_S \left(\sum_{i \in S} (x_i; y_i) \right) \right] \\ &= \left\langle \{G_S\}_{S \subseteq [d]} \right\rangle_{LU_d}, \end{aligned}$$

where G_S is simply the product of all the functions g_a such that $a \in \mathrm{Odd} \cup E'$ and $a \cap [d] = S$.

Since $(1, \ldots, d)$ is maximal, all the other edges in E' do not contain $(1, \ldots, d)$ as a subset. Thus $G_{[d]} = g_{[d]}$ and $\mathbb{E}\, G_{[d]} = 0$. By Lemma 3.1.3, the linear Gowers inner product of a family of functions $\{G_S\}$ being positive implies that two functions from

the family must share a variable with positive influence. More precisely, there exist $S \neq T \subseteq [d]$, $i \in [2n]$, $\tau > 0$, such that $\mathrm{I}_i(G_S), \mathrm{I}_i(G_T) \geq \tau$, where $\tau = \frac{\epsilon^4}{2^{O(d)}}$.

Note that $G_\emptyset$ is the product of all the functions g'_a that are indexed by vertices or edges outside of $[d]$. So $G_\emptyset$ is a constant function, and all of its variables clearly have influence 0. Thus neither S nor T is empty. Since G_S and G_T are products of at most 2^k functions, by Lemma 3.1.2 there must exist some $a \neq b \in [d] \cup E'$ such that $\mathrm{I}_i(g_a), \mathrm{I}_i(g_b) \geq \frac{\tau}{2^{2k}}$. Recall that we have defined $g_a(x; y)$ to be $\mathbb{E}_z\, f_a(c'_a + x + (c_a + y) \wedge z)$. Thus we can apply Proposition 3.3.4 to obtain

$$\begin{aligned}
\mathrm{I}_i(g_a) &= \sum_{(\alpha,\beta)\in\{0,1\}^{2n};\, i\in(\alpha,\beta)} \widehat{g}_a(\alpha;\beta)^2 \\
&= \sum_{\alpha\in\{0,1\}^n;\, i\in\alpha} \sum_{\beta\subseteq\alpha} \hat{f}_a(\alpha)^2\, 4^{-|\alpha|} \\
&= \sum_{\alpha\in\{0,1\}^n;\, i\in\alpha} \hat{f}_a(\alpha)^2\, 2^{-|\alpha|}.
\end{aligned}$$

Let w be the least positive integer such that $2^{-w} \leq \frac{\tau}{2^{2k+1}}$. Then it is easy to see that $\mathrm{I}_i^{\leq w}(f_a) \geq \frac{\tau}{2^{2k+1}}$. Similarly, $\mathrm{I}_i^{\leq w}(f_b) \geq \frac{\tau}{2^{2k+1}}$ as well. Hence this completes the proof. □

Chapter 4

Linear-invariant non-linear properties

In this chapter, we consider boolean functions on boolean cube written in the form $f : \{0,1\}^n \to \{0,1\}$, and at times, it may be helpful to view a function f as a subset $A \subseteq \{0,1\}^n$, where $A = \{x : f(x) = 1\}$. If H is a subspace of $\{0,1\}^n$, for any $g \in \{0,1\}^n$, we write $f_{g+H} : H \to \{0,1\}$ to denote $f_{g+H}(h) = f(g+h)$. In other words, f_{g+H} is the restriction of f to the affine subspace $g + H$.

4.1 Problem statement

We begin by giving a more general definition of freeness of a single forbidden pattern.

Definition 4.1.1. Let M be an r by k matrix over $\{0,1\}$ and $\sigma \in \{0,1\}^k$. We say that a function $f : \{0,1\}^n \to \{0,1\}$ is (M,σ)-free if there are no $x_1, \ldots, x_k \in \{0,1\}^n$ such that $\langle f(x_1), \ldots, f(x_k)\rangle = \sigma$ and $M\vec{x} = \vec{0}$, where $\vec{x} = (x_1, \ldots, x_k)$.

We define formally the property of freeness of a single forbidden pattern.

Definition 4.1.2. For $\epsilon > 0$, we define the property of (M,σ)-free to be the pair $(\mathcal{YES}_{M,\sigma}, \mathcal{NO}_\epsilon)$, where $\mathcal{YES}_{M,\sigma}$ consists of functions $f : \{0,1\}^n \to \{0,1\}$ that are (M,σ)-free, and $\mathcal{NO}_\epsilon$ consists of functions that are ϵ-far from $\mathcal{YES}_{M.\sigma}$.

To help the readers understand our exposition, we observe that the property of freeness is linear-invariant.

Observation 4.1.1. *Let M be an r by k matrix over $\{0,1\}$, $\sigma \in \{0,1\}^k$, and L a linear transformation on $\{0,1\}^n$. If f is (M,σ)-free, then $f \circ L$ is also (M,σ)-free.*

Proof. Suppose not, and there exist some $x_1, \ldots, x_k \in \{0,1\}^n$ such that for each $i \in [k]$, $f(L(x_i)) = \sigma_i$, and $M\vec{x} = \vec{0}$, where $\vec{x} = (x_1, \ldots, x_k)$. Let $\ell_1, \ldots, \ell_r$ be linear functionals corresponding to the r rows of M. By definition, for each $i \in [r]$, $\ell_i(x_1, \ldots, x_k) = 0$. Since L is a linear transformation and $L\vec{0} = \vec{0}$, for each $i \in [r]$, we have $\vec{0} = L(\ell_i(x_1, \ldots, x_k)) = \ell_i(L(x_1), \ldots, L(x_k))$, thus contradicting the assumption that f is (M,σ)-free.

□

We shall show that the property of being $(M,\vec{1})$-free is testable when M is of a special type of matrix as described below.

Definition 4.1.3. Let M be an r by k matrix over $\{0,1\}$. We say that M is *graphic* if there exists a graph on k edges, each edge associated with an integer from $\{1, \ldots, k\}$, such that any nontrivial linear combination of the rows of M corresponds to a cycle in the graph. More specifically, for each cycle in the graph, its indicator vector (on the edge set) lies in the span of the rows of M.

Remark. We assume without loss of generality that the matrix M has full rank. We can do this because if the nullspaces of M and another matrix M' are equal, then the set of functions that are M-free is the same set of functions that are M'-free. So we can always replace a matrix M by another matrix M' that has the same nullspace as of M and is full rank. This makes our definition of graphic matrix cleaner as we do not worry about degenerate cycles.

We consider the definition of graphic matrix as a natural and simple generalization of Green's definition of triangle-freeness. Recall that a function f is triangle-free if f has no $x, y \in \{0,1\}^n$ such that $f(x)$, $f(y)$, and $f(x+y)$ are all 1. Equivalently, triangle-freeness can be re-stated as $(M, 1^3)$-freeness, where the matrix M is simply the vector

$(1,1,1)$, represented by a triangle (or a cycle of length 3). In our exposition, when M is graphic, we may sometimes use the phrases M-free and G-free interchangeably where G is the underlying graph representing M.

Example 4.1.1. k–cycle freeness: The matrix M is the $1 \times k$ vector $(1, \ldots, 1)$, and the graph is a cycle on k edges, where each edge is labeled from 1 through k.

Example 4.1.2. K_4 freeness: The matrix

$$M = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

can be represented by K_4, the complete graph on 4 vertices. Consider a spanning tree of K_4 and label its edges as $\{1, 2, 3\}$. Label the edge that completes a cycle with edges 1 and 2 as edge 4, label the edge that completes a cycle with edges 1 and 3 as edge 5, and label the edge that completes a cycle with edges 2 and 3 as edge 6. It can be checked that every nontrivial linear combination of M corresponds to a cycle in K_4.

4.2 Green's regularity lemma

Now we describe Green's arithmetic regularity lemma, the crux of the analysis of our testing algorithm. Green's regularity lemma over $\{0,1\}^n$ is a structural theorem for boolean functions. It asserts that for every boolean function, there is some decomposition of the boolean cube into cosets, such that the function restricted to most of these cosets has small Fourier transform. An alternate and equivalent way is that no matter where we cut the boolean cube by a hyperplane, the densities of f on the two halves of the cube separated by the hyperplane does not differ greatly. Formally, we say that a function is *regular* if all of its nonzero Fourier coefficients are small.

Definition 4.2.1 (regularity)**.** For every $0 < \epsilon < 1$, we say that a function $f : \{0,1\}^n \to \{0,1\}$ is *ϵ-regular* if for every $\alpha \neq 0 \in \{0,1\}^n$, $\left|\widehat{f}(\alpha)\right| \leq \epsilon$.

Remark. Recall that by Proposition 2.2.3, if f is ϵ-regular, then $\|f - \mathbb{E} f\|_{U_2} \leq \sqrt{\epsilon}$.

We note that $W(t)$ indicates a tower of twos with height $\lceil t \rceil$. To obtain a partition of the boolean cube that satisfies the required regularity requirement, the number of cosets in the partition may be rather large. More precisely,

Lemma 4.2.1 (Green's regularity lemma)**.** *Let $f : \{0,1\}^n \to \{0,1\}$. For every $0 < \epsilon < 1$, there exists a subspace H of $G = \{0,1\}^n$ of co-dimension at most $W(\epsilon^{-3})$, such that $\Pr_{g \in G}[f_{g+H}$ is ϵ-regular$] \geq 1 - \epsilon$.*

The proof of Green's regularity lemma is not difficult, and we say a few words about how the lemma is proved. Suppose a boolean function f is not ϵ-regular. Then f must have an nonzero Fourier coefficient with magnitude at least ϵ, say $\left|\widehat{f}(\alpha)\right| \geq \epsilon$ for some $\alpha \neq 0$. Consider the subspace $H = \{x : \langle x, \alpha \rangle = 0\}$. Then the difference between the density of f on the subspace H and the density of f on $\{0,1\}^n \setminus H$ must be at least 2ϵ. In other words, one can verify that the average of the densities of f restricted to the two (affine) subspaces must be more than the density of f on $\{0,1\}^n$ by some constant depending on ϵ. Now repeat this argument on the two functions (f restricted to the two subspaces) and see if they are ϵ-regular. The number of iteration must be finite, since the average densities of a set of functions cannot be more than 1.

Example 4.2.1. By definition, the constant functions $f = 0$ and $f = 1$ are 0-regular.

Example 4.2.2. Consider $f(x) = \sum_{i \in S} x_i$ for some $S \subseteq [n]$. The function f is not ϵ–regular for $\epsilon < \frac{1}{2}$. Now take the subspace $H = \{x : \sum_{i \in S} x_i = 0\}$. $f_H = 0$ and $f_{\bar{H}} = 1$ are constant functions, and these two functions are 0-regular.

4.3 Our result

Our discussion of freeness of patterns suggests the following simple and natural test.

TEST T: On inputs M, an r by k binary matrix, and a vector $\sigma \in \{0,1\}^k$, with oracle access to f,

1. Pick $\vec{x} = (x_1, \ldots, x_k)$ uniformly at random among all vectors such that $M\vec{x} = \vec{0}$.

2. Reject iff $\langle f(x_1), \ldots, f(x_k) \rangle = \sigma$.

4.3.1 Non-linear properties

Consider the case when $\sigma = 1^k$. The property $\mathcal{YES}_{M,1^k}$ is easily seen to be non-linear. We show that when M is graphic, the property $\mathcal{YES}_{M,1^k}$ is testable.

Theorem 4.3.1 (Theorem 1.1.3 restated)**.** *Let $r \geq 1$, $k \geq 3$ be integers. Let M be an r by k graphic matrix over $\{0,1\}$ and $\epsilon > 0$. Then there exists a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\left(\mathcal{YES}_{M,1^k}, \mathcal{NO}_\epsilon\right)$ is testable with k queries, completeness 1, and soundness $1 - \tau(\epsilon)$.*

Proof Heuristic: We now specify the distribution $\mathcal{D}$, predicate ϕ, extension Φ, and the notion of pseudorandomness that we use to analyze T. The distribution $\mathcal{D}$ and the predicate ϕ should be clear from the test T itself. $\mathcal{D}$ is the uniform distribution on k-tuples $\vec{x} = (x_1, \ldots, x_k)$ such that $M\vec{x} = \vec{0}$. $\psi : \{0,1\}^k \to \{0,1\}$ is simply $\psi(x) = 1$ if $x \neq 1^k$ and 0 otherwise. A function f is considered pseudorandom if $\|f - \mathbb{E} f\|_{U_2} = o(1)$, and a collection of functions is pseudorandom if most of the functions in the collection are pseudorandom. The "extension" Φ is the decomposition guaranteed by Green's regularity lemma. In other words, $\Phi(f)$ sends f to a collection of functions $\{f_{g+H}\}_{g \in \{0,1\}^n}$.

If f is a random function with density δ, the test rejects with probability close to δ^k. As shown in Lemma 4.3.2, a graphic matrix gives rise to a system of linear equations of complexity 1. Thus if f is a pseudorandom function with density δ, then the test also rejects with probability close to δ^k. The hard part is to analyze the soundness of the test when $f \in \mathcal{NO}_\epsilon$. To this end, we apply Φ to f to obtain a collection of functions, and by Green's regularity lemma, $\Phi(f)$ is pseudorandom. Since $f \in \mathcal{NO}_\epsilon$, we show that 1) k of the functions in $\Phi(f)$ are pseudorandom and dense, and 2) the rejection probability of T on f can be bounded below (by a scaling

factor) by the rejection probability of T on these k functions. Condition 1 and 2 together imply that the test T rejects with positive probability when $f \in \mathcal{NO}_\epsilon$. We now begin with the formal proof.

Lemma 4.3.2. *Let M be an r by k graphic matrix over $\{0,1\}$. Let $f_1, \ldots, f_k : \{0,1\}^n \to [-1,1]$. Then*

$$\mathop{\mathbb{E}}_{\vec{x}:M\vec{x}=\vec{0}} \left[\prod_{i\in[k]} f_i(x_i) \right] \leq \min_{i\in[k]} \|f_i\|_{U_2} .$$

Proof. Since the set of vectors $\vec{x} = (x_1, \ldots, x_k)$ such that $M\vec{x} = \vec{0}$ forms a subspace of $\mathbb{F}_{2^n}^k$, for each $i \in [k]$, we can view x_i as a linear combination of x_j, where $j \in [k]$. In other words, there exists some underlying set of variables, and each x_i is a linear combination of these variables. We argue that the system of linear forms $(x_1, \ldots, x_k)$ has complexity at most 1. With M being graphic, consider its associated graph G on k edges, each labeled with an integer from $[k]$. Fix $i \in [k]$, and let $(u, v) \in E(G)$ be the edge labeled i. We partition the set of linear forms $\{x_j\}_{j\neq i}$ into two disjoint classes $(\mathcal{C}, \bar{\mathcal{C}})$ as follows. We let $x_j \in \mathcal{C}$ if the edge in G labeled j contains the vertex u, and $\bar{\mathcal{C}}$ is simply the complement of $\mathcal{C}$. The linear form x_i does not lie in the span of $\mathcal{C}$ since the set of edges incident to u cannot form a cycle with the edge (u, v). Similarly, x_i cannot lie in the span of $\bar{\mathcal{C}}$ as well. The lemma follows now by applying Proposition 2.2.4.

□

We can now finish the proof of Theorem 1.1.3.

Proof of Theorem 1.1.3. Clearly the test T makes k queries and always accepts whenever f is $(M, 1^k)$-free. Suppose f is ϵ-far from being $(M, 1^k)$-free, and we want to show that the test rejects with probability at least $\tau(\epsilon)$ for some function τ. Let $a(\epsilon)$ and $b(\epsilon)$ be two functions of ϵ that satisfy the constraint $a(\epsilon) + b(\epsilon) < \epsilon$, and along with τ, we shall specify these two functions at the end of the proof. Let G denote $\{0,1\}^n$. We now apply Lemma 4.2.1 to f to obtain a subspace H of G of co-dimension at

most $W(a(\epsilon)^{-3})$. Consequently, f restricted to all but at most $a(\epsilon)$ fraction of the cosets of H are $a(\epsilon)$-regular. We define a reduced function $f^R : \{0,1\}^n \to \{0,1\}$ as follows.

For each $g \in G$, if f restricted to the coset $g + H$ is $a(\epsilon)$-regular, then define

$$f^R_{g+H}(x) = \begin{cases} 0 & \text{if } \mathbb{E}\, f_{g+H} \leq b(\epsilon) \\ f_{g+H} & \text{otherwise.} \end{cases}$$

Else, define $f^R_{g+H} = 0$.

Note that at most $a(\epsilon)+b(\epsilon)$ fraction of modification has been made to f to obtain f^R. Since f is ϵ-far from being $(M, 1^k)$-free, there exists $\vec{x} = (x_1, \ldots, x_k)$ such that $M\vec{x} = 0$ and $f^R(x_i) = 1$ for every $i \in [k]$. Now consider the k cosets $x_i + H$. By our construction of f^R, we know that f restricted to each coset $x_i + H$ is $a(\epsilon)$-regular and at least $b(\epsilon)$-dense. We will count the number of "M-patterns" across these k cosets.

Notice that the probability the test rejects is at least

$$2^{-k \cdot W(a(\epsilon)^{-3})} \cdot \Pr_{\vec{z}: M\vec{z} = \vec{0}, \vec{z} \in H^k} \left[\forall i, f_{x_i+H}(z_i) = 1\right],$$

where $\vec{z} \in H^k$ indicates that $\vec{z} = (z_1, \ldots, z_k)$ and $z_i \in H$ for each $i \in [k]$. To bound this rejection probability from below, it suffices to show that the probability

$$\Pr_{\vec{z}: M\vec{z} = \vec{0}, \vec{z} \in H^k} \left[\forall i, f_{x_i+H}(z_i) = 1\right] \tag{4.1}$$

is bounded below by some constant depending only on ϵ. To this end, write $f_i = f_{x_i+H}$. We rewrite Equation 4.1 as

$$\mathop{\mathbb{E}}_{\vec{z}: M\vec{z} = \vec{0}, \vec{z} \in H^k} \left[\prod_{i \in [k]} f_i(z_i)\right]. \tag{4.2}$$

By replacing each function f_i by $\mathbb{E}\, f_i + (f_i - \mathbb{E}\, f_i)$, it is easy to see that Equation 4.2 can be expanded into a sum of 2^k terms, one of which is $\prod_{i \in [k]} \mathbb{E}\, f_i$, which is at least

$b(\epsilon)^k$. For the other $2^k - 1$ terms, since H is isomorphic to a boolean cube, we can applying Proposition 2.2.4 and Proposition 2.2.3 to show that each of these terms can be bounded above by $a(\epsilon)^{1/2}$. So Equation 4.2 is at least $b(\epsilon)^k - (2^k - 1)a(\epsilon)^{1/2}$. To finish the analysis, we need to specify $a(\epsilon), b(\epsilon)$ such that $b(\epsilon)^k - (2^k - 1)a(\epsilon)^{1/2} > 0$ and $a(\epsilon) + b(\epsilon) < \epsilon$. Both are satisfied by setting $a(\epsilon) = (\frac{\epsilon}{4})^{2k}$, $b(\epsilon) = \frac{\epsilon}{2}$. Thus, the rejection probability is at least $\tau(\epsilon) \geq 2^{-k(W((4/\epsilon)^{6k})+2)} \cdot \epsilon^k$, completing the proof.

□

4.3.2 Linear properties

In this section, we show that the analysis of forbidden patterns can be used to derive alternate proofs for linearity testing [8, 6] and affinity testing [33]. To do so, we consider "non-monotone" properties, the case when the pattern vector σ is not 1^k or 0^k. The analysis is more difficult, and we employ a different "rounding" scheme inspired by the testability of non-monotone graph properties in [2]. Unlike Szemerédi's regularity lemma, a "strong form" of Green's regularity lemma is not known, so we restrict our attention to the case when M is specified by a cycle and exploit the additive structure of the pattern. Specifically, the matrix M is of the form $(1, \ldots, 1)$, i.e., $r = 1$ and all entries of M are ones.

Theorem 4.3.3. *Let $r \geq 1$, $k \geq 3$ be integers. Let M be the all ones k-dimensional vector, and suppose $\sigma \in \{0,1\}^k$ and $\sigma \neq 1^k, 0^k$. Then there exists some function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ such that the property $(\mathcal{YES}_{M,\sigma}, \mathcal{NO}_\epsilon)$ is testable with k queries, completeness 1, and soundness $1 - \tau(\epsilon)$.*

Proof Heuristic: The notion of pseudorandomness and the extension Φ are the same as in the discussion preceding the proof of Theorem 1.1.3. In the "rounding" scheme to obtain f^R in the proof of Theorem 1.1.3, if f_{g+H} is not regular, f^R_{g+H} is simply set to 0. Since f^R is close to f, if f is far from being $(M, \vec{1})$-free, then f^R must have a $(M, \vec{1})$ pattern. Thus, for some $x_1, \ldots, x_k$, the k functions f_{x_i+H} are regular. (We only need one regular function for the analysis to go through.)

When the pattern vector σ is not all ones, we can no longer set $f^R_{g+H} = 0$ when f_{g+H} is not regular since f_{g+H} might not be dense in the symbol 0. To remedy this, we change the rounding scheme so that if none of the k functions f_{x_i+H} are regular, the densities of f_{x_i+H} will ensure that the pattern vector σ appears frequently. We now proceed with a formal proof.

Proof of Theorem 4.3.3. The test T makes k queries and is easily seen to have completeness 1. We proceed to analyze the soundness of the test. Suppose f is ϵ-far from being (M, σ)-free. We need to show that the test rejects with probability at least $\tau(\epsilon)$ such that $\tau(\epsilon) > 0$ whenever $\epsilon > 0$.

To this end, let $\frac{1}{2} < \eta < 1$ be any constant, and $a(\epsilon)$ and $b(\epsilon)$ be functions of epsilons that satisfy the constraints $a(\epsilon) + b(\epsilon) < \epsilon$ and $1 - \eta > b(\epsilon)$. We shall specify these two functions at the end of the proof.

Now let G denote $\{0,1\}^n$. We apply Lemma 4.2.1 to f to obtain a subspace H of G of co-dimension at most $W(a(\epsilon)^{-3})$. We define a reduced function $f^R : \{0,1\}^n \to \{0,1\}$ as follows. We assume that σ has at least two occurrences of 1. (Otherwise it has at least two occurrences of 0, and in the construction of f^R, we flip the roles of 1 and 0 when f_{g+H} is non-regular. The rest of the proof will proceed analogously, and we leave its verification to the readers.)

For each $g \in G$, if f restricted to the coset $g + H$ is $a(\epsilon)$-regular, then define

$$f^R_{g+H} = \begin{cases} 0 & \text{if } \mathbb{E}\, f_{g+H} < b(\epsilon) \\ 1 & \text{if } \mathbb{E}\, f_{g+H} > 1 - b(\epsilon) \\ f_{g+H} & \text{otherwise.} \end{cases}$$

Else, define

$$f^R_{g+H} = \begin{cases} 1 & \text{if } \mathbb{E}\, f_{g+H} \geq \eta \\ 0 & \text{otherwise.} \end{cases}$$

Note that at most $a(\epsilon) + b(\epsilon)$ fraction of modification has been made to f to obtain f^R, so f^R is ϵ-close to f. Since f is ϵ-far from being (M, σ)-free, there exist

$x_1, \ldots, x_k \in \{0,1\}^n$ such that $\sum_{i\in[k]} x_i = \vec{0}$ and for each $i \in [k]$, $f^R(x_i) = \sigma_i$. Consider the cosets $x_i + H$. By our choice of rounding, f restricted to each coset $x_i + H$ is dense in the symbol σ_i. In other words, define $\mu_b(f)$ to be $\Pr_x[f(x) = b]$. Then for each $i \in [k]$, $\mu_{\sigma_i}(f_{x_i+H}) \geq b(\epsilon)$, since $1 - \eta \geq b(\epsilon)$. We want to show that there are many "cyclic" σ patterns across these k cosets. It is easy to see that the test rejects with probability at least

$$2^{-(k-1)W(a(\epsilon)^{-3})} \cdot \Pr_{z_1,\ldots,z_k \in H; \sum_i z_i = 0} [\forall i \in [k], f_{x_i+H}(z_i) = \sigma_i] .$$

To lower bound this rejection probability, it suffices to show that the probability

$$\Pr_{z_1,\ldots,z_k \in H; \sum_i z_i = 0} [\forall i \in [k], f_{x_i+H}(z_i) = \sigma_i] \tag{4.3}$$

can be bounded below by some constant depending only on ϵ. To this end, we divide our analysis into two cases, based on whether there is some $j \in [k]$ such that f_{x_j+H} is $a(\epsilon)$-regular or not.

Case 1: There exists some $j \in [k]$ such that f_{x_j+H} is $a(\epsilon)$-regular.

For each $i \in [k]$, define $f_i : H \to \{0,1\}$ to be $f_i = f_{x_{i+H}} + \sigma_i + 1$. Note that by definition, $\mathbb{E}\, f_i \geq b(\epsilon)$. We begin by arithmetizing Equation 4.3 as

$$\mathop{\mathbb{E}}_{z_1,\ldots,z_k \in H : \sum_i z_i = 0} \left[\prod_{i\in[k]} f_i(z_i) \right] .$$

By Fourier expansion, it is not hard to see that

$$\mathop{\mathbb{E}}_{z_1,\ldots,z_k \in H : \sum_i z_i = 0} \left[\prod_{i\in[k]} f_i(z_i) \right] = \sum_{\alpha \in H} \prod_{i\in[k]} \widehat{f_i}(\alpha).$$

Using the facts that $\mathbb{E}\, f_i \geq b(\epsilon)$, f_j is $a(\epsilon)$-regular, there exist two distinct indices $i_1, i_2 \neq j \in [k]$ (since $k \geq 3$), Cauchy-Schwarz, and Parseval's Identity, respectively, we have

$$\begin{aligned}
\sum_{\alpha\in H}\prod_{i\in[k]}\widehat{f_i}(\alpha) &\geq b(\epsilon)^k - \sum_{\alpha\neq 0\in H}\prod_{i\in[k]}\left|\widehat{f_i}(\alpha)\right| \\
&\geq b(\epsilon)^k - a(\epsilon)\sum_{\alpha\neq 0\in H}\prod_{i\in[k]\setminus\{j\}}\left|\widehat{f_i}(\alpha)\right| \\
&\geq b(\epsilon)^k - a(\epsilon)\sum_{\alpha\neq 0\in H}\left|\widehat{f_{i_1}}(\alpha)\right|\left|\widehat{f_{i_2}}(\alpha)\right|, \\
&\geq b(\epsilon)^k - a(\epsilon)\left(\sum_{\alpha\neq 0\in H}\widehat{f_{i_1}}(\alpha)^2\right)^{1/2}\left(\sum_{\alpha\neq 0\in H}\widehat{f_{i_2}}(\alpha)^2\right)^{1/2} \\
&\geq b(\epsilon)^k - a(\epsilon).
\end{aligned}$$

To finish the analysis, we need to specify $a(\epsilon), b(\epsilon)$ such that the constraints $a(\epsilon) + b(\epsilon) < \epsilon$ and $1-\eta > b(\epsilon)$ are satisfied. To this end, we set $b(\epsilon) = (1-\eta)\cdot\epsilon$ and $a(\epsilon) = \frac{1}{2}(1-\eta)^k\epsilon^k$, and the rejection probability is at least $\tau(\epsilon) \geq 2^{-(k-1)W(a(\epsilon)^{-3})}\cdot(\epsilon-\eta\epsilon)^k/2$.

Case 2: No $j \in [k]$ exists such that f_{x_j+H} is $a(\epsilon)$-regular.

Since σ contains at least two ones, we may assume without loss of generality that $\sigma_{k-1}, \sigma_k = 1$. For each $i \in [k-2]$, since f_{x_i+H} is dense in σ_i, we may fix some $z_i \in H$ such that $f_{x_i+H}(z_i) = \sigma_i$. Now set $Z = \sum_{i=1}^{k-2} z_i$. Since $\eta > \frac{1}{2}$, by union bound we have

$$\begin{aligned}
\Pr_{z\in H}[f_{x_{k-1}+H}(z), f_{x_k+H}(Z+z) = 1] &= 1 - \Pr_{z\in H}[f_{x_{k-1}+H}(z) = 0 \text{ or } f_{x_k+H}(Z+z) = 0] \\
&\geq 1 - 2(1-\eta) \\
&> 0.
\end{aligned}$$

Since for each $i \in [k]$, f_{x_i+H} is not $a(\epsilon)$-regular, by our choice of rounding, $\Pr_{z\in H}[f_{x_i+H}(z) = \sigma_i]$ is at least $1-\eta$. By picking $z_1, \ldots, z_{k-2}$ uniformly at random from H, it is not hard to see that

$$\Pr_{z_1,\ldots,z_k\in H;\sum_i z_i=0}[\forall i \in [k], f_{x_i+H}(z_i) = \sigma_i] \geq (1-\eta)^{k-2}(2\eta-1).$$

Thus the rejection probability of the test is at least

$$\tau(\epsilon) \geq 2^{-(k-1)W(a(\epsilon)^{-3})} \cdot (1-\eta)^{k-2}(2\eta - 1),$$

where $a(\epsilon) = \frac{1}{2}(1-\eta)^k \epsilon^k$, completing the proof.

□

We now use Theorem 4.3.3 to derive alternate proofs that the properties of affinity and linearity are testable. We say that f is an affine subspace function if $f^{-1}(1)$ is some affine subspace of $\{0,1\}^n$. (We assume that the zero function $f = 0$ is also an affine subspace function.) We say that a boolean function $f : \{0,1\}^n \to \{0,1\}$ is linear if $f(x) = \sum_{i \in S} x_i$ for some $S \subseteq [n]$. (We also consider the constant functions $f = 0$ and $f = 1$ as linear.)

Corollary 4.3.4. *Let $\epsilon > 0$. There exists a test that determines if a function is an affine subspace function or ϵ-far from being one, and the test makes 4 queries, has completeness 1, and soundness $1 - \tau(\epsilon)$, for some function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$.*

Proof. Recall the following well-known characterization of affine subspace: A is an affine subspace if and only if for any triple $x, y, z \in A$, $x + y + z \in A$. So a function f is $(M, 1110)$-free if and only if $f^{-1}(1)$ is an affine subspace. Now apply Theorem 4.3.3 by setting $M = (1,1,1,1)$ (a cycle of length of 4) and $\sigma = (1,1,1,0)$, and thus, the property of being an affine subspace function is testable.

□

Corollary 4.3.5. *Let $\epsilon > 0$. There exists a test that determines if a function is linear or ϵ-far from being linear, and the test makes 5 queries, has completeness 1, and soundness $1 - \tau(\epsilon)$, for some function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$.*

Proof. It suffices to show that a function is linear if and only if it is $(M, 11100)$-free, where $M = (1,1,1,1,1)$, corresponding to the cycle of length 5. The forward direction is trivial. Suppose a function f is linear. For any $x_1, \ldots, x_4 \in \{0,1\}^n$, $\sum_i f(x_i) = f\left(\sum_i x_i\right)$. Thus $(f(x_1), \ldots, f(x_4), f(\sum_i x_i))$ cannot equal $(1,1,1,0,0)$.

For the other containment, suppose f is $(M, 11100)$-free. Let $S = f^{-1}(0)$. If $S = \emptyset$ or $S = \{0,1\}^n$, then f is a constant function, and the proof is finished. So suppose there exist $x, y, z \in \{0,1\}^n$ (x, y not necessarily distinct) such that $f(x), f(y) = 0$ and $f(z) = 1$. Then $f(x+y) = 0$, otherwise $(f(z), f(z), f(x+y), f(x), f(y))$ forms a $(1,1,1,0,0)$ pattern. Thus S is a linear subspace of $\{0,1\}^n$. Suppose the dimension of S is k with $k \geq 1$. Then there are k linearly independent vectors $a_1, \ldots, a_k \in \{0,1\}^n$ such that $z \in S$ iff $(\langle a_1, z\rangle = 0) \bigwedge \cdots \bigwedge (\langle a_k, z\rangle = 0)$. Therefore, by De Morgan's law, $f(z) = 1$ iff $z \in \bar{S}$ iff $(\langle a_1, z\rangle = 1) \bigvee \cdots \bigvee (\langle a_k, z\rangle = 1)$. If $k > 1$, since a_1 and a_2 are linearly independent, there exist $x, y \in \{0,1\}^n$ such that $\langle a_1, x\rangle = \langle a_2, y\rangle = 1$ while $\langle a_1, y\rangle = \langle a_2, x\rangle = 0$. However, this implies that $(f(x), f(y), f(x+y), f(0), f(0))$ forms a $(1,1,1,0,0)$ pattern, a contradiction. Thus, $k = 1$, and $f(z) = \langle a_1, z\rangle$ is a linear function.

□

Chapter 5

Open problems

We list several possible directions from the work in this thesis.

5.1 Dictatorship testing

One may view Theorem 1.1.1 as a self-contained result on property testing. However, the ultimate goal is to use Theorem 1.1.1 to construct a PCP system with perfect completeness and amortized query complexity $1 + O\left(\frac{\log q}{q}\right)$. At present, our test does not "lift" into a new PCP construction for a number of reasons. The first of which is that a dictatorship test without "consistency checks" is most easily composed with Khot's unique label cover [26] (as opposed to the standard label cover [34]) as the outer verifier in a PCP reduction. As the conjectured NP-hardness of the unique label cover cannot have perfect completeness, the obvious approach in combining our test with the unique games-based outer verifier does not imply a new PCP result. However, there are variants of the unique label cover (e.g., Khot's d to 1 Conjecture) [26] that do have conjectured perfect completeness, and these variants are used to derive hardness of coloring problems in [12]. In addition, O'Donnell and Wu [31] have recently designed an optimal three bit dictatorship test with perfect completeness, and in a followup work, the same authors [32] have constructed a conditional PCP system using Khot's d to 1 Conjecture. With these recent works, we believe that using similar ideas one may establish some conditional evidence toward the following

as well:

Conjecture 5.1.1. *For every $q \geq 3$, there exists a* PCP *system that makes q queries, has completeness* 1*, and soundness $\frac{\text{poly}(q)}{2^q}$, and in particular has amortized query complexity $1 + O\left(\frac{\log q}{q}\right)$.*

It will also be interesting to remove the adaptivity in our dictatorship test since the well-known correspondence between PCP constructions and hardness of approximation needs a fully nonadaptive test. Nevertheless, the hardness of satisfying of a constraint satisfaction problem (CSP) may differ depending on whether the perfect completeness condition is imposed or not. For instance, when the constraints are linear equations, one may solve the system of linear equations by Gaussian elimination if a solution exists (see e.g. [22]). Whether or not Conjecture 5.1.1 implies a conditional hardness result for some CSP is another possible direction to explore.

5.2 Green's Conjecture

Theorem 1.1.3 has now been subsumed by Král', Serra, and Vena [29] and Shapira [40] as they have independently proved Green's Conjecture (Conjecture 1.1.2). However, there are still several avenues to investigate. The first of which is to understand the behavior of the soundness parameter $\tau(\epsilon)$ in Theorem 1.1.3. When $f \in \mathcal{NO}_\epsilon$, we can only guarantee that the tester rejects with probability at least $\tau(\epsilon)$, where $\tau(\epsilon) = W\left(\text{poly}(1/\epsilon)\right)^{-1}$. It is possible that the rejection probability is higher, but our proof only says that this is bounded away from 0. In fact, in both [29] and [40] where Green's Conjecture is proved, the bounds also depend on towers of $\frac{1}{\epsilon}$. In fact their bounds are even worse than ours. As we mentioned, the tower dependency is inherent in the regularity lemma of Szemerédi and Green; proving that the tower dependency is necessary for property testing seems very difficult. Still, it is tempting to conjecture that in Green's triangle-free test, $\tau(\epsilon)$ cannot be some polynomial in ϵ.

Conjecture 5.2.1 (Green [17])**.** *Let $\epsilon > 0$. For every n, there exists a function $f : \{0,1\}^n \to \{0,1\}$ such that*

- *f is ϵ-far from being triangle-free, and*
- $\Pr_{x,y\in\{0,1\}^n}[f(x), f(y), f(x+y) = 1] = o\left(\epsilon^d\right)$ *for every positive integer d.*

We remark that there is some evidence toward this conjecture. Alon [1] has constructed a *graph* that is ϵ-far from being triangle-free but has only $o\left(\epsilon^d\right)$ triangles, for any positive integer d. Using similar ideas, Green [17] remarks that for functions over $\mathbb{Z}_N$ for large N, Conjecture 5.2.1 holds.

Another direction is as follows. There is a close connection between the d-th dimension Gowers uniformity norm of a function and quasirandomness of a d-uniform hypergraph. The fact that Theorem 1.1.3 is also proved independently by Král', Serra, and Vena [28] is not so surprising. For each matrix M that has complexity 1, we apply the U_2 Gowers norm to count the number of copies of M in a function. Král', Serra, and Vena identify a graph H with M and give a reduction from a function f to a graph G so that the number of copies of M in f is equal (up to scaling) to the number of copies of the induced subgraph H in G. Then they use Szmerédi's regularity lemma to count the number of copies of H. In [29, 40], this method is extended to hypergraph, and higher degrees of quasirandomness are used to analyze an arbitrary matrix M. An alternate proof of Green's Conjecture using the Gowers norm to count the number of M without going through a reduction to hypergraph should also be of interest.

5.3 Gowers Inverse Conjecture

The inverse Gowers conjecture states that if a bounded function has positive d-th dimension Gowers norm, then it correlates with a degree $d-1$ polynomial. Over finite fields with low characteristic, the conjecture was refuted independently by Green and Tao [18] and Lovett, Meshulam, and Samorodnitsky [30]. Specifically, over $\{0,1\}^n$, the function $(-1)^{S_4}$ (where S_4 is the symmetric function of degree 4), has positive 4-th dimension Gowers norm yet does not correlate with any cubic polynomial.

Recently, Bergelson, Tao, Ziegler [9] have established a form of the inverse Gowers conjecture over low characteristic. However, over characteristic two, their work does

not imply the distance property in the usual sense in property testing. So some form of inverse for the Gowers norm is still lacking.

Formulating the Gowers Inverse Conjecture in the property testing language, the conjecture states that the property of being a degree d polynomial is testable with 2^d queries, completeness 1, and soundness $\frac{1}{2}$, where the set of $\mathcal{YES}$ instances consists of degree d polynomials, and the set of $\mathcal{NO}$ instances consists of functions that are at least $\big(\frac{1}{2} - o(1)\big)$-far from $\mathcal{YES}$ instances. The counterexample $(-1)^{S_4}$ is $\big(\frac{1}{2} - o(1)\big)$-far from cubic polynomials yet passes the cubic test with high probability (the test simply estimates the U_4 norm).

To circumvent the counterexample, one may relax the property testing criterion by showing that a restricted subset of low-degree polynomials is testable with soundness $\frac{1}{2}$. In particular, it seems reasonable to conjecture the following.

Conjecture 5.3.1 (Samorodnitsky [36])**.** *Let $d \in \mathbb{Z}^+$. $\mathcal{YES}_d$ denotes "some non-trivial" subset of polynomials degree at most d, and $\mathcal{NO}$ denotes the set of boolean functions that are at least $\big(\frac{1}{2} - o(1)\big)$-far from any functions in $\mathcal{YES}_d$. There exists a constant $0 \leq \delta < \frac{1}{2}$ such that the property $(\mathcal{YES}_d, \mathcal{NO})$ is testable with 2^d queries, completeness $1 - \delta$, and soundness $\frac{1}{2}$.*

Bibliography

[1] Noga Alon. Testing subgraphs in large graphs. *Random Struct. Algorithms*, 21(3-4):359–370, 2002.

[2] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 251–260, New York, NY, USA, 2006. ACM.

[3] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing low-degree polynomials over GF(2). In *Proceedings of Random 2003*, pages 188–199, 2003.

[4] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[5] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *J. ACM*, 45(1):70–122, 1998.

[6] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos A. Kiwi, and Madhu Sudan. Linearity testing in characteristic two. In *IEEE Symposium on Foundations of Computer Science*, pages 432–441, 1995.

[7] Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, PCPs, and nonapproximability—towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998.

[8] Vitaly Bergelson, Bernard Host, and Bryna Kra. Multiple recurrence and nilsequences. *Inventiones Mathematicae*, 160(2):261–303, 2005.

[9] Vitaly Bergelson, Terence Tao, and Tamar Ziegler. An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}^\omega$. *arXiv/0901.2602*, 2009.

[10] Arnab Bhattacharyya, Victor Chen, Madhu Sudan, and Ning Xie. Testing linear-invariant non-linear properties. In Susanne Albers and Jean-Yves Marion, editors, *26th International Symposium on Theoretical Aspects of Computer Science (STACS 2009)*, Dagstuhl, Germany, 2009. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

[11] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993.

[12] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 344–353, New York, NY, USA, 2006. ACM.

[13] Irit Dinur and Shmuel Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162(1):439–485, 2005.

[14] Lars Engebretsen and Jonas Holmerin. More efficient queries in PCPs for NP and improved approximation hardness of maximum CSP. In *STACS*, pages 194–205, 2005.

[15] Timothy Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[16] Timothy Gowers and Julia Wolf. The true complexity of a system of linear equations. *Proceedings of the London Mathematical Society*, To appear.

[17] Ben Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376, 2005.

[18] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *arXiv/0711.3191*, 2007.

[19] Ben Green and Terence Tao. An inverse theorem for the Gowers U_3 norm. *Proceedings of the Edinburgh Mathematical Society*, To appear.

[20] Ben Green and Terence Tao. Linear equations in primes. *Annals of Mathematics*, To appear.

[21] Venkatesan Guruswami, Daniel Lewin, Madhu Sudan, and Luca Trevisan. A tight characterization of NP with 3 query PCPs. In *FOCS '98: Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, page 8, Washington, DC, USA, 1998. IEEE Computer Society.

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

[23] Johan Håstad and Subhash Khot. Query efficient PCPs with perfect completeness. *Theory of Computing*, 1(7):119–148, 2005.

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

[25] Tali Kaufman and Madhu Sudan. Algebraic property testing: The role of invariance. In *STOC '08: Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 403–412, 2008.

[26] Subhash Khot. On the power of unique 2-prover 1-round games. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, New York, NY, USA, 2002. ACM.

[27] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable CSPs?. *SIAM J. Comput.*, 37(1):319–357, 2007.

[28] Daniel Král', Oriol Serra, and Lluís Vena. A combinatorial proof of the removal lemma for groups. *Preprint*, to appear.

[29] Daniel Král', Oriol Serra, and Lluís Vena. A removal lemma for systems of linear equations over finite fields. *Preprint*, to appear.

[30] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. In *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 547–556, New York, NY, USA, 2008. ACM.

[31] Ryan O'Donnell and Yi Wu. 3-bit dictator testing: 1 vs. 5/8. In *SODA '09: Proceedings of the Nineteenth Annual ACM -SIAM Symposium on Discrete Algorithms*, pages 365–373, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.

[32] Ryan O'Donnell and Yi Wu. Conditional hardness for satisfiable-3CSPs. In *STOC '09: Proceedings of the forty-first annual ACM symposium on Theory of computing*, page To appear, New York, NY, USA, 2009. ACM.

[33] Michal Parnas, Dana Ron, and Alex Samorodnitsky. Testing basic boolean formulae. *SIAM Journal on Discrete Mathematics*, 16(1):20–46, 2002.

[34] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[35] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[36] Alex Samorodnitsky. Personal communication.

[37] Alex Samorodnitsky. Low-degree tests at large distances. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 506–515, New York, NY, USA, 2007. ACM.

[38] Alex Samorodnitsky and Luca Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 191–199, New York, NY, USA, 2000. ACM.

[39] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 11–20, New York, NY, USA, 2006. ACM.

[40] Asaf Shapira. A proof of Green's conjecture regarding the removal properties of sets of linear equations. In *STOC '09: Proceedings of the forty-first annual ACM symposium on Theory of computing*, page To appear, New York, NY, USA, 2009. ACM.

[41] Madhu Sudan and Luca Trevisan. Probabilistically checkable proofs with low amortized query complexity. In *FOCS '98: Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, page 18, Washington, DC, USA, 1998. IEEE Computer Society.

[42] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

[43] Terence Tao. Structure and randomness in combinatorics. In *FOCS '07: Proceedings of the forty-eighth annual ACM symposium on Foundations of computer science*, pages 3–15, New York, NY, USA, 2007. ACM.

[44] Luca Trevisan. Recycling queries in PCPs and in linearity tests (extended abstract). In *STOC '98: Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 299–308, New York, NY, USA, 1998. ACM.