Testability of Linear-Invariant Properties

by

Arnab Bhattacharyya

Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2011

Author ....

Department of Electrical Engineering and Computer Science

September 1, 2011

Certified by ....

Ronitt Rubinfeld

Professor of Electrical Engineering and Computer Science

Thesis Supervisor

Accepted by ....

Professor Leslie A. Kolodziejski

Chairman, Department Committee on Graduate Students

Testability of Linear-Invariant Properties

by

Arnab Bhattacharyya

Submitted to the Department of Electrical Engineering and Computer Science
on September 1, 2011, in partial fulfillment of the
requirements for the degree of
Doctor Of Philosophy

Abstract

Property Testing is the study of super-efficient algorithms that solve "approximate decision problems" with high probability. More precisely, given a property $\mathcal{P}$, a testing algorithm for $\mathcal{P}$ is a randomized algorithm that makes a small number of queries into its input and distinguishes between whether the input satisfies $\mathcal{P}$ or whether the input is "far" from satisfying $\mathcal{P}$, where "farness" of an object from $\mathcal{P}$ is measured by the minimum fraction of places in its representation that needs to be modified in order for it to satisfy $\mathcal{P}$. Property testing and ideas arising from it have had significant impact on complexity theory, pseudorandomness, coding theory, computational learning theory, and extremal combinatorics.

In the history of the area, a particularly important role has been played by linear-invariant properties, i.e., properties of Boolean functions on the hypercube which are closed under linear transformations of the domain. Examples of such properties include linearity, homogeneousness, Reed-Muller codes, and Fourier sparsity. In this thesis, we describe a framework that can lead to a unified analysis of the testability of all linear-invariant properties, drawing on techniques from additive combinatorics and from graph theory. We also show the first nontrivial lowerbound for the query complexity of a natural testable linear-invariant property.

Thesis Supervisor: Ronitt Rubinfeld
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

Thanks to the members of my thesis committee: Madhu Sudan, Piotr Indyk, and Ronitt Rubinfeld. I first felt the magic of theoretical computer science when taking Madhu's class on complexity theory. In fact by the end of that class, it seemed inevitable to me that theory and I would have a future together. Madhu is also the reason that I started the work contained in this thesis. Piotr captures what I love most about people in this field: intelligent in an uncontained way, unassuming, easy-going, trusting, free, funny, and nocturnal. To my fabulous advisor, Ronitt, I owe thanks not only for the many hours of technical and non-technical discussions but also for the effortless way in which she inspired and cheered me up after every one of those meetings.

Thanks to my friends for sharing the journey with me so far. More specifically, thanks to Jelani for his unbelievably ridiculous stories, to Krzysztof for the late-night table tennis matches, to David for showing me the importance of persistence in research, to Shubhangi for the marathon `zephyr` sessions in our undergrad days, to Ankur for his enthusiasm about anything related to theory, to Olek for keeping me from veering too far off EST, to Elena for her quick smile and her great collaboration on several papers, to Sofya for inviting me to my first real research project, to Alex for secretly being an inspiration, to Sesh for repeatedly demonstrating perfect talks, to Arvind for sharing the pain and joy in cooking for 150 people, to Nasser for symmetry, to Manohar for the surprise invitation, to Pristene for starting storms, ... I cannot possibly complete this list. Also, a special acknowledgment to Tarun, Punya, Ryan, Oleg, Fernando, and Armando; we have gone our separate ways since 2005 but our Ω dream is still quite alive and well.

Much of my research was sparked by visits. I'm indebted to Eldar Fischer for the very enjoyable trips to Haifa, to Alex Samorodnitsky for inviting me to Hebrew University and for the many other stimulating discussions, to Nikhil Bansal for the short but productive trip to IBM Watson, and to Ravi Kannan and Neeraj Kayal for the great stay at MSR India. Thanks to all my coauthors, and most specially, to Asaf and Elena for the most enjoyable collaboration I've been part of.

The first four years of my graduate school were funded by the Computational Science Graduate Fellowship from the Department of Energy and the Krell Institute. The support was indispensable.

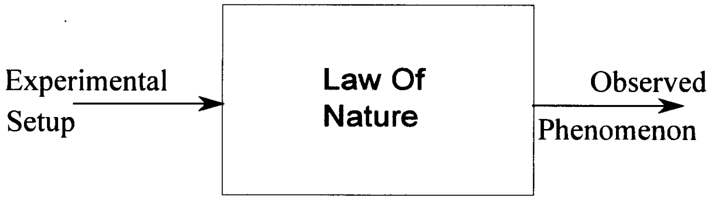
Lastly and most importantly, I want to thank Mamma, Boo and Papai. Simply put, I would be nowhere near here if it weren't for their unconditional love, patience and support. My parents have always been my role models, and my brother has always been my best friend. This thesis is dedicated with love to them.

Chapter 1

Introduction

1.1 An Invitation to the Testability Question

Science is the systematic attempt to discover the laws of nature. One can schematically depict a scientific experiment as follows:



Once the experiment is run a few times, or in other words, once it becomes known what output phenomena are produced by a few different experimental setups, the scientist tries to use the data to discover some property of the law of nature under consideration. For instance, Galileo famously discovered that gravitational acceleration is independent of mass by dropping two different weights from the Leaning Tower of Pisa.

A fundamental problem of science then is understanding how to design good experiments. Usually, a scientist has a hypothesis in mind (for instance, that a law of motion satisfies conservation of energy), and she wants to test if the hypothesis is satisfied by nature. What is the most "efficient" way to test the hypothesis? Unless the setting is trivial, it is infeasible to run the experiment for all possible experimental setups. The scientist must then cleverly choose particular setups so as to yield the most information about her hypothesis. Sometimes, she is helped by prior information already acquired about nature. Sometimes, she is helped by the fact that any violations to the hypothesis are easy to detect. Sometimes, she is helped by her willingness to admit a small possibility of error in her findings.

In theoretical computer science, we study these problems in a rigorous and abstract fashion. The formal setting is as follows:

$$x \in \mathcal{D} \longrightarrow \boxed{f : \mathcal{D} \to R} \longrightarrow f(x) \in R$$

The object at the center of attention is a function f mapping a finite domain $\mathcal{D}$ to a finite range R. In the above discussion, $\mathcal{D}$ is the set of all possible experimental setups and R the set of all possible observed phenomena. The function f is not known exactly, although there may be some prior information available. Additional information about f can only be obtained by *querying*, i.e., observing the value of f on elements of $\mathcal{D}$. The goal is to determine if f satisfies some particular property $\mathcal{P}$ or not, by making as few queries as possible.

Note that a more ambitious goal might be to *learn* the function f itself, instead of merely deciding whether it satisfies a particular property. This task is formally studied in the field of computational learning theory. In learning theory, one has a priori knowledge of a nontrivial property satisfied by f and then one wants to determine the function. The stronger the property known to be satisfied by f, the easier it usually is to learn f. In this thesis, our concern is with obtaining the prior knowledge, that is, determining whether f satisfies a given property. For instance, given an unknown natural phenomenon, the scientist might first want to check whether the system produces net positive entropy before treating it as a closed system and finding the laws governing it. Or, an economist might want to determine whether stock market values increase with consumer confidence before making a detailed economic model.

We define the *query complexity* of a property $\mathcal{P}$ to be the minimal number of queries needed by an algorithm to determine whether a given f satisfies $\mathcal{P}$ or not. The query complexity is defined with respect to: (1) the computational model for the algorithm that chooses the values in $\mathcal{D}$ to query and that makes the final decision, and (2) any prior conditions that the unknown f is known to satisfy. We consider three settings below, each more restrictive than the previous. We will see that the first two require high query complexity for interesting properties, whereas the third setting does not and, at the same time, is sufficient for many purposes.

- **Deterministic Query Complexity.** The deterministic query complexity of a property $\mathcal{P}$ is the minimum number of queries that a deterministic algorithm has to make to be able to determine whether a given f satisfies $\mathcal{P}$ or not. For instance, let $\mathcal{P}$ be the property that is satisfied by a function $f : \mathcal{D} \to \{0, 1\}$ exactly when f is constant on all of $\mathcal{D}$. The deterministic query complexity of $\mathcal{P}$ is $|\mathcal{D}|$ because otherwise, the function could be non-constant just due to the element of $\mathcal{D}$ not queried. Properties with query complexity $|\mathcal{D}|$ are called *evasive.*

  Unfortunately, evasiveness has been shown to hold for many interesting properties. One important family of properties to which we will refer repeatedly and which will serve as a reference throughout this thesis is the class of *graph properties.* A graph property is any isomorphism-invariant property of a graph, such as planarity or connectivity or bipartiteness. Now, in the above described query model, suppose $\mathcal{D} = \binom{[n]}{2}$, the set of unordered pairs from $[n] \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$, and interpret any function $f : \mathcal{D} \to \{0, 1\}$ as describing the adjacency matrix of a graph on n vertices. Then, each query, or experiment in the physical metaphor, reveals whether there is an edge between a pair of vertices.

  Many graph properties are known to be evasive. Examples include containing a clique of a given size, k-colorability [Bol76], planarity [BvEBL74], and perfectness [HW04]. Chakrabarti, Khot and Shi [CKS01] showed that any minorclosed graph property is evasive. Also, Aanderaa, Karp and Rosenberg [Ros73] famously conjectured that evasiveness is immediate for any non-trivial graph property $\mathcal{P}$ that is[1] preserved under deleting edges (or preserved under adding edges). For any such property, Rivest and Vuillemin [RV76] proved the weaker fact that at least $\Omega(n^2)$ queries are necessary, while Yao proved [Yao88] the conjecture for any such property of bipartite graphs. In short then, it turns out that almost every interesting graph property is provably either evasive or nearly evasive.

- **Randomized Query Complexity.** Given that deterministic algorithms often require too many queries, it is natural to ask what happens if the computational model is made probabilistic. The randomized query complexity of a property is the minimum number of queries needed by a randomized algorithm which is allowed to err with some small probability (where the probability is over the randomness of the algorithm, not over choice of f).

  However, even with randomness, the query complexity of many natural properties remains high. Consider again the class of non-trivial properties of n-vertex graphs which are preserved under addition of edges. Recall that Rivest and Vuillemin showed that the deterministic query complexity of such properties is $\Omega(n^2)$. It turns out that even with randomness, there is no such property known with randomized query complexity less than $n^2/4$, and Karp conjectured that $\Omega(n^2)$ queries are necessary here as well. The current best proof [Haj91, CK07] gives a lower bound of $\tilde{\Omega}(n^{4/3})$ queries and for specific graph properties, such as connectivity, Hamiltonicity, absence of isolated vertices, and containing a triangle, [FKW02] have proved the randomized complexity to be indeed at least $\tilde{\Omega}(n^2)$ queries.

- **Approximate Randomized Query Complexity.** Going back to the scenario described earlier, suppose the scientist does not care much if nature doesn't

[1] Such graph properties are often called *monotone* but we reserve the term for a different meaning to be given later on.

conform to her hypothesis for a small fraction of possible experimental setups. Perhaps, she knows there are lower-order effects that could lead to the hypothesis being violated a small fraction of times. It is only when there is significant inconsistency with *any* law satisfying the hypothesis that she deems it necessary to reject the hypothesis. In this very reasonable setting, where one only needs to solve an "approximate decision problem" instead of an exact one, it turns out that the query complexity often decreases dramatically.

More precisely, the setting is as follows. It is guaranteed that if the given function f does not satisfy property $\mathcal{P}$, then in fact, it is going to disagree on at least 5% of the domain $\mathcal{D}$ with with *any* function satisfying $\mathcal{P}$. The *approximate randomized query complexity* of $\mathcal{P}$ is the minimum number of queries needed by a randomized algorithm to determine membership in $\mathcal{P}$ assuming this guarantee about f. For functions f which do not satisfy this assumption, the testing algorithm can make an arbitrary decision.

As a reference, let us again consider the problem of testing a graph property by querying entries from an adjacency matrix. In stark contrast to the discussion above, nearly every natural graph property has *constant* approximate randomized query complexity, meaning no dependence on the size of the graph whatsoever! This phenomenon was first unearthed by Goldreich, Goldwasser and Ron [GGR98] in a seminal work. They showed that many properties such as k-colorability, containing a large clique as a subgraph, and having a large cut, have constant query complexity. Their work was substantially generalized in a series of works, ultimately resulting in the important theorem of Alon and Shapira [AS08a] that every hereditary graph property, meaning every graph property preserved by taking induced subgraphs, has constant query complexity!

So we see that query complexity in the approximate randomized model can behave very differently from more traditional models. The conventional term for the approximate randomized model is **property testing**, and its study is the focus of our work. We will formally describe the model and our results soon, but before we do so, let us state the main question that motivates our work:

> For what properties $\mathcal{P}$ of functions mapping $\mathcal{D}$ to R is the approximate randomized query complexity of $\mathcal{P}$ a *constant*, independent of the size of the domain $\mathcal{D}$ and range R?

Such properties are called **testable**. As we described above, the testability of graph properties has been very well studied. Here, we will describe work towards a complete characterization of testability for another important family of properties, the linear-invariant properties.

1.2 Property Testing and Linear-Invariant Properties

1.2.1 Boolean Functions

Our primary concern in this thesis will be properties of functions $f : \{0,1\}^n \rightarrow \{0,1\}$. So, $\mathcal{D} = \{0,1\}^n$ and $R = \{0,1\}$ in the above. This is a very common setting in computer science. We will interpret $\mathcal{D} = \{0,1\}^n$ as $\mathbb{F}_2^n$, the n-dimensional vector space over the field of two elements $\mathbb{F}_2$.

It should be possible to extend the work described here to vector spaces over larger fields of constant characteristic, but we will not attempt to do so here. While it is true that larger characteristic (and zero characteristic) is of considerable interest, we restrict overselves to $\mathbb{F}_2$ right now since it is the simplest setting in which to carry out our program of characterizing testability.

1.2.2 Property testing definitions and historical background

Let $\mathcal{P}$ be a property of Boolean functions over the hypercube. In other words, $\mathcal{P} = \bigcup_{n\in\mathbb{Z}^+} \mathcal{P}_n$ where $\mathcal{P}_n$ is a subset of the set of functions $f : \{0,1\}^n \rightarrow \{0,1\}$. Two functions $f, g : \{0,1\}^n \rightarrow \{0,1\}$ are ϵ-far if they differ on at least $\epsilon 2^n$ of the inputs.

We say that f is ϵ-far from satisfying a property $\mathcal{P}$ if it is ϵ-far from any function g satisfying $\mathcal{P}$.

A *tester* for the property $\mathcal{P}$ is a randomized algorithm which distinguishes between the case that an input function f satisfies $\mathcal{P}$ from the case that it is ϵ-far from satisfying $\mathcal{P}$. Here we assume that the function f is given to the tester as an oracle that can be queried. $\mathcal{P}$ is said to be *testable* if there is a function $q : (0, 1) \to \mathbb{Z}^+$ and an algorithm T that, given as input a parameter $\epsilon \in (0, 1)$ and oracle access to a function $f : \{0, 1\}^n \to \{0, 1\}$, makes at most $q(\epsilon)$ queries, accepts with probability at least 2/3 if $f \in \mathcal{P}$ and rejects with probability at least 2/3 if f is ϵ-far from $\mathcal{P}$. Thus, the query complexity of a testable property is a constant, independent of n. Finally, we say that a testing algorithm has *one-sided* error if it always accepts (i.e., with probability 1) functions satisfying $\mathcal{P}$.

The study of testing of Boolean functions began with the work of Blum, Luby and Rubinfeld [BLR93] on testing linearity of Boolean functions. This work was further extended by Rubinfeld and Sudan [RS96]. Around the same time, Babai, Fortnow and Lund [BFL91] also studied similar problems as part of their work on MIP=NEXP. These works are all related to the PCP Theorem, and an important part of it involves tasks which are similar in nature to testing properties of Boolean functions. The work of Goldreich, Goldwasser and Ron [GGR98] extended these results to more combinatorial settings, and initiated the study of similar problems in various areas. More recently, numerous testing questions in the Boolean functions settings have sparked great interest: testing dictators [PRS02], low-degree polynomials [AKK$^+$05, Sam07], juntas [FKR$^+$04, Bla09], concise representations [DLM$^+$07], halfspaces [MORS09], codes [KS07, KS09]. These are documented in several surveys [Fis04, Rub06, Ron08, Sud10], and we refer the reader to these surveys for more background and references on property testing.

1.2.3 Linear Invariance

What features of a property make it testable? On the one hand, Goldreich, Goldwasser, and Ron [GGR98] showed that with high probability, a random property of

Boolean functions is not testable[2]. In fact, nearly all of the domain needs to be queried with high probability. On the other hand, we have already mentioned that a variety of mathematically natural properties of Boolean functions are testable. Can we isolate the ingredients in natural properties that make them testable?

Kaufman and Sudan in [KS08] suggested that the large number of symmetries usually exhibited by properties occurring in mathematics might play a crucial role in explaining their testability. They initiated the study of the relationship between a property's testability and its invariance group. A very common invariance shown by properties of Boolean functions is linear invariance. Formally, a property of Boolean functions $\mathcal{P}$ is said to be *linear-invariant* if for every function $f : \mathbb{F}_2^n \to \{0,1\}$ satisfying $\mathcal{P}$ and for any linear transformation $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$, the composition $f \circ L$ satisfies $\mathcal{P}$ as well, where we define $(f \circ L)(x) = f(L(x))$. Note that here we explicitly identify $\{0,1\}^n$ with $\mathbb{F}_2^n$, and we will use this convention from now on throughout.

For a thorough discussion of the importance of linear-invariance, we refer the reader to Sudan's recent survey on the subject [Sud10] and to the paper of Kaufman and Sudan which initiated this line of work [KS08]. As should be apparent from the title of this thesis, the main thrust of our work will be to classify the set of testable linear-invariant properties. Before we do so, though, let us describe some specific linear-invariant properties and what is known about their testability.

1.2.4 Examples of Linear-Invariant Properties

We will refer to the following four linear-invariant properties for reference throughout this work:

- **Linearity:** We say a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is *linear* if $f(x+y) = f(x) + f(y)$ for all $x, y \in \mathbb{F}_2^n$. It is clearly linear-invariant as compositions of linear functions are linear. The testability of linearity was shown in the seminal paper [BLR93] which introduced property testing as a line of inquiry in computer science. They

[2] By a random property, we mean a random subcollection of the collection of functions mapping $\{0,1\}^n$ to $\{0,1\}$.

proved that if f is ϵ-far from linear, then with probability at least ϵ, uniformly chosen x, y from $\mathbb{F}_2^n$ does not satisfy the condition $f(x+y) = f(x) + f(y)$. Repeating this process independently for $O(1/\epsilon)$ times ensures that such an ϵ-far f is rejected with probability $2/3$ as desired.

- **Low-degree polynomials:** The property of a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ *being a polynomial of degree at most d* is also clearly linear-invariant. [AKK+05] showed that one can test the property using $O(d4^d/\epsilon)$ queries. Subsequently, [BKS+09] improved the query complexity to $O(2^d + 1/\epsilon)$ (not included in this thesis). For constant d, these works show that the property of being of degree at most d is testable. The testability of low-degree polynomials was also studied much earlier in the context of probabilistically checkable proofs, but these works addressed the problem when the field characteristic is very large, growing with n.

- **Fourier dimensionality[3] and sparsity:** A function $f : \mathbb{F}_2^n \to \{0,1\}$ is said to have *Fourier dimension* k if the Fourier spectrum[4] of f is supported in a subspace of dimension k, while it is said to have *Fourier sparsity* k if the number of nonzero Fourier coefficients of f is at most k. It is easy to check that both of these properties are linear-invariant. Namely, one has to observe that if $F = f \circ L$ for a linear transformation L, then the spectrum of F is a subset of the image of a linear transformation applied to the spectrum of f. The testability of Fourier dimensionality and Fourier sparsity for constant k was established by [GOS+09].

- **Odd-Cycle-Freeness:** A function $f : \mathbb{F}_2^n \to \{0,1\}$ is said to be *odd-cycle-free* if there exists some $\alpha \in \mathbb{F}_2^n$ for which $\hat{f}(\alpha) = -\mathbb{E}_x[f(x)]$. Note that $-\mathbb{E}_x[f(x)]$ is the minimum value a Fourier coefficient can attain. One can check the linear-invariance of odd-cycle-freeness starting from its definition. We will give one proof of the testability of odd-cycle-freeness in this thesis. Two different

[3] Another term used for the same property is *subspace junta* [VX11].

[4] We define Fourier coefficients later on in the context of our work. But to be self-contained, we recall that for any $\alpha \in \mathbb{F}_2^n$, $\hat{f}(\alpha)$ is defined to be $\mathbb{E}_x[f(x)(-1)^{\langle\alpha,x\rangle}]$ and the Fourier spectrum of f refers to the set $\{\alpha \mid \hat{f}(\alpha) \neq 0\}$.

proofs, both much better quantitatively than the proof given here, were found subsequently in [BGRS11] (not included in this thesis). Those same works also show that the minimum Fourier coefficient can be estimated efficiently, by using a modification of the testing algorithm.

1.2.5 Subspace Hereditary Properties

If $\mathcal{P}$ is a linear-invariant property of Boolean functions on $\mathbb{F}_2^n$, then it does not depend on the basis used to represent the coordinates of points in $\mathbb{F}_2^n$. This is a hallmark of many natural properties of Boolean functions, as illustrated in the previous section. But the properties described in the previous section have another common feature: they are defined uniformly, independently of n. For arbitrary linear-invariant properties, it might be that for different values of n (log of the domain size), membership in the property is defined in completely different ways. That is, the *description* of a linear-invariant property might heavily depend on n. Intuitively, this makes it implausible that an arbitrary linear-invariant property is testable because knowledge about the function restricted to a smaller space (the space of queried points) may not yield much information about whether the function satisfies the property on the entire space. Indeed, a variant[5] of an argument in [GGR98] shows that there exist linear-invariant properties that are not only non-testable but require $\Omega(2^n)$ queries.

Subspace-hereditary properties gets around the possible obstruction to testability mentioned above.

Definition 1 (Subspace-Hereditary Properties) *A linear-invariant property $\mathcal{P}$ is said to be* subspace-hereditary *if it is closed under restriction to subspaces. That is, if f is in $\mathcal{P}_n$ and H is a m-dimensional linear subspace of $\mathbb{F}_2^n$, then $f|_H \in \mathcal{P}_m$ also,*

[5] Proposition 4.1 of [GGR98] shows that for every n, there exists a property of Boolean functions that contains $2^{\frac{1}{10}2^n}$ of the Boolean functions over $\mathbb{F}_2^n$ and cannot be tested with less than $\frac{1}{20}2^n$ queries. This family of functions is not necessarily linear-invariant, so we just "close" it under linear transformation, by adding to the property all the linear-transformed such functions. Since the number of these linear transformation is bounded by 2^{n^2} (corresponding to all possible $n \times n$ matrices over $\mathbb{F}_2$) we get that the new property contains at most $2^{n^2}2^{\frac{1}{10}2^n} \leq 2^{\frac{1}{5}2^n}$ Boolean functions. One can verify that since this new family contains a small fraction of all possible functions, the argument of [GGR98] caries over, and the new property cannot be tested with $o(2^n)$ queries.

where[6] $f|_H : \mathbb{F}_2^m \to \{0,1\}$ *is the restriction of* f *to* H.

Subspace-hereditary properties include all the properties mentioned in the previous section: linearity, low-degree polynomials, Fourier sparsity, and odd-cycle-freeness.

One of the main contributions of this thesis is the following conjecture.

Conjecture 2 (Main Testability Conjecture) *Every subspace-hereditary linear-invariant property is testable with one-sided error.*

The truth of the conjecture would unify testability results for all the properties from Section 1.2.4, as well as lead towards an *exact* characterization of testable linear-invariant properties (see Section 1.3.3 below). In this thesis, among other things, we develop tools for establishing the conjecture for a limited class of subspace-hereditary properties that we hope is useful in the future. Before we describe our results though, we need one additional piece of terminology.

1.2.6 Local Constraints for Linear-Invariant Properties

The notion of a local constraint turns out to be crucial in describing the class of linear-invariant properties for which we show testability as well as for their analysis.

First, let us define what we mean by local constraints for arbitrary properties of Boolean functions. For a positive integer k, a *k-local constraint* $C = (a_1, \ldots, a_k; \sigma)$ is given by k elements $a_1, \ldots, a_k \in \{0,1\}^n$ and a string $\sigma \in \{0,1\}^k$. A function $f : \{0,1\}^n \to \{0,1\}$ is said to satisfy the constraint C if $(f(a_1), \ldots, f(a_k)) \neq \sigma$, and a property $\mathcal{P}$ satisfies C if every function $f \in \mathcal{P}$ satisfies C. For instance, the property of linearity (as defined in Section 1.2.4) satisfies the constraints $(a_1, a_2, a_1 + a_2; 111)$ and $(a_1, a_2, a_1 + a_2; 001)$ for every choice of $a_1, a_2 \in \mathbb{F}_2^n$, since a function f will violate the identity $f(x+y) = f(x) + f(y)$ if $f(a_1) = 1, f(a_2) = 1, f(a_1 + a_2) = 1$ or if $f(a_1) = 0, f(a_2) = 0, f(a_1 + a_2) = 1$ for some choice of $a_1, a_2 \in \mathbb{F}_2^n$. In fact, these constraints suffice to completely define the linearity property. We will see next that

[6] Note that we are implicitly composing $f|_H$ with a linear transformation so that it is now defined on $\mathbb{F}_2^m$. Here, we are using the fact that $\mathcal{F}$ is linear-invariant.

any subspace-hereditary linear-invariant property can be defined using such local constraints.

For linear-invariant properties, local constraints have an especially nice structure. To see this, observe that specifying a property to be linear-invariant also enforces a symmetry among the local constraints satisfied by the property. If a linear-invariant property $\mathcal{P}$ satisfies the constraint $C = (a_1, \dots, a_k; \sigma)$, then it must also satisfy the constraint $C \circ L = (L(a_1), \dots, L(a_k); \sigma)$ for any linear map $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$. Thus, $\mathcal{P}$ must satisfy all constraints in the *orbit* of C, i.e. the family of constraints $\{C \circ L \mid \text{linear } L : \mathbb{F}_2^n \to \mathbb{F}_2^n\}$. It is straightforward to verify that one can encode the orbit of a constraint $C = (a_1, \dots, a_k; \sigma)$ by a tuple $(v_1, \dots, v_k; \sigma)$ for vectors v_i in the smaller space $\mathbb{F}_2^r$ (for some $r \leq k$) such that the orbit of C equals $\{(L(v_1), \dots, L(v_k); \sigma) : \text{linear } L : \mathbb{F}_2^r \to \mathbb{F}_2^n\}$. Here, the exact identity of the elements $v_1, \dots, v_k$ is not important – the only thing that matters is the linear dependencies between them. Hence, it is convenient to think of them as the representation of a *linear matroid.*[7]. To make the discussion more concrete, consider again the property of linearity. It can be defined as the property which satisfies the orbit of the following two constraints: $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2; 111)$ and $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2; 001)$, where $\mathbf{e}_1, \mathbf{e}_2$ are two linearly independent vectors in $\mathbb{F}_2^2$.

We thus arrive at the following definition:

Definition 3 (Induced Matroid freeness) *Given integers $k \geq r \geq 1$, a set $\mathcal{M} = \{v_1, \dots, v_k\}$ of k vectors in $\mathbb{F}_2^r$, and a string $\sigma \in \{0,1\}^k$, we say that a function $f : \mathbb{F}_2^n \to \{0,1\}$ is* $(\mathcal{M}, \sigma)$-free *if there does not exist any linear map $L : \mathbb{F}_2^r \to \mathbb{F}_2^n$ such that $f(L(v_i)) = \sigma_i$ for all $i \in [k]$. Otherwise, if such an L exists, we say f* contains $(\mathcal{M}, \sigma)$ at L.

Any property that is equivalent to $(\mathcal{M}, \sigma)$-freeness for some $\mathcal{M}$ and σ is called an induced matroid freeness property*, and furthermore if $\sigma = \mathbf{1}$, then it is called a* matroid freeness property.

[7] The formal definition of a matroid is not too important in this context. In the rest, the reader can just think of a matroid as a set of elements in a vector space over $\mathbb{F}_2$.

Notice that we needed *two* induced matroid freeness constraints to define linearity. Thus, it is natural to formulate the condition of a property satisfying multiple constraints.

Definition 4 *Given a collection $\mathcal{F} = \{(\mathcal{M}^i, \sigma^i) : i \in \mathbb{Z}^+\}$, where each $\mathcal{M}^i$ is a set of k_i vectors in $\mathbb{F}_2^{r_i}$, for some integers $k_i \geq r_i \geq 1$, and each $\sigma^i \in \{0,1\}^{k_i}$, we say that a function $f : \mathbb{F}_2^n \to \{0,1\}$ is $\mathcal{F}$-free if f is $(\mathcal{M}^i, \sigma^i)$-free for every $i \in \mathbb{Z}^+$.*

We will show later on (Chapter 3) that the subspace-hereditary properties are exactly the properties described by Definition 4.

Proposition 5 *A linear-invariant property $\mathcal{P}$ is subspace-hereditary if and only if it is an $\mathcal{F}$-freeness property as in Definition 4 or Definition 7.*

This formulation of subspace-hereditary properties using local constraints will be essential in what follows.

In this thesis, we will sometimes switch between the notation used in Definitions 3 and 4 to an alternate but entirely equivalent notation. Let us quickly define this alternate notation now. To motivate it, observe that if the vectors $v_1, \ldots, v_k$ satisfy a linear dependency, then $L(v_1), \ldots, L(v_k)$ also satisfy the same linear dependency for every linear transformation L. In fact, the only information that is needed about $\mathcal{M}$ in Definition 3 above is the linear dependencies between the vectors $v_1, \ldots, v_k$.

Given $\mathcal{M} = \{v_1, \ldots, v_k\} \subseteq \mathbb{F}_2^r$, let V be the k-by-r matrix whose ith row is the vector v_i. Now, let M be the matrix over $\mathbb{F}_2$ whose kernel is exactly the column-space of V. Immediately, $MV = 0$, and also if, for a linear transformation $L : \mathbb{F}_2^r \to \mathbb{F}_2^n$, V_L is the matrix formed out of the rows $L(v_1), \ldots, L(v_k)$, then $MV_L = 0$. Combining these facts, we can reformulate Definition 3 as follows:

Definition 6 (**(M, σ)-free**) *Given an $m \times k$ matrix M over $\mathbb{F}_2$ and $\sigma \in \{0,1\}^k$, we say that a function $f : \mathbb{F}_2^n \to \{0,1\}$ is* (M, σ)-free *if there is no $X = (x_1, \ldots, x_k) \in (\mathbb{F}_2^n)^k$ such that $MX = 0$ and for all $1 \leq i \leq k$ we have $f(x_i) = \sigma_i$.*

Such a constraint (M, σ) is said to an induced system of linear equations.

And similarly, we also have an alternate notation for $\mathcal{F}$-freeness in Definition 4.

Definition 7 ($\mathcal{F}$-free) *Let $\mathcal{F} = \{(M^1, \sigma^1), (M^2, \sigma^2), \ldots\}$ be a (possibly infinite) set of induced systems of linear equations. A function f is said to be $\mathcal{F}$-free if it is (M^i, σ^i)-free for all i.*

To return to the example of linearity, it is equivalent to the property defined by $\{([1\ 1\ 1], 111), ([1\ 1\ 1], 001)\}$-freeness, as can be verified directly. Also, Proposition 5 implies that subspace-hereditariness is equivalent to being an $\mathcal{F}$-freeness property in the sense of Definition 7.

1.3 Our Results

1.3.1 Testability of some Subspace-Hereditary Properties

Our first main result in this thesis is that a large subclass of subspace-hereditary linear-invariant properties of Boolean functions is testable with one-sided error. Recall that a property is said to be testable if its query complexity does not depend on n.

Theorem 8 (Main Testability Result) *Let $\mathcal{F} = \{(M^1, \sigma^1), (M^2, \sigma^2), \ldots\}$ be a (possibly infinite) set, where each M^i is a matrix of size 1-by-k_i for some integer k_i and each σ_i is an arbitrary string in $\{0,1\}^{k_i}$. Then the property of $\mathcal{F}$-freeness is testable with one-sided error.*

Linearity and odd-cycle-freeness meet the conditions of this theorem and are thus testable by Theorem 8.

One can view our work as paralleling previous work on testing graph properties. The correspondence is informal but useful to keep in mind. Given a function $f : \mathbb{F}_2^n \to \{0,1\}$, consider the set $S_f = \{x \in \mathbb{F}_2^n : f(x) = 1\}$. Then, f is $([1\ 1\ 1], 111)$-free if and only if S_f contains no x_1, x_2, x_3 such that $x_1 + x_2 + x_3 = 0$. The notion of $(M, \mathbf{1})$-freeness is analogous to the graph property of being H-free for some fixed graph H, where a graph is said to be H-free if and only if its edge set does not contain a copy of H. Observe that in both cases, the property is *monotone* in the sense that if f is $(M, \mathbf{1})$-free, then removing elements from S_f results in a set that contains no solution to $Mx = 0$. Similarly if G is H-free, then removing edges from G results in an H-free graph.

Let us now go back to considering arbitrary $\sigma \in \{0,1\}^k$ in Definition 6, where again the intuition comes from graph properties. Observe that a natural variant of the monotone graph property of being H-free is the property of being induced H-free for some fixed graph H, where a graph is said to be induced H-free if it contains no set of $|H|$ vertices that induces a copy of H and no other edges. Note that being induced H-free is no longer a monotone property since if G is induced H-free, then removing an edge can actually create induced copies of H. Getting back to the property of being (M, σ)-free, observe that we can think of this as requiring S_f to contain no *induced* solution to the system of equations $Mx = 0$. That is, the requirement is that there should be no solution vector $V = (v_1, \dots, v_k)$ satisfying $MV = 0$, where $v_i \in S_f$ if $\sigma_i = 1$ and $v_i \in \mathbb{F}_2^n \setminus S_f$ if $\sigma_i = 0$. So we can think of σ as encoding which elements of a potential solution vector V should belong to S_f and which should belong to its complement.

Keeping in mind this informal correspondence with graph properties, let us compare Theorem 8 with what is known for testability of graph properties. H-freeness for a fixed graph H was shown to be testable by Alon (cited as private communication in [GGR98]). Testability of induced H-freeness came a few years later [AFKS00]. Subsequenly, Alon and Shapira [AS08b] showed that (non-induced) freeness from any fixed collection of subgraphs is testable and, finally, the same two authors [AS08a] established testability for induced freeness from any fixed collection of subgraphs. The last result shows that any *hereditary* graph property is testable, meaning any property $\mathcal{P}$ for which if a graph G satisfies $\mathcal{P}$, then every induced subgraph of G also does. The result of [AS08a] has been later extended to hereditary hypergraph properties by Austin and Tao [AT08] and Rödl and Schacht [RS09].

Now, let us return to our linear-invariant properties. One work that initiated the results motivating this thesis was by Green [Gre05]. His result can be formulated as saying that for any rank one matrix M, the property of being $(M, \mathbf{1})$-free can be tested with one-sided error. Green conjectured that the same result holds for matrices of higher rank also. This conjecture was confirmed by Shapira [Sha09] and Král', Serra and Vena [KSV08]. In our language, the results of [Sha09, KSV08] can be stated as saying that for any matrix M, the property of being $(M, \mathbf{1})$-free is testable with one-sided error. The work described here is incomparable to these results; recall that we show testability of (M, σ)-freeness when σ is arbitrary but M is of rank one. Also, we show testability of an infinite collection of such properties whereas previous works did not address the possibility of having an infinite number of constraints.

Comparing Theorem 8 to the results on graph properties, it seems that we have achieved the parallel of the result of Alon and Shapira [AS08a] that hereditary graph properties are testable. We believe that when M is of high rank, it is actually the *hypergraph* property of being induced H-free (for some fixed sub-hypergraph H) that is analogous to being (M, σ)-free. Since as mentioned, the result of [AS08a] has been later extended to hypergraphs, it is natural to expect that one could also handle an infinite number of forbidden induced systems of equations in the functional case as well. All the above provides inspiration for Conjecture 2 which we now reformulate in terms of local constraints.

Conjecture 2 (restated) *For every (possibly infinite) set of systems of induced equations $\mathcal{F}$, the property of being $\mathcal{F}$-free is testable with one-sided error.*

We mention that while the notions of graph properties being hereditary and functions being subspace-hereditary are somewhat more natural than the equivalent notions of being free of induced subgraphs and equations respectively, it is actually easier to think about these properties using the latter notion when proving theorems about them. This was the case for graphs in [AS08a], and it will be the case in the present work as well.

Let us now recall the four examples of testable linear-invariant properties described in Section 1.2.4: linearity, low-degree polynomials, Fourier dimensionality/sparsity, and odd-cycle-freeness. One can observe that all these properties are subspace-hereditary. Thus, if our Conjecture 2 is true, as we strongly believe, then we could explain the testability of all these properties through a unified perspective. Note that our main result already shows (yet again!) that linearity is testable but from a completely different viewpoint than used in previous analysis. Furthermore, to show the testability of low degree polynomials (a.k.a., Reed-Muller codes) and Fourier dimensionality/sparsity, we would only need to resolve Conjecture 2 for a *finite* family of forbidden induced systems of equations.

1.3.2 Lower Bound for Triangle-Freeness

Our second main result gives a non-trivial lower bound for an explicit testable linear-invariant property. Observe that for any property $\mathcal{P}$ of functions, a tester that makes $o(1/\epsilon)$ queries will not be able to distinguish with constant probability between functions satisfying $\mathcal{P}$ and functions ϵ-far from satisfying $\mathcal{P}$. So, $\Omega(1/\epsilon)$ is a trivial lower-bound that holds for any one-sided tester of a non-trivial property. We give the first super-linear (in $1/\epsilon$) lower-bound for a linear-invariant property of Boolean functions.

We analyze the testability of *triangle-freeness*: a function $f : \mathbb{F}_2^n \to \{0, 1\}$ is said to be triangle-free if there are no $x, y \in \mathbb{F}_2^n$ such that $f(x) = f(y) = f(x + y) = 1$.

Theorem 9 (Main Lower Bound Result) *The one-sided query complexity for triangle-freeness is at least $\Omega((1/\epsilon)^{2.423})$.*

Green [Gre05] showed that triangle-freeness is testable with query complexity that is a tower-of-exponentials of height polynomial in $1/\epsilon$. Thus, while we are quite far away from understanding the right query complexity, our work shows that one cannot expect an $O(1/\epsilon)$ or an $O(1/\epsilon^2)$ algorithm, such as those for the properties described in Section 1.2.4.

It is interesting to compare the testability of triangle-freeness for Boolean functions and triangle-freeness for graphs. Using Szemerédi's regularity lemma, triangle-freeness in graphs is also known to be testable with a tower-type query complexity upper bound. Alon [Alo02] gave a super-polynomial query complexity lower bound, and it is the strongest query lower bound for a natural testable property known to date. However, the proof technique in [Alo02] does not seem to directly apply to the our setting due to the algebraic structure of the Boolean cube.

Let us briefly describe the main thrust of the analysis. Call a 3-element set $\{x, y, x+y\}$ for some $x, y \in \mathbb{F}_2^n$ a *triangle in a function* $f : \mathbb{F}_2^n \to \{0,1\}$ if $f(x) = f(y) = f(x+y) = 1$. The *canonical tester* for triangle-freeness repeatedly picks x and y uniformly and independently at random and checks if $f(x) = f(y) = f(x+y) = 1$. Note that the canonical tester is inded a one-sided testing algorithm for triangle-freeness. Moreover, if the number of triangles is N_Δ for a function that is ϵ-far from being triangle-free, then for the canonical tester to reject such functions with constant probability, it needs to make at least $\Omega(\frac{2^{2n}}{N_\Delta})$ queries. Thus, in order to show that the canonical tester has constant query complexity $O(Q(\epsilon))$ for some function $Q : (0,1) \to \mathbb{Z}^+$, one would want to show that $N_\Delta = 2^{2n}/Q(\epsilon)$. Green [Gre05] showed this fact, albeit for $Q(\epsilon)$ that was enormous, upper-bounded by a tower of 2's of height polynomial in $1/\epsilon$. The question of obtaining a better bound for N_Δ was explicitly left open in [Gre05].

In our work, we essentially show the existence of a function $f : \mathbb{F}_2^n \to \{0,1\}$ which is ϵ-far from being triangle-free but for which $N_\Delta = O(\epsilon^{4.847\cdots}) \cdot 2^{2n}$. Thus, we get a lower-bound of $\Omega((1/\epsilon)^{4.847\cdots})$ for the query complexity of the canonical tester. Ultimately though, the goal is to lower bound the query complexity for an arbitrary testing algorithm. To this end, we show that if there is a one-sided, possibly adaptive tester for triangle-freeness with query complexity q, then one can transform that tester into a canonical one with query complexity at most $O(q^2)$. A more naïve argument blows up the query complexity to 2^q. The main fact used in our proof is the pairwise independence of linear subspaces. Combining with our results for canonical testers, this implies a query complexity lower bound of $(\frac{1}{\epsilon})^{2.423\cdots}$ for triangle-freeness, with respect to one-sided testers. A lower bound for 2-sided testers remains an open problem.

1.3.3 One-sided Testability and Subspace Hereditariness

We now turn to discuss our third result, in which we explore the converse direction to the results of Section 1.3.1. Namely, we roughly show that one-sided testability using "natural" testers implies that the property is subspace-hereditary. Let us start with formally defining the types of "natural" testers we consider here.

Definition 10 (Oblivious Tester) *An* oblivious tester *for a property $\mathcal{P} = \{\mathcal{P}_n\}_n$ is a (possibly 2-sided error) non-adaptive, probabilistic algorithm, which, given a distance parameter ϵ, and oracle access to an input function $f : \mathbb{F}_2^n \to \{0,1\}$, performs the following steps:*

1. *Computes an integer $d = d(\epsilon)$. If $d(\epsilon) > n$, let $H = \mathbb{F}_2^n$. Otherwise, let $H \leq \mathbb{F}_2^n$ be a subspace of dimension $d(\epsilon)$ chosen uniformly at random.*
2. *Queries f on all elements $x \in H$.*
3. *Accepts or rejects based only on the outcomes of the received answers, the value of ϵ, and its internal randomness.*

We now discuss the motivation for considering the above type of algorithms. We prove the fact that we can assume the tester is non-adaptive and queries a random linear subspace without loss of generality; this is analogous to the fact [AFKS00, GT03] that one can assume a graph property tester makes its decision only by inspecting a randomly chosen induced subgraph. The only essential restriction we place on oblivious testers is that their behavior cannot depend on the value of n, the domain size of the input function. If we allow the testing algorithm to make its decisions based on n, then it can do very strange and unnatural things. For example, we can now consider properties that depend on the parity of n. As was shown in [AS08c], the algorithm can use the size of the input in order to compute the optimal query complexity. All these abnormalities will not allow us to give any meaningful characterization. As observed in [AS08a] by restricting the algorithm to make its decisions while not considering the size of the input, we can still test any (natural) property while at the same time avoid annoying technicalities. We finally note that all the testing algorithms for testable properties of Boolean functions in prior works were indeed oblivious, and that furthermore many of them implicitly consider only oblivious testers. In particular, these types of testers were considered in [Sud10].

We first show that if Conjecture 2 is true, meaning subspace-hereditary properties are testable, then actually, oblivious testers can test properties which are slightly more general than subspace-hereditary properties. This larger class of properties is defined as follows.

Definition 11 (Semi Subspace-Hereditary Property) *A property $\mathcal{P} = \{\mathcal{P}_n\}_n$ is* semi subspace-hereditary *if there exists a subspace-hereditary property $\mathcal{H}$ such that*

1. *Any function f satisfying $\mathcal{P}$ also satisfies $\mathcal{H}$.*
2. *There exists a function $M : (0, 1) \to \mathbb{N}$ such that for every $\epsilon \in (0, 1)$, if $f : \mathbb{F}_2^n \to \{0, 1\}$ is ϵ-far from satisfying $\mathcal{P}$ and $n \geq M(\epsilon)$, then $f|_V$ does not satisfy $\mathcal{H}$, for some subspace $V \leq \mathbb{F}_2^n$.*

The intuition behind the above definition is that a semi subspace-hereditary property can only deviate from being "truly" subspace-hereditary on functions over a finite domain, where the finiteness is controlled by the function M in the definition. Our next theorem connects the notion of oblivious testing and semi subspace-hereditary properties. Assuming Conjecture 2, it essentially characterizes the linear-invariant properties that are testable with one-sided error, thus resolving Sudan's problem raised in [Sud10].

Theorem 12 *If Conjecture 2 holds, then a linear-invariant property $\mathcal{P}$ is testable by a one-sided error oblivious tester* **if and only if** *$\mathcal{P}$ is semi subspace-hereditary.*

Getting back to the similarity to graph properties, we note that [AS08a] obtained a similar characterization for the graph properties that are testable with one-sided error. Let us close by mentioning two points. The first is that most linear-invariant properties are known to be testable with one-sided error, and hence the question of characterizing these properties is well motivated. In fact, for the subclass of linear-invariant properties which also themselves form a linear subspace, [BHR05] showed that the optimal tester is always one-sided and non-adaptive. Our second point is that it is natural to ask if there are linear-invariant properties which are not testable. A linear-invariant property with query complexity $\Omega(2^n)$ arises implicitly from the arguments of [GGR98]. A second, more natural, example comes from Reed-Muller codes. [BKS$^+$09] shows that for any $1 \ll q(n) \ll n$ the linear-invariant property of being a $\log_2(q(n))$-Reed-Muller code cannot be tested with $o(q(n))$ queries. We also conjecture that the property of two functions being isomorphic up to linear transformations of the variables is not a testable property. Lower bounds for isomorphism testing have been studied both in the Boolean function model [FKR$^+$04, BO10] and in the dense graph model [Fis05], but our problem specifically does not seem to have been examined in a property testing setting.

1.4 Organization

This thesis has four chapters subsequent to this one with technical content. Chapter 2 establishes combinatorial results regarding Boolean functions defined on $\mathbb{F}_2^n$ that are instrumental in our testability work. In particular, we establish strengthened arithmetic regularity lemmas, in the style of those developed by Alon et al. [AFKS00] for graphs. In Chapter 3, we prove Theorem 8 on the testability of properties described by freeness from a collection of induced linear equations. At the end of this chapter, we also show a stronger version of the theorem which shows testability for properties described by freeness from induced systems of equations of complexity 1. The matrices describing such equations can have rank greater than 1. Chapter 4 gives the proof of Theorem 9. In the course of doing so, we actually prove a more general result about the structure of testers of (non-induced) matroid-freeness properties. In Chapter 5, we prove the claims in Section 1.3.3, leading up to Theorem 12, the conditional characterization of the testable linear-invariant properties. This chapter also includes the proof of Proposition 5.

The content of Chapters 2, 3 and 5 appeared in [BGS10]. The content of Chapter 4 appeared in [BX10].

Chapter 2

Regular Partitionings of the Hypercube

In this chapter, we prove several facts about Boolean functions of the hypercube. In particular, we will see a few different arithmetic regularity lemmas that are extensions of Green's regularity lemma discussed in the Introduction. These extensions will be important for proving the main testability result in this thesis.

2.1 Fourier Coefficients and Subspace Restrictions

The *support* of a Boolean function f refers to the subset of the domain on which f evaluates to 1. If H is a subspace of $\mathbb{F}_2^n$ and given function $f : H \to \{0,1\}$, let $\rho(f)$, the *density* of f, denote $\frac{\sum_{x \in H} f(x)}{|H|}$. Recall that the Fourier coefficients of f, defined for each $\alpha \in H^*$, are:

$$\widehat{f}(\alpha) = \mathop{\mathbb{E}}_{x \in H}\left[f(x) \cdot (-1)^{\langle x,\alpha \rangle}\right]$$

For a parameter $\epsilon \in (0,1)$, we say f is *ϵ-uniform* if $\max_{\alpha \neq 0} |\widehat{f}(\alpha)| < \epsilon$. This definition captures the notion that the function does not agree with any linear function on H and is hence "pseudorandom" against the class of linear functions.

Given a function $f : \mathbb{F}_2^n \to \{0,1\}$, a subspace $H \leq \mathbb{F}_2^n$ and an element $g \in \mathbb{F}_2^n$, define the function $f_H^{+g} : H \to \{0,1\}$ to be $f_H^{+g}(x) = f(x+g)$ for $x \in H$. The support of f_H^{+g} represents the intersection of the support of f with the coset $g + H$. The following lemma shows that if a uniform function is restricted to a coset of a subspace of low codimension, then the restriction does not become too non-uniform and its density stays roughly the same.

Lemma 13 *Let $f : \mathbb{F}_2^n \to \{0,1\}$ be an ϵ-uniform function of density ρ, and let $H \leq \mathbb{F}_2^n$ be a subspace of codimension k. Then for any $c \in \mathbb{F}_2^n$, the function $f_H^{+c} : H \to \{0,1\}$ is $(2^k\epsilon)$-uniform and of density ρ_c satisfying $|\rho_c - \rho| < 2^k\epsilon$.*

Proof: Let $H^\perp = \{\alpha \in \mathbb{F}_2^n | \ \langle \alpha, h\rangle = 0 \ \forall h \in H\}$ be the dual to the vector space H, and let $H' = \mathbb{F}_2^n/H$ be the quotient of H in $\mathbb{F}_2^n$. We wish to show that, for every $c \in H'$, the Fourier coefficients of f_H^{+c} are small.

For every $\beta \in \mathbb{F}_2^n/H^\perp$ and $\alpha \in H^\perp$:

$$\begin{aligned}\widehat{f}(\beta+\alpha) = \mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n}[f(x)\chi_{\beta+\alpha}(x)] &= \mathop{\mathbb{E}}_{c'\in H'}\mathop{\mathbb{E}}_{h\in H} f_H^{+c'}(h)\chi_{\beta+\alpha}(c'+h)\\ &= \mathop{\mathbb{E}}_{c'\in H'}\chi_{\beta+\alpha}(c')\mathop{\mathbb{E}}_{h\in H} f_H^{+c'}(h)\chi_\beta(h)\\ &= \frac{1}{2^k}\sum_{c'\in H'}\chi_{\beta+\alpha}(c')\widehat{f_H^{+c'}}(\beta)\end{aligned}$$

Recall that $\sum_{\alpha\in H^\perp}\chi_\alpha(c') = \begin{cases}0, \text{ if } c'\neq 0\\ 1, \text{ if } c' = 0.\end{cases}$ Fixing $\beta \in \mathbb{F}_2^n/H^\perp$ and $c \in H'$ and summing up the quantity computed above over all $\alpha \in H^\perp$, we obtain

$$\begin{aligned}2^k\left(\sum_{\alpha\in H^\perp}\chi_{\beta+\alpha}(c)\widehat{f}(\beta+\alpha)\right) &= \sum_{c'\in H'}\sum_{\alpha\in H^\perp}\chi_{\beta+\alpha}(c+c')\widehat{f_H^{+c'}}(\beta)\\ &= \sum_{\alpha\in H^\perp}\chi_{\beta+\alpha}(0)\widehat{f_H^{+c}}(\beta) + \sum_{c'\in H'-\{c\}}\sum_{\alpha\in H^\perp}\chi_{\beta+\alpha}(c+c')\widehat{f_H^{+c'}}(\beta)\\ &= 2^k\widehat{f_H^{+c}}(\beta) + \sum_{c'\in H'-\{0\}}\sum_{\alpha\in H^\perp}\chi_{\beta+\alpha}(c')\widehat{f_H^{+c'+c}}(\beta)\\ &= 2^k\widehat{f_H^{+c}}(\beta) + \sum_{c'\in H'-\{0\}}\chi_\beta(c')\left(\sum_{\alpha\in H^\perp}\chi_\alpha(c')\right)\widehat{f_H^{+c'+c}}(\beta)\\ &= 2^k\widehat{f_H^{+c}}(\beta).\end{aligned}$$

Furthermore,

$$\left|\widehat{f_H^{+c}}(\beta)\right| = \left|\sum_{\alpha \in H^\perp} \chi_{\beta+\alpha}(c)\widehat{f}(\beta+\alpha)\right| \leq \sum_{\alpha \in H^\perp} \left|\chi_{\beta+\alpha}(c)\widehat{f}(\beta+\alpha)\right| = \sum_{\alpha \in H^\perp} \left|\widehat{f}(\beta+\alpha)\right|$$

Since f is ϵ-uniform, setting $\beta = 0$ in the above inequality shows that $|\rho_c - \rho| \leq \sum_{0 \neq \alpha \in H^\perp} |\widehat{f}(\alpha)| < 2^k\epsilon$. For nonzero β in $\mathbb{F}_2^n/H^\perp$, it follows again from ϵ-uniformity that $|\widehat{f_H^{+c}}(\beta)| < 2^k\epsilon$. ■

2.2 Regularity Lemmas

For a subspace $H \leq \mathbb{F}_2^n$, the *H-based partition* refers to the partitioning of $\mathbb{F}_2^n$ into the cosets in $\mathbb{F}_2^n/H$. If $H' \leq H$, then the H'-based partition is called a *refinement* of the H-based partition. The *order* of the H-based partition is defined to be $[G : H]$, i.e., the index of H as a subgroup or the dimension of the quotient space $\mathbb{F}_2^n/H$. Using this notation, Green's regularity lemma can be stated as follows.

Lemma 14 (Green's Regularity Lemma [Gre05]) *For every m and $\epsilon > 0$, there exists $T = T_{14}(m,\epsilon)$ such that the following is true. Given function $f : \mathbb{F}_2^n \to \{0,1\}$ with $n > T$ and H-based partition of $\mathbb{F}_2^n$ with order at most m, there exists a refined H'-based partition of order k, with $m \leq k \leq T$, for which $f_{H'}^{+g}$ is not ϵ-uniform for at most $\epsilon 2^n$ many $g \in \mathbb{F}_2^n$.*

Our main tool in this work is a functional variant of Green's regularity lemma, in which the uniformity parameter ϵ is not a constant but rather an arbitrary function of the order of the partition. It is quite analogous to a similar lemma, first proved in [AFKS00], in the graph property testing setting. The recent work [GT10] shows a (very strong) functional regularity lemma in the arithmetic setting but it applies over the integers and not $\mathbb{F}_2$.

Lemma 15 (Functional regularity lemma) *For integer m and function $\mathcal{E} : \mathbb{Z}^+ \to (0,1)$, there exists $T = T_{15}(m,\mathcal{E})$ such that the following is true. Given function $f : \mathbb{F}_2^n \to \{0,1\}$ with $n \geq T$, there exist subspaces $H' \leq H \leq \mathbb{F}_2^n$ that satisfy:*

- *Order of H-based partition is $k \geq m$, and order of H'-based partition is $\ell \leq T$.*
- *There are at most $\mathcal{E}(0) \cdot 2^n$ many $g \in \mathbb{F}_2^n$ such that f_H^{+g} is not $\mathcal{E}(0)$-uniform.*
- *For every $g \in \mathbb{F}_2^n$, there are at most $\mathcal{E}(k) \cdot 2^{n-k}$ many $h \in H$ such that $f_{H'}^{+g+h}$ is not $\mathcal{E}(k)$-uniform.*
- *There are at most $\mathcal{E}(0) \cdot 2^n$ many $g \in \mathbb{F}_2^n$ for which there are more than $\mathcal{E}(0) \cdot 2^{n-k}$ many $h \in H$ such that $|\rho(f_H^{+g}) - \rho(f_{H'}^{+g+h})| > \mathcal{E}(0)$.*

Proof: Let us first give an informal overview of the proof. The basic idea is to repeatedly apply Lemma 14, at each step refining the partition obtained in the previous step. At each step, Lemma 14 is applied with a uniformity parameter that depends on the order of the partition obtained in the previous step. We stop when the *index* of the partitions stop increasing substantially. Given a subspace H, the index of the H-based partition is defined to be the variance of the densities in the cosets:

$$\mathsf{ind}(f, H) \overset{\text{def}}{=} \frac{1}{2^n} \sum_{g \in \mathbb{F}_2^n} \rho^2(f_H^{+g})$$

We show that when the indexes of two successive partitions are close, then on average, each coset of the finer partitioning has roughly the same density as the coset of the coarser partitioning it is contained in.

To implement the above ideas, we need the following two claims about the index of partitions. Their proofs are essentially identical to those for the corresponding Lemmas 3.6 and 3.7 respectively in [AFKS00], and so we are a bit brief in the following.

Claim 16 *Given subspace $H \leq \mathbb{F}_2^n$ and function $f : \mathbb{F}_2^n \to \{0, 1\}$, suppose that there are at least $\epsilon 2^n$ many $g \in \mathbb{F}_2^n$ such that $|\rho(f) - \rho(f_H^{+g})| > \epsilon$. Then:*

$$\mathsf{ind}(f, H) > \rho^2(f) + \frac{\epsilon^3}{2}$$

Proof: Observe that the average of $\rho(f_H^{+g})$ over all $g \in \mathbb{F}_2^n$ equals $\rho(f)$. From our assumptions, either there are $\frac{\epsilon}{2}2^n$ many $g \in \mathbb{F}_2^n$ such that $\rho(f) - \rho(f_H^{+g}) > \epsilon$ or there are $\frac{\epsilon}{2}2^n$ many $g \in \mathbb{F}_2^n$ such that $\rho(f) - \rho(f_H^{+g}) < -\epsilon$. For either case, we can use the defect form of the Cauchy-Schwarz inequality to prove our claim. ■

Claim 17 *For function $f : \mathbb{F}_2^n \to \{0,1\}$ and subspaces $H' \leq H \leq \mathbb{F}_2^n$, suppose the H-based partition of order k and its refinement, the H'-based partition, of order ℓ satisfy $\mathsf{ind}(f, H') - \mathsf{ind}(f, H) \leq \frac{\epsilon^4}{2}$ for some ϵ. Then, there are at most $\epsilon 2^n$ many $g \in \mathbb{F}_2^n$ for which there are more than $\epsilon 2^{n-k}$ many $h \in H$ satisfying $|\rho(f_H^{+g}) - \rho(f_{H'}^{+g+h})| > \epsilon$.*

Proof: Suppose that there are $> \epsilon 2^n$ many $g \in \mathbb{F}_2^n$ such that there are $> \epsilon 2^{n-k}$ many $h \in H$ satisfying $|\rho(f_H^{+g}) - \rho(f_{H'}^{+g+h})| > \epsilon$. Use Claim 16 to obtain a contradiction:

$$\begin{aligned}\mathsf{ind}(f, H') = \frac{1}{2^\ell} \sum_{u \in \mathbb{F}_2^n / H'} \rho^2(f_{H'}^{+u}) &= \frac{1}{2^k} \sum_{v \in \mathbb{F}_2^n / H} \frac{1}{2^{\ell-k}} \sum_{h \in H/H'} \rho^2(f_{H'}^{+v+h}) \\ &= \frac{1}{2^k} \sum_{v \in \mathbb{F}_2^n / H} \mathsf{ind}(f_H^{+v}) \\ &> \frac{1}{2^k} \left(\sum_{v \in \mathbb{F}_2^n / H} \rho^2(f_H^{+v}) + \epsilon \cdot 2^k \frac{\epsilon^3}{2} \right) \\ &= \mathsf{ind}(f, H) + \frac{\epsilon^4}{2}\end{aligned}$$

■

Now we have the pieces needed to prove the lemma. We can assume $\mathcal{E}(\cdot)$ is monotone non-increasing. Let $\epsilon = \mathcal{E}(0)$. We define T inductively as follows. Let $T^{(1)} = T_{14}(m, \epsilon)$, and for $i > 1$, let:

$$T^{(i)} = T_{14}\left(T^{(i-1)}, \mathcal{E}\left(T^{(i-1)}\right) \cdot 2^{-T^{(i-1)}}\right)$$

Set $T = T_{15}(m, \mathcal{E}) \overset{\text{def}}{=} T^{(2\epsilon^{-4}+1)}$.

We now show that this choice of T suffices. Given function $f : \mathbb{F}_2^n \to \{0,1\}$, apply Lemma 14 with m and ϵ to get a subspace H_1, and thereafter repeatedly apply it to get a sequence of finer subspaces $H_2, H_3, H_4, \ldots$, with $H_1 \geq H_2 \geq H_3 \geq H_4 \geq \cdots$, by invoking Lemma 14 at each step $i > 1$ with $T^{(i-1)}$ and $\mathcal{E}\left(T^{(i-1)}\right) \cdot 2^{-T^{(i-1)}}$ as the two input parameters. Stop when $\mathsf{ind}(f, H_{i+1}) - \mathsf{ind}(f, H_i) < \frac{\epsilon^4}{2}$. This happens when i is at most $2\epsilon^{-4} + 1$ because the index of any partition is less than 1. Let $H = H_i$ and $H' = H_{i+1}$. It's clear that the codimension k of H at least m and that the codimension ℓ of H' is at most T. The second item in the lemma follows from the uniformity guarantee of Lemma 14 and from the fact that $\mathcal{E}(T^{(i-1)}) < \mathcal{E}(0)$. For the third, note that Lemma 14 guarantees that there are at most $\mathcal{E}(k)2^{-k}2^n = \mathcal{E}(k)2^{n-k}$ values of $g \in \mathbb{F}_2^n$ such that $f_{H'}^{+g}$ is not $(\mathcal{E}(k)2^{-k})$-uniform and, hence, not $\mathcal{E}(k)$-uniform. So, clearly, there are at most so many g contained in any coset of H. Finally, the fourth item follows from Claim 17. This completes the proof of Lemma 15. ■

We use Lemma 15 in two main ways. For one of them, we use the lemma directly. For the other, we use the following simple but extremely useful corollary which allows us to say that there are many cosets in a partitioning which, on the one hand, are *all* uniform, and on the other hand, are arranged in an algebraically nice structure.

Corollary 18 *For every m and $\mathcal{E} : \mathbb{Z}^+ \to (0,1)$, there exist $T = T_{18}(m, \mathcal{E})$ and $\delta = \delta_{18}(m, \mathcal{E})$ such that the following is true. Given function $f : \mathbb{F}_2^n \to \{0,1\}$ with $n \geq T$, there exist subspaces $H' \leq H \leq \mathbb{F}_2^n$ and an injective linear map $I : \mathbb{F}_2^n/H \to \mathbb{F}_2^n/H'$ such that:*

- *The H-based partition is of order k, where $m \leq k \leq T$. Additionally, $|H'| \geq \delta 2^n$.*
- *For each $u \in \mathbb{F}_2^n/H$, $I(u) + H'$ lies inside the coset $u + H$. Note that $I(\mathbf{0}) = 0$ since I is linear.*
- *For every nonzero $u \in \mathbb{F}_2^n/H$, the set $f_{H'}^{+I(u)}$ is $\mathcal{E}(k)$-uniform.*
- *There are at most $\mathcal{E}(0)2^n$ many $g \in \mathbb{F}_2^n$ for which $|\rho(f_H^{+g}) - \rho(f_{H'}^{+I(u)})| > \mathcal{E}(0)$ where $u = g \pmod{H}$.*

Proof: We can assume $\mathcal{E}$ is a nonincreasing function. Denote $\mathcal{E}(0)$ as ϵ, and set $\mathcal{E}'(r) = \min(\mathcal{E}(r), \frac{\epsilon}{6}, \frac{1}{2^{r+1}})$. We will show that $T = T_{18}(m, \mathcal{E}) \overset{\text{def}}{=} T_{15}(m, \mathcal{E}')$ and $\delta = \delta_{18}(m, \mathcal{E}) \overset{\text{def}}{=} 1/2^T$ suffice for our proof.

Apply Theorem 15 with m and the function $\mathcal{E}'$ as inputs. Let H and H' be the subspaces obtained there, for the given $f : \mathbb{F}_2^n \to \{0,1\}$. We find I satisfying the conditions of the claim exists using the probabilistic method.

Fix k linearly independent elements $u_1, \ldots, u_k \in \mathbb{F}_2^n/H$ (viewing $\mathbb{F}_2^n/H$ as a vector space over $\mathbb{F}_2$). For every $i \in [k]$, choose independently and uniformly at random an element v from H/H' and let $I(u_i)$ equal $u_i + v + H'$. The value of I over the rest of $\mathbb{F}_2^n/H$ is determined by linearity, as the u_i's form a basis for $\mathbb{F}_2^n/H$. It's immediate that $I(u) + H'$ lies inside $u + H$ for every $u \in \mathbb{F}_2^n/H$.

Observe that unless $u = \mathbf{0}$, each $I(u) + h'$ is uniformly distributed among the cosets of H' lying in $u + H$. Hence, for any nonzero u, the probability that $f_{H'}^{+I(u)}$ is not $\mathcal{E}(k)$-uniform is at most $1/2^{k+1}$, by our choice of parameters. Applying the union bound, the probability that there exists nonzero $u \in \mathbb{F}_2^n/H$ such that $f_{H'}^{+I(u)}$ is not $\mathcal{E}(k)$-uniform is at most $1/2$. Also, the expected number of $g \in \mathbb{F}_2^n$, with $u = g \pmod{H}$, for which $|\rho(f_H^{+g}) - \rho(f_{H'}^{+I(u)})| > \epsilon$ is at most $\frac{\epsilon}{6}2^n + \frac{\epsilon}{6}2^n + 1 \leq \frac{\epsilon}{2}2^n$, and hence by the Markov inequality, with probability at least $\frac{1}{2}$, the number of $g \in \mathbb{F}_2^n$ satisfying this condition is at most $\epsilon 2^n$. Therefore, there must exist a choice of I making both the third and fourth claims true. ■

The next lemma is in a similar spirit to Corollary 18. It also obtains a set of uniform cosets which are structured algebraically, but in this case, all of them are contained inside the same subspace.

Lemma 19 *For every positive integer d and $\gamma \in (0,1)$, there exists $\delta = \delta_{19}(d, \gamma)$ such that the following is true. Given $f : \mathbb{F}_2^n \to \{0,1\}$, there exists a subspace $H \leq \mathbb{F}_2^n$ and a subspace K of dimension d in the quotient space $\mathbb{F}_2^n/H$ with the following properties:*

- *$|H| \geq \delta 2^n$.*

- *For every nonzero $u \in K$, f_H^{+u} is γ-uniform.*

- *Either $\rho(f_H^{+u}) \geq \frac{1}{2}$ for every nonzero $u \in K$ or $\rho(f_H^{+u}) < \frac{1}{2}$ for every nonzero $u \in K$.*

We need a different set of tools to prove this lemma. Specifically, we use linear algebraic variants of the classic theorems of Turán and Ramsey. We note that the (classic) Turán and Ramsey Theorems are key tools in many applications of the graph regularity lemma, for example in the well known bound on the Ramsey numbers of bounded degree graphs [CRSW83]. Hence, the variants that we use of these classic results may be useful in other applications of Green's regularity lemma.

Proposition 20 (Turán theorem for subspaces) *For positive integers n, if S is a subset of $\mathbb{F}_2^n$ with density greater than $1-\frac{1}{2^{d-1}}$, then there exists a subspace $H \leq \mathbb{F}_2^n$ of dimension d such that $H-\{0\}$ is contained in S. Moreover, there is a subset of $\mathbb{F}_2^n$ with density $\left(1-\frac{1}{2^{d-1}}\right)$ which does not contain $H-\{0\}$ for any subspace $H \leq \mathbb{F}_2^n$.*

Proof: Let $S \subseteq \mathbb{F}_2^n$ be a maximal set that does not contain $H-\{0\}$ for any d-dimensional subspace H. Since S is maximal, it must contain $K-\{0\}$ for some $(d-1)$-dimensional subspace K (if not, we can simply add it to S without introducing points of $H-\{0\}$ for any d-dimensional subspace H). Let K' be an $(n-d+1)$-dimension subspace that intersects K only at $\{0\}$.

Now, observe that for any nonzero $\alpha \in K'$, at least one of the elements of $\{\alpha+k : k \in K\}$ must not belong to S. Otherwise, S would contain $(K-\{0\}) \cup \{\alpha+k : k \in K\} = H-\{0\}$ for a d-dimensional subpace $H = \text{span}(K \cup \{\alpha\})$, contradicting our assumption for S. Thus, we can upper-bound the number of points in S by:

$$|S| \leq |K'-\{0\}|\cdot(|K|-1)+|K-\{0\}| = (2^{n-d+1}-1)\cdot(2^{d-1}-1)+(2^{d-1}-1) = 2^n-2^{n-d+1}$$

To see that the above bound is tight, let $S = \mathbb{F}_2^n - K'$ for any $(d-1)$-dimensional subspace $K \leq \mathbb{F}_2^n$ and K' as above. It is easy to check that this S does not contain $H-\{0\}$ for any $H \leq \mathbb{F}_2^n$ with $\dim(H) = d$. ■

Theorem 21 (Ramsey theorem for subspaces) *For every positive integer d, there exists $N = N_{21}(d)$ such that for any subset $S \subseteq \mathbb{F}_2^N$, there exists a subspace $H \leq \mathbb{F}_2^N$ of dimension d such that $H-\{0\}$ is contained either in S or in $\bar{S}$.*

Proof: We will show a stronger statement, which we describe in the following lemma.

Lemma 22 *For every positive integer d_1, d_2, there exists $N(d_1, d_2)$ such that for any subset $S \subseteq \mathbb{F}_2^{N(d_1,d_2)}$, either there exists a subspace $H_1 \leq \mathbb{F}_2^{N(d_1,d_2)}$ of dimension d_1 such that $H_1 - \{0\}$ is contained in S or there exists a subspace $H_2 \leq \mathbb{F}_2^{N(d_1,d_2)}$ of dimension d_2 such that $H_2 - \{0\}$ is contained in $\bar{S}$.*

One can immediately deduce the statement of the theorem by taking $d = d_1 = d_2$ in Lemma 22. To prove Lemma 22 we first prove the following helpful result. For a subspace $H \leq \mathbb{F}_2^n$ we say that an affine subspace $a + H$ is *strict* if $a \in \mathbb{F}_2^n/H - \{0\}$.

Lemma 23 *For every positive integer d, there exists $N_a = N_a(d)$ such that for any subset $S \subseteq \mathbb{F}_2^{N_a}$, there exists a strict affine subspace $A \leq \mathbb{F}_2^{N_a}$ of dimension d such that A is contained either in S or in $\bar{S}$.*

Proof: Notice that $N_a(1) = 1$. Assume, by induction that the lemma holds for dimension $d-1$, and let $N_a(d) = 2^{N_a(d-1)+1} + N_a(d-1)$. Let $S \subseteq \mathbb{F}_2^{N_a(d)}$ be an arbitrary set, let $H = \mathbb{F}_2^{N_a(d-1)}$, and $H' = \mathbb{F}_2^{N_a(d)}/H$. Notice that $|H'| = 2^{2^{N_a(d-1)}+1}$. For each $c \in H' - \{0\}$ consider the set $f_H^{+c} \subset H$. Since there are $2^{2^{N_a(d-1)+1}} - 1$ possible such sets, and each set has size at most $2^{N_a(d-1)}$ it follows that there exists $c_1 \neq c_2 \in H' - \{0\}$ such that $f_H^{+c_1} = f_H^{+c_2}$. By the induction hypothesis, either $f_H^{+c_1}$ or its complement contains a $d-1$ dimensional affine subspace. Assume w.l.o.g. that $f_H^{+c_1}$ contains an affine subspace $\alpha + f_{d-1}$ of dimension $d-1$ (otherwise replace S by $\bar{S}$), for some $\alpha \in H - f_{d-1}$. Then the affine subspaces $\alpha + c_1 + f_{d-1}$ and $\alpha + c_2 + f_{d-1}$ are both contained in S. Let $A_d = (\alpha + c_1 + f_{d-1}) \cup (\alpha + c_2 + f_{d-1}) \subset S$. To conclude the proof, notice that $A_d = \alpha + c_1 + \mathrm{span}(c_2 - c_1, f_{d-1})$ is a strict affine subspace of dimension d, since $\alpha \neq c_1$ and $c_2 - c_1 \notin f_{d-1}$. ■

Proof of Lemma 22: The proof follows by induction on d_1 and d_2, with the base cases $N(0,1) = N(1,0 = 1$. Assume that there exists $N(d_1 - 1, d_2)$ and $N(d_1, d_2 - 1)$

satisfying the conditions of the lemma. Define

$$N(d_1, d_2) = N_a(\max(N(d_1 - 1, d_2), N(d_1, d_2 - 1))),$$

where $N_a(d)$ is the quantity defined in Lemma 23. We show that for any arbitrary set $S \subseteq \mathbb{F}_2^{N(d_1,d_2)}$ either it contains a subspace of dimension d_1 (except 0) or its complement contains a subspace of dimension d_2 (except 0). Suppose $N(d_1 - 1, d_2) \geq N(d_1, d_2 - 1)$. By definition and by Lemma 23, there exists a strict affine subspace $A \subseteq \mathbb{F}_2^{N(d_1,d_2)}$ such that $A = a + H \subseteq S$ or $A \subseteq \bar{S}$ (where H is the subspace underlining A). Assume for now that the former holds. Since $H \cap S \subseteq \mathbb{F}_2^{N(d_1-1,d_2)}$, by the induction hypothesis, either $H \cap S$ contains a subspace of dimension $d_1 - 1$ or $H - S$ contains a subspace of dimension d_2, in which case we are done. If $H \cap S$ contains a subspace $f_{d_1-1} - \{0\}$ of dimension $d_1 - 1$, then define $f_{d_1} = f_{d_1-1} \cup a + f_{d_1-1} = \text{span}(a, f_{d_1-1})$. Clearly $f_{d_1} \in S$ and it has dimension d_1, which completes the proof of this case. It remains to deal with the case when $A \subseteq \bar{S}$. Since $N(d_1 - 1, d_2) \geq N(d_1, d_2 - 1)$, there exists another affine subspace $A' = a' + H' \subset A \subseteq \bar{S}$ of dimension $N(d_1, d_2 - 1)$. Again, by the induction hypothesis, the set $H' \cap S$ either contains a subspace of dimension d_1, in which case we are done, or $H' - S$ contains a subspace f_{d_2-1} of dimension $d_2 - 1$. In the latter case define $f_{d_2} = f_{d_2-1} \cup a' + f_{d_2-1} = \text{span}(a', f_{d_2-1})$. Finally, notice that $f_{d_2} \in \bar{S}$ and it has dimension d_2. ■

This concludes the proof of Theorem 21.■

Given these results, Lemma 19 follows fairly readily.

Proof of Lemma 19: Set $\delta = \delta_{19}(d, \gamma) \overset{\text{def}}{=} 2^{-T_{14}(r, \min(2^{-r-2}, \gamma))}$ with $r = N_{21}(d)$. Given $f : \mathbb{F}_2^n \to \{0, 1\}$, apply Lemma 14 with inputs r and $\min(2^{-r-2}, \gamma)$ to obtain a subspace H such that restrictions of S to at most 2^{-r-2} fraction of the cosets of the H-based partition are not γ-uniform. Using Proposition 20, there exists a subspace $L \leq \mathbb{F}_2^n / H$ of dimension r such that for every nonzero $u \in L$, the set f_H^{+u} is γ-uniform. Furthermore, since L is of dimension $N_{21}(d)$, by Theorem 21, there exists a subspace $K \leq L \leq \mathbb{F}_2^n / H$ satisfying the final condition of the lemma. ■

Chapter 3

Testability of Non-Monotone Properties

In this chapter, we prove the result (Theorem 8) that properties characterized by infinitely many forbidden induced equations are testable. To begin, let us fix some notation. Given a matrix M over $\mathbb{F}_2$ of size m-by-k, a string $\sigma \in \{0,1\}^k$, and a function $f : \mathbb{F}_2^n \to \{0,1\}$, if there exists $x = (x_1, \ldots, x_k) \in (\mathbb{F}_2^n)^k$ such that $Mx = 0$ and $f(x_i) = \sigma_i$ for all $i \in [k]$, we say that f *induces* (M, σ) *at* x and denote this by $(M, \sigma) \mapsto f$.

The following theorem is the core of the proof of Theorem 8.

Theorem 24 *For every infinite family of equations* $\mathcal{F} = \{(E^1, \sigma^1), (E^2, \sigma^2), \ldots, (E^i, \sigma^i), \ldots\}$ *with each* E^i *being a row vector* $[1\ 1\ \cdots\ 1]$ *of size* k_i *and* $\sigma^i \in \{0,1\}^{k_i}$ *a* k_i*-tuple, there are functions* $N_{\mathcal{F}}(\cdot)$, $k_{\mathcal{F}}(\cdot)$ *and* $\delta_{\mathcal{F}}(\cdot)$ *such that the following is true for any* $\epsilon \in (0,1)$. *If a function* $f : \mathbb{F}_2^n \to \{0,1\}$ *with* $n > N_{\mathcal{F}}(\epsilon)$ *is* ϵ*-far from being* $\mathcal{F}$*-free, then* f *induces* $\delta \cdot 2^{n(k_i-1)}$ *many copies of some* (E^i, σ^i), *where* $k_i \leq k_{\mathcal{F}}(\epsilon)$ *and* $\delta \geq \delta_{\mathcal{F}}(\epsilon)$.

Armed with Theorem 24 our main theorem becomes now a straightforward consequence.

Proof of Theorem 8: Theorem 24 allows us to devise the following tester T for $\mathcal{F}$-freeness. T, given input $f : \mathbb{F}_2^n \to \{0,1\}$, first checks if $n \leq N_{\mathcal{F}}(\epsilon)$, and in this case, it queries f on the entire domain and decides accordingly. Otherwise, T selects independently and uniformly at random a set D of d elements from $\mathbb{F}_2^n$, where we will specify d at the end of the argument. It then queries all points in the linear subspace spanned by the elements of D and then accepts or rejects based on whether f restricted to this subspace is $\mathcal{F}$-free or not.

Clearly, if f is $\mathcal{F}$-free, then the tester always accepts because the property is subspace-hereditary. Also, if $n \leq N_{\mathcal{F}}(\epsilon)$, then the correctness of the algorithm is trivial. So, suppose f is ϵ-far from $\mathcal{F}$-free and $n > N_{\mathcal{F}}(\epsilon)$. For the M^i guaranteed to exist from Theorem 24, let K be a $k_i \times c$ matrix over $\mathbb{F}_2$, where $c = k_i - m_i \leq k_{\mathcal{F}}(\epsilon)$, such that the columns of K form a basis for the kernel of M^i. Then, every $y = (y_1, \dots, y_c) \in (\mathbb{F}_2^n)^c$ yields a distinct vector $x = (x_1, \dots, x_k) \in (\mathbb{F}_2^n)^k$ formed by letting $x = Ky$ that satisfies $M^i x = M^i K y = 0$. Therefore, because of Theorem 24, the probability that uniformly chosen $y_1, \cdots, y_c \in \mathbb{F}_2^n$ yield $x = (x_1, \dots, x_k)$ such that f induces (M^i, σ^i) at x is at least $\delta_{\mathcal{F}}(\epsilon)$. The probability that D does not contain such $y_1, \dots, y_c$ is at most $(1-\delta)^{d/c} < e^{\delta_{\mathcal{F}}(\epsilon)d/c} < 1/3$ if we choose $d = O(c/\delta_{\mathcal{F}}(\epsilon)) = O(k_{\mathcal{F}}(\epsilon)/\delta_{\mathcal{F}}(\epsilon))$. Thus with probability at least 2/3, $\mathsf{span}(D)$ contains $x_1, \dots, x_k$ such that f induces (M^i, σ^i) at $x = (x_1, \dots, x_k)$, making the tester reject. ■

To start the proof of Theorem 24, let us relate pseudorandomness (uniformity) of a function to the number of solutions to a single equation induced by it. Similar and more general statements have been shown previously, but we need only the following claim for what follows.

Lemma 25 (Counting Lemma) *For every $\eta \in (0,1)$ and integer $k > 2$, there exist $\gamma = \gamma_{25}(\eta, k)$ and $\delta = \delta_{25}(\eta, k)$ such that the following is true. Suppose E is the row vector $[1\ 1 \cdots 1]$ of size k, $\sigma \in \{0,1\}^k$ is a tuple, H is a subspace of $\mathbb{F}_2^n$, and $f : \mathbb{F}_2^n \to \{0,1\}$ is a function. Furthermore, suppose there are k not necessarily distinct elements $u_1, \dots, u_k \in \mathbb{F}_2^n/H$ such that $Mu = 0$ where $u = (u_1, \dots, u_k)$, $f_H^{+u_i} : H \to \{0,1\}$ is γ-uniform for all $i \in [k]$, and $\rho(f_H^{+u_i})$ is at least η if $\sigma(i) = 1$ and at most $1-\eta$ if $\sigma(i) = 0$ for all $i \in [k]$. Then, there are at least $\delta|H|^{k-1}$ many k-tuples $x = (x_1, x_2, \dots, x_k)$, with each $x_i \in u_i + H$, such that f induces (E, σ) at x.*

Proof: Fix $v_1 \in u_1 + H$, $v_2 \in u_2 + H, \ldots, v_k \in u_k + H$ such that $v_1 + v_2 + \cdots + v_k = 0$; there exist such v_i's because $u_1 + u_2 + \cdots + u_k = 0$ in the quotient space $\mathbb{F}_2^n/H$. Define Boolean functions $f_1, \ldots, f_k : H \to \{0, 1\}$ so that $f_i(x) = f_H^{+v_i}(x)$ if $\sigma(i) = 1$ and $f_i(x) = 1 - f_H^{+v_i}(x)$ if $\sigma(i) = 0$. By our assumptions, $\widehat{f_i}(0) \geq \eta$ and each $|\widehat{f_i}(\alpha)| < \gamma$ for all $\alpha \neq 0$. Now, observe that, using γ-uniformity and Cauchy-Schwarz, we have:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_1,\ldots,x_{k-1} \in H} [f_1(x_1) f_2(x_2) \cdots f_{k-1}(x_{k-1}) f_k(x_1 + x_2 + \cdots + x_{k-1})] \\
&\quad = \sum_{\alpha \in H^*} \widehat{f_1}(\alpha) \widehat{f_2}(\alpha) \cdots \widehat{f_k}(\alpha) \\
&\quad \geq \eta^k - \sum_{\alpha \neq 0} |\widehat{f_1}(\alpha) \widehat{f_2}(\alpha) \cdots \widehat{f_k}(\alpha)| \\
&\quad \geq \eta^k - \gamma^{k-2} \sqrt{\sum_{\alpha} |\widehat{f_1}(\alpha)|^2} \sqrt{\sum_{\alpha} |\widehat{f_2}(\alpha)|^2} \\
&\quad \geq \eta^k - \gamma^{k-2}
\end{aligned}$$

Setting $\gamma = \gamma_{25}(\eta, k) \stackrel{\text{def}}{=} (\eta^k/2)^{1/(k-2)}$ makes the above expectation at least $\eta^k/2$. Now note that every $x_1, \ldots, x_k \in H$ such that $x_1 + \cdots + x_k = 0$ gives $y = (y_1, \ldots, y_k)$, where $y_i = v_i + x_i$ for all $i \in [k]$, such that f induces (E, σ) at y. Thus, we have from above that there are at least $\delta |H|^{k-1}$ many such y's, where $\delta = \delta_{25}(\eta, k) \stackrel{\text{def}}{=} \eta^k/2$. ■

3.0.1 Proof of Theorem 24

Before seeing the full technical details of the proof of Theorem 24 we proceed with a more intuitive overview.

In light of Lemma 25, our strategy will be to partition the domain into uniform cosets, using Green's regularity lemma (Lemma 14) in some fashion, and then to use the above counting lemma to count the number of induced solutions to some equation in $\mathcal{F}$. But one issue that immediately arises is that, because $\mathcal{F}$ is an infinite family of equations, we do not know the size of the equation we would want the input function to induce. Since Lemma 25 needs different uniformity parameters to count equations of different lengths, it is not *a priori* clear how to set the uniformity parameter in applying the regularity lemma. (If $\mathcal{F}$ was finite, one could set the uniformity parameter to correspond to the size of the largest equation in $\mathcal{F}$.)

To handle the infinite case, our basic approach will be to classify the input function into one of a finite set of classes. For each such class c, there will be an associated number k_c such that it is guaranteed that any function classified as c must induce an equation in $\mathcal{F}$ of size at most k_c. If there is such a classification scheme, then we know that *any* input function must induce an equation of size at most $\max_c k_c$. How do we perform this classification? We use the regularity lemma. Consider the following idealized situation. Fix an integer r. Suppose we could modify the input $f : \mathbb{F}_2^n \to \{0,1\}$ at a small fraction of the domain to get a function $F : \mathbb{F}_2^n \to \{0,1\}$ and then could apply Lemma 14 to get a partition of order r so that the restrictions of F to each coset was exactly 0-uniform. F is then a constant function (either 0 or 1) on each of the 2^r cosets, and so, we can classify F by a Boolean function $\mu : \mathbb{F}_2^r \to \{0,1\}$ where $\mu(x)$ is the value of F on the coset corresponding to x. Notice that there are only finitely many such μ's. Since F differs from f at only a small fraction of the domain and since f is far from $\mathcal{F}$-free, F must also induce some equation in $\mathcal{F}$. Then, for every such μ and corresponding F, there is a smallest equation in $\mathcal{F}$ that is induced by F. We can let $\Psi_{\mathcal{F}}(r)$ be the maximum over all such μ of the size of the smallest equation in $\mathcal{F}$ that is induced by the F corresponding to μ. We then might hope that this function $\Psi_{\mathcal{F}}(\cdot)$ can be used to tune the uniformity parameter by using the functional variant of the regularity lemma (Lemma 15).

There are a couple of caveats. First, we will not be able to get the restrictions to every coset to look perfectly uniform. Second, if F induces solutions to an equation, it does not necessarily follow that f also does. To get around the first problem, we use the fact that Lemma 25 is not very restrictive on the density conditions. We think of the uniform cosets which have density neither too close to 0 nor 1 as "wildcard" cosets at which both the restriction of f and its complement behave pseudorandomly and have non-negligible density. Thus, the μ in the above paragraph will map into $\{0,1,*\}^r$, where a '$*$' denotes a wildcard coset. For the second problem, note that it is not really a problem if $\mathcal{F}$-freeness is known to be monotone. In this case, F inducing an equation automatically means f also induces an equation, if we obtained F by removing elements from the support of f. For induced freeness properties, though, this is not the case. Using ideas from [AFKS00] and the tools from Chapter 2, we structure the modifications from f to F in such a way so as to force f to induce solutions of an equation if F induces a solution to the same equation. We elaborate much more on this issue during the course of the proof.

The observations described in the proof sketch above motivate the following definitions.

Definition 26 *Given function $\mu : \mathbb{F}_2^r \to \{0,1,*\}$, a m-by-k matrix M and a k-tuple $\sigma \in \{0,1\}^k$, suppose there exist $x_1, \ldots, x_k \in \mathbb{F}_2^r$ such that $Mx = 0$ where $x = (x_1, \ldots, x_k)$, and for every $i \in [k]$, $\mu(x_i)$ equals either $\sigma(i)$ or $*$. In this case, we say μ* partially induces *(M, σ)* at *x and denote this by $(M, \sigma) \mapsto_* \mu$.*

Definition 27 *Given a positive integer r and an infinite family of systems of equations $\mathcal{F} = \{(M^1, \sigma^1), (M^2, \sigma^2), \ldots\}$ with M^i being a m_i-by-k_i matrix of rank m_i and $\sigma^i \in \{0,1\}^{k_i}$ a k_i-tuple, define $\mathcal{F}_r$ to be the set of functions $\mu : \mathbb{F}_2^r \to \{0,1,*\}$ such that there exists some $(M^i, \sigma^i) \in \mathcal{F}$ with $(M^i, \sigma^i) \mapsto_* \mu$. Given $\mathcal{F}$ and integer r for which $\mathcal{F}_r \neq \emptyset$, define the following function:*

$$\Psi_{\mathcal{F}}(r) \stackrel{\text{def}}{=} \max_{\mu \in \mathcal{F}_r} \min_{\{(M^i,\sigma^i):(M^i,\sigma^i)\mapsto_* \mu\}} k_i$$

Proof of of Theorem 24: Define the function $\mathcal{E}$ by setting $\mathcal{E}(0) = \epsilon/8$ and for any $r > 0$:

$$\mathcal{E}(r) = \delta_{19}(\Psi_{\mathcal{F}}(r), \gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(r))) \cdot \min(\epsilon/8, \gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(r)))$$

Additionally, let $T(\epsilon) = T_{18}(8/\epsilon, \mathcal{E})$, and set $N_{\mathcal{F}}(\epsilon) \stackrel{\text{def}}{=} T(\epsilon)$. Also, set $k_{\mathcal{F}}(\epsilon) \stackrel{\text{def}}{=} \Psi_{\mathcal{F}}(T(\epsilon))$ and

$$\delta_{\mathcal{F}}(\epsilon) \stackrel{\text{def}}{=} (\delta_{19}(\Psi_{\mathcal{F}}(r), \gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(r))) \cdot \delta_{18}(8/\epsilon, \mathcal{E}))^{\Psi_{\mathcal{F}}(\epsilon)} \cdot \delta_{25}(\epsilon/8, \Psi_{\mathcal{F}}(T(\epsilon)))$$

We proceed to show that these parameter settings suffice.

Suppose we are given input function $f : \mathbb{F}_2^n \to \{0,1\}$ with $n > N_{\mathcal{F}}(\epsilon) = T_{18}(8/\epsilon, \mathcal{E})$. As mentioned in the paragraphs preceding the proof, our strategy will be to partition the domain in such a way that we can find cosets in the partition satisfying the conditions of Lemma 25. To this end, we apply Corollary 18 with $8/\epsilon$ and the function $\mathcal{E}$ as inputs. This yields subspaces $H' \leq H \leq \mathbb{F}_2^n$ and linear map $I : \mathbb{F}_2^n/H \to \mathbb{F}_2^n/H'$, where the order of the H-based partition, which we denote ℓ, satisfies $8/\epsilon \leq \ell \leq T_{18}(8/\epsilon, \mathcal{E})$. Recall that $I(u) + H'$ is contained in $u + H$ for every coset $u \in \mathbb{F}_2^n/H$. Observe that from our setting of parameters, we have that for every *nonzero* $u \in \mathbb{F}_2^n/H$, the restriction $f_{H'}^{+I(u)}$ is $(\delta_{19}(\Psi_{\mathcal{F}}(\ell), \gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(\ell))) \cdot \gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(\ell)))$-uniform.

But we have no such uniformity guarantee for $f_{H'}^{+0}$. This would not pose an obstacle if $\mathcal{F}$-freeness were a monotone property (i.e., if each σ^i equalled 1^{k_i}). If that were the case, we could simply make f zero on all elements of H. Since H is still only a small fraction of the domain, the modified function would still be far from $\mathcal{F}$-free, and we would be guaranteed that remaining solutions to equations of $\mathcal{F}$ induced by f would only use elements from cosets of H for which we have a guarantee about the corresponding coset of H'. But if $\mathcal{F}$-freeness is not monotone, such a scheme would not work, since it's not clear at all how to change the value of f on H so that any solution to an equation from $\mathcal{F}$ would only involve elements from nonzero shifts of H.

To resolve this issue, we further partition H' to find affine subspaces within H' on which we can guarantee that the restriction of f is uniform. The idea is that once we know that there is a solution involving H, we are going to look not at H' itself but at the smaller affine subspace within H' on which f is known to be uniform. Specifically, apply Lemma 19 to $f_{H'}^{+0}$ with input parameters $\Psi_{\mathcal{F}}(\ell)$ and $\gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(\ell))$. This yields subspaces H'' and W, both of which contained in H', such that $|H''| \geq \delta_{19}(\Psi_{\mathcal{F}}(\ell), \gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(\ell)))|H'|$ and $\dim(W/H'') = \Psi_{\mathcal{F}}(\ell)$. We further know that for every nonzero $v \in W/H''$, the function $f_{H''}^{+v}$ is $\gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(\ell))$-uniform.

Now, let's "copy" W on cosets $I(u) + H'$ for every $u \in \mathbb{F}_2^n/H$. We do this by specifying[1] another linear map $J : \mathbb{F}_2^n/H \to \mathbb{F}_2^n$ so that for any $u \in \mathbb{F}_2^n/H$, the coset[2] $J(u)+W$ lies inside $I(u)+H'$ (which itself lies inside $u+H$). Each coset $J(u)+W$ also has an H''-based partition of order $\Psi_{\mathcal{F}}(\ell)$, just as W itself does. Consider $v \in \mathbb{F}_2^n/H''$ such that $v + H''$ lies inside $J(u) + W$ for some nonzero $u \in \mathbb{F}_2^n/H$. Then, because we know the uniformity of $f_{H'}^{+I(u)}$ and we have a lower bound on the size of H'', it follows from Lemma 13 that $f_{H''}^{+v}$ is $\gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(\ell))$-uniform. Thus, for any nonzero $v \in \mathbb{F}_2^n/H''$ such that $v + H''$ lies inside $J(u) + W$ for some $u \in \mathbb{F}_2^n/H$, it is the case that $f_{H''}^{+v}$ is $\gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(\ell))$-uniform.

In the following, we will show how to apply Lemma 25 on some of these cosets $f_{H''}^{+v}$. We have already argued their uniformity above. We now need to make sure that the pattern of their densities allow Lemma 25 to infer many induced copies of some equation in $\mathcal{F}$. To this end, we modify f to construct a new function $F : \mathbb{F}_2^n \to \{0,1\}$. F is initially identical to f on the entire domain, but is then modified in the following order:

1. For every nonzero $u \in \mathbb{F}_2^n/H$ such that $|\rho(F_H^{+u}) - \rho(F_{H'}^{+I(u)})| > \epsilon/8$, do the following. If $\rho(F_{H'}^{+I(u)}) \geq \frac{1}{2}$, then make $F(x) = 1$ on all $x \in u + H$. Otherwise, make $F(x) = 0$ on all $x \in u + H$.

2. For every nonzero $u \in \mathbb{F}_2^n/H$ such that $\rho(F_{H'}^{+I(u)}) > 1 - \epsilon/4$, make $F(x) = 1$ for all $x \in u + H$. On the other hand, if $u \in \mathbb{F}_2^n/H$ is nonzero and $\rho(F_{H'}^{+I(u)}) < \epsilon/4$, make $F(x) = 0$ for all $x \in u + H$.

3. If for all nonzero $v \in W/H''$, $\rho(F_{H''}^{+v}) \geq \frac{1}{2}$, then make $F(x) = 1$ for all $x \in H$. On the other hand, if for all nonzero $v \in W/H''$, $\rho(F_{H''}^{+v}) < \frac{1}{2}$, them make $F(x) = 0$ for all $x \in H$. (One of these two conditions is true by construction.)

The following observation shows that F also must induce solutions to some equation from $\mathcal{F}$, since F is ϵ-far from being $\mathcal{F}$-free.

[1] One way to accomplish this is to define J appropriately for ℓ linearly independent elements of $\mathbb{F}_2^n/H$ and then use linearity to define it on all of $\mathbb{F}_2^n/H$.

[2] Note that the image of J is to elements of $\mathbb{F}_2^n$ and not $\mathbb{F}_2^n/W$, even though we think of the output as denoting a coset of W. The reason is that we will find it convenient to fix the shift and not make it modulo W.

Claim 28 *F is ϵ-close to f.*

Proof: We count the number of elements added or removed at each step of the modification. For the first step, Corollary 18 guarantees that at most $\mathcal{E}(0) \leq \epsilon/8$ fraction of cosets $u + H$ have $|\rho(F_H^{+u}) - \rho(F_{H'}^{+I(u)})| > \epsilon/8$. So, F is modified in at most $\frac{\epsilon}{8}2^n$ locations in the first step. In the second step, if $1 > \rho(F_{H'}^{+I(u)}) > 1 - \epsilon/4$, then $\rho(F_H^{+u}) > 1 - 3\epsilon/8$ because the first step has been completed. Similarly, if $0 < \rho(F_{H'}^{+I(u)}) < \epsilon/4$, then $\rho(F_H^{+u}) < 3\epsilon/8$. So, F is modified in at most $\frac{3\epsilon}{4}2^n$ locations in the second step. As for the third step, H contains at most $2^{n-\ell} \leq 2^{n-8/\epsilon} < \frac{\epsilon}{8}2^n$ elements for $\epsilon \in (0, 1)$. So, in all, F is ϵ-close to f. ■

Now, we define a function $\mu : \mathbb{F}_2^\ell \to \{0, 1, *\}$ based on F and argue that it must partially induce solutions to some equation in $\mathcal{F}$. Since H is of codimension ℓ, $\mathbb{F}_2^n/H \cong \mathbb{F}_2^\ell$ and we identify the two spaces. For $u \in \mathbb{F}_2^n/H$, if $F(x) = 1$ on the entire coset $u + H$, let $\mu(u) = 1$. On the other hand, if $F(x) = 0$ on the entire coset $u + H$, then let $\mu(u) = 0$. In any other case, let $\mu(u) = *$.

Claim 29 *There exists $(E^i, \sigma^i) \in \mathcal{F}$ such that $(E^i, \sigma^i) \mapsto_* \mu$.*

Proof: As already observed, F is not $\mathcal{F}$-free, and let $(E^i, \sigma^i) \in \mathcal{F}$ be some equation whose solution is induced by F at $(x_1, \ldots, x_{k_i}) \in (\mathbb{F}_2^n)^{k_i}$. Now let $y = (y_1, \ldots, y_{k_i}) \in (\mathbb{F}_2^\ell)^{k_i}$ where for each $j \in [k_i]$, $y_j = x_j \pmod{H}$. It's clear that $E^i y = 0$. To argue that F partially induces μ at y, suppose for contradiction that for some $j \in [k_i]$, $\mu(y_j) = 0$ but $\sigma_j^i = 1$. But if $\mu(y_j) = 0$, then F is the constant function 0 on all of $y_j + H$, contradicting the existence of $x_j \in y_j + H$ with $F(x) = 1$. We get a similar contradiction if $\mu(y_j) = 1$ but $\sigma_j^i = 0$. ■

Using Definition 27, we immediately get that there is some $(E^i, \sigma^i) \in \mathcal{F}$ of size at most $\Psi_{\mathcal{F}}(\ell)$ such that $(E^i, \sigma^i) \mapsto_* \mu$. Fix $x_1, \ldots, x_{k_i} \in \mathbb{F}_2^n$ where F induces (E^i, σ^i), and as in the above proof, let $y_1, \ldots, y_{k_i} \in \mathbb{F}_2^n/H$ where each $y_j = x_j \pmod{H}$. Also, pick $k_i - 1$ linearly independent elements $\tilde{v}_1, \ldots, \tilde{v}_{k_i-1}$ from W/H'', which is possible since $\dim(W/H'') = \Psi_{\mathcal{F}}(\ell) > k_i - 1$, and choose $v_1 \in \tilde{v}_1 +$

$H'', \ldots, v_{k_i - 1} \in \tilde{v}_{k_i-1} + H''$ such that $v_1, \ldots, v_{k_i}$ are linearly independent. Additionally set $v_{k_i} = \sum_{j=1}^{k_i-1} v_j$. Notice that none of $v_1, \ldots, v_{k_i}$ are in H''. Now, consider the sets $f_{H''}^{+J(y_1)+v_1}, f_{H''}^{+J(y_2)+v_2}, \ldots, f_{H''}^{+J(y_{k_i})+v_{k_i}}$. (Notice these are restrictions of f, not F!) We will show that these sets respect the density and uniformity conditions for Lemma 25 to apply.

As for uniformity, we have already argued that each of these sets is $\gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(\ell))$-uniform, since $J(y_j) + v_j$ is not in H'' for every $j \in [k_i]$. For density, we argue as follows. For every $j \in [k_i]$, there are three cases: $\mu(y_j) = 1$, $\mu(y_j) = 0$, and $\mu(y_j) = *$. Consider the first case. If $y_j + H$ was affected by the first modification from f to F, then, $\rho(f_{H'}^{+I(y_j)}) \geq \frac{1}{2}$, and using the $\mathcal{E}(\ell)$-uniformity of $f_{H'}^{+I(y_j)}$ along with Lemma 13, we get that $\rho(f_{H''}^{+J(y_j)+v_j}) \geq \frac{1}{2} - \mathcal{E}(\ell) \cdot \delta_{19}^{-1}(\Psi_{\mathcal{F}}(r), \gamma_{25}(\epsilon/8, \Psi_{\mathcal{F}}(r))) \geq \frac{1}{2} - \frac{\epsilon}{8} \geq \frac{\epsilon}{8}$. If $y_j + H$ was affected by the second modification, then, by the same argument, we get that $\rho(f_{H''}^{+J(y_j)+v_j}) \geq 1 - \frac{\epsilon}{4} - \frac{\epsilon}{8} \geq \frac{\epsilon}{8}$. Else, if $y_j + H$ was affected by the third modification from S to S', we are automatically guaranteed that $\rho(f_{H''}^{+J(y_j)+v_j}) \geq \frac{1}{2}$ since $J(y_j) + v_j \notin H''$. The case $\mu(y_j) = 0$ is similar, and the analysis shows that $\rho(f_{H''}^{+J(y_j)+v_j}) \geq 1 - \frac{\epsilon}{8}$. Finally, consider the "wildcard" case, $\mu(y_j) = *$. This case arises only if $y_j \neq 0$ and $\epsilon/4 \leq \rho(f_{H'}^{+I(y_j)}) \leq 1 - \epsilon/4$. Again using $\mathcal{E}(\ell)$-uniformity of $f_{H'}^{+I(y_j)}$ along with Lemma 13, we get that $\epsilon/8 \leq \rho(f_{H''}^{+J(y_j)+v_j}) \leq 1 - \epsilon/8$.

Thus, we can apply Lemma 25 with $\epsilon/8$ and $\Psi_{\mathcal{F}}(\ell)$ as the parameters to get that there are at least $\delta_{25}(\epsilon/8, \Psi_{\mathcal{F}}(\ell))|H''|^{k_i-1}$ tuples $z = (z_1, \ldots, z_{k_i})$ with each $z_j \in J(y_j) + v_j + H''$ at which (E^i, σ^i) is induced . Finally, each such $z_1, \ldots, z_{k_i}$ leads to a distinct $z' = (z'_1, \ldots, z'_{k_i}) \in (\mathbb{F}_2^n)^{k_i}$ at which (E^i, σ^i) is induced by f, by setting each z'_j to $J(y_j) + v_j + z_j$ and observing that $\sum_{j=1}^{k_i} J(y_j) + v_j = J\left(\sum_{j=1}^{k_i} y_j\right) + \sum_{j=1}^{k_i} v_j = 0$. This completes the proof of Theorem 24. ■

3.0.2 Extending to Complexity 1 Systems of Equations

As mentioned in the introduction, the result we actually prove is stronger than Theorem 8. To describe the full set of properties for which we can show testability, we first need to make the following definition.

Definition 30 (Complexity of linear system [GT08]) *An $m \times k$ matrix M over $\mathbb{F}_2$ is said to be of* (Cauchy-Schwarz) complexity *c, if c is the smallest positive integer for which the following is true. For every $i \in [k]$, there exists a partition of $[k] \backslash \{i\}$ into $c+1$ subsets $S_1, \cdots, S_{c+1}$ such that for every $j \in [c+1]$, $\left(\mathbf{e}_i + \sum_{i' \in S_j} \mathbf{e}_{i'}\right) \notin$ $\mathsf{rowspace}(M)$, where $\mathsf{rowspace}(M)$ is the linear subspace of $\mathbb{F}_2^k$ spanned by the rows of M.*

In other words, if we view the rowspace of the matrix M as specifying a collection of linear dependencies on k variables $x_1, \dots, x_k$, then M has complexity c if for every variable x_i, the rest of the variables $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k$ can be partitioned into $c+1$ sets $S_1, \dots, S_{c+1}$ such that x_i is not linearly dependent on the variables of just a single S_j. Let us make a few remarks to illustrate the definition. Green and Tao show (Lemma 1.6 in [GT08]) that if each of these linear dependencies involves more than two variables, then the complexity of M is at most $\mathsf{rank}(M) = m$. In particular then, if M has one row and is nonzero on more than two coordinates, M has complexity 1. This is the setting we discussed in the introduction. We slightly extend this observation in the claim below. Before we state it, we observe that in the context of property testing, it is only natural to exclude matrices which yield linear dependencies involving less than three variables. If the rowspace of the matrix M contains a vector which is nonzero only at one coordinate i, then for any string σ of length k, the property of (M, σ)-freeness must contain all functions f such that $f(0) = 1 - \sigma_i$, and so *every* function is exponentially close to such a property. Similarly, if $\mathsf{rowspace}(M)$ contains a vector nonzero only at two coordinates i and j, then for any $\sigma \in \{0,1\}^k$, either (M, σ)-freeness is trivial (if $\sigma_i \neq \sigma_j$) or it is equivalent to (M', σ')-freeness where σ' is the string obtained by removing coordinate j and M' is the matrix obtained by removing column j, adding 1 (mod 2) to every element in column i and removing any resulting all-zero rows.

Claim 31 *If $M \in \mathbb{F}_2^{m \times k}$ is a matrix with two rows such that every vector in its rowspace has at least three nonzero coordinates, then M has complexity* 1.

Proof: Let $R_1 \subseteq [k]$ be the set of coordinates for which the first row is nonzero, and $R_2 \subseteq [k]$ those for which the second row is nonzero. We can assume that $R_1 \not\subseteq R_2$ and $R_2 \not\subseteq R_1$, because if, say, $R_1 \subseteq R_2$, we could replace the second row by the sum of the first and second, making R_1 and R_2 disjoint but preserving the rowspace of the matrix. Also, we we can assume w.l.o.g. that $R_1 \cup R_2 = [k]$.

Fix $i \in [k]$. We want to show a partition of $[k]\backslash\{i\}$ into sets S_1, S_2 such that $\mathbf{e}_i + \sum_{i' \in S_1} \mathbf{e}_{i'} \notin \mathsf{rowspace}(M)$ and similarly for S_2. If $i \in R_1\backslash R_2$, let S_1 consist of two elements, one from $R_2\backslash R_1$ and one from $R_1\backslash\{i\}$, and let S_2 be the rest. If $i \in R_2\backslash R_1$, let S_1 consist of one element from $R_1\backslash R_2$ and one from $R_2\backslash\{i\}$, and let S_2 be the rest. And finally, if $i \in R_1 \cap R_2$, let S_1 consist of one element from $R_1\backslash R_2$ and one from $R_2\backslash R_1$, and let S_2 be the rest. It is straightforward to check that the definition of complexity 1 is satisfied by these choices. ■

More generally, an infinitely large class of complexity 1 linear systems is generated by *graphic matroids*. We refer the reader to [BCSX09] for definition and details. That this class contains the class of matrices proved to be of complexity 1 in Claim 31 is easy to show. We proved the claim separately above only to be self-contained without introducing matroid notation. One final remark is that if M is the matrix in the characterization of Reed-Muller codes of order d, then M has complexity exactly d; see Example 3 of [GT08].

Our main result in this section is the extension of Theorem 8 to complexity 1 systems of equations.

Theorem 32 *Let $\mathcal{F} = \{(M^1, \sigma^1), (M^2, \sigma^2), \dots\}$ be a possibly infinite set of induced systems of equations, with each M^i of complexity 1. Then, the property of being $\mathcal{F}$-free is testable with one-sided error.*

We next describe how to modify the previous proof to the new settings. The following analogue to Theorem 24 is the core of the proof of Theorem 32.

Theorem 33 *For every infinite family $\mathcal{F} = \{(M^1, \sigma^1), (M^2, \sigma^2), \dots, (M^i, \sigma^i), \dots\}$, where each M^i is a $m_i \times k_i$ matrix over $\mathbb{F}_2$ of complexity 1, there are functions $N_{\mathcal{F}}(\cdot)$,*

$k_{\mathcal{F}}(\cdot)$ and $\delta_{\mathcal{F}}(\cdot)$ such that the following is true for any $\epsilon \in (0,1)$. If a function $f : \mathbb{F}_2^n \to \{0,1\}$ with $n > N_{\mathcal{F}}(\epsilon)$ is ϵ-far from being $\mathcal{F}$-free, then f induces $\delta \cdot 2^{n(k_i - m_i)}$ many copies of some (M^i, σ^i), where $k_i \le k_{\mathcal{F}}(\epsilon)$ and $\delta \ge \delta_{\mathcal{F}}(\epsilon)$.

The proof of Theorem 33 follows exactly the same argument as before as soon as a result analogous to Lemma 25 can be established. We state this result formally next.

Lemma 34 (Counting Lemma) *For every $\eta \in (0,1)$ and integer $k > 2$, there exist $\gamma = \gamma_{25}(\eta, k)$ and $\delta = \delta_{25}(\eta, k)$ such that the following is true. Suppose M is an $m \times k$ matrix of complexity 1 and rank $m < k$, $\sigma \in \{0,1\}^k$ is a tuple, H is a subspace of $\mathbb{F}_2^n$, and $f : \mathbb{F}_2^n \to \{0,1\}$ is a function. Furthermore, suppose there are k not necessarily distinct elements $u_1, \dots, u_k \in \mathbb{F}_2^n/H$ such that $Mu = 0$ where $u = (u_1, \dots, u_k)$, $f_H^{+u_i} : H \to \{0,1\}$ is γ-uniform for all $i \in [k]$, and $\rho(f_H^{+u_i})$ is at least η if $\sigma(i) = 1$ and at most $1 - \eta$ if $\sigma(i) = 0$ for all $i \in [k]$. Then, there are at least $\delta |H|^{k-m}$ many k-tuples $x = (x_1, x_2, \dots, x_k)$, with each $x_i \in u_i + H$, such that f induces (M, σ) at x.*

Lemma 34 is an immediate consequence of the Generalized von Neumann Theorem (Proposition 7.1 in [GT08]).

Chapter 4

Lower Bound for Triangle-Freeness

In this chapter, we prove Theorem 9, an $\Omega((1/\epsilon)^{2.423})$ query complexity lower bound for testing triangle-freeness. Recall that given a function $f : \mathbb{F}_2^n \to \{0, 1\}$, a *triangle* in f refers to a 3-element set $\{x, y, x+y\}$ for some $x, y \in \mathbb{F}_2^n$ such that $f(x) = f(y) = f(x + y) = 1$, and f is said to be *triangle-free* if there are no triangles in f. As we mentioned in the introduction, the lower bound is proved by first analyzing the query complexity of the canonical tester for triangle-freeness and then bounding the price one pays by testing using the canonical tester instead of some other algorithm.

4.1 Lower Bound for the Canonical Tester

4.1.1 Proof Overview

The canonical tester, recall, repeatedly and independently chooses uniformly at random two elements $x, y \in \mathbb{F}_2^n$, checks if the pair forms a triangle in f, and rejects if so. It accepts only when none of the chosen pairs forms a triangle in f. From a combinatorial point of view, proving a lower bound for the query complexity of the canonical tester for triangle-freeness amounts to constructing a function $F : \mathbb{F}_2^n \to \{0, 1\}$ (for every large enough n) which is far from being triangle-free but contains only a small number of triangles.

Our first observation (also independently due to Eli Ben-Sasson) is that we can construct such an F by using a seemingly weaker construction. Namely, it is enough to construct three, not necessarily identical, functions $F_1, F_2, F_3 : \mathbb{F}_2^n \to \{0,1\}$ such that the the function-triple (F_1, F_2, F_3) is far from triangle-free but contains a small number of triangles. Here, a triangle in a function-triple (f_1, f_2, f_3) refers to a triple $(x, y, x+y)$ for some $x, y \in \mathbb{F}_2^n$ such that $f_1(x) = f_2(y) = f_3(x+y) = 1$, and a triangle-free function-triple is one which does not contain a triangle. The distance of (f_1, f_2, f_3) to triangle-freeness is the minimum over all triangle-free (g_1, g_2, g_3) of $\frac{1}{3}(\Pr_x[f_1(x) \neq g_1(x)] + \Pr_x[f_2(x) \neq g_2(x)] + \Pr[f_3(x) \neq g_3(x)])$.

Claim 35 *If there are functions $F_1, F_2, F_3 : \mathbb{F}_2^n \to \{0,1\}$ so that the triple (F_1, F_2, F_3) is ϵ-far from triangle-free and contains k triangles, then there is a function $F : \mathbb{F}_2^{n+2} \to \{0,1\}$ that is $3\epsilon/4$-far from triangle-free and contains k triangles.*

Proof: For $x \in \mathbb{F}_2^{n+2}$, write x as $x' \circ x''$ where $x' \in \mathbb{F}_2^n$ and $x'' \in \mathbb{F}_2^2$. Define $F : \mathbb{F}_2^{n+2} \to \{0,1\}$ in the following way: $F(x) = F_1(x')$ if $x'' = 01$, $F(x) = F_2(x')$ if $x'' = 10$, $F(x) = F_3(x')$ if $x'' = 11$, and $F(x) = 0$ if $x'' = 00$. It is easy to see that $\{x, y, x+y\}$ is a triangle in F for some $x, y \in \mathbb{F}_2^n$ if and only if one of the six permutations of $(x', y', x'+y')$ is a triangle in (F_1, F_2, F_3). So, F contains the same number of triangles as (F_1, F_2, F_3).

To argue about distance, consider the function $g : \mathbb{F}_2^n \to \{0,1\}$ that is the closest triangle-free function to F. Observe that $g(x) = F(x)$ when $x'' = 00$ because changing the value of F from 0 to 1 will never remove a triangle. Thus, we can obtain a triangle-free function-triple (g_1, g_2, g_3) from g. Because (F_1, F_2, F_3) is ϵ-far from triangle-free, the bound on F's distance immediately follows. ■

We obtain our desired (F_1, F_2, F_3) by constructing a *vertex-disjoint* function-triple, meaning that for no two triangles (x_1, y_1, x_1+y_1) and (x_2, y_2, x_2+y_2) in the function-triple is it the case that $x_1 = x_2$ or $y_1 = y_2$ or $x_1 + y_1 = x_2 + y_2$. The property of being vertex-disjoint makes it simple to calculate the function-triple's distance from triangle-freeness as well as counting the number of triangles within the function-triple.

We start our construction of a vertex-disjoint function-triple from three sets, each of cardinality m, of k-bit binary vectors, $\{a_i\}_{i=1}^m$, $\{b_j\}_{j=1}^m$ and$\{c_\ell\}_{\ell=1}^m$, where k and

m are fixed integers. Next we define three sets, $\{A_I\}$, $\{B_J\}$ and$\{C_L\}$, of mk-bit vectors, each consisting of the vectors obtained by concatenating $\{a_i\}$, $\{b_j\}$ and$\{c_\ell\}$, respectively, in all possible orders. Finally we define our function-triple (f_A, f_B, f_C) to be the characteristic functions of the three sets $\{A_I\}$, $\{B_J\}$ and$\{C_L\}$. In order to make the triangles in this function-triple pairwise disjoint, we impose the constraint that $\{a_i\}$, $\{b_j\}$ and$\{c_\ell\}$ satisfy the 1-perfect-matching-free (1-PMF for short) property which we define soon. To make this construction work for arbitrarily small ϵ, we make some $n' \geq 1$ copies of each $\{a_i\}$, $\{b_j\}$ and$\{c_\ell\}$, take the multiset of all the copies, and require them to satisfy the n'-PMF property for all $n' \geq 1$. It turns out that $\{a_i\}$, $\{b_j\}$ and$\{c_\ell\}$ being PMF is equivalent to a (small) set of homogeneous Diophantine linear equations having no non-trivial solution, which in turn can be checked by linear programming.

Numerical computation indicates the existence of a PMF family of vectors for $k = 3, 4$, and 5. (Unfortunately, it was computationally infeasible to search for PMF families of vectors for $k \geq 6$.) The PMF family with $k = 5$ yields a vertex disjoint function triple (f_A, f_B, f_C), each of which are Boolean functions on constant sized domains. To get a function-triple (F_1, F_2, F_3) so that the functions are defined on $\mathbb{F}_2^n$ for arbitrary large n, we use a blow-up operation which does not affect the distance to triangle-freeness or the density of triangles. We show that (F_1, F_2, F_3) is ϵ-far from trinagle-free but contains $O(\epsilon^{4.847}) \cdot 2^{2n}$ many triangles.

4.1.2 Perfect-matching-free Families of Vectors

In this section, we show how to build vertex-disjoint function-triples using constructions of *perfect-matching free* families of vectors.

Definition 36 (PERFECT-MATCHING-FREE FAMILIES OF VECTORS) *Let k and m be integers such that $0 < k < m < 2^k$. Let $\{a_i\}_{i=1}^m$ and $\{b_i\}_{i=1}^m$ be two families of vectors, with $a_i, b_i \in \{0,1\}^k$ for every $1 \leq i \leq m$. Let $c_i = a_i + b_i$.*

1. *Let $\{A_I\}_I$ be the set of (mk)-bit vectors formed by concatenating the m vectors in $\{a_i\}$ in all possible orders (there are $m!$ such vectors), where $I =$*

$(i_1, i_2, \ldots, i_m)$ is a permutation of $[m]$. Similarly define $\{B_J\}_J$ and $\{C_L\}_L$ as the concatenations of vectors in $\{b_i\}$ and $\{c_i\}$ with $J = (j_1, j_2, \ldots, j_m)$ and $L = (\ell_1, \ell_2, \ldots, \ell_m)$, respectively. We say the set of vectors $\{a_i, b_i, c_i\}$ is a (k, m) 1-perfect-matching-free *(abbreviated as 1-PMF) family of vectors if $A_I + B_J = C_L$ necessarily implies that $I = J = L$ (i.e., $i_s = j_s = \ell_s$ for every $1 \leq s \leq m$).*

2. *Let $n' \geq 1$ be an integer and now let $\{A_I\}_I$, $\{B_J\}_J$ and $\{C_L\}_L$ be the sets of $n'mk$-bit vectors by concatenating n' copies of $\{a_i\}$, $\{b_i\}$ and $\{c_i\}$, respectively, in all possible orders (two concatenations are regarded the same if they give rise to two identical strings in $\{0,1\}^{n'mk}$). We say the set of vectors $\{a_i, b_i, c_i\}$ is a (k, m) n'-*PMF *family of vectors if $A_I + B_J = C_L$ necessarily implies that $I = J = L$.*

3. *Finally we say $\{a_i, b_i, c_i\}$ is a (k, m)-*PMF *family of vectors if it is n'-PMF for all $n' \geq 1$.*

In other words, suppose we color all the $3m$ vectors in $\{a_i, b_i, c_i\}$ with m different colors so that a_i, b_i and c_i are assigned the same color. Suppose further we are given equal number of copies of $\{a_1, b_1, c_1; \ldots;\ a_m,\ b_m,\ c_m\}$ and we wish to arrange them in three aligned rows such that all the a_i's are in the first row, all the b_i's are in the second row and all the c_i's are in the third row. Then the only way of making every column summing to 0^k is to take the trivial arrangement in which every column is monochromatic.

Construction Based on PMF Families of Vectors

Let $\{a_i, b_i, c_i\}$ be a (k, m)-PMF family of vectors. Let n be an integer such that $mk|n$ and let $n' = \frac{n}{mk}$. let $\{A_I\}_I$, $\{B_J\}_J$ and $\{C_L\}_L$ be the sets of n-bit vectors by concatenating n' copies of $\{a_i\}$, $\{b_i\}$ and $\{c_i\}$ respectively. Note that $|\{A_I\}| = |\{B_J\}| = |\{C_L\}| = \frac{(n'm)!}{(n'!)^m}$. Now let $f_A, f_B, f_C : \mathbb{F}_2^n \to \{0,1\}$ be three Boolean functions which are the characteristic functions of sets $\{A_I\}_I$, $\{B_J\}_J$ and $\{C_L\}_L$ respectively. That is, $f_A(x) = 1$ iff $x \in \{A_I\}$, $f_B(x) = 1$ iff $x \in \{B_J\}$ and $f_C(x) = 1$ iff $x \in \{C_L\}$.

Proposition 37 *All the triangles in the function-triple (f_A, f_B, f_C) are pairwise disjoint.*

Proof: Immediate. ■

Lemma 38 *If (k, m)-PMF family of vectors exists, then for infinitely many ϵ that can be made arbitrarily small, there is a $n_0 = n_0(\epsilon)$ and functions $f_A, f_B, f_C : \mathbb{F}_2^{n_0} \to \{0, 1\}$ such that (f_A, f_B, f_C) is ϵ-far from being triangle-free and the number of triangles in (f_A, f_B, f_C) is $O(\epsilon^{\alpha - o(1)}) \cdot 2^{2n_0}$, where $\alpha = \frac{2 - \frac{\log m}{k}}{1 - \frac{\log m}{k}}$ and the "$o(1)$" goes to zero as ϵ goes to zero.*

Proof: Suppose $\epsilon = \left(\frac{(n'm)!}{(n'!)^m}\right) / 2^{n'mk}$ for a positive integer n'. Let f_A, f_B and f_C be the characteristic functions of $\{A_I\}_I$, $\{B_J\}_J$ and $\{C_L\}_L$ respectively defined above. Set $n_0 = n'mk$ and then f_A, f_B and f_C are Boolean functions on n_0 variables. Let N_Δ be the number of triangles in (f_A, f_B, f_C). Then by Stirling's formula, for all small enough ϵ (meaning for all large enough n'),

$$\begin{aligned} N_\Delta &= \frac{(n'm)!}{(n'!)^m} \\ &= \frac{\sqrt{2\pi mn'}(\frac{mn'}{e})^{mn'}(1 + O(\frac{1}{n'}))}{\left(\sqrt{2\pi n'}(\frac{n'}{e})^{n'}(1 + O(\frac{1}{n'}))\right)^m} \\ &= 2^{(m \log m)n' - -o(1)} \\ &= 2^{(\beta - o(1))n_0}, \end{aligned}$$

where $\beta = \frac{\log m}{k}$. Since $\epsilon = N_\Delta / 2^{n_0}$, it follows that $2^{n_0} = (1/\epsilon)^{1/(1-\alpha) + o(1)}$.

By Proposition 37, all the triangles in (f_A, f_B, f_C) are pairwise disjoint. Therefore modifying the function-triple at one point in the domain can remove at most one triangle. Hence $\text{dist}((f_A, f_B, f_C), \text{T-FREE}) \geq \frac{N_\Delta}{2^{n_0}} = \epsilon$. And the triangle density is $N_\Delta / 2^{2n_0} = 2^{(\beta - 2 - o(1))n_0} = \epsilon^{\frac{2-\beta}{1-\beta} - o(1)} = \epsilon^{\alpha - o(1)}$. ■

One can construct f_A, f_B, f_C to be Boolean functions on $\mathbb{F}_2^n$ for any $n \geq n_0$, by simply making the functions ignore the last $n - n_0$ bits and behave as defined above on the first n_0 bits. In Theorem 47, we give a construction by tensoring with bent functions so that the resulting functions depend on all n bits.

We conjecture the following to be true.

Conjecture 39 *There are infinitely many (k,m)-PMF families of vectors with $m \geq 2^{k(1-o(1))}$ where "$o(1)$" goes to zero as k goes to infinity.*

By Lemma 38, Conjecture 39 would imply a super-polynomial query lower bound for testing triangle-freeness in function-triples using the canonical tester. To be more specific, if there exists a (k,m)-PMF family of vectors with $m \geq 2^{k(1-o(1))}$, then the query complexity of the canonical tester is at least $\Omega((\frac{1}{\epsilon})^{\frac{1}{o(1)}})$. Moreover, when composed with Theorem 52 it would also give a super-polynomial lower bound for *any* one-sided triangle-freeness tester.

Existence of PMF Families of Vectors

In this section we present an efficient algorithm which, given a family of vectors $\{a_i, b_i, c_i\}_{i=1}^m$, checks if it is PMF. We will use this algorithm to find an explicit PMF family.

Let $\{a_i, b_i, c_i\}_{i=1}^m$ be a family of vectors such that $a_i, b_i, c_i \in \mathbb{F}_2^k$ and $c_i = a_i + b_i$ for every $1 \leq i \leq m$. First we observe that if $\{a_i, b_i, c_i\}$ is PMF, then all the vectors in $\{a_i\}$ must be distinct. The same distinctness condition holds for vectors in $\{b_i\}$ and $\{c_i\}$. From now on, we assume these to be true. Next we define a set of "collision blocks".

Definition 40 (Collision Blocks) *Let $\{a_i, b_i, c_i\}_{i=1}^m$ be a family of vectors satisfying the distinctness condition. We say (i,j,ℓ) is a* collision block *if $a_i + b_j = c_\ell$, and for simplicity will just call it a block. We denote the set of all blocks by $\mathcal{B}$. We will call a block* trivial *if $i = j = \ell$ and* non-trivial *otherwise.*

Since $\{a_i, b_i, c_i\}$ satisfies the distinctness condition, clearly $|\mathcal{B}| < m^2$. Let r be the number of non-trivial blocks, and let $\{\mathsf{bl}_1, \ldots, \mathsf{bl}_r\}$ be the set of non-trivial blocks. For a collision block bl_s, we use $\mathsf{bl}_s^a, \mathsf{bl}_s^b$ and bl_s^c to denote the three indices of the colliding vectors. That is, if $\mathsf{bl}_s = (i,j,\ell)$ is a block, then $\mathsf{bl}_s^a = i$, $\mathsf{bl}_s^b = j$ and $\mathsf{bl}_s^c = \ell$.

Now suppose $\{a_i, b_i, c_i\}_{i=1}^m$ is not PMF. Then by the definition of PMF, there exists an integer n' such that $A_I, B_J, C_L \in \{0,1\}^{n'mk}$, $A_I + B_J = C_L$ and I, J, and L are not the same sequence of indices. We consider the equation $A_I + B_J = C_L$ as a tiling of $3 \times (n'm)$ k-bit vectors: the first row consists of the $n'm$ vectors from $\{a_i\}$ with each a_i appearing exactly n' times and the ordering is consistent with that of A_I. Similarly we arrange the second row with vectors from $\{b_i\}$ according to B_J and the third row with vectors from $\{c_i\}$ according to C_L. Observe that when we look at the columns of the tiling, each column corresponds to a block in $\mathcal{B}$. Now we remove all the trivial blocks, then because I, J, and L are not identical sequences of indices, there are some non-trivial blocks left in the tiling. Since all the blocks removed are trivial blocks, the remaining tiling still has equal number of a_i, b_i and c_i for every $1 \le i \le m$. We denote these numbers by $y_1, \dots, y_m$. Note that y_i's are non-negative integers and not all of them are zero. Let the number of blocks bl_i left in the tiling be x_i, $1 \le i \le r$. Again x_i's are non-negative integers and not all zero. Moreover, we have the following constraints when counting the number of a_i, b_i and c_i vectors, respectively, left in the tiling:

$$\begin{cases} \sum_{j \in [r]:\mathsf{bl}_j^a = i} x_j - y_i = 0 \\ \sum_{j \in [r]:\mathsf{bl}_j^b = i} x_j - y_i = 0 \qquad \text{(for every } 1 \le i \le m) \\ \sum_{j \in [r]:\mathsf{bl}_j^c = i} x_j - y_i = 0 \end{cases} \tag{4.1}$$

where

$$x_j \quad = \quad \text{number of type } j \text{ blocks left after removing the trivial blocks}$$

and

$$y_i = \text{number of vectors } a_i \text{ (equiv. } b_i \text{ or } c_i) \text{ left after removing the trivial blocks.}$$

Lemma 41 *$\{a_i, b_i, c_i\}_{i=1}^m$ is not PMF if and only there is a non-zero integral solution* *to the system of linear equations (4.1).*

Proof: We only need to show that if there is a non-zero solution to (4.1), then $\{a_i, b_i, c_i\}_{i=1}^{m}$ is not PMF. Let $\{x_i, y_j\}$ be a set of non-zero integer solution. Note that the solution corresponds to a partial tiling with equal number of a_i, b_i and c_i for every $1 \leq i \leq m$. Set $n' = \max_i y_i$. Since the solution is non-trivial, $n' \geq 1$. Now for each $1 \leq i \leq m$, add $(n' - y_i)$ number of trivial blocks (i, i, i) to the tiling. Then the resulting tiling gives $A_I, B_J, C_L \in \{0,1\}^{n'mk}$ and $A_I + B_J = C_L$ such that I, J and L are not identical. ■

Writing equations (4.1) in matrix form, we have

$$\mathbf{MZ} = \mathbf{0},$$

where

$$\mathbf{M} = \begin{bmatrix} 1 & & \cdots & 1 & -1 & & \\ & 1 & \cdots & & & \ddots & \\ & & \cdots & & & & -1 \\ & & \cdots & 1 & -1 & & \\ & 1 & \cdots & & & \ddots & \\ 1 & & \cdots & & & & -1 \\ & 1 & \cdots & & -1 & & \\ & & \cdots & & & \ddots & \\ 1 & & \cdots & 1 & & & -1 \end{bmatrix}$$

is a $3m \times (r + m)$ integer-valued matrix (actually all entries are in the set $\{-1, 0, 1\}$) and

$$\mathbf{Z} = [x_1, \ldots, x_r, y_1, \ldots, y_m]^T$$

is an $(r + m) \times 1$ non-negative integer-valued column vector. Note that each of first r columns of $\mathbf{M}$ has exactly three 1s and all other entries are zero, and the last m columns of $\mathbf{M}$ consist of three $-I_{m \times m}$ matrices.

The following observation of Domenjoud [Dom91], which essentially follows from

Carathéodory's theorem, gives an exact characterization of when the system of equations (4.1) has a non-zero integral solution. We provide a proof below for completeness.

Theorem 42 **([Dom91])** *Let* $\mathbf{M}$ *be an* $s \times t$ *integer matrix, then the Diophantine linear system of equations* $\mathbf{MZ} = \mathbf{0}$ *with* $\mathbf{Z} \in \mathbb{N}^t$ *has a non-zero solution if and only if* $\mathbf{0} \in Conv(M_1, \ldots, M_t)$*, where* M_i*'s are the column vectors of* $\mathbf{M}$ *and* $Conv(M_1, \ldots, M_t)$ *denotes the convex hull of vectors* $M_1, \ldots, M_t$*.*

Proof: If there exists a non-zero vector $\mathbf{Z} \in \mathbb{N}^t$ such that $\mathbf{MZ} = \mathbf{0}$, the vector $\mathbf{z} = \frac{\mathbf{Z}}{\|\mathbf{Z}\|_1}$ also satisfies $M\mathbf{z} = \mathbf{0}$. But then, $\mathbf{0} \in \text{Conv}(M_1, \ldots, M_t)$ because $\sum_i z_i M_i = \mathbf{0}$ and each $z_i \geq 0$ with $\sum_i z_i = 1$.

In the other direction, suppose $\mathbf{0} \in \text{Conv}(M_1, \ldots, M_t)$. Let $\{M_{i_1}, \ldots, M_{i_k}\}$ be a minimal subset of $\{M_1, \ldots, M_t\}$ which contains $\mathbf{0}$ in its convex hull. We now need the following well-known theorem of Carathéodory in convex geometry (see, e.g., [Gru07]).

Theorem 43 (Carathéodory's Theorem) *Suppose* V *is a subset of* $\mathbb{R}^n$ *that contains a point* $X \in \mathbb{R}^n$ *in its convex hull. Then there exists a set* $V' \subseteq V$ *such that* $|V'| \leq n+1$ *and* X *is contained in the convex hull of* V'*. An implication is that if* V *contains* $\mathbf{0}$ *in its convex hull and there is no strict subset* V' *containing* $\mathbf{0}$ *in its convex hull, then* $rank(V) = |V| - 1$*.*

Carathéodory's theorem implies that the rank of $\{M_{i_1}, \ldots, M_{i_k}\}$ from above is $k - 1 \leq s$. Let $\mathbf{M}'$ be the s-by-k matrix with columns $\{M_{i_1}, \ldots, M_{i_k}\}$. Then [1] there exists a unimodular (that is, the determinant of the matrix is either 1 or -1) s-by-s matrix $\mathbf{U}$ such that

$$\mathbf{UM}' = \begin{bmatrix} \mathbf{N} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

[1] See, for example, Theorem 2.4.3 in [Coh00].

where $\mathbf{N}$ is a $(k-1)$-by-k integer matrix of rank $(k-1)$ in row-echelon form. It follows that the nullspace of $\mathbf{N}$ is spanned by a single non-zero vector in $\mathbb{R}^k$. Since $\mathbf{0}$ is in the convex hull of $\{M_{i_1}, \ldots, M_{i_k}\}$, there exists a non-zero vector $\mathbf{X} \in (\mathbb{R}^{\geq 0})^k$ such that $\mathbf{NX} = \mathbf{0}$. It follows that all the vectors in the nullspace of $\mathbf{N}$ have the same sign at each coordinate. But the vector consists of the cofactors of $\mathbf{N}$, namely, $\mathbf{Y} = (\left|N_2 \cdots N_k\right|, \ldots, (-1)^{k-1}\left|N_1 \cdots N_{k-1}\right|)$ is a solution to $\mathbf{NX} = \mathbf{0}$. Furthermore, all the entries in $\mathbf{Y}$ are non-zero since the rank of $\mathbf{N}$ is $k-1$. Hence either $\mathbf{Y}$ or $-\mathbf{Y}$ is a positive integer solution to $\mathbf{NX} = \mathbf{0}$, and because $\mathbf{U}$ is invertible, the same positive integer vector satisfies $\mathbf{M'X} = \mathbf{0}$. Appending 0 entries to $\mathbf{X}$ at all the remaining $(t-k)$ coordinates gives a non-negative integer solution to $\mathbf{MZ} = \mathbf{0}$. ■

It is well known that checking point-inclusion in a convex hull can be solved by Linear Programming, see e.g. [BC87]. In particular, following the definition of convex hulls, $\mathbf{0} \in \mathrm{Conv}(M_1, \ldots, M_t)$ if and only if there exist real numbers $\theta_1 \geq 0, \ldots, \theta_t \geq 0$ such that

$$\sum_{i=1}^{t} \theta_i M_i = \mathbf{0}$$

and

$$\sum_{i=1}^{t} \theta_i = 1.$$

After introducing additional slack variables and plugging in our collision matrix $\mathbf{M}$ into the formalism, we finally have the following characterization of a family of vectors being PMF.

Lemma 44 *The family of vectors $\{a_i, b_i, c_i\}_{i=1}^m$ is PMF if and only if the following LP*

$$\begin{aligned} &\textit{Maximize } W = \mathbf{c} \cdot \theta \\ &\textit{Subject to } \mathbf{M}'\theta = \mathbf{b} \\ &\qquad\qquad \theta \geq \mathbf{0} \end{aligned}$$

has no feasible solution with $W \geq 0$.

Here

$$\mathbf{M}' = \left[\begin{array}{ccc|c} & \mathbf{M} & & \\ 1 & \cdots & 1 & \mathbf{I}_{(3m+1)} \end{array} \right]$$

is a $(3m+1) \times (4m+r+1)$ *integer matrix with* $\mathbf{M}$ *being the collision matrix of the family of vectors* $\{a_i, b_i, c_i\}_{i=1}^{m}$,

$$\mathbf{b} = [0, \ldots, 0, 1]^T$$

is a $3m+1$*-dimensional integer vector and*

$$\mathbf{c} = [\underbrace{0, \ldots, 0}_{r+m}, \underbrace{-1, \ldots, -1}_{3m+1}]^T$$

is the objective function vector of dimension $4m+r+1$.

Using this procedure for checking if a family of vectors $\{a_i, b_i, c_i\}_{i=1}^{m}$ is PMF or not, we find the following (k,m)-PMF families of vectors.

Theorem 45 *There are* $(3,4)$*-PMF,* $(4,7)$*-PMF and* $(5,13)$*-PMF families of vectors.*

Proof: By numerical calculation, the following set of vectors is $(3,4)$-PMF:

$$\begin{array}{ll} a_1 = 110 & b_1 = 001 \\ a_2 = 010 & b_2 = 100 \\ a_3 = 101 & b_3 = 111 \\ a_4 = 011 & b_4 = 011. \end{array}$$

The following set of vectors is $(4,7)$-PMF:

$$\begin{array}{ll} a_1 = 1101 & b_1 = 0011 \\ a_2 = 0001 & b_2 = 1011 \\ a_3 = 0010 & b_3 = 0111 \\ a_4 = 0110 & b_4 = 1001 \end{array}$$

$a_5 = 0000$	$b_5 = 0000$
$a_6 = 0111$	$b_6 = 0100$
$a_7 = 1001$	$b_7 = 0101.$

The following set of vectors is $(5, 13)$-PMF:

$a_1 = 11101$	$b_1 = 01101$
$a_2 = 11001$	$b_2 = 11101$
$a_3 = 11000$	$b_3 = 10011$
$a_4 = 00101$	$b_4 = 10001$
$a_5 = 10010$	$b_5 = 00101$
$a_6 = 11110$	$b_6 = 10100$
$a_7 = 10000$	$b_7 = 10000$
$a_8 = 01000$	$b_8 = 01111$
$a_9 = 00011$	$b_9 = 01010$
$a_{10} = 11100$	$b_{10} = 00111$
$a_{11} = 00010$	$b_{11} = 11010$
$a_{12} = 01100$	$b_{12} = 10010$
$a_{13} = 01010$	$b_{13} = 11111.$

■

We were unable to check the cases $k \geq 6$ since they are too large to do numerical calculations. However, our best findings for $k = 3, 4, 5$ indicates that the exponent α defined in Lemma 38 increases as k increases, which we view as a supporting evidence for Conjecture 39.

Now using the $(5, 13)$-PMF family of vectors as the building block, Lemma 38 implies the following.

Theorem 46 *For infinitely many ϵ that approach zero arbitrarily closely, there is an $n_0 = n_0(\epsilon)$ and functions $f_A, f_B, f_C : \mathbb{F}_2^{n_0} \to \{0,1\}$ such that (f_A, f_B, f_C) is ϵ-far from being triangle-free and contains $O(\epsilon^{4.847\cdots}) \cdot 2^{2n_0}$ triangles.*

A simple blow-up procedure on appropriate number of bits with the function-triples constructed in Theorem 46 yields the following Theorem.

Theorem 47 *For all small enough ϵ, there is an integer $n_0(\epsilon)$ such that the following holds. For all integers $n \geq n_0$, there are functions $F_1, F_2, F_3 : \mathbb{F}_2^n \to \{0,1\}$ such that the function-triple (F_1, F_2, F_3) is ϵ-far from being triangle-free and contains $O(\epsilon^{4.847\cdots})\cdot 2^{2n}$ many triangles.*

Proof: First, apply Theorem 46 to get functions $f_A, f_B, f_C : \mathbb{F}_2^{n_0} \to \{0,1\}$ so that the function-triple (f_A, f_B, f_C) is ϵ-far from triangle-free but has $O(\epsilon^{4.847\cdots}) \cdot 2^{2n_0}$ many triangles, where n_0 is a function of ϵ. Next, for any $n \geq n_0$, define $F_1, F_2, F_3 : \mathbb{F}_2^n \to \{0,1\}$ by $F_1(x) = f_A(x|_{n_0}), F_2(x) = f_B(x|_{n_0})$, and $F_3(x) = f_C(x|_{n_0})$, where $x|_{n_0}$ denotes the first n_0 bits of a longer string x. It is easy to see that the number of triangles in (F_1, F_2, F_3) is $O(\epsilon^{4.847\cdots}) \cdot 2^{2n}$.

To argue that (F_1, F_2, F_3) is ϵ-far from being triangle-free, recall that the triangles in (f_A, f_B, f_C) are disjoint. Suppose $(x, y, x+y)$ is a triangle in (F_1, F_2, F_3) for some x, y. If we change $F_1(x)$ to zero, we remove exactly 2^{n-n_0} many triangles, all corresponding to one triangle in (f_A, f_B, f_C). The number of triangles in (F_1, F_2, F_3) is $2^{2(n-n_0)}$ times the number of triangles in (f_A, f_B, f_C). So, it follows that (F_1, F_2, F_3) is also ϵ-far from being triangle-free. ■

Applying Claim 35 finally yields our desired lower bound.

Corollary 48 *The query complexity of the canonical tester for triangle-freeness is $\Omega((1/\epsilon)^{4.847\ldots})$.*

4.2 Query Complexities of the Canonical Tester and General One-sided Testers

In this section, we prove a connection between the query complexities of an arbitrary one-sided tester and the canonical tester, for a large class of algebraic properties. This class includes triangle-freeness, and our result will show that if the query complexity of the optimal one-sided tester for triangle-freeness is q, then the query complexity of the canonical tester for triangle-freeness is $O(q^2)$. Combining with Corollary 48 yields the lower bound of $\Omega((1/\epsilon)^{2.423\cdots})$ for the one-sided query complexity of triangle-freeness.

The more general class of properties that we study is related to the class of $(\mathcal{M}, \mathbf{1})$-freeness properties as defined in Definition 3.

Definition 49 ($\mathcal{M}^*$-free) *Given a rank-r matroid $\mathcal{M} = (v_1, \ldots, v_k)$ with each $v_i \in \mathbb{F}_2^r$, a Boolean function $f : \mathbb{F}_2^n \to \{0, 1\}$ is said to be* $\mathcal{M}^*$-free *if there is no full-rank linear transformation $L : \mathbb{F}_2^r \to \mathbb{F}_2^n$ such that $f(L(v_i)) = 1$ for every $i \in [k]$. Otherwise, if such an L exists, f is said to* contain $\mathcal{M}$ at L, *or equivalently, L is called a* violating linear transformation *of $\mathcal{M}$.*

Remark: Let $(\mathbf{e}_1, \ldots, \mathbf{e}_r)$ be a set of basis vectors in $\mathbb{F}_2^r$. Each linear map L in the above definition is then specified by r vectors $z_1, \ldots, z_r$ in $\mathbb{F}_2^n$ such that $L(\mathbf{e}_i) = z_i$ for every $1 \leq i \leq r$. The linear map L is *full rank* if $(z_1, \ldots, z_r)$ are linearly independent.

To see that this generalizes the triangle-freeness property, let $\mathbf{e}_1$ and $\mathbf{e}_2$ be the two unit vectors in $\mathbb{F}_2^2$ and consider the matroid $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2)$. Then the three elements of the matroid will be mapped to all triples of the form $(x, y, x + y)$ by the set of full-rank linear transformations, where x and y are two distinct non-zero elements in $\mathbb{F}_2^n$. Also note that in this case, $r = 2$ and $k = 3$.

The property of being $\mathcal{M}^*$-free is not linear-invariant. The original notion of $\mathcal{M}$-freeness (shorthand for $(\mathcal{M}, 1)$-freeness in Definition 3) allows L in the above definition to be arbitrary linear transformations, not just the full-rank ones, and is hence truly linear-invariant. However, from a conceptual level, for a fixed matroid

$\mathcal{M}$, the property of being $\mathcal{M}$-free and being $\mathcal{M}^*$-free are very similar. It is analogous to the distinction between a graph being free of H as a subgraph and being free of homomorphic images of H, for a fixed graph H.

In terms of testability, we have some evidence that the distinction is unimportant, although we are unable to prove a formal statement at this time. For the case when $\mathcal{M} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2)$, we can show that a tester for triangle-freeness can be converted to one for triangle*-freeness. Consider a function-triple (f_1, f_2, f_3) that is promised to be either triangle*-free or ϵ-far from being triangle*-free, where the distance parameter ϵ is a constant. Define a new function-triple (f_1', f_2', f_3') by setting, for $i = 1, 2, 3$, $f_i'(0) = 0$ and $f_i'(x) = f_i(x)$ for all $x \neq 0$. Observe that if (f_1, f_2, f_3) is triangle*-free, then (f_1', f_2', f_3') is triangle-free because setting $f_i'(0) = 0$ removes all degenerate triangles. On the other hand, if (f_1, f_2, f_3) is ϵ-far from triangle*-free, then (f_1', f_2', f_3') is still $\epsilon' \geq \epsilon - 3/2^n$ far from triangle*-free and, hence, also from triangle-free. Since ϵ' approaches ϵ as n goes to infinity, assuming the continuity of the query complexity as a function of the distance parameter, the query complexity of triangle-freeness is therefore lower-bounded [2] by the query-complexity of triangle*-freeness.

For general binary matroids $\mathcal{M} = (v_1, \ldots, v_k)$ with each $v_i \in \mathbb{F}_2^r$, observe that if a function is far from being $\mathcal{M}$-free, then almost all the linear maps where $\mathcal{M}$ is contained are full-rank. This is because the main theorems of [Sha09] and [KSV08] show that if a function is $\Omega(1)$-far from $\mathcal{M}$-free, then $\mathcal{M}$ is contained at $\Omega(2^{nr})$ many linear maps, while there are only $o(2^{nr})$ many linear maps $L : \mathbb{F}_2^r \to \mathbb{F}_2^n$ of rank less than r. Therefore, in fact, any $\mathcal{M}^*$-free function is $o(1)$-close to $\mathcal{M}$-free. If there were a more query efficient one-sided tester for $\mathcal{M}$-freeness than for $\mathcal{M}^*$-freeness, it must be the case that the few linear maps with rank less than r where $\mathcal{M}$ is contained can somehow be discovered more efficiently than the full-rank maps. But on the other hand, we know of a large class of matroids $\mathcal{M}$ for which there exist functions that are far from $\mathcal{M}$-free but do not contain $\mathcal{M}$ at *any* non-full-rank linear map. More

[2] The other direction is easy to show in general: for any binary matroid $\mathcal{M}$ and constant ϵ, an ϵ-tester for $\mathcal{M}^*$-freeness can be used to ϵ-test $\mathcal{M}$-freeness (again assuming continuity of the query complexity function).

precisely, letting $C_k = (\mathbf{e}_1, \ldots, \mathbf{e}_{k-1}, \mathbf{e}_1 + \cdots + \mathbf{e}_{k-1})$ be the graphic matroid of the k-cycle, Theorem 1.3 in [BCSX09] proves that for any odd $k \geq 5$, there exist functions which are far from C_k-free but contain C_k only at full-rank linear maps (by showing a separation between the classes C_k-free and C_{k-2}-free). So, for these reasons, it seems unlikely that the query complexities of testing $\mathcal{M}^*$-freeness properties are very different from those of testing $\mathcal{M}$-freeness properties. We conjecture that the query complexities of testing $\mathcal{M}$-freeness and $\mathcal{M}^*$-freeness properties are the same [3] and leave this as an open problem.

We first observe a simple fact about the behavior of any *one-sided* tester for $\mathcal{M}^*$-freeness.

Lemma 50 *Let $\mathcal{M}$ be a matroid of k vectors. Then any one-sided tester T for $\mathcal{M}^*$-freeness rejects if and only if it detects a violating full-rank linear transformation L of $\mathcal{M}$.*

Proof: Let $f : \mathbb{F}_2^n \to \{0,1\}$ be the given Boolean function. If T finds a violating full-rank linear transformation L, clearly it should reject. For the other direction, suppose that T rejects f without seeing any violating linear maps from the points it queried. Since $\mathcal{M}^*$-freeness is a monotone property, we can set all the points of the function-tuple that have not been queried by T to 0, thus making f $\mathcal{M}^*$-free. Therefore T errs on this function-tuple. But this contradicts our assumption that T is a one-sided tester for $\mathcal{M}^*$-freeness. ■

Next, we define the canonical tester for $\mathcal{M}^*$-freeness, which naturally extends the previously described canonical tester for triangle-freeness.

Definition 51 (Canonical Tester) *Let $\mathcal{M} = (v_1, \ldots, v_k)$, with each $v_i \in \mathbb{F}_2^r$, be a rank-r matroid of k vectors. A tester $\mathcal{T}$ for $\mathcal{M}^*$-freeness is* canonical *if $\mathcal{T}$ operates as follows. Given as input a distance parameter ϵ and oracle access to Boolean function*

[3] It seems possible that some functions may have quite different query complexities for these two properties. However, the query complexities in our conjecture are measured as (non-increasing) functions of the distance parameter ϵ, which are *worst-case* query complexities among all input functions that are ϵ-far from the corresponding properties.

$f : \mathbb{F}_2^n \to \{0,1\}$, the tester $\mathcal{T}$ repeats the following process independently $\ell(\epsilon)$ times: select uniformly at random a rank-r linear transformation $L : \mathbb{F}_2^r \to \mathbb{F}_2^n$ and check if f contains $\mathcal{M}$ at L. If so, $\mathcal{T}$ rejects and halts. If $\mathcal{T}$ does not reject after $\ell(\epsilon)$ iterations, then $\mathcal{T}$ accepts. The query complexity of the canonical tester is therefore at most $\ell(\epsilon) \cdot k$.

Our main theorem in this section is the following.

Theorem 52 *For a given rank-r matroid $\mathcal{M} = (v_1, \ldots, v_k)$ with each $v_i \in \mathbb{F}_2^r$, suppose there is a one-sided tester for $\mathcal{M}^*$-freeness with query complexity $q(\mathcal{M}, \epsilon)$. Then the canonical tester for $\mathcal{M}^*$-freeness has query complexity at most $O(k \cdot q(\mathcal{M}, \epsilon)^r)$.*

Proof: Since the rank of $\mathcal{M}$ is r, without loss of generality, we assume that $v_1, \ldots, v_r$ are the r basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r$. Thus, any linear transformation $L : \mathbb{F}_2^r \to \mathbb{F}_2^n$ is uniquely determined by $L(v_1), \ldots, L(v_r)$.

Suppose we have a one-sided, possibly adaptive, tester T for $\mathcal{M}$-freeness with query complexity $q(\mathcal{M}, \epsilon)$. We say T operates in *steps*, where at each step $i \in [q(\mathcal{M}, \epsilon)]$, T selects an element y_i from $\mathbb{F}_2^n$ (based on a distribution that depends arbitrarily on internal coin tosses and oracle answers in previous steps) and then queries the oracle for the value of $f(y_i)$.

We convert the tester T into another tester T' that operates as follows. Given oracle access to function $f : \mathbb{F}_2^n \to \{0,1\}^n$, T' first selects, uniformly at random, a *non-singular* linear map $\Pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$, and then invokes the tester T, providing it with $f(\Pi(y))$ whenever it queries for $f(y)$. For convenience the linear map may be generated on-the-fly in the following sense. Suppose in the first $i-1$ queries, T queries $(y_1, \ldots, y_{i-1})$ and T' queries $(x_1, \ldots, x_{i-1})$. Now if T chooses a new point y_i to query, tester T' picks a Π uniformly at random from all non-singular maps that are consistent with all the points queried previously, that is, maps satisfying $\Pi(y_1) = x_1, \ldots, \Pi(y_{i-1}) = x_{i-1}$, and feeds the query result at $\Pi(y_i)$ to the original tester T.

Claim 53 *T' is also a tester for $\mathcal{M}^*$-freeness with the same query complexity as T.*

Proof: This is immediate, since for any function f, f and $f \circ \Pi$ have exactly the same distance from $\mathcal{M}^*$-freeness. ■

For convenience, let us fix the following notation. At a step $i \in [q(\mathcal{M}, \epsilon)]$, the element whose value is requested by T is denoted y_i, and the element of $\mathbb{F}_2^n$ queried by T' (and whose value is supplied to T) is denoted x_i. Both x_i and y_i are of course random variables, and also $x_i = \Pi(y_i)$. We now make the simple observation that at each step, no matter how cleverly T selects the y_i's, each x_i is either uniformly distributed outside or lies inside the span of elements selected at previous steps. More precisely:

Lemma 54 *Fix an integer $i \in [q(\mathcal{M}, \epsilon)]$. Let $y_1, \ldots, y_i$ be the elements in $\mathbb{F}_2^n$ requested by T in the first i stages, and elements $x_1, \ldots, x_{i-1}$ be the points queried by T' in the first $i-1$ steps. Then, conditioned on these, x_i, the element queried by T' at the i^{th} step, is either an element in* $\text{span}(x_1, \ldots, x_{i-1})$ *or is uniformly distributed in* $\mathbb{F}_2^n - \text{span}(x_1, \ldots, x_{i-1})$.

Due to Lemma 54, we may divide the queries of T into two types: *staying query* if the newly queried point is in the span of the previously queried points, and *expanding query* if the newly queried point is a random point outside the span of previously queried points. Let the number of expanding queries of T' be t, $t \leq q(\mathcal{M}, \epsilon)$ and let the subspace spanned by $(x_1, \ldots, x_{q(\mathcal{M},\epsilon)})$ be $V_{T'}$, then clearly $\dim(V_{T'}) = t$ and the expanding query points generate $V_{T'}$ (i.e., the set of expanding queries $(x_{i_1}, \ldots, x_{i_t})$ form a basis for $V_{T'}$). Therefore, as a corollary to Lemma 54, we have the following property of $V_{T'}$.

Corollary 55 *The subspace $V_{T'}$ spanned by the query points of tester T' is a random subspace of dimension t in $\mathbb{F}_2^n$.*

Next, suppose we remove the conditioning on the elements selected by T' in Lemma 54. Then, the algebraic structure of the domain allows us to prove the following:

Lemma 56 *For any integer $i \in [q_{\mathcal{M}}(\epsilon)]$, the element x_i queried by T' at the i'th step, conditioned on being nonzero, is uniformly distributed on the nonzero elements of $\mathbb{F}_2^n$.*

Proof: The proof is by induction on i. We show that for each value of i, not only is x_i uniformly distributed on the non-zero elements of $\mathbb{F}_2^n$, but also any linear combination of $x_1, \ldots, x_i$ is uniformly distributed on the non-zero elements of $\mathbb{F}_2^n$. For $i = 1$, Lemma 54 directly implies that x_1 is uniform on $\mathbb{F}_2^n - \{0\}$. Now consider $i > 1$ and assume our conclusion holds for smaller i. Fix a choice of the elements $y_1, \ldots, y_i$ selected by T in the first i steps. Consider some linear combination $z = \sum_{j=1}^{i} c_j x_j$ which we need to show is uniformly distributed on $\mathbb{F}_2^n - \{0\}$. Assume $c_i = 1$ (otherwise, we are done directly by the induction hypothesis). If y_i is a linear combination of $y_1, \ldots, y_{i-1}$, then x_i, and so z, is also a linear combination of $x_1, \ldots, x_{i-1}$, which is then uniformly distributed in $\mathbb{F}_2^n - \{0\}$ by the induction hypothesis. Otherwise, y_i is not in the span of $y_1, \ldots, y_{i-1}$ and because the only randomness remaining is in choosing L, x_i is chosen uniformly at random from $\mathbb{F}_2^n - \mathsf{span}(x_1, \ldots, x_{i-1})$. By Corollary 55, $\mathsf{span}(x_1, \ldots, x_{i-1})$ is a uniformly chosen subspace of $\mathbb{F}_2^n$ of dimension d, for $d = \dim(\mathsf{span}(y_1, \ldots, y_{i-1}))$. Therefore, x_i itself is uniformly distributed over $\mathbb{F}_2^n - \{0\}$. Moreover, z is uniformly distributed over $\mathbb{F}_2^n - \{0\}$ because z, like x_i, is also a uniformly chosen element of $\mathbb{F}_2^n - \mathsf{span}(x_1, \ldots, x_{i-1})$. ■

We will actually need the following more general lemma.

Lemma 57 *For any r-tuple $(i_1, \ldots, i_r) \in [q_{\mathcal{M}}(\epsilon)]^r$, if the r-tuple $(x_{i_1}, \ldots, x_{i_r})$, where x_i is the element queried by T' at the i'th step, is conditioned on being linearly independent, then it is uniformly distributed among the linearly independent r-tuples in $(\mathbb{F}_2^n)^r$.*

Proof: We can prove a stronger claim. Let $\ell = q(\mathcal{M}, \epsilon)$, let $L_1, \ldots, L_r : (\mathbb{F}_2^n)^\ell \to \mathbb{F}_2^n$ be arbitrary $\mathbb{F}_2$-linear maps and let $x = (x_1, \ldots, x_\ell)$. Then, we show that the tuple $(L_1(x), \ldots, L_r(x))$ is uniformly distributed over linearly independent r-tuples in $(\mathbb{F}_2^n)^r$.

Fix the choice of internal randomness ρ for T and the elements $y_1, \ldots, y_\ell$ selected by T in the ℓ steps. We can represent the linear maps $L_1, \ldots, L_r$ as an r-by-ℓ matrix

A over $\mathbb{F}_2$, where the i'th row contains the coefficients corresponding to L_i. Next, find a minimal subset $S \subseteq [\ell]$ with size s such that $\mathsf{span}(\{y_j : j \in S\}) = \mathsf{span}(\{y_1, \dots, y_\ell\})$; so, the elements of $\{y_j : j \in S\}$ are linearly independent. For any $i \notin S$, one must be able to express y_i as a linear combination of elements from $\{y_j : j \in S\}$, and this same linear combination suffices to express x_i in terms of elements from $\{x_j : j \in S\}$. Let B be the ℓ-by-ℓ matrix with entries $b_{i,j}$ where $x_i = \sum_{j \in S} b_{i,j} x_j$. Let $C = AB$ with entries $c_{i,j}$. By removing zero columns, we can make sure that C is an r-by-s matrix over $\mathbb{F}_2$.

From Lemma 54, we know that the s-tuple $x' = (x_i : i \in S)$ is a uniformly chosen random linearly independent s-tuple in $(\mathbb{F}_2^n)^s$. If C is not full rank, then Cx' is a linearly dependent r-tuple. Otherwise, because C is full-rank, Cx' is also a uniformly chosen linearly independent r-tuple in $(\mathbb{F}_2^n)^r$, proving our claim. ■

By Lemma 50, T' rejects if and only if it detects a violating full-rank linear transformation. In other words, T' rejects iff it finds a linearly independent r-tuple $z = (z_1, \dots, z_r)$ such that $f(\langle v_i, z \rangle) = 1$ for all $i \in [k]$. Furthermore, because $v_1 = \mathbf{e}_1, \dots, v_r = \mathbf{e}_r$, the elements $z_1, \dots, z_r$ must lie in the set of samples made by T'. Then, since T' makes $q(\mathcal{M}, \epsilon)$ queries, the total number of linearly independent r-tuples T' can check is at most $q(\mathcal{M}, \epsilon) \cdot (q(\mathcal{M}, \epsilon) - 1) \cdots (q(\mathcal{M}, \epsilon) - r + 1) < q(\mathcal{M}, \epsilon)^r$. Let δ be the fraction of violating linearly independent r-tuples $z = (z_1, \dots, z_r) \in (\mathbb{F}_2^n)^r$. By Lemma 57, each linearly independent r-tuple checked by T' is drawn uniformly at random from the set of all linearly independent r-tuples in $(\mathbb{F}_2^n)^r$. That is, the probability that T' rejects after checking any non-singular linear transformation it inspects is exactly δ. By union bound, the probability that T' rejects $(f_1, \dots, f_k)$ after $q(\mathcal{M}, \epsilon)$ queries is at most $\delta q(\mathcal{M}, \epsilon)^r$. In order to reject with probability at least $2/3$, the query complexity of T' is at least $q(\mathcal{M}, \epsilon) \geq (\frac{2}{3\delta})^{1/r}$. Now consider the canonical tester T'' that runs in ℓ independent stages which, at each stage, selects uniformly at random a linearly independent r-tuple $(z_1, \dots, z_r)$ and checks for violation of $\mathcal{M}^*$-freeness. How many queries does T'' need to make to achieve the same rejection probability on $(f_1, \dots, f_k)$ as T' does after $q(\mathcal{M}, \epsilon)$ queries? Clearly the probability that T'' rejects $(f_1, \ldots, f_k)$ after ℓ stages is $1 - (1-\delta)^\ell \geq 2/3$, for all $\ell \geq \ell_0 = \frac{2}{\delta} = O(q(\mathcal{M}, \epsilon)^r)$. Since T'' makes k queries in each stage, the total number of queries T'' makes is at most $k\ell_0 = O(k \cdot q(\mathcal{M}, \epsilon)^r)$. ■

Chapter 5

One-sided Testability and Subspace Hereditariness

5.1 Oblivious Testability

We now turn to showing Theorem 12 which states that for linear-invariant properties, testability with a one-sided error oblivious tester is equivalent to the property being semi subspace-hereditary (recall here Definition 11).

First we formalize the discussion from the introduction regarding the fact that it is always possible to assume that the testing algorithm for a one-sided testable linear-invariant property makes its decision only by querying the input function on a random linear subspace of constant dimension.

Proposition 58 *Let $\mathcal{P}$ be a linear invariant property, and let T be an arbitrary one-sided tester for $\mathcal{P}$ with query complexity $d(\epsilon, n)$. Then, there exists a one-sided tester T' for $\mathcal{P}$ that selects a random subspace H of dimension $d(\epsilon, n)$, queries the input on all points of H, and decides based on the oracle answers, the value of ϵ and n, and internal randomness*[1]*. Note that T' is non-adaptive and has query complexity $2^{d(\epsilon,n)}$.*

[1] Note here, we leave open the possibility that the decision of the tester may not be based only on properties of the selected subspace. This gap can be resolved using the same techniques as used by [GT03] for the graph case, but this point is not relevant for our purposes and so we do not elaborate more here.

Proof: Consider a tester T_2 that acts as follows. If the tester T on the input makes queries $x_1, \ldots, x_d$, then T_2 queries all points in $\text{span}(x_1, \ldots, x_d)$ but makes its decision based on $x_1, \ldots, x_d$ just as T does. Clearly, T_2 is also a one-sided tester for $\mathcal{P}$ and with query complexity at most $2^{d(\epsilon)}$.

Now, define a tester T' as follows. Given oracle access to a function $f : \mathbb{F}_2^n \to \{0,1\}$, T' first selects uniformly at random a non-singular linear transformation $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$, and then invokes T_2 providing it with oracle access to the function $f \circ L$. That is, when T_2 makes query x, then algorithm T' makes query $L(x)$. We argue that the sequence of queries made by T' are the elements of a uniformly chosen random subspace of dimension at most $d(\epsilon)$. To see this, fix the input f and the randomness of T_2. Then, for each $i \in [2^{d(\epsilon)}]$ for which the i'th query, x_i, made by T_2 is linearly independent of the previous $i-1$ queries, $x_1, \ldots, x_{i-1}$, it's the case that $L(x_i)$ is a uniformly chosen random element from outside $\text{span}(L(x_1), \ldots, L(x_{i-1}))$. So, for every fixing of the random coins of T_2, the queries made by T' span a uniformly chosen subspace of dimension at most $d(\epsilon)$, and hence, this is also the case when the coins are not fixed. T' is a one-sided tester for $\mathcal{P}$ because if $f \in \mathcal{P}$, then $f \circ L \in \mathcal{P}$ by linear invariance, and if f is ϵ-far from $\mathcal{F}$, then $f \circ L$ is also ϵ-far from $\mathcal{P}$ because L is a permutation on $\mathbb{F}_2^n$. ■

An oblivious tester, as defined in Definition 10, differs from the tester T' of the above proposition in that the dimension of the selected subspace and the decision made by the tester are not allowed to depend on n. As argued there, it is very reasonable to expect natural linear-invariant properties to have such testers, and indeed, prior works have already implicitly restricted themselves in this way.

We can now proceed with the proof of Theorem 12.

Proof of Theorem 12: Let us first prove the forward direction of the theorem. Note that for this direction, we do not need to assume the truth of Conjecture 2. Given a linear-invariant property $\mathcal{P}$ that can be tested with one-sided error by an oblivious tester, we will build a subspace-hereditary property $\mathcal{H}$ containing $\mathcal{P}$, by identifying a (possibly infinite) collection of matrices M^i and binary strings σ^i such that $\mathcal{H}$ is equivalent to the property of being $\{(M^i,\sigma^i)\}_i$- free.

Let $\mathcal{S}$ consist of the pairs (H,S), where H is a subspace of $\mathbb{F}_2^n$ and $S \subseteq H$ is a subset, that satisfy the following two properties: (1) $\dim(H) = d(\epsilon)$ for some ϵ, and (2) if for this ϵ, the tester rejects its input with some positive probability when the evaluation of its input on the sampled subspace is $\mathbf{1}_S$. For $(H,S) \in \mathcal{S}$ let $d = \dim(H)$. Consider the matrix A_H over $\mathbb{F}_2$ with each row representing an element of H in some fixed basis. Notice that A_H is a $(2^\ell \times \ell)$-sized matrix. Define M_H, a matrix over $\mathbb{F}_2$ of size $(2^\ell - \ell) \times 2^\ell$, such that $M_H A_H = 0$. Finally, for each $i \in [2^\ell]$ define $\sigma_S(i) = \mathbf{1}_S(x_i)$, where x_i is the element represented in the i'th row of A_H. Let $\mathcal{M}$ be the set of pairs (M_H, σ_S) obtained in this way from every $(H,S) \in \mathcal{S}$.

We now proceed to verify that $\mathcal{H}$ satisfies the conditions of Definition 11. To show that $\mathcal{P}$ is $\mathcal{M}$-free, let $f \in \mathcal{P}_n$, and suppose that there exists $(M_H, \sigma_S) \in \mathcal{M}$ such that $(M_H, \sigma_S) \mapsto f$, for some ϵ, and for some H with $\dim(H) = d(\epsilon)$ and $S \subseteq H$. We show that f is rejected with some positive probability, a contradiction to the fact that the test is one-sided. If (M_H, σ_S) is induced by f at $(x_1, \ldots, x_{2^{d(\epsilon)}})$, then these elements necessarily span a $d(\epsilon)$-dimensional subspace so that the function restricted to that subspace is $\mathbf{1}_S \circ L$ for some linear transformation $L : \mathbb{F}_2^n \to \mathbb{F}_2^{d(\epsilon)}$ (determined by the choice of basis that was used to represent H). Thus, this immediately implies by the definition of (M_H, σ_S) that the tester rejects f with positive probability.

To verify the second part of the Definition 11, let $M(\epsilon) = d(\epsilon)$. Suppose $f : \mathbb{F}_2^n \to \{0,1\}$, with $n > M(\epsilon)$ is ϵ-far from satisfying $\mathcal{P}$. In this case, in order for the tester to reject f with positive probability, it must select a $d(\epsilon)$-dimensional subspace H so that the restriction to H equals the indicator function on S (upto a linear transformation), for some $(H,S) \in \mathcal{S}$. Therefore T is not $\mathcal{M}$-free, and thus $T \notin \mathcal{H}$.

It remains to show the opposite direction of Theorem 12. We here assume Conjecture 2 that every subspace-hereditary property $\mathcal{P}$ is testable by a one-sided tester. Our first observation that, in this case, it is actually testable by an *oblivious* one-sided tester. Namely, we show that the clearly oblivious tester, which checks whether the input function restricted to a random linear subspace satisfies $\mathcal{P}$ or not, is a valid tester. We need to argue that if a non-oblivious tester rejects input f that is ϵ-far from $\mathcal{P}$ by querying its values on a random $d(\epsilon)$-dimensional subspace (we already know the tester is of this type from Proposition 58), then with high probability, the input function restricted to a random $3d(\epsilon)$-dimensional subspace does not satisfy the property $\mathcal{P}$. Suppose it did. But then, if the original tester first uniformly selected a $3d(\epsilon)$-dimensional subspace H and then uniformly selected a $d(\epsilon)$-dimension subspace H' inside it, and ran its decision based on $f|_{H'}$, it will accept the input with large probability, which is a contradiction to the soundness of the tester since H'' is a uniformly distributed $d(\epsilon)$-dimensional subspace. Thus, for a testable subspace-hereditary property, we can assume that the tester simply checks for $\mathcal{P}$ on the sampled subspace, and is hence, oblivious to the value of n. This argument is analogous to one of Alon for graph properties, reported in [GT03].

Now, assuming that every subspace-hereditary property is testable by an oblivious one-sided tester (Conjecture 2), we wish to show that every semi subspace-hereditary property is testable by an oblivious one-sided tester. Let $\mathcal{P}$ be a a semi subspace-hereditary property and let $\mathcal{H}$ be the subspace-hereditary property associated to $\mathcal{P}$ in Definition 11. By our assumption, $\mathcal{H}$ has a one-sided tester T', which on input ϵ makes $Q'(\epsilon)$ queries and rejects inputs ϵ-far from $\mathcal{H}$ with probability 2/3. The tester T for $\mathcal{P}$ makes $Q(\epsilon) = \max(Q'(\epsilon/2), 2^{M(\epsilon/2)})$ queries (where $M(\cdot)$ comes from Definition 11) and proceeds as follows. If the size of the input is at most $Q(\epsilon)$, then by definition, T receives the evaluation of the function all of the input and in this case, it simply checks if the input belongs to $\mathcal{P}$. Otherwise T emulates T' with distance parameter $\epsilon/2$ and accepts if and only if T' accepts.

Notice that T is one-sided. Indeed, if the input f satisfies $\mathcal{P}$ then $f \in \mathcal{H}$ and thus T' always accepts, causing T to always accept. To prove soundness, we first argue that if f is ϵ-far from $\mathcal{P}$ then it is $\epsilon/2$-far from $\mathcal{H}$. Suppose otherwise, and modify f in at most an $\epsilon/2$ fraction of the domain in order to obtain a function $g \in \mathcal{H}$. Thus g is still $\epsilon/2$-far from $\mathcal{P}$, and by Definition 11 $g \notin \mathcal{H}$, a contradiction. Finally, since f is $\epsilon/2$-far from $\mathcal{H}$ and since T' mistakenly accepts such inputs with probability at most 1/3 so does T'. ■

5.2 Representing Subspace-Hereditary Properties by Local Constraints

Below, we prove Proposition 5 that subspace-hereditary properties exactly coincide with the $\mathcal{F}$-freeness properties from Definition 7.

Proof of Proposition 5: In one direction, it is easy to check that $\mathcal{F}$-freeness is a subspace-hereditary linear-invariant property, for any fixed family $\mathcal{F}$.

Now, we show the other direction. For a subspace-hereditary linear-invariant property $\mathcal{P}$, let Obs denote the collection of pairs (d, S), where $d \geq 1$ is an integer and $S \subseteq \mathbb{F}_2^d$ is a subset, such that $\mathbf{1}_S$ does not have property $\mathcal{P}$ and is minimal with respect to restriction to subspaces. In other words, (d, S) is contained in Obs iff $\mathbf{1}_S \notin \mathcal{P}_d$ but for any vector subspace $U \subseteq \mathbb{F}_2^d$ of dimension $d' < d$, $\mathbf{1}_{S|_U} \in \mathcal{P}_{d'}$ where $S|_U \subseteq U$ is the restriction of S to U.

For every $(d, S) \in \mathsf{Obs}$, we construct a matrix M_d and a tuple σ_S such that any f with property $\mathcal{P}$ is (M_d, σ_S)-free. Define A_d to be the 2^d-by-d matrix over $\mathbb{F}_2$, where each of the 2^d rows corresponds to a distinct element of $\mathbb{F}_2^d$ represented using some choice of bases. Now, define M_d to be a $(q^d - d)$-by-q^d matrix over $\mathbb{F}$, such that $M_d A_d = 0$ and $\mathsf{rank}(M_d) = q^d - d$. Define σ_S as $(\sigma(1), \sigma(2), \ldots, \sigma(2^d))$ where $\sigma(i) = \mathbf{1}_S(x_i)$ with x_i being the element of $\mathbb{F}_2^d$ represented in the ith row of A_d. We observe now that any $f : \mathbb{F}_2^n \to \{0, 1\}$ having property $\mathcal{P}$ is (M_d, σ_S)-free. Suppose the opposite, so that there exists $x = (x_1, \ldots, x_{q^d}) \in (\mathbb{F}_2^n)^d$ satisfying $Mx = 0$ and $f(x_i) = \sigma(i)$. Then, by definition of M_d, the $x_1, \ldots, x_{2^d}$ are the elements of a d-dimensional subspace V over $\mathbb{F}_2$, and by definition of σ_S, $S_f|_V = S$ where S_f is the support of f. Thus $f|_V \notin \mathcal{P}$ which is a contradiction to the fact that f has property $\mathcal{P}$ because $\mathcal{P}$ is subspace-hereditary.

Finally, define $\mathcal{F}_\mathcal{P} = \{(M_d, \sigma_S)\}$. We have just seen that any f having property $\mathcal{P}$ is $\mathcal{F}_\mathcal{P}$-free. On the other hand, suppose f does not have property $\mathcal{P}$. Then, because of heredity, there must be a d-dimensional subspace V such that the support of $f|_V$ is isomorphic to S for some $(d, S) \in \mathsf{Obs}$ under linear transformations, which means by the same argument as above, that f will not be (M_d, σ_S)-free. ■

Bibliography

[AFKS00] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[AKK$^+$05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Transactions on Information Theory*, 51(11):4032–4039, 2005.

[Alo02] Noga Alon. Testing subgraphs in large graphs. *Random Structures and Algorithms*, 21(3-4):359–370, 2002.

[AS08a] Noga Alon and Asaf Shapira. A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. on Comput.*, 37(6):1703–1727, 2008.

[AS08b] Noga Alon and Asaf Shapira. Every monotone graph property is testable. *SIAM J. on Comput.*, 38(2):505–522, 2008.

[AS08c] Noga Alon and Asaf Shapira. A separation theorem in property testing. *Combinatorica*, 28:261–281, 2008.

[AT08] Tim Austin and Terence Tao. On the testability and repair of hereditary hypergraph properties. *Random Structures and Algorithms (to appear)*, 2008. Preprint available at http://arxiv.org/abs/0801.2179.

[BC87] Thomas Bailey and John Cowles. A convex hull inclusion test. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2):312–316, 1987.

[BCSX09] Arnab Bhattacharyya, Victor Chen, Madhu Sudan, and Ning Xie. Testing linear-invariant non-linear properties. In *STACS*, pages 135–146, 2009. Full version at http://www.eccc.uni-trier.de/report/2008/088/.

[BFL91] László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991.

[BGRS11] Arnab Bhattacharyya, Elena Grigorescu, Prasad Raghavendra, and Asaf Shapira. Testing odd-cycle-freeness of boolean functions. *Electronic Colloquium in Computational Complexity*, TR11-075, May 2011.

[BGS10] Arnab Bhattacharyya, Elena Grigorescu, and Asaf Shapira. A unified framework for testing linear-invariant properties. In *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 478–487. IEEE Computer Society, 2010.

[BHR05] Eli Ben-Sasson, Prahladh Harsha, and Sofya Raskhodnikova. Some 3cnf properties are hard to test. *SIAM J. on Comput.*, 35(1):1–21, 2005.

[BKS$^+$09] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of Reed-Muller codes. *Electronic Colloquium in Computational Complexity*, TR09-086, October 2009.

[Bla09] Eric Blais. Testing juntas nearly optimally. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 151–158, 2009.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comp. Sys. Sci.*, 47:549–595, 1993. Earlier version in STOC'90.

[BO10] Eric Blais and Ryan O'Donnell. Lower bounds for testing function isomorphism. In *Proc. 25th Annual IEEE Conference on Computational Complexity (to appear)*, 2010.

[Bol76] Béla Bollobás. Complete subgraphs are elusive. *J. Comb. Thy. Ser. B*, 21(1):1 – 7, 1976.

[BvEBL74] M.R. Best, P. van Emde Boas, and H.W. Lenstra. A Sharpened Version of the Aanderaa-Rosenberg Conjecture. Technical Report ZW 30/74, Mathematisch Centrum, 1974. Amsterdam, The Netherlands.

[BX10] Arnab Bhattacharyya and Ning Xie. Lower bounds for testing triangle-freeness in boolean functions. In *Proc. 21st ACM-SIAM Symposium on Discrete Algorithms*, pages 87–98, Philadelphia, PA, USA, 2010. Society for Industrial and Applied Mathematics.

[CK07] Amit Chakrabarti and Subhash Khot. Improved lower bounds on the randomized complexity of graph properties. *Random Structures and Algorithms*, 30(3):427–440, May 2007.

[CKS01] Amit Chakrabarti, Subhash Khot, and Yaoyun Shi. Evasiveness of subgraph containment and related properties. *SIAM J. on Comput.*, 31(3):866–875, 2001.

[Coh00] Henri Cohen. *A Course in Computational Algebraic Number Theory.* Springer, 2000.

[CRSW83] C. Chvatál, V. Rödl, E. Szemerédi, and W. T. Trotter Jr. The Ramsey number of a graph with bounded maximum degree. *Journal of Combinatorial Theory, Series B*, 34(3):239–243, 1983.

[DLM$^+$07] Ilias Diakonikolas, Homin K. Lee, Kevin Matulef, Krzysztof Onak, Ronitt Rubinfeld, Rocco A. Servedio, and Andrew Wan. Testing for concise representations. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 549–558, 2007.

[Dom91] Eric Domenjoud. Solving systems of linear diophantine equations: an algebraic approach. In *In Proc. 16th Mathematical Foundations of Computer Science, Warsaw, LNCS 520*, pages 141–150. Springer-Verlag, 1991.

[Fis04] Eldar Fischer. The art of uninformed decisions: A primer to property testing. In G. Paun, G. Rozenberg, and A. Salomaa, editors, *Current Trends in Theoretical Computer Science: The Challenge of the New Century*, volume 1, pages 229–264. World Scientific Publishing, 2004.

[Fis05] Eldar Fischer. The difficulty of testing for isomorphism against a graph that is given in advance. *SIAM J. on Comput.*, 34(5):1147–1158, 2005.

[FKR$^+$04] Eldar Fischer, Guy Kindler, Dana Ron, Shmuel Safra, and Alex Samorodnitsky. Testing juntas. *J. Comp. Sys. Sci.*, 68(4):753–787, 2004.

[FKW02] Ehud Friedgut, Jeff Kahn, and Avi Wigderson. Computing graph properties by randomized subcube partitions. In *APPROX-RANDOM*, pages 105–113, 2002.

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998.

[GOS$^+$09] Parikshit Gopalan, Ryan O'Donnell, Rocco A. Servedio, Amir Shpilka, and Karl Wimmer. Testing Fourier dimensionality and sparsity. In *ICALP (1)*, pages 500–512, 2009.

[Gre05] Ben Green. A Szemerédi-type regularity lemma in abelian groups. *Geometric and Functional Analysis*, 15(2):340–376, 2005.

[Gru07] Peter Gruber. *Convex and Discrete Geometry.* Springer, New York, 2007.

[GT03] Oded Goldreich and Luca Trevisan. Three theorems regarding testing graph properties. *Random Structures and Algorithms*, 23(1):23–57, 2003.

[GT08] Ben Green and Terence Tao. Linear equations in primes. Preprint available at http://arxiv.org/abs/math/0606088v2, April 2008.

[GT10] Ben Green and Terence Tao. An arithmetic regularity lemma, associated counting lemma, and applications. Preprint available at http://arxiv.org/abs/1002.2028, February 2010.

[Haj91] Péter Hajnal. An $\Omega(n^{4/3})$ lower bound on the randomized complexity of graph properties. *Combinatorica*, 11(2):131–143, 1991.

[HW04] Stefan Hougardy and Annegret Wagler. Perfectness is an elusive graph property. *SIAM J. on Comput.*, 34(1):109–117, 2004.

[KS07] Tali Kaufman and Madhu Sudan. Sparse random linear codes are locally decodable and testable. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 590–600, 2007.

[KS08] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 403–412, New York, NY, USA, 2008. ACM.

[KS09] Swastik Kopparty and Shubhangi Saraf. Tolerant linearity testing and locally testable codes. In *APPROX-RANDOM*, pages 601–614, 2009.

[KSV08] Daniel Král', Oriol Serra, and Lluís Vena. A removal lemma for systems of linear equations over finite fields. *Israel Journal of Mathematics (to appear)*, 2008. Preprint available at http://arxiv.org/abs/0809.1846.

[MORS09] Kevin Matulef, Ryan O'Donnell, Ronitt Rubinfeld, and Rocco A. Servedio. Testing halfspaces. In *Proc. 20th ACM-SIAM Symposium on Discrete Algorithms*, pages 256–264, 2009.

[PRS02] Michal Parnas, Dana Ron, and Alex Samorodnitsky. Testing basic boolean formulae. *SIAM J. Discrete Math.*, 16(1):20–46, 2002.

[Ron08] Dana Ron. Property Testing: A Learning Theory Perspective. In *Foundations and Trends in Machine Learning*, volume 1, pages 307–402. 2008.

[Ros73] Arnold L. Rosenberg. On the time required to recognize properties of graphs: a problem. *SIGACT News*, 5:15–16, October 1973.

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. on Comput.*, 25:252–271, 1996.

[RS09] Vojtěch Rödl and Mathias Schacht. Generalizations of the removal lemma. *Combinatorica*, 29(4):467–502, 2009.

[Rub06] Ronitt Rubinfeld. Sublinear time algorithms. In *Proceedings of International Congress of Mathematicians 2006*, volume 3, pages 1095–1110, 2006.

[RV76] Ronald L. Rivest and Jean Vuillemin. On recognizing graph properties from adjacency matrices. *Theoretical Computer Science*, 3(3):371 – 384, 1976.

[Sam07] Alex Samorodnitsky. Low-degree tests at large distances. In *Proc. 37th Annual ACM Symposium on the Theory of Computing*, pages 506–515, 2007.

[Sha09] Asaf Shapira. Green's conjecture and testing linear-invariant properties. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 159–166, 2009.

[Sud10] Madhu Sudan. Invariance in property testing. *Electronic Colloquium in Computational Complexity*, TR10-051, March 2010.

[VX11] Santosh Vempala and Ying Xiao. Structure from local optima: Learning subspace juntas via higher-order PCA. July 2011.

[Yao88] Andrew Chi-Chih Yao. Monotone bipartite graph properties are evasive. *SIAM J. on Comput.*, 17(3):517–520, 1988.