Electronic Colloquium on Computational Complexity, Report No. 88 (2008)

Testing Linear-Invariant Non-Linear Properties

Arnab Bhattacharyya* Victor Chen† Madhu Sudan‡ Ning Xie§

September 16, 2008

Abstract

We consider the task of testing properties of Boolean functions that are invariant under linear transformations of the Boolean cube. Previous work in property testing, including the linearity test and the test for Reed-Muller codes, has mostly focused on such tasks for linear properties. The one exception is a test due to Green for "triangle freeness": A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ satisfies this property if $f(x), f(y), f(x+y)$ do not all equal 1, for any pair $x, y \in \mathbb{F}_2^n$.

Here we extend this test to a more systematic study of testing for linear-invariant non-linear properties. We consider properties that are described by a single forbidden pattern (and its linear transformations), i.e., a property is given by k points $v_1, \ldots, v_k \in \mathbb{F}_2^k$ and $f : \mathbb{F}_2^n \to \mathbb{F}_2$ satisfies the property that if for all linear maps $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ it is the case that $f(L(v_1)), \ldots, f(L(v_k))$ do not all equal 1. We show that this property is testable if the underlying matroid specified by $v_1, \ldots, v_k$ is a graphic matroid. This extends Green's result to an infinite class of new properties.

Our techniques extend those of Green and in particular we establish a link between the notion of "1-complexity linear systems" of Green and Tao, and graphic matroids, to derive the results.

*MIT CSAIL. abhatt@csail.mit.edu. Research supported in part by a DOE Computational Science Graduate Fellowship.

†MIT CSAIL. victor@math.mit.edu. Research supported in part by NSF Award CCR-0514915.

‡MIT CSAIL. madhu@csail.mit.edu. Research supported in part by NSF Award CCR-0514915.

§MIT CSAIL. ningxie@csail.mit.edu. Research supported in part by an Akamai Presidential Fellowship and NSF grant 0514771. Keywords: Property Testing, additive combinatorics, Fourier analysis

ISSN 1433-8092

1 Introduction

Property testing considers the task of testing, "super-efficiently", if a function $f : D \to R$ mapping a finite domain D to a finite range R essentially satisfies some desirable property. Letting $\{D \to R\}$ denote the set of all functions from D to R, a *property* is formally specified by a family $\mathcal{F} \subseteq \{D \to R\}$ of functions. A *tester* has oracle access to the function f and should accept with high probability if $f \in \mathcal{F}$ and reject (also with high probability) functions that are *far* from $\mathcal{F}$, while making very few queries to the oracle for f. Here, distance between functions $f, g : D \to R$, denoted $\delta(f, g)$, is simply the probability that $f(x) \neq g(x)$ when x is chosen uniformly at random from D and $\delta(f, \mathcal{F}) = \min_{g \in \mathcal{F}}\{\delta(f, g)\}$. We say f is δ-far from $\mathcal{F}$ if $\delta(f, \mathcal{F}) \geq \delta$ and δ-close otherwise. The central parameter associated with a tester is the number of oracle queries it makes to the function f being tested. In particular, a property is called *(locally) testable* if there is a tester with query complexity that is a constant depending only on the distance parameter δ. Property testing was initiated by the works of Blum, Luby and Rubinfeld [12] and Babai, Fortnow and Lund [9] and was formally defined by Rubinfeld and Sudan [23]. The systematic exploration of property testing was initiated by Goldreich, Goldwasser, and Ron [14] who expanded the scope of property testing to combinatorial and graph-theoretic properties (all previously considered properties were algebraic). In the subsequent years, a rich collection of properties have been shown to be testable [5, 4, 1, 13, 22, 3, 2, 19, 18] and many property tests have ended up playing a crucial role in constructions of probabilistically checkable proofs [8, 7, 11, 17, 24].

The rich collection of successes in property testing raises a natural question: Why are so many different properties turning out to be locally testable? Are there some broad "features" of properties that make them amenable to such tests? Our work is part of an attempt to answer such questions. Such questions are best understood by laying out broad (infinite) classes of properties (hopefully some of them are new) and showing them to be testable (or characterizing the testable properties within the class). In this paper we introduce a new such class of properties, and show that (1) they are locally testable, and (2) that they contain infinitely many new properties that were not previously known to be testable.

The properties, and our results:

The broad scope of properties we are interested in are properties that view their domain D as a vector space and are invariant under linear transformations of the domain. Specifically, we consider the domain $D = \mathbb{F}_2^n$, the vector space of n-dimensional Boolean vectors, and the range $R = \mathbb{F}_2$. In this setting, a property $\mathcal{F}$ is said to be *linear-invariant* if for every $f \in \mathcal{F}$ and linear map $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$ we have that $f \circ L \in \mathcal{F}$. Specific examples of linear-invariant properties that were previously studied (esp. in the Boolean setting) include that of linearity, studied by Blum et al. [12] and Bellare et al. [10], and the property of being a "moderate-degree" polynomial (aka Reed-Muller codeword) studied by Alon et al. [2][1]. While the tests in the above mentioned works potentially used all features of the property being tested, Kaufman and Sudan [20] show that the testability can be attributed principally to the linear-invariance of the property. However their setting only considers *linear* properties, i.e., $\mathcal{F}$ itself is a vector space over $\mathbb{F}_2$ and this feature plays a key role in their results: It lends an algebraic flavor to all the properties being tested and plays a central role in their analysis.

We thus ask the question: Does linear-invariance lead to testability even when the property $\mathcal{F}$ is not linear? The one previous work in the literature that gives examples of non-linear linear-invariant properties is Green [15] where a test for the property of being "triangle-free" is described. A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is said to be *triangle-free* if for every $x, y \in \mathbb{F}_2^n$ it is the case that at least one of $f(x), f(y), f(x + y)$ does not equal 1. The property of being triangle-free is easily seen to be linear-invariant and yet not linear. Green [15]

[1] In the literature, the term low-degree polynomial is typically used for polynomials whose degree is smaller than the field size. In the work of [2] the degrees considered are larger than the field size, but are best thought of as large constants. The phrase "moderate-degree" above describes this setting of parameters.

shows that the natural test for this property does indeed work correctly, though the analysis is quite different from that of typical algebraic tests and is more reminiscent of graph-property testing. In particular, Green develops an algebraic regularity lemma to analyze this test. (We note that the example above is not the principal objective of Green's work, which is directed mostly at abelian groups D and R. The above example with $D = \mathbb{F}_2^n$ and $R = \mathbb{F}_2$ is used mainly as a motivating example.)

Motivated by the above example, we consider a broad class of properties that are linear-invariant and non-linear. A property in our class is given by k vectors $v_1, \ldots, v_k$ in the k-dimensional space $\mathbb{F}_2^k$. (Throughout this paper we think of k as a constant.) These k vectors uniformly specify a family $\mathcal{F} = \mathcal{F}_{n;v_1,\ldots,v_k}$ for every positive integer n, containing all functions that, for every linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ take on the value 0 on at least one of the points $L(v_1), \ldots, L(v_k)$. (In Appendix C we consider an even more generalized class of properties where the forbidden pattern of values for f is not 1^k but some other string and show a limited set of cases where we can test such properties.) To see that this extends the triangle-freeness property, note that triangle-freeness is just the special case with $k = 3$ and $v_1 = \langle 100 \rangle$, $v_2 = \langle 010 \rangle$, $v_3 = \langle 110 \rangle$. Under different linear transforms, these three points get mapped to all the different triples of the form $x, y, x + y$ and so $\mathcal{F}_{n;v_1,v_2,v_3}$ equals the class of triangle-free functions.

Before giving a name to our class of functions, we make a quick observation. Note that the property specified by $v_1, \ldots, v_k$ is equivalent to the property specified by $T(v_1), \ldots, T(v_k)$ where T is a non-singular linear map from $\mathbb{F}_2^k \to \mathbb{F}_2^k$. Thus the property is effectively specified by the dependencies among $v_1, \ldots, v_k$ which are in turn captured by the matroid[2] underlying $v_1, \ldots, v_k$. This leads us to our nomenclature:

Definition 1.1 *Given a (binary, linear) matroid $\mathcal{M}$ represented by vectors $v_1, \ldots, v_k \in \mathbb{F}_2^k$, the property of being $\mathcal{M}$-*free *is given by, for every positive integer n, the family*

$$\mathcal{F}_{\mathcal{M}} = \{f : \mathbb{F}_2^n \to \mathbb{F}_2 | \forall \text{ linear } L : \mathbb{F}_2^k \to \mathbb{F}_2^n, \langle f(L(v_1)), \ldots, f(L(v_k)) \rangle \neq 1^k\}.$$

The property of being $\mathcal{M}$-free has a natural k-local test associated with it: Pick a random linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ and test that $\langle f(L(v_1)), \ldots, f(L(v_k)) \rangle \neq 1^k\}$. Analyzing this test turns out to be non-trivial, and indeed we only manage to analyze this in special cases.

Recall that a matroid $\mathcal{M} = \{v_1, \ldots, v_k\}$, $v_i \in \mathbb{F}_2^k$, forms a *graphic matroid* if there exists a graph G on k edges with the edges being associated with the elements $v_1, \ldots, v_k$ such that a set $S \subset \{v_1, \ldots, v_k\}$ has a linear dependency if and only if the associated set of edges contains a cycle. In this paper, we require that the graph G be simple, that is, without any self-loops or parallel edges. Our main theorem shows that the property $\mathcal{F}$ associated with a graphic matroid $v_1, \ldots, v_k \in \mathbb{F}_2^k$ is testable.

Theorem 1.1 *For a graphic matroid $\mathcal{M}$, the property of being $\mathcal{M}$-free is locally testable. Specifically, let $\mathcal{M} = \{v_1, \ldots, v_k\}$ be a graphic matroid. Then, there exists a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ and a k-query tester that accepts members of $\mathcal{M}$-free functions with probability one and rejects functions that are ϵ-far from being $\mathcal{M}$-free with probability at least $\tau(\epsilon)$.*

Our bound on τ is quite weak. We let $W(t)$ denote a tower of twos with height $\lceil t \rceil$. Our proof only guarantees that $\tau(\epsilon) \geq \frac{1}{W(poly(1/\epsilon))}$, a rather fast vanishing function. We do not know if such a weak bound is required for any property we consider.

We describe the techniques used to prove this theorem shortly (which shed light on why our bound on τ is so weak) but first comment on the implications of the theorem . First, note that for a graphic matroid it is more natural to associate the property with the underlying graph. We thus use the phrase G-free to denote

[2]For the sake of completeness we include a definition of matroids in Appendix A. However a reader unfamiliar with this notion may just use the word matroid as a synonym for a finite collection of binary vectors, for the purposes of reading this paper.

the property of being $\mathcal{M}$-free where $\mathcal{M}$ is the graphic matroid of G. This terminology recovers the notion of being triangle-free, as in [15], and extends to cover the case of being k-cycle free (also considered in [15]). But it includes every other graph too!

Syntactically, Theorem 1.1 seems to include infinitely many new properties (other than being k-cycle free). However, this may not be true semantically. For instance the property of being triangle-free is essentially the same as being G-free for every G whose biconnected components are triangles. Indeed, prior to our work, it was not even explicitly noted whether being C_k-free is essentially different from being triangle-free. (By "essentially", we ask if there exist triangle-free functions that are *far* from being C_k-free.) It actually requires careful analysis to conclude that the family of properties being tested include (infinitely-many) new ones. Our second theorem addresses this point.

Theorem 1.2 *The class of G-free properties include infinitely many distinct ones. In particular:*

1. *For every odd k, if f is C_{k+2}-free, then it is also C_k-free. Conversely, there exist functions g that are C_k-free but far from being C_{k+2}-free.*
2. *If $k \leq \ell$ and f is K_k-free, then it is also K_ℓ-free. On the other hand, if $k \geq 3$ and $\ell \geq \binom{k}{2} + 2$ then there exists a function g that is K_ℓ-free but far from being K_k-free.*

Techniques:

Our proof of Theorem 1.1 is based on Green [15]'s analysis of the triangle-free case. To analyze the triangle-free case, Green develops a "regularity" lemma for groups, which is analogous to Szemerédi's regularity lemma for graphs. In our setting, Green's regularity lemma shows how, given any function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, one can find a subgroup H of $\mathbb{F}_2^n$ such that the restriction of f to almost all cosets of H is "regular", where "regularity" is defined based on the "Fourier coefficients" of f. (These notions are made precise in Section 3.1.)

This lemma continues to play a central role in our work as well, but we need to work further on this. In particular, a priori it is not clear how to use this lemma to analyze $\mathcal{M}$-freeness for *arbitrary* matroids $\mathcal{M}$. To extract a large feasible class of matroids we use a notion from a work of Green and Tao [16] of the complexity of a linear system (or matroids, as we refer to them). The "least complex" matroids have complexity 1, and we show that the regularity lemma can be applied to all matroids of complexity 1 to show that they are testable (see Section 3).

The notion of a 1-complex matroid is somewhat intricate, and a priori it may not even be clear that this introduces new testable properties. We show (in Section 4) that these properties actually capture all graphic matroids which is already promising. Yet this is not a definite proof of novelty, and so in Section 5 we investigate properties of graphic matroids and give some techniques to show that they are "essentially" different. Our proofs show that if two (binary) matroids are not "homomorphically" equivalent (in a sense that we define) then there is an essential difference between the properties represented by them.

Significance of problems/results:

We now return to the motivation for studying $\mathcal{M}$-free properties. Our interest in these families is mathematical. We are interested in broad classes of properties that are testable; and invariance seems to be a central notion in explaining the testability of many interesting properties. Intuitively, it makes sense that the symmetries of a property could lead to testability, since this somehow suggests that the value of a function at any one point of the domain is no more important than its values at any other point. Furthermore this intuition is backed up in many special cases like graph-property testing (where the family is invariant under all permutations of the domain corresponding to relabelling the vertex names). Indeed this was what led

Kaufman and Sudan [20] to examine this notion explicitly in the context of algebraic functions. They considered families that were linear-invariant and *linear*, and our work is motivated by the quest to see if the latter part is essential.

In contrast to other combinatorial settings, linear-invariance counts on a (quantitatively) very restricted collection of invariances. Indeed the set of linear transforms is only quasi-polynomially large in the domain (which may be contrasted with the exponentially large set of invariances that need to hold for graph-properties). So ability to test properties based on this feature is mathematically interesting and leads to the question: what kind of techniques are useful in these settings. Our work manages to highlight some of those (in particular, Green's regularity lemma).

Parallel Works:

After completing our work, we learned from Asaf Shapira that, independently of us, $\mathcal{M}$-freeness for an arbitrary matroid $\mathcal{M}$ has been shown to be testable in Shapira's recent preprint [25]. His result solves a question that we posed as open in an earlier version of this paper. His result is built on the work of Král', Serra, and Vena in [21], where an alternate proof of Green's cycle-freeness result is provided. Essentially the authors in [21] demonstrate a reduction from testing freeness of the cycle matroid in a function to testing freeness of the cycle subgraph in a graph, and then they apply regularity lemmas for graphs to analyze the number of cycles in a function far from being cycle-free. In this manner, the authors show that Theorem 1.1 holds as well. By extending this method and utilizing hypergraph regularity lemmas, Shapira [25] shows that arbitrary monotone matroid-freeness properties are testable.

We remark that our proofs are very different from [21] and [25], and in particular, our view on invariance leads us to develop techniques to show that syntactically different properties are indeed distinct.

Organization of this paper:

In the following section (Section 2) we define a slightly broader class of properties that we can consider (including some non-monotone properties). We also define the notion of 1-complexity matroids which forms a central tool in our analysis of the tests. In Section 3 we show that for any 1-complexity matroid $\mathcal{M}$, $\mathcal{M}$-freeness is testable. In Section 4 we show that graphic matroids are 1-complexity matroids. Theorem 1.1 thus follows from the results of Section 3 and 4. In Section 5 we prove that there are infinitely many distinct properties among G-free properties. Finally, in Appendix C, we include results on testing some non-monotone properties, along with some "collapse" results showing that many non-monotone properties collapse to some simple ones in Appendix D.

2 Additional Definitions, Results, and Overview of Proofs

In this section, we describe some further results that we present in the paper and give an outline of proofs.

2.1 Extensions to Non-Monotone families

We start with a generalization of Definition 1.1 to a wider collection of forbidden patterns.

Definition 2.1 *Given $\Sigma \in \mathbb{F}_2^k$ and a binary matroid $\mathcal{M}$ represented by vectors $v_1, \ldots, v_k \in \mathbb{F}_2^k$, the property of being $(\mathcal{M}, \Sigma)$-*free *is given by, for every positive n, the family $\mathcal{F}_{(\mathcal{M},\Sigma)} = \{f : \mathbb{F}_2^n \to \mathbb{F}_2 | \forall \text{ linear } L : \mathbb{F}_2^k \to \mathbb{F}_2^n, \langle f(L(v_1)), \ldots, f(L(v_k))\rangle \neq \Sigma\}$.*

If for some linear $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$, $\langle f(L(v_1)), \ldots, f(L(v_k))\rangle = \Sigma$, then we say f *contains* $(\mathcal{M}, \Sigma)$ *at* L. Also, to be consistent with Definition 1.1, we suppress mention of Σ when $\Sigma = 1^k$.

Recall that a property $\mathcal{P} \subseteq \{D \to \{0,1\}\}$ is said to be *monotone* if $f \in \mathcal{P}$ and $g \prec f$ implies $g \in \mathcal{P}$, where $g \prec f$ means that $g(x) \leq f(x)$ for all $x \in D$.

Observation 2.1 *For a binary matroid $\mathcal{M}$, $(\mathcal{M}, \Sigma)$-freeness is a monotone property if and only if $\Sigma = 1^k$.*

In addition to our main results (Theorems 1.1 and 1.2) on monotone properties, we also obtain local testability results for a limited class of non-monotone properties.

Theorem 2.2 *Let C_k denote the cycle on k vertices and let Σ be an arbitrary element of $\mathbb{F}_2^k$. Then, there exists a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ and a k-query tester that accepts members of $\mathcal{F}_{(C_k,\Sigma)}$ with probability 1 and rejects f that are ϵ-far from $\mathcal{F}_{(C_k,\Sigma)}$ with probability at least $\tau(\epsilon)$.*

However, in strong contrast to Theorem 1.2, we show that unless Σ equals 0^k or 1^k, the class of (C_k, Σ)-freeness properties is not at all very rich semantically.

Theorem 2.3 *The class of properties $\{\mathcal{F}_{(C_k,\Sigma)} : k \geq 3, \Sigma \neq 0^k, \Sigma \neq 1^k\}$ is only finitely large.*

The goal of Theorem 2.2 is not to introduce new testable properties but rather to illustrate possible techniques for analyzing local tests that may lead to more classes of testable non-monotone properties.

2.2 Overview of Proofs

We now give an outline of the proofs of our main theorems (Theorems 1.1 and 1.2), and also the extensions (Theorems 2.2 and 2.3).

Our claim in Theorem 1.1, that graphic matroid freeness properties are locally testable, is based on analyzing the structure of dependencies among elements of a graphic matroid. To this end, we first recall the classification of linear forms due to Green and Tao in [16]. We require a minor reformulation of their definition since, for us, the structure of the linear constraints is described by elements of a matroid.

Definition 2.2 *Given a binary matroid $\mathcal{M}$ represented by $v_1, \ldots, v_k \in \mathbb{F}_2^k$, we say that $\mathcal{M}$ has* complexity c at coordinate i *if we can partition $\{v_j\}_{j\in[k]\setminus\{i\}}$ into $c+1$ classes such that v_i is not in the span of any of the classes. We say that $\mathcal{M}$ has* complexity c *if c is the minimum such that $\mathcal{M}$ has complexity c at coordinate i for all $i \in [k]$.*

The above definition makes sense because the span of a set of elements is not dependent on the specific basis chosen to represent the matroid. As a motivating example, consider the graphic matroid of C_k studied by Green in [15]. It can be represented by $v_1 = e_1, v_2 = e_2, \ldots, v_{k-1} = e_{k-1}$ and $v_k = e_1 + \cdots + e_{k-1}$. We see then that the graphic matroid of C_k has complexity 1 because for every $i < k$, the rest of the matroid elements can be partitioned into two sets $\{e_j\}_{j\neq i}$ and $\left\{\sum_{j\in[k]} e_j\right\}$ such that v_i is not contained in the span of either set, and for $i = k$, any nontrivial partition of the remaining elements ensures that v_k does not lie in the span of either partition. In Section 4, we extend this observation about C_k to all graphs.

Lemma 2.4 *For all graphs G, the graphic matroid of G has complexity* 1.

Green and Tao in [16] showed that if a matroid $\mathcal{M}$ has complexity c and if A is a subset of $\mathbb{F}_2^n$, then the number of linear maps $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ such that $L(v_i) \in A$ for all $i \in [k]$ is controlled by the $(c+1)$'th Gowers uniformity norm of A. Previously, Green proved in [15] an arithmetic regularity lemma, which essentially states that any set $A \subseteq \mathbb{F}_2^n$ can be partitioned into subsets of affine subspaces such that nearly every partition is nearly uniform with respect to linear tests. We show in Section 3 how to combine these two results to obtain the following:

Lemma 2.5 *Given any binary matroid $\mathcal{M}$ represented by $v_1, \ldots, v_k \in \mathbb{F}_2^k$, if $\mathcal{M}$ has complexity 1, then there exists a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$ and a k-query tester that accepts members of $\mathcal{F}_\mathcal{M}$ with probability 1 and rejects f that are ϵ-far from $\mathcal{F}_\mathcal{M}$ with probability at least $\tau(\epsilon)$.*

Theorem 1.1 directly follows from combining Lemma 2.4 and Lemma 2.5. In fact, Lemma 2.5 implies testability of all matroids that have complexity one, not only those that are graphic. In Section 4, we give examples of binary matroids that have complexity 1 and yet are provably not graphic.

Theorem 1.2 provides a proper hierarchy among the graphical properties. Moreover, the containments $\mathcal{P}_1 \subsetneq \mathcal{P}_2$ in this hierarchy are shown to be "statistically proper" in the sense that we demonstrate functions f that are ϵ-far from $\mathcal{P}_1$ but are in $\mathcal{P}_2$. The theorem implies the following hierarchy:

$$\cdots \subsetneq C_{k+2}\text{-free} \subsetneq C_k\text{-free} \subsetneq \cdots \subsetneq C_3\text{-free} = K_3\text{-free} \subsetneq \cdots \subsetneq K_k\text{-free} \subsetneq K_{\binom{k}{2}+2}\text{-free} \subsetneq \cdots$$

Thus, the class of properties $\mathcal{F}_G$ does indeed contain infinitely many more properties than the cycle freeness properties considered by Green in [15].

Both the hierarchy among the cyclic freeness properties and among the clique freeness properties are derived in Section 5 using a general technique. In order to show a statistically proper containment $\mathcal{M}_1$-free $\subsetneq \mathcal{M}_2$-free, we construct a function f that, by its definition, contains $\mathcal{M}_1$ at a large number of linear maps and so is far from being $\mathcal{M}_1$-free. On the other hand, the construction ensures that if f is also not $\mathcal{M}_2$-free, then there is a *matroid homomorphism* from $\mathcal{M}_2$ to $\mathcal{M}_1$. We define a matroid homomorphism from a binary matroid $\mathcal{M}_2$ to a binary matroid $\mathcal{M}_1$ to be a map from the ground set of $\mathcal{M}_2$ to the ground set of $\mathcal{M}_1$ which maps cycles to cycles. The separation between $\mathcal{M}_2$-freeness and $\mathcal{M}_1$-freeness is then obtained by proving that there do not exist any matroid homomorphisms from $\mathcal{M}_2$ to $\mathcal{M}_1$. This proof framework suffices for both the claims in Theorem 1.2 and is reminiscent of proof techniques involving graph homomorphisms in the area of graph property testing (see [6] for a survey).

Theorem 2.2 is the result of a more involved application of the regularity lemma. To deal with non-monotone properties, we employ a different "rounding" scheme inspired by the testability of non-monotone graph properties in [1]. Unlike Szemerédi's regularity lemma, a "strong form" of the arithmetic regularity lemma is not known, so we restrict our attention to cyclic matroids and exploit the additive structure of the pattern. Theorem 2.3 is based on a characterization theorem in Appendix D that classifies (C_k, Σ)-freeness properties into 9 classes when $\Sigma \neq 0^k, 1^k$.

3 Freeness of Complexity 1 Matroids is Testable

In this section we prove Lemma 2.5. Before doing so, we fix our notation and provide a quick background on Fourier analysis. If H is a subgroup of G, the cosets of H are indicated by $g + H$, with g in G. Let $f_{g+H} : H \to \mathbb{F}_2$ denote f restricted to the coset $g + H$, defined by sending h to $f(g + h)$; that is, for every $h \in H, g \in G$, $f_{g+H}(h) := f(g + h)$. For $\sigma \in \mathbb{F}_2$, we define $\mu_\sigma(f_{g+H}) := \Pr_{h\in H}[f_{g+H}(h) = \sigma]$ to be the density of σ in f restricted to coset $g + H$.

3.1 Fourier Analysis and Green's Regularity Lemma

Definition 3.1 *[Fourier transform] If $f : \mathbb{F}_2^n \to \mathbb{F}_2$, then we define its Fourier transform $\widehat{f} : \mathbb{F}_2^n \to \mathbb{R}$ to be $\widehat{f}(\alpha) = \mathbb{E}_{x\in\mathbb{F}_2^n}[f(x)\chi_\alpha(x)]$, where $\chi_\alpha(x) = (-1)^{\sum_{i\in[n]} \alpha_i x_i}$. $\widehat{f}(\alpha)$ is called the Fourier coefficient of f at α, and the $\{\chi_\alpha\}_\alpha$ are the characters of $\mathbb{F}_2^n$.*

It is easy to see that for $\alpha, \beta \in \mathbb{F}^n$, $\langle \chi_\alpha, \chi_\beta \rangle := \mathbb{E}_{x \in \mathbb{F}_2^n}[\chi_\alpha(x)\chi_\beta(x)]$ is 1 if $\alpha = \beta$ and 0 otherwise. So the characters form an orthonormal basis for $\mathbb{F}_2^n$, and we have the Fourier inversion formula $f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)\chi_\alpha(x)$ and Parseval's Identity $\sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 = \mathbb{E}_x[f(x)^2] = \widehat{f}(0)$.
Next we turn to Green's arithmetic regularity lemma, the crux of the analysis of our local testing algorithm. Green's regularity lemma over $\mathbb{F}_2^n$ is a structural theorem for Boolean functions. It asserts that for every Boolean function, there is some decomposition of the Hamming cube into cosets, such that the function restricted to most of these cosets are uniform and pseudorandom with respect to the linear functions. An alternate and equivalent way is that no matter where we slice the Hamming cube by a hyperplane, the density of f on these cosets of the hyperplane is what we expect a random function looks like. Formally, we say that a function is uniform if all of its nonzero Fourier coefficients are small.

Definition 3.2 *[Uniformity] For every $0 < \epsilon < 1$, we say that a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is ϵ-uniform if for every $\alpha \neq 0 \in \mathbb{F}_2^n, |\hat{f}(\alpha)| \leq \epsilon$.*

Recall that we let $W(t)$ denote a tower of twos with height $\lceil t \rceil$. To obtain a partition of the Hamming cube that satisfies the required uniformity requirement, the number of cosets in the partition may be rather large. More precisely,

Lemma 3.1 *[Green's Regularity Lemma over $\mathbb{F}_2^n$] Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Let $\epsilon \in (0, 1)$. Then there exists a subspace H of $G = \{0,1\}^n$ of co-dimension at most $W(\epsilon^{-3})$, such that $\Pr_{g \in G}[f_{g+H}$ is ϵ-uniform$] \geq 1 - \epsilon$.*

3.2 Testability of Complexity 1 Matroid Freeness

The proposition below is proved in [16]. Collectively, statements capturing the phenomenon that expectation over certain forms are controlled by varying degrees of the Gowers norm are termed *generalized von-Neumann type Theorems* in the additive combinatorics literature. In particular, as we only require the degree 2 Gowers norm of a function, which is the sum of its Fourier coefficients raised to the fourth power, the following holds:

Proposition 3.2 *[[16]] Suppose a binary matroid $\mathcal{M} = \{v_1 \ldots, v_k\}$ has complexity 1 and let $f_1, \ldots, f_k : \mathbb{F}_2^n \to \mathbb{F}_2$. Then*

$$\mathop{\mathbb{E}}_{L:\mathbb{F}_2^k \to \mathbb{F}_2^n} \left[\prod_{i=1}^{k} f_i(L(v_i)) \right] \leq \min_{i \in [k]} \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f_i}(\alpha)^4.$$

It is an easy deduction from Proposition 3.2 to see that if f is uniform, then the number of linear maps L where f has a $\mathcal{M}$-pattern is close to $\mathbb{E}[f]^m N^d$, where $N = 2^n$. Combining this observation with the regularity lemma, we prove Lemma 2.5.

Proof: [of Lemma 2.5] Consider a test that picks a linear map L uniformly at random from all linear maps from $\mathbb{F}_2^k \to \mathbb{F}_2^n$ and rejects iff for all $i \in [k]$, $f(L(v_i)) = 1$. Clearly the test has completeness one.
Now we analyze the soundness of this test. Suppose f is ϵ-far from being $\mathcal{M}$-free. We want to show that the test rejects with probability at least $\tau(\epsilon)$, such that $\tau(\epsilon) > 0$ whenever $\tau > 0$. Let $a(\epsilon)$ and $b(\epsilon)$ be two functions of ϵ that satisfy the constraint $a(\epsilon) + b(\epsilon) < \epsilon$, we shall specify these two functions at the end of the proof. We now apply Lemma 3.1 to f to obtain a subspace H of G of co-dimension at most $W(a(\epsilon)^{-3})$. Consequently, f restricted to all but at most $a(\epsilon)$ fraction of the cosets of H are $a(\epsilon)$-uniform. We define a reduced function $f^R : \mathbb{F}_2^n \to \mathbb{F}_2$ as follows.

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

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

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

Note that at most $a(\epsilon) + b(\epsilon)$ fraction of modification has been made to f to obtain f^R. Since f is ϵ-far from being $\mathcal{M}$-free, f^R has a $\mathcal{M}$-pattern at some linear map L. More precisely, for every $i \in [k]$, $f^R(L(v_i)) = 1$. Now consider the cosets $L(v_i) + H$. By our choice of rounding, we know that f restricted to each of these cosets is $a(\epsilon)$-uniform and at least $b(\epsilon)$ dense. We will count the number of linear maps $\phi : \mathbb{F}_2^k \to H$ such that f has a $\mathcal{M}$ pattern at $L + \phi$. Notice that the probability the test rejects is at least

$$2^{-k \cdot W(a(\epsilon)^{-3})} \Pr_{\phi:\mathbb{F}_2^k \to H} \left[\forall i, f_{L(v_i)+H}(\phi(v_i)) = 1\right],$$

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

$$\Pr_{\phi:\mathbb{F}_2^k \to H} \left[\forall i, f_{L(v_i)+H}(\phi(v_i)) = 1\right]$$

is bounded below by at least some constant depending on ϵ. To this end, we rewrite this probability as $\mathbb{E}_{\phi:\mathbb{F}_2^k \to H} \left[\prod_{i\in[k]} f_i(\phi(v_i))\right]$, where $f_i = f_{L(v_i)+H}$. By replacing each function f_i by $\widehat{f_i}(0) + (f_i - \widehat{f_i}(0))$, it is easy to see that the above expression can be expanded into the sum of 2^k terms, one of which is $\prod_{i\in[k]} \widehat{f_i}(0)$, which is at least $b(\epsilon)^k$. For the other $2^k - 1$ terms, by applying Proposition 3.2 and using Parseval's Identity, each of these terms is bounded above by $a(\epsilon)^2$. So the expression is at least $b(\epsilon)^k - (2^k - 1)a(\epsilon)^2$. To finish the analysis, we need to specify $a(\epsilon), b(\epsilon)$ such that $b(\epsilon)^k - (2^k - 1)a(\epsilon)^2 > 0$ and $a(\epsilon) + b(\epsilon) < \epsilon$. Both are satisfied by setting $b(\epsilon) = \frac{\epsilon}{2}$, $a(\epsilon) = (\frac{\epsilon}{2})^k$. Thus, the rejection probability is at least $\tau(\epsilon) \geq 2^{-kW((\frac{2}{\epsilon})^{3k})} 2^{-k}(\epsilon^k - \epsilon^{2k})$, completing the proof. ∎

4 Graphic Matroids have Complexity 1

Here we prove that graphic matroids have complexity 1. While the proof is simple, we believe it sheds insight into the notion of complexity and shows that even the class of 1-complexity matroids is quite rich.

Proof: [of Lemma 2.4] Recall that throughout we are assuming G to be a simple graph. Fix an arbitrary edge e in G with vertices v_1 and v_2 as its two ends. We partition the remaining edges of G into two sets S_1 and S_2 such that, if an edge is incident to v_1 then it is in S_1 and otherwise, it is in S_2. Because G is simple, a cycle in G containing e must include an edge (apart from edge e) which is incident to v_1 and another edge (other than e) which is not incident to v_1. Therefore e is not in the span of either S_1 or S_2. ∎

As we have seen earlier, Lemma 2.5 holds for any matroid of complexity 1. Hence, it is a natural question to ask whether there exist non-graphic matroids which have complexity 1. In Appendix B we show that such matroids do exist. It is an open question to come up with a natural characterization of matroids having complexity 1.

5 Infinitely many Monotone Properties

In this section we prove Theorem 1.2, that there are infinitely many matroids for which the property of being $\mathcal{M}$-free are pairwise very different.

To do so we consider a pair of target matroids $\mathcal{M}_1$ and $\mathcal{M}_2$. Based on just the first matroid $\mathcal{M}_1$, we create a canonical function $f = f_{\mathcal{M}_1} : \mathbb{F}_2^n \to \mathbb{F}_2$. We show, using a simple analysis, that this canonical function is far from being $\mathcal{M}_1$ free. We then show that if this function has an instance of $\mathcal{M}_2$ inside, then there is a "homomorphism" (in a sense we define below) from $\mathcal{M}_2$ to $\mathcal{M}_1$. Finally we show two different ways in which one can rule out homomorphisms between pairs of graphic matroids; one based on the odd girth of the matroids, and the other based on the maximum degree of $\mathcal{M}_1$. Together these ideas lead to proofs of distinguishability of many different matroids.

Definition 5.1 *Given a binary matroid $\mathcal{M}$ represented by vectors $v_1, \ldots, v_k \in \mathbb{F}_2^k$, and integer $n \geq k$, let the canonical function $f = f_{\mathcal{M}} : \mathbb{F}_2^n \to \mathbb{F}_2$ be given by $f(x, y) = 1$ if $x \in \{v_1, \ldots, v_k\}$ and 0 otherwise; where $x \in \mathbb{F}_2^k$ and $y \in \mathbb{F}_2^{n-k}$.*

Claim 5.2 *Let $\mathcal{M}$ be a binary matroid with $v_i \neq 0$ for all $i \in \{1, \ldots, k\}$. Then $f_{\mathcal{M}}$ is $\frac{1}{2^k}$-far from being $\mathcal{M}$-free.*

Proof: Note that if we consider the linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ that sends x to $\langle x, 0\rangle$, then f contains $\mathcal{M}$ at L. So f is not $\mathcal{M}$-free. However we wish to show that any function that is 2^{-k}-close to f contains $\mathcal{M}$ somewhere. Fix a function g such that $\delta(f, g) = \delta < 2^{-k}$. We will show that g contains $\mathcal{M}$ somewhere.

For $i \in [k]$ let $\delta_i = \Pr_{y \in \mathbb{F}_2^{n-k}}[f(v_i, y) \neq g(v_i, y)]$. Note that $\sum_{i=1}^k \delta_i \leq 2^k \cdot \delta < 1$. Now consider a random linear map $L_1 : \mathbb{F}_2^k \to \mathbb{F}_2^{n-k}$, and its extension $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ given by $L(x) = \langle x, L_1(x)\rangle$. For every non-zero x and in particular for $x \in \{v_1, \ldots, v_k\}$, we have $L_1(x)$ is distributed uniformly over $\mathbb{F}_2^{n-k}$. Thus, for any fixed $i \in [k]$, we have $\Pr_{L_1}[g(L(v_i)) \neq 1] \leq \delta_i$. By the union bound, we get that $\Pr_{L_1}[\exists i \,\text{s.t.}\, g(L(v_i)) \neq 1] \leq \sum_i \delta_i < 1$. In other words there exists a linear map L_1 (and thus L) such that for every i, $g(L(v_i)) = 1$ and so g contains $\mathcal{M}$ at L. ∎

We now introduce our notion of a "homomorphism" between binary matroids. (We stress that the phrase homomorphism is conjured up here and we are not aware of either this notion, or the phrase being used in the literature. We apologize for confusion if this phrase is used to mean something else.)

Definition 5.3 *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be binary matroids given by $v_1, \ldots, v_k \in \mathbb{F}_2^k$ and $w_1, \ldots, w_\ell \in \mathbb{F}_2^\ell$. We say that $\mathcal{M}_2$ has a homomorphism to $\mathcal{M}_1$ if there is a map $\phi : \{w_1, \ldots, w_\ell\} \to \{v_1, \ldots, v_k\}$ such that for every set $T \subseteq [\ell]$ such that $\sum_{i \in T} w_i = 0$, it is the case that $\sum_{i \in T} \phi(w_i) = 0$.*

For graphic matroids, the matroid-homomorphism from G to H is a map from the edges of G to the edges of H that ensures that cycles are mapped to even degree subgraphs of H.

Lemma 5.4 *If the canonical function $f_{\mathcal{M}_1}$ contains an instance of $\mathcal{M}_2$ somewhere, then $\mathcal{M}_2$ has a homomorphism to $\mathcal{M}_1$.*

Proof: Let $f = f_{\mathcal{M}_1}$ contain $\mathcal{M}_2$ at L. So $L : \mathbb{F}_2^\ell \to \mathbb{F}_2^n$ is a linear map satisfying $f(L(w_i)) = 1$ for every $i \in [\ell]$. Now consider the projection map $\pi : \mathbb{F}_2^n \to \mathbb{F}_2^k$ which sends $\langle x, y\rangle$ to x (where $x \in \mathbb{F}_2^k$ and $y \in \mathbb{F}_2^{n-k}$).

We claim that the map ϕ which sends x to $\pi(L(x))$ gives a homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$. On the one hand ϕ is linear and so if $\sum_{i\in T} w_i = 0$, then we have $\sum_{i\in T}\phi(w_i) = \phi(\sum_{i\in T} w_i) = \phi(0) = 0$. On the other hand, we also have that $\phi(w_i) \in \{v_1, \ldots, v_k\}$. This is true since $f(L(w_i)) = 1$, which implies, by the definition of the canonical function f that $\pi(L(w_i)) \in \{v_1, \ldots, v_k\}$. Thus ϕ satisfies the requirements of a homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$. ∎

The above lemma now motivates the search for matroids $\mathcal{M}_2$ that are not homomorphic to $\mathcal{M}_1$. Proving non-homomorphism in general may be hard, but we give a couple of settings where we can find simple proofs. Each addresses a different case of Theorem 1.2.

For a matroid $\mathcal{M}$, let its *odd girth*, denoted $\mathrm{og}(\mathcal{M})$, be the size of the smallest dependent set of odd cardinality, i.e. the size of the smallest odd set $T \subseteq [\ell]$ such that $\sum_{i\in T} w_i = 0$.

Lemma 5.5 *If $\mathcal{M}_2$ has a homomorphism to $\mathcal{M}_1$, then $\mathrm{og}(\mathcal{M}_2) \geq \mathrm{og}(\mathcal{M}_1)$.*

Proof: Let ϕ be a homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$ and let $T \subset [\ell]$ denote the smallest odd dependent set of $\mathcal{M}_2$. Now let $T' \subseteq [k]$ be the set $T' = \{j \in [k] | \#\{i \in T | \phi(w_i) = v_j\} \text{ is } \mathrm{odd}\}$. On the one hand, we have T' has odd cardinality; and on the other, we have $0 = \sum_{i\in T}\phi(w_i) = \sum_{j\in T'} v_j$. So T' is an odd dependent set in $\mathcal{M}_1$. The lemma follows since $|T| \geq |T'|$. ∎

For graphic matroids constructed from the odd cycle graph C_k, we have that its odd girth is just k and so the above lemmas combine to give that C_k-freeness is distinguishable from C_{k+2}-freeness, and this suffices to prove Part (1) of Theorem 1.2.

However the odd girth criterion might suggest that G-freeness for any graph containing a triangle might be equivalent. Below we rule this possibility out.

Lemma 5.6 *Let $\mathcal{M}_1$ be the graphic matroid of the complete graph K_a on a vertices, and let $\mathcal{M}_2$ be the graphic matroid of K_b. Then, if $b \geq \binom{a}{2} + 2$, there is no homomorphism from $\mathcal{M}_2$ to $\mathcal{M}_1$.*

Proof: Assume otherwise and let ϕ be such a homomorphism. Fix any vertex of K_b and let $e_1, \ldots, e_{b-1}$ denote the $b-1$ edges incident to this vertex. By the pigeonhole principle, (since $b - 1 > \binom{a}{2}$) there must exist a pair of incident edges e_i and e_j such that $\phi(e_i) = \phi(e_j)$. But now let f denote the edge which forms a triangle with e_i and e_j. Since in K_b we have $e_i + e_j + f = 0$ (viewing these elements as vectors over $\mathbb{F}_2$), it must be that $\phi(f) = \phi(e_i) + \phi(e_j) = 0$ which is not an element of the ground set of $\mathcal{M}_1$. This yields the desired contradiction. ∎

We are now ready to prove Theorem 1.2.

Proof: [of Theorem 1.2] First note that C_{k+2}-free functions are also C_k-free. Informally, suppose a function f has a k cycle at point $x_1, \ldots, x_k$, i.e., $f(x_i) = 1$ at these points and $\sum_i x_i = 0$. Then f has a $k+2$ cycle at the points $x_1, x_1, x_1, x_2, \ldots, x_k$. (This informal argument can obviously be converted to a formal one once we specify the graphic matroids corresponding to C_k and C_{k+2} formally.)

On the other hand, if we take $\mathcal{M}_1$ to be the graphic matroid corresponding to C_k and f to be the canonical function corresponding to $\mathcal{M}_1$, then by Claim 5.2 it is 2^{-k}-far from $\mathcal{M}_1$-free, and by Lemmas 5.4 and 5.5 it does not contain $\mathcal{M}_2$, the graphic matroid of C_{k+2}.

For the second part of the theorem, note that every property that is G-free is also H-free if G is a subgraph of H. Thus K_k-free is contained in K_ℓ free is $k \leq \ell$. The proper containment can now be shown as above, now using Claim 5.2 and Lemmas 5.4 and 5.6. ∎

6 Conclusions and Future Work

We introduced an infinite family of properties of Boolean functions and showed them to be testable. These properties were specified by a matroid $\mathcal{M}$ on k elements and a pattern $\Sigma \subseteq \{0,1\}^k$. However to capture the full range of linear-invariant non-linear properties that allow one-sided error local tests, we should also allow the conjunction of a constant number of constraints. We believe this could lead to a characterization of all linear-invariant non-linear properties that allow one-sided error local tests.

In a different direction, we feel that it would also be nice to develop richer techniques to show the distinguishability of syntactically different properties. For instance, even for the graphic case we don't have a good understanding of when two different graphs represent essentially the same properties, and when they are very different.

Acknowledgments

We are grateful to Kevin Matulef for suggesting this research direction. We thank Tali Kaufman and Swastik Kopparty for helpful discussions. We thank Asaf Shapira for drawing our attention to his preprint [25].

References

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

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

[3] Noga Alon, Michael Krivelevich, Ilan Newman, and Mario Szegedy. Regular languages are testable with a constant number of queries. *SIAM Journal on Computing*, 30(6):1842–1862, 2000.

[4] Noga Alon and Asaf Shapira. A characterization of the (natural) graph properties testable with one-sided error. In *FOCS*, pages 429–438. IEEE Computer Society, 2005.

[5] Noga Alon and Asaf Shapira. Every monotone graph property is testable. In Harold N. Gabow and Ronald Fagin, editors, *STOC '05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 128–137. ACM, 2005.

[6] Noga Alon and Asaf Shapira. Homomorphisms in graph property testing - a survey. Technical Report 085, Electronic Colloquium on Computational Complexity (ECCC), 2005.

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

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

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

[10] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos A. Kiwi, and Madhu Sudan. Linearity testing over characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.

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

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

[13] Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós, Balázs Szegedy, and Katalin Vesztergombi. Graph limits and parameter testing. In *STOC '06: Proceedings of the Thirtyeighth Annual ACM Symposium on Theory of Computing*, pages 261–270, 2006.

[14] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.

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

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

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

[18] C.S. Jutla, A.C. Patthak, A. Rudra, and D. Zuckerman. Testing low-degree polynomials over prime fields. In *FOCS '04: Proceedings of the forty-fifth annual ACM symposium on Foundations of computer science*, pages 423–432, 2004.

[19] T. Kaufman and D. Ron. Testing polynomials over general fields. In *FOCS '04: Proceedings of the forty-fifth annual ACM symposium on Foundations of computer science*, pages 413–422, 2004.

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

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

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

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

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

[25] A. Shapira. A proof of Green's conjecture regarding the removal properties of sets of linear equations. *Preprint*, to appear.

[26] W.T. Tutte. Matroids and graphs. *Transactions of the American Mathematical Society*, 90:527–552, 1959.

[27] D.J.A. Welsh. *Matroid Theory*. Academic Press Inc., London, 1976.

A Matroids

For background on matroid theory, we refer the reader to [27].

Definition A.1 *A matroid $\mathcal{M}$ is a finite set S (called* ground set*) and a collection $\mathcal{F}$ of subsets of S (called* independent sets*) such that the following hold:*

1. $\emptyset \in \mathcal{F}$.
2. *If $X \in \mathcal{F}$ and $Y \subseteq X$, then $Y \in \mathcal{F}$.*
3. *If X and Y are both in $\mathcal{F}$ with $|X| = |Y| + 1$, then there exists an $x \in X \setminus Y$ such that $Y \cup x \in \mathcal{F}$.*

A matroid $\mathcal{M}$ on a ground set $S = \{x_1, \ldots, x_k\}$ is said to be *linear* if there exists a field $\mathbb{F}$ and vectors $v_1, \ldots, v_k \in \mathbb{F}^k$ such that some subset $\{x_i | i \in T\}$ indexed by $T \subseteq \{1, \ldots, k\}$ is independent if and only if the corresponding vectors $\{v_i | i \in T\}$ are linearly independent. A linear matroid is *binary* if $\mathbb{F} = \mathbb{F}_2$.

B Non Graphic Matroids of Complexity 1

First, we make the following claim that follows immediately from the definition of cographic matroids and the notion of complexity.

Claim B.1 *A cographic matroid $\mathcal{M}^*(G)$ has complexity 1 if and only if, for every edge $e \in E(G)$, there is a partition of $E(G) \setminus \{e\}$ into two disjoint sets A and B such that both of the subgraphs $(V(G), A)$ and $(V(G), B)$ are connected.*

Proposition B.2 *There is a matroid with complexity one that is not graphic.*

Proof: Consider the cographic matroid of K_5. Embed K_5 in the plane as a pentagon and all its diagonals. Fix an outer edge e and partition the remaining 9 edges into two sets. One is the 4 outer edges and the other is the remaining 5 diagonal edges. Clearly both outer-edge set and diagonal-edge set make the five vertices connected. Therefore by Claim B.1, the cographic matroid of K_5 is of complexity one. On the other hand, by a theorem of Tutte [26], a matroid cannot be graphic if it contains $\mathcal{M}^*(K_5)$ as a minor, which $\mathcal{M}^*(K_5)$ clearly does. So, $\mathcal{M}^*(K_5)$ is an example of a non-graphic matroid that has complexity 1. ∎

We remark that not all cographic matroids have complexity 1. For example, the cographic matroid of $K_{3,3}$ cannot have complexity 1 because if we remove an edge from $K_{3,3}$, there do not remain enough edges to form two edge-disjoint connected graphs on 6 vertices, violating Claim B.1.

C Testing Non-monotone Properties

In this section we prove Theorem 2.2. (Readers may find it useful to recall the background material in Section 3.1.) We show that for non-monotone properties, i.e., when $\Sigma \neq 0^k$ or 1^k, the property of $(\mathcal{M}, \Sigma)$-free is testable when the underlying graph is a cycle. However, as opposed to Section 5, the number of non-monotone properties associated with cycles is finite. In fact we give a complete characterization of these non-monotone properties in Appendix D.

Proof: [of Theorem 2.2] Suppose we have oracle access to a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Consider the following k-query test T, which selects a linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$ uniformly at random from all such possible linear

maps. T has oracle access to f and queries f at the points $L(v_1), \ldots, L(v_k)$. T rejects iff all of these points are evaluated to 1. If f is $(\mathcal{M}, \Sigma)$-free, T never rejects and has completeness one.

Now we analyze the soundness of T. Suppose that f is ϵ-far from being $(\mathcal{M}, \Sigma)$-free. We want to show that T rejects with probability at least $\tau(\epsilon)$, such that $\tau(\epsilon) > 0$ whenever $\epsilon > 0$.

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

Now let G denote $\mathbb{F}_2^n$. We apply Lemma 3.1 to f to obtain a subspace H of G of co-dimension at most $W(a(\epsilon)^{-3})$. We define a reduced function $f^R : \mathbb{F}_2^n \to \mathbb{F}_2$ as follows. We assume that Σ has at least two occurrences of 1. (Otherwise it has at least two occurrences of 0, and in the construction of f^R, we flip the role of 1 and 0 for non-uniform cosets. The rest of the proof will proceed analogously, and we leave its verification to the readers.)

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

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

Else, define

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

Note that at most $a(\epsilon) + b(\epsilon)$ fraction of modification has been made to f to obtain f^R, so f^R is ϵ-close to f.

By assumption, f is ϵ-far from $(\mathcal{M}, \Sigma)$-free, so f^R has a $(\mathcal{M}, \Sigma)$ pattern at some linear map $L : \mathbb{F}_2^k \to \mathbb{F}_2^n$, i.e., for each $i \in [k]$, $f^R(L(v_i)) = \sigma_i$, where $\Sigma =< \sigma, \ldots, \sigma_k >$, and $\sigma_i \in \mathbb{F}_2$. Now consider the cosets $L(v_i) + H$. By our choice of rounding, f restricted to each $L(v_i) + H$ is dense in the symbol σ_i, i.e., $\mu_{\sigma_i}(f_{L(v_i)+H}) \geq b(\epsilon)$. We want to show that there are many $(\mathcal{M}, \Sigma)$ patterns spanning across these cosets. In particular, we restrict our attention to the relative number of $(\mathcal{M}, \Sigma)$-patterns at linear maps of the form $L + \phi$, where ϕ maps linearly from $\mathbb{F}_2^k$ to H. Notice that the probability the test T rejects is at least

$$2^{-(k-1)W(a(\epsilon)^{-3})} \cdot \Pr_{\phi:\mathbb{F}_2^k \to H}[\text{for each } i, f_{L(v_i)+H}(\phi(v_i)) = \sigma_i].$$

It suffices to show that the probability

$$\Pr_{\phi:\mathbb{F}_2^k \to H}[\text{for each } i, f_{L(v_i)+H}(\phi(v_i)) = \sigma_i] \tag{1}$$

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

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

We begin by arithmetizing Equation 1 as

$$\mathop{\mathbb{E}}_{\phi:\mathbb{F}^k \to H}\left[\prod_{i \in [k]} f_i(\phi(v_i))\right],$$

where $f_i = f_{L(v_i)+H} + \sigma_i + 1$. Since $\mathcal{M}$ is a cyclic matroid, it is not hard to show, by Fourier expansion, that

$$\mathop{\mathbb{E}}_{\phi:\mathbb{F}^k \to H} \left[\prod_{i\in[k]} f_i(\phi(v_i)) \right] = \sum_{\alpha \in H} \prod_{i\in[k]} \hat{f}_i(\alpha).$$

Using the facts that each f_i has density at least $b(\epsilon)$, some $j \in [k]$ such that f_j is $a(\epsilon)$-uniform, there exist distinct $i_1, i_2 \in [k]$ not equal to j (since $k \geq 3$), Cauchy-Schwarz Inequality, and Parseval's Identity, respectively, we have

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

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

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

Since $\mathcal{M}$ is a cyclic matroid, it is not hard to see that Equation 1 is equal to

$$\Pr_{x_1,\ldots,x_k\in H;\sum_i x_i=0} \left[\forall i \in [k], f_{L(v_i)}(x_i) = \sigma_i\right]. \tag{2}$$

Recall that Σ contains two occurrences of the symbol 1. Without loss of generality we assume that $\sigma_{k-1} = \sigma_k = 1$. Fix $x_1, \ldots, x_{k-2} \in H$ such that $f_{L(v_i)}(x_i) = \sigma_i$. Let $z = \sum_{i=1}^{k-2} x_i$. By union bound and the fact that $\eta > \frac{1}{2}$, we have

$$\begin{aligned}
\Pr_{x\in H}[f_{L(v_{k-1})}(x) = f_{L(v_k)}(x+z) = 1] &= 1 - \Pr_{x\in H}[f_{L(v_{k-1})}(x) = 0 \text{ or } f_{L(v_k)}(x+z) = 0] \\
&\geq 1 - 2(1-\eta) \\
&> 0.
\end{aligned}$$

For each $i \in [k]$, $\Pr_{x\in H}[f_{L(v_i)}(x_i) = \sigma_i]$ is at least $1-\eta$. By picking x_i uniformly at random from H in sequence, it is not hard to see that the rejection probability of the test is at least

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

where $a(\epsilon) = \frac{1}{2}(1-\eta)^k\epsilon^k$. ∎

D Characterization of Cycle Free Functions

In this section we consider the property of being $(\mathcal{M}, \Sigma)$-free, where $\mathcal{M}$ is the matroid of the k-cycle. Syntactically these appear to be infinitely many different properties. We show that there are only finitely many distinct properties here when Σ is not equal to 0^k or 1^k. (As noted in Section 5, when $\Sigma = 1^k$, we do get infinitely many distinct properties.)

We start with some terminology that describes the distinct families we get.

Definition D.1

- *Let* $\mathbf{Const}$ *denote the set of constant functions (i.e., the zero function and the one function).*
- *Let* $\mathbf{Lin}$ *denote the set of all linear functions, including the constant functions. (We note that throughout we think of the constant functions as linear, affine etc.). Let* $\overline{\mathbf{Lin}}$ *denote the complementary family, i.e., all functions whose complements are in* $\mathbf{Lin}$*.*
- *Let* $\mathbf{Aff}$ *denote the set of all affine functions, i.e., the linear functions and their complements. Let* $\overline{\mathbf{Aff}}$ *denote the complementary family.*
- *We use the notation* $\mathcal{F}_{\mathbf{lin}}$ *to denote the family of linear subspace functions and the* 0 *function, i.e.,* $\mathcal{F}_{\mathbf{lin}} = \{0\} \cup \{f : \mathbb{F}_2^n \to \mathbb{F}_2 | f^{-1}(1)$ *is a linear subspace of* $\mathbb{F}_2^n\}$*.* $\overline{\mathcal{F}_{\mathbf{lin}}}$ *is the complementary family.*
- *We use the notation* $\mathcal{F}_{\mathbf{aff}}$ *to denote the family of affine subspace functions and the* 0 *function, i.e.,* $\mathcal{F}_{\mathbf{aff}} = \{0\} \cup \{f : \mathbb{F}_2^n \to \mathbb{F}_2 | f^{-1}(1)$ *is an affine subspace of* $\mathbb{F}_2^n\}$*.* $\overline{\mathcal{F}_{\mathbf{aff}}}$ *is the complementary family.*

It turns out that for every $k \geq 3$ and every $\Sigma \neq 0^k, 1^k$, a (C_k, Σ)-free family is one of the nine families $\mathbf{Const}, \mathbf{Lin}, \overline{\mathbf{Lin}}, \mathbf{Aff}, \overline{\mathbf{Aff}}, \mathcal{F}_{\mathbf{lin}}, \overline{\mathcal{F}_{\mathbf{lin}}}, \mathcal{F}_{\mathbf{aff}}, \overline{\mathcal{F}_{\mathbf{aff}}}$. To give further details, let $Z(\Sigma)$ denote the number of zeroes in Σ and $O(\Sigma)$ denote the number of ones in Σ. We have:

Theorem D.2 *For every* $k \geq 3$ *and every* $\Sigma \neq 0^k, 1^k$*, a* (C_k, Σ)*-free family is one of*

$$\mathbf{Const}, \mathbf{Lin}, \overline{\mathbf{Lin}}, \mathbf{Aff}, \overline{\mathbf{Aff}}, \mathcal{F}_{\mathbf{lin}}, \overline{\mathcal{F}_{\mathbf{lin}}}, \mathcal{F}_{\mathbf{aff}}, \overline{\mathcal{F}_{\mathbf{aff}}}.$$

Specifically:

1. *If* $Z(\Sigma)$ *and* $O(\Sigma)$ *are even, the* $\mathcal{F}_{C_k,\Sigma} = \mathbf{Const}$*.*
2. *If* $O(\Sigma) > 1$ *is odd and* $Z(\Sigma)$ *is even, then* $\mathcal{F}_{C_k,\Sigma} = \mathbf{Lin}$*. Complementarily, if* $Z(\Sigma) > 1$ *is odd and* $O(\Sigma)$ *is even, then* $\mathcal{F}_{C_k,\Sigma} = \overline{\mathbf{Lin}}$*.*
3. *If* $O(\Sigma) = 1$ *and* $Z(\Sigma)$ *is even, then* $\mathcal{F}_{C_k,\Sigma} = \mathcal{F}_{\mathbf{lin}}$*. Complementarily, if* $Z(\Sigma) = 1$ *and* $O(\Sigma)$ *is even, then* $\mathcal{F}_{C_k,\Sigma} = \overline{\mathcal{F}_{\mathbf{lin}}}$*.*
4. *If* $O(\Sigma), Z(\Sigma) > 1$ *are odd, then* $\mathcal{F}_{C_k,\Sigma} = \mathcal{F}_{\mathbf{aff}}$*.*
5. *If* $Z(\Sigma) = 1$ *and* $O(\Sigma) > 1$ *is odd, then* $\mathcal{F}_{C_k,\Sigma} = \mathcal{F}_{\mathbf{aff}}$*. Complementarily, if* $O(\Sigma) = 1$ *and* $Z(\Sigma) > 1$ *is odd, then* $\mathcal{F}_{C_k,\Sigma} = \overline{\mathcal{F}_{\mathbf{aff}}}$*.*

We begin with some simple facts and observations.

Fact D.3 *[[22]] Let* S *be an affine subspace. Then* x, y *and* z *are all in* S *implies* $x + y + z$ *is also in* S*. Conversely, if for any triple* x, y *and* z *in* S *implying* $x + y + z$ *in* S*, then* S *is an affine subspace.*

This fact immediately gives the following Observation.

Observation D.4 *A function $f : \{0,1\}^n \to \{0,1\}$ is $(C_4, 1110)$-free if and only if f is in $\mathcal{F}_{\mathbf{aff}}$.*

Observation D.5 *A function $f : \{0,1\}^n \to \{0,1\}$ is $(C_3, 100)$-free if and only if f is the disjunction (OR) of linear functions (or the all 1 function). Consequently, f is $(C_3, 110)$-free if and only if f is in $\mathcal{F}_{\mathbf{lin}}$.*

Proof: Let $S = \{x \in \{0,1\}^n : f(x) = 0\}$. If S is empty, then f is the all 1 function. Otherwise let x and y be any two elements in S (not necessarily distinct). Then if f is $(C_3, 100)$-free, it must be the case that $x + y$ is also in S. Thus S is a linear subspace of $\{0,1\}^n$. Suppose the dimension of S is k with $k \geq 1$. Then there are k linearly independent vectors $a_1, \ldots, a_k \in \{0,1\}^n$ such that $z \in S$ iff $(\langle z, a_1\rangle = 0) \bigwedge \cdots \bigwedge (\langle z, a_k\rangle = 0)$. Therefore, by De Morgan's law, $f(z) = 1$ iff $z \in \bar{S}$ iff $(\langle z, a_1\rangle = 1) \bigvee \cdots \bigvee (\langle z, a_k\rangle = 1)$, which is equivalent to the claim. ∎

Observation D.6 *If $\Sigma \neq 1^k$ for some $k > 2$, then $(C_{k+2}, \Sigma \circ 00)$-free $\subseteq (C_k, \Sigma)$-free. Similarly, if $\Sigma \neq 0^k$ for some $k > 2$, then $(C_{k+2}, \Sigma \circ 11)$-free $\subseteq (C_k, \Sigma)$-free.*

Proof: By symmetry, we only need to prove the first part. Let $f \in (C_{k+2}, \Sigma \circ 00)$-free. Suppose $f \notin (C_k, \Sigma)$-free, then there exists a violating tuple, say, $\langle x_1, x_2, \ldots, x_j\rangle$ such that $\sum_{i=1}^{j} x_i = 0$ and

$$\langle f(x_1), f(x_2), \ldots, f(x_j)\rangle = \Sigma.$$

Since Σ is not an all 1 vector, there exists some k such that $f(x_k) = 0$. But then $\langle x_1, x_2, \ldots, x_j, x_k, x_k\rangle$ would be a violation tuple of pattern $(C_{k+2}, \Sigma \circ 00)$, contradicting our assumption that the function f is in $(C_{k+2}, \Sigma \circ 00)$-free. ∎

Observation D.7 *$(C_4, 0011)$-free equals the set of constant functions.*

Proof: Clearly a constant function has no 0011 pattern. For the reverse inclusion, suppose f is $(C_4, 0011)$-free but not a constant function. Then there exist x and y such that $f(x) = 0$ $f(y) = 1$. Then

$$\langle f(x), f(x), f(y), f(y)\rangle = 0011.$$

∎

Proof: [of Theorem D.2]

1. Follows from Observation D.6 and Observation D.7.
2. We only need to prove the first half of the claim, the second half will then follow by symmetry. It is easy to check that, if $O(\Sigma) > 1$ is odd and $Z(\Sigma)$ is even, then $\mathbf{Lin} \subseteq \mathcal{F}_{C_k,\Sigma}$. Furthermore, by Observation D.5 and Observation D.6, in this case $\mathcal{F}_{C_k,\Sigma}$ is contained in the disjunction of linear functions. Now suppose f is $(C_5, 00111)$. So $f(x) = (\langle a_1, x\rangle = 1) \bigvee (\langle a_2, x\rangle = 1) \bigvee \cdots$, where $a_1, a_2, \ldots,$ are non-zero, distinct and linearly independent vectors. Since a_1 and a_2 are linearly independent, there exist x_1, x_2 such that $\langle a_1, x_1\rangle = \langle a_2, x_2\rangle = 1$ while $\langle a_1, x_2\rangle = \langle a_2, x_1\rangle = 0$. Then $\langle 0, 0, x_1, x_2, x_1 + x_2\rangle = 00111$. Therefore f can not be the disjunction of more than one linear function, making it linear. Finally note that $\mathbf{Lin} \subseteq (C_{|\Sigma|}, \Sigma)$-free $\subseteq (C_5, 00111)$-free $\subseteq \mathbf{Lin}$.
3. This follows from Observation D.5 and Observation D.6.

4. Let i and j be odd integers. If $f(x)$ is linear, then it is $(C_{i+j}, 0^i1^j)$-free since j is odd. If $f(x)$ is the complement of linear, then it is $(C_{i+j}, 0^i1^j)$-free since i is odd. So if f is an affine function, then it is $(C_{i+j}, 0^i1^j)$-free. Now consider $(C_6, 000111)$-free. If f is $(C_6, 000111)$-free then the set $f^{-1}(1)$ form an affine subspace (since f is also $(C_4, 0111)$-free.). Similarly the set $f^{-1}(0)$ form an affine subspace (since f is also $(C_4, 0001)$-free) and so f is an affine function.

5. This follows from Observation D.4 and Observation D.6.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/