Testing Linear-Invariant Non-Linear Properties

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

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 work 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 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 $D = \{0,1\}^n$, the vector space of n-dimensional Boolean vectors, and the range $R = \{0,1\}$. In this setting, a property $\mathcal{F}$ is said to be *linear-invariant* if for every $f \in \mathcal{F}$ and linear map $L : \{0,1\}^n \to \{0,1\}^n$ 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. [4] and Bellare et al. [2], and the property of being a "moderate-degree" polynomial (aka Reed-Muller codeword) studied by Alon et al. [1]. While the tests in the above mentioned works potentially used all features of the property being tested, Kaufman and Sudan [7] 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 $\{0,1\}$ 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 gave examples of non-linear linear-invariant properties is Green [5] where a test for the property of being "triangle-free" was described. A function $f : \{0,1\}^n \to \{0,1\}$ is said to be *triangle-free* if for every $x, y \in \{0,1\}^n$ it is the case that at least one of $f(x), f(y), f(x+y)$ does not equal 1. The property of being triangle-free is easily seen to be linear-invariant and yet not linear. Green [5] showed 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 $D = \{0,1\}^n$ and $R = \{0,1\}$ 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 $\{0,1\}^k$. (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

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

†MIT CSAIL. victor@csail.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.

every positive integer n, containing all functions that, for every linear map $L : \{0,1\}^k \to \{0,1\}^n$ take on the value 0 on at least one of the points $L(v_1), \ldots, L(v_k)$. (In our work, we also 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 $\{0,1\}^k \to \{0,1\}^k$. Thus the property is effectively specified by the dependencies among $v_1, \ldots, v_k$ which are in turn captured by the matroid underlying $v_1, \ldots, v_k$. This leads us to our nomenclature:

Definition ($\mathcal{M}$-freeness)**.** Given a (binary, linear) matroid $\mathcal{M}$ represented by vectors $v_1, \ldots, v_k \in \{0,1\}^k$, the property of being *$\mathcal{M}$-free* is given by, for every positive integer n, the family

$$\mathcal{F}_{\mathcal{M}} = \{f : \{0,1\}^n \to \{0,1\} \mid \forall \text{ linear } L : \{0,1\}^k \to \{0,1\}^n, \langle f(L(v_1)), \ldots, f(L(v_k)) \rangle \neq 1^k\}.$$

The property of being $\mathcal{M}$-free has a natural k-local test associated with it: Pick a random linear map $L : \{0,1\}^k \to \{0,1\}^n$ 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 $\mathcal{M} = \{v_1, \ldots, v_k\}$, $v_i \in \{0,1\}^k$, 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 $S \subseteq \{v_1, \ldots, v_k\}$ 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 $v_1, \ldots, v_k \in \{0,1\}^k$ is testable.

Theorem. *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 $\tau(\epsilon) \geq W(\mathrm{poly}(1/\epsilon))^{-1}$, 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 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 [5], and extends to cover the case of being k-cycle free (also considered in [5]). But it includes every other graph too!

Syntactically, the above theorem 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. *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 testability proof is based on Green's analysis of the triangle-free case [5]. 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 $f : \{0,1\}^n \to \{0,1\}$, one can find a subgroup H of $\{0,1\}^n$ 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.

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 [6] 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.

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 that these properties actually capture all graphic matroids which is already promising. Yet this is not a definite proof of novelty, and so in 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 [7] 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 paper [10]. His

result solved a question that we posed as open in an earlier version of this paper [3]. His result was built on the work of Král', Serra, and Vena in [8], where an alternate proof of Green's cycle-freeness result was provided. Essentially the authors in [8] demonstrated 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 applied regularity lemmas for graphs to analyze the number of cycles in a function far from being cycle-free. In this manner, the authors showed that our main testability theorem holds as well. By extending this method and utilizing hypergraph regularity lemmas, Shapira [10] and Král', Serra, and Vena in a followup work [9] showed that arbitrary monotone matroid-freeness properties are testable.

We remark that our proofs are very different from those in [8], [9], and [10], and in particular, our view on invariance leads us to develop techniques to show that syntactically different properties are indeed distinct.

Future work It would be nice to extend our results to the case where the pattern Σ is an arbitrary binary string, as opposed to being monotone. We did manage to extend this in the special case where $\mathcal{M}$ is a cyclic matroid, but in this case the extension is not very interesting. We do feel that our proof techniques already capture some non-trivial other cases, but are far from capturing all cases, even for graphic matroids.

Extending the patterns further, there is no real reason to view the range as a field element, so a major generalization would be to consider matroids over arbitrary fields, and letting the range be some arbitrary finite set R where the forbidden pattern $\Sigma \in R^k$. (We don't believe there should be any major technical barriers in this step, once we are able to handle arbitrary 0/1 patterns Σ.) Finally, all the above problems consider the case of a single forbidden pattern (and its linear transformations).

These properties were specified by a matroid $\mathcal{M}$ on k elements and a pattern $\Sigma \in \{0,1\}^k$. 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 interesting 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.

References

[1] 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.

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

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

[4] 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.

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

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

[7] 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.

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

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

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