Electronic Colloquium on Computational Complexity, Report No. 166 (2013)

On testing affine-invariant properties over finite fields*

Arnab Bhattacharyya

Indian Institute of Science
Bangalore, India
arnabb@csa.iisc.ernet.in

Abstract

An affine-invariant property over a finite field is a property of functions over $\mathbb{F}_p^n$ that is closed under all affine transformations of the domain. This class of properties includes such well-known beasts as low-degree polynomials, polynomials that nontrivially factor, and functions of low spectral norm. The last few years has seen rapid progress in characterizing the affine-invariant properties which are testable with a constant number of queries. We survey the current state of this project.

1 What Is Property Testing?

A scientific experiment takes as input an unknown object[1], and it aims to determine whether or not a certain statement about the object holds true. Often, it's not feasible to decide whether the statement is exactly true or not, and the experimenter is satisfied with knowing whether the statement is "approximately true" for the object. The experiment consists of making certain kinds of measurements on the object, and one usually wants to minimize the number of measurements needed to reach a conclusion. ***Property testing*** is an algorithmic formalization of this basic scientific endeavor.

Let $\mathcal{O}$ be a set of objects, and let $\mathcal{P}$ be the subset of $\mathcal{O}$ that satisfies a certain desirable property. Given an unknown object $o \in \mathcal{O}$, we wish to know whether o is "close to being a member of $\mathcal{P}$" by making a small number of "measurements" on o. To make this precise, introduce a distance function dist between pairs of objects in $\mathcal{O}$ and a query model that specifies what the possible queries (measurements) into o are and how each query reveals information about o. Then, for a given parameter $\varepsilon > 0$ that parameterizes the level of approximation and a positive integer q, an (ε, q)-*tester for* $\mathcal{P}$ is an algorithm[2] $\mathcal{A}$ that makes q queries into an unknown input o, outputs YES if o is in $\mathcal{P}$ and NO if $\mathsf{dist}(o, o') > \varepsilon$ for every o' in $\mathcal{P}$. That is, if the algorithm $\mathcal{A}$ outputs YES, then there is a guarantee that $\mathsf{dist}(o, o') \leqslant \varepsilon$ for some o' in $\mathcal{P}$, and hence, $\mathcal{P}$ approximately holds true for o.

An example will help illustrate the notion. Suppose we are presented with an unknown graph (e.g., the Facebook friend graph), and we want to test the hypothesis $\mathcal{P}$ that there are no six nodes which induce $K_{3,3}$ as a subgraph. Say the query model is that we can ask if there exists an edge between any two nodes. And the distance function dist is defined as: $\mathsf{dist}(G, G') = |E(G)\Delta E(G')|/\binom{n}{2}$, where G and G' are two graphs on n vertices and $E(G)\Delta E(G')$ is the symmetric difference between the edge sets of G and G'. Now suppose $\mathcal{A}$ is an (ε, q)-tester for $\mathcal{P}$, and we run $\mathcal{A}$ on the input graph G. We have that if $\mathcal{A}$ outputs YES, then G is approximately $K_{3,3}$-free, in the sense that only $\varepsilon\binom{n}{2}$ edges can be added to/removed from G to make it free of $K_{3,3}$ induced subgraphs. Also, if $\mathcal{A}$ outputs NO, then there exist six nodes which induce a $K_{3,3}$. Moreover, $\mathcal{A}$ only examines q pairs of nodes before making its decision.

*A version of this article will appear in the SIGACT News Complexity Theory Column.

[1] Well, not completely unknown. The experimenter usually already has some partial information about the object.

[2] The algorithm is randomized, and the output guarantees hold with constant probability over the randomness of the algorithm. We give a precise definition later.

ISSN 1433-8092

Property testing as a subject in this generality was introduced by Goldreich, Goldwasser and Ron [GGR98]. However, it was preceded by some very influential works that treated algebraic properties of functions, such as linear functions and low-degree multivariate polynomials. This line of study began with the seminal work of Blum, Luby and Rubinfeld [BLR93] on testing linearity. This work was extended by Rubinfeld and Sudan [RS96], where the testing question was first explicitly asked in the context of program checking. 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 technical component in these works is the analysis of testers for algebraic properties. We recommend to the reader the surveys [Fis04, Rub06, Ron09, RS11] for a general overview of property testing.

2 Testability and Invariances

A property $\mathcal{P}$ is called *testable* if for every ε, there exists an (ε, q)-tester for $\mathcal{P}$ such that the query complexity q is independent of the size of the input. Thus, if a property is testable, even when the input grows unbounded in size, the complexity of the tester is bounded.

The most classical example of a testable property perhaps is majority. Here, the input is an arbitrary function[3] $f : [n] \to \{0, 1\}$, and the property $\mathcal{P}$ to be tested is satisfied by f only when $f(x) = 1$ for at least $n/2$ values of x. A tester can query f on any element of $[n]$. Also, the distance between two functions f and g is defined to just be the fraction of $[n]$ on which they differ. A standard analysis then shows that the algorithm which samples $q = O(1/\varepsilon^2)$ random points of $[n]$, evaluates f on each, and accepts if at least $q(1-\varepsilon)/2$ evaluations is 1 is an (ε, q)-tester for $\mathcal{P}$. Notice that q is independent of n.

The most significant and surprising achievement of property testing has been showing that not only are classical statistical properties like majority testable but that this is true for many more complicated (indeed, NP-hard) properties under natural query models. One way to appreciate the modern work in property testing is through a formulation in terms of *invariances*. For the majority property $\mathcal{P}$, note that if f satisfies $\mathcal{P}$, then for any permutation $\pi : [n] \to [n]$, the function $f \circ \pi$ also satisfies $\mathcal{P}$. We say then that $\mathcal{P}$ is invariant under all permutations. Indeed, one can show that any property $\mathcal{P}$ that is invariant under all permutations is testable (using essentially the same argument needed to argue testability for majority). What modern work has shown is that properties which are invariant under a much smaller set of permutations can still be testable, under natural query models!

To illustrate this, consider any property $\mathcal{G}$ of graphs on n vertices. A graph property is invariant under any permutation of the vertices. As in the example in the previous section, assume the input graph G is represented as a function $G : [n] \times [n] \to \{0, 1\}$, each query evaluates G on one pair in $[n]^2$, and the distance between two graphs is the Hamming distance between their adjacency matrices. Then, if $N = n^2$ is the size of the domain of G, the size of the group of invariances of $\mathcal{G}$ is $\sqrt{N}! = 2^{\tilde{O}(\sqrt{N})}$, exponentially smaller than the full set of $N!$ permutations. On the other hand, since the landmark work of Goldreich, Goldwasser and Ron [GGR98], it's known that many important graph properties are testable (using analysis that's significantly more non-trivial than that used for majority). In particular, the example property of freeness from $K_{3,3}$ induced subgraphs is testable [AS08a].

In fact, for graph properties, an exact characterization of the testable graph properties was found by Alon et al. [AFNS06] and Borgs et al. [BCL+06] independently. This motivates the following ***characterization question for invariant properties***:

> Given a subgroup Π of all permutations on $[n]$, characterize the properties of functions on $[n]$ which are invariant under Π and are testable[4]?

As discussed, the question is settled for the invariance group being the full symmetric group (trivial) and for graph properties [AFNS06, BCL+06]. For all other groups Π, the question is open as far as we know. Kaufman and Sudan [KS08] first suggested the possibility that a property's invariances might be enough

[3] $[n]$ denotes the set $\{1, 2, \ldots, n\}$.

[4] Here, we assume that the tester queries by evaluating the input function at an arbitrary point and that the distance between two functions is the fraction of $[n]$ on which they differ.

to analyze the performance of testers. We refer the reader to [GK11] for more detailed discussion about testability and invariance.

3 Affine-Invariant Properties

This survey will focus on a natural group of symmetries for algebraic properties: ***affine invariance***. Fix a prime $p \geqslant 2$ and an integer $R \geqslant 2$. Let $\mathbb{F} = \mathbb{F}_p$, the finite field of order p. We consider properties of functions $f : \mathbb{F}^n \to [R]$. Formally, for every $n > 0$, $\mathcal{P}_n$ is a subset of the functions $\mathbb{F}^n \to [R]$, and $\mathcal{P} = \cup_n \mathcal{P}_n$. In many examples, $R = p$, and the range $[R]$ is identified with $\mathbb{F}$. But our exposition will be no harder if we assume the general setting of R-colorings of $\mathbb{F}^n$.

Definition 3.1. *A property $\mathcal{P}$ is said to be affine-invariant if for every $f : \mathbb{F}^n \to [R]$ in $\mathcal{P}$ and for every affine*[5] *map $A : \mathbb{F}^n \to \mathbb{F}^n$, the function $f \circ A$ is also in $\mathcal{P}$.*

Well-studied examples of affine-invariant properties include Reed-Muller codes (in other words, bounded degree polynomials) [BFL91, BFLS91, FGL+96, RS96, AKK+05], Fourier sparsity [GOS+09], spectral norm and homogeneous polynomials. In fact, affine invariance seems to be a common feature of most interesting properties that one would classify as "algebraic". Kaufman and Sudan in [KS08] made explicit note of this phenomenon and initiated a general study of the testability of affine-invariant properties (see also [GK11] and [Sud10]). In particular, they asked for necessary and sufficient conditions for the testability of affine-invariant properties.

Note that the group of affine transformations has size $2^{O(\log^2 N)}$, where $N = p^n$ is the size of the domain of input functions. Since the invariance group for graph properties has size $2^{\tilde{O}(\sqrt{N})}$, if the characterization question is successfully resolved for affine-invariant properties, then the question would be answered for a smaller group of invariances than any other known so far.

Historically, a lot of interest has centered around testing affine-invariant properties because of a connection to coding theory (and from there, to probabilistically checkable proofs and beyond). Consider the *Reed-Muller code of order d* which one can view as the subset of functions $f : \mathbb{F}^n \to \mathbb{F}$ that are polynomials of degree at most d. The testability of Reed-Muller codes means that one can quickly test whether a string is a significant corruption of a codeword by querying only a constant number of symbols of the string. In other words, Reed-Muller is a *locally testable code*. This fact has played a central role in several complexity theoretic applications of coding theory [Lip89, GL89, STV01, TSZ03]. Affine-invariant properties are in some sense a very rich generalization of low-degree polynomials, and it offers the tantalizing possibility that there exists a locally testable code corresponding to an affine-invariant property with much better parameters than Reed-Muller. For a recent example of an application that uses testability of the Reed-Muller code (over $\mathbb{F}_2$) and would benefit from locally testable codes with better parameters, see [BGH+12].

Affine-invariant properties include many natural properties other than those directly related to low-degree polynomials and locally testable codes. Here are some examples of natural affine-invariant properties that arise frequently in learning theory, complexity, algebra and additive combinatorics:

- ***Fourier sparsity***: A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is said to be k-Fourier sparse if it has at most k nonzero coefficients in its Fourier representation.
- ***Spectral norm***: A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ has spectral norm $\leqslant k$ if $\sum_\alpha |\hat{f}(\alpha)| \leqslant k$, where $\hat{f}$ is the Fourier transform of f.
- ***Splittability***: A function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ splits into d factors if it is a product of $\leqslant d$ linear functions.
- ***d-Rank***: A function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ has d-rank at most r if $f(x) = \Gamma(P_1(x), P_2(x), \ldots, P_r(x))$ for some function $\Gamma : \mathbb{F}^r \to \mathbb{F}$ and for some r polynomials each of degree $< d$.

[5] We say $A : \mathbb{F}^n \to \mathbb{F}^n$ is affine if for every $x, y, z \in \mathbb{F}^n$, $A(x + y + z) - A(x + y) - A(x + z) + A(x) = 0$. Equivalently, $A(x) = Mx + b$ for $A \in \mathbb{F}^{n \times n}$ and $b \in \mathbb{F}^n$.

There have been two lines of work since the question about the testability of affine-invariant properties was explicitly asked by Kaufman and Sudan [KS08]. One line of work that directly stems from the results of [KS08] focuses on *linear*, affine-invariant properties. Here, attention is restricted to functions that map to a finite field, and the properties are themselves vector spaces over this field. Such properties are directly connected to locally testable codes and probabilistically checkable proofs. Linearity is a strong constraint to put on a property, and indeed, it allows for a beautiful characterization [BSMSS11] of linear, affine-invariant properties in terms of the monomials that appear in polynomials satisfying the property. Consult the excellent Complexity Theory Column survey by Sudan [Sud11] for more details and references.

The fact that linearity is such a strong structural condition however obscures our understanding of how affine invariance alone affects testability. The second line of work, which we focus on in this survey, attempts to characterize testability based on affine invariance alone. Aside from shining more light on the relationship between invariance and testability, it also allows us to handle non-linear affine-invariant properties, such as Fourier sparsity, spectral norm, splittability and rank. This research direction was first explicitly formulated by Bhattacharyya, Chen, Sudan and Xie [BCSX11], although it had precursors in a work of Green [Gre05]. Of course, there is a price to pay for more generality. In contrast to the research for linear properties, many of the results in this area are "qualitative" in nature, in the sense that either very poor or no quantitative relationships are known between different parameters. Also, current techniques require that the domain be restricted to a vector space over a fixed finite field (whereas for linear properties, much emphasis is placed on the case when the vector space is over a growing field). However, these drawbacks also demonstrate how remarkable it is that we can successfully address the characterization question just assuming affine invariance. The rest of this survey hopes to convey that excitement.

4 Towards a Characterization

Thanks to recent result, we are now much of the way towards a characterization of testable affine-invariant properties over finite fields. In order to state our results precisely, it is time we clarify the definition of testability. Let us say that a function f is *ε-far from $\mathcal{P}$* if $\min_{g\in\mathcal{P}} |\{x : f(x) \neq g(x)\}| > \varepsilon p^n$.

Definition 4.1. *A property $\mathcal{P}$ is said to be* testable *if there are functions $q : (0,1) \to \mathbb{Z}_{>0}$, $\delta : (0,1) \to (0,1)$, a real number $c \in (0,1]$, and an algorithm T that, given as input a parameter $\varepsilon > 0$ and oracle access to a function $f : \mathbb{F}^n \to [R]$, makes at most $q(\varepsilon)$ queries to the oracle for f, accepts with probability at least c if $f \in \mathcal{P}$ and accepts with probability at most $c - \delta(\varepsilon)$ if f is ε-far from $\mathcal{P}$. The function $q(\cdot)$ is called the* query complexity *and $1 - c + \delta(\cdot)$ is called the* rejection probability.

If $c = 1$, then $\mathcal{P}$ is said to be one-sided testable. *Otherwise, $\mathcal{P}$ is* two-sided testable. *If $c = 1$ and q is a constant function, then $\mathcal{P}$ is said to be* proximity-obliviously testable (PO testable).

The term proximity-oblivious testing was coined by Goldreich and Ron in [GR11][6]. As an example of a testable (in fact, PO testable) property, let us recall a famous result by Blum, Luby and Rubinfeld [BLR93]. They showed that linearity of a function $f : \mathbb{F}^n \to \mathbb{F}$ is testable by a test which makes 3 queries. This test accepts with probability 1 if f is linear and rejects with probability $\Omega(\varepsilon)$ if f is ε-far from linear.

Let us concentrate on one-sided and PO testable properties for now. A different way to view one-sided testers is that if the input f is ε-far from $\mathcal{P}$, then the tester discovers a set S of $q(\varepsilon)$ many points such that the evaluation of f on S forms a witness that f does not satisfy $\mathcal{P}$. The reason is that if there were a function g in $\mathcal{P}$ for which $g(x) = f(x)$ for all $x \in S$, the tester could not distinguish between f and g and so, would reject g with positive probability, contrary to the requirement in Definition 4.1 that g should always be accepted. So, we see that a requirement for one-sided testability is that there should exist $q(\varepsilon)$-sized set S such that $\{(x, f(x)) : x \in S\}$ is a proof for $f \notin \mathcal{P}$. Without loss of generality, we can assume S is an affine subspace of dimension $q(\varepsilon)$ (simply by putting into S everything in its linear span). Now, because $\mathcal{P}$

[6] Later on, Goldreich and Shinkar [GS12] introduced two-sided proximity oblivious testing, but here, we assume PO testability implies one-sided testability. Also, the definitions in [GR11] and [GS12] are stronger in that not only is q independent of ε, the behavior of the tester is also; only the rejection probability depends on ε. We will only need the weaker definition here.

is affine-invariant, one can argue[7] that S can be assumed to be a uniformly chosen random affine subspace of dimension $q(\varepsilon)$. So we have:

Lemma 4.2. [8] *If a property $\mathcal{P}$ is one-sided testable with query complexity $q(\varepsilon)$, there is a tester that, given input function $f : \mathbb{F}^n \to [R]$, decides only based on n and the evaluations of f on a random affine subspace of dimension $q(\varepsilon)$.*

What more can we say about how the tester acts once it makes its queries? Well, for an obvious reason, we cannot make a general statement about this. For instance, suppose that $\mathcal{P}$ is satisfied by affine forms $f : \mathbb{F}^n \to \mathbb{F}$ when n is even and satisfied by degree-2 polynomials $f : \mathbb{F}^n \to \mathbb{F}$ when n is odd. In this case, how the tester decides depends on what n is for the input function. But such properties are somewhat artificial. For all natural affine-invariant properties known to be testable, once the tester obtains its queries, its behavior is agnostic of n. We call a tester *oblivious* that accepts or rejects solely based on its queries.

Assuming obliviousness allows us to strengthen Lemma 4.2 considerably, especially for PO testable properties:

Theorem 4.3 ([BGS10])**.** *If a property $\mathcal{P}$ is obliviously PO testable with query complexity q, there is a tester that, given input function f, uniformly chooses a random subspace H of dimension q and accepts if and only if the restriction of f to H satisfies $\mathcal{P}$.*

Notice that now, the test itself just checks for membership in $\mathcal{P}$, albeit for a function on a constant-sized domain. This fact motivates the following definition of ***local characterization***:

Definition 4.4. *For an integer $K > 0$, an affine-invariant property $\mathcal{P}$ is said to be K-*locally characterized *if both of the following hold:*

- *For every function $f : \mathbb{F}^n \to [R]$ in $\mathcal{P}$ and every affine subspace H of $\mathbb{F}^n$, the restriction of f to H, denoted $f|_H$, is also in $\mathcal{P}$.*
- *If a function $f : \mathbb{F}^n \to [R]$ is not in $\mathcal{P}$ and $n > K$, then there exists a subspace H of dimension K such that the restriction $f|_H$ is also not in $\mathcal{P}$.*

The constant K is said to be the locality *of $\mathcal{P}$.*

Note that local characterization is a purely combinatorial condition on a property. Theorem 4.3 amounts to saying that every obliviously PO testable property is locally characterized. Now, almost all affine-invariant properties of interest known to be testable are also known to be (obliviously) PO testable, and indeed, PO testability seems to often be taken for granted when discussing algebraic property testing. We are then led to ask whether the converse of Theorem 4.3 is true, that is whether every locally characterized property is also obliviously PO testable. This question was answered in the affirmative in [BFH+13]:

Theorem 4.5 ([BFH+13])**.** *If $\mathcal{P}$ is a K-locally characterized property, then $\mathcal{P}$ is obliviously PO testable. The tester uniformly chooses a random subspace H of dimension K and accepts exactly when the restriction of the input function to H satisfies $\mathcal{P}$.*

We will sketch how this theorem is proved in the next section. Theorem 4.3 and Theorem 4.5 together give a combinatorial characterization of natural PO testable properties. This answers a question asked by Sudan in [Sud10].

Such a characterization still remains open for one-sided testable properties. The only known natural affine-invariant property that is one-sided testable but not PO testable is *odd-cycle-freeness*, introduced in [BGRS12]. This is the property of whether for a function $f : \mathbb{F}_2^n \to \{0, 1\}$, there exists $\alpha \in \mathbb{F}_2^n$ such that $\hat{f}(\alpha) = -\rho$ where $\rho = \mathbf{E}_x[f(x)]$. [BGRS12] showed that odd-cycle-freeness is testable (with $\mathrm{poly}(1/\varepsilon)$ query complexity!). However, a general result analogous to Theorem 4.5 remains elusive. For that, as Bhattacharyya, Grigorescu and Shapira [BGS10] show, the following question needs to be resolved.

[7] Think about what happens when you compose the input function f with a random bijective affine transformation.

[8] In fact, this lemma also holds for 2-sided testable properties by the techniques of [GT03] used to show a similar statement for graph properties.

Question 4.6. *An affine-invariant property $\mathcal{P}$ is said to be* **subspace-hereditary** *if whenever a function f satisfies $\mathcal{P}$, the restriction of f to any affine subspace also satisfies $\mathcal{P}$. Is it true that every subspace-hereditary property is obliviously one-sided testable?*[9] *In particular, what is the performance of the test that chooses a random affine subspace H of dimension $q_{\mathcal{P}}(\varepsilon)$ and checks whether $\mathcal{P}$ is satisfied for the restriction of the input to H, where $q_{\mathcal{P}}$ is a positive monotone decreasing function determined by the property $\mathcal{P}$?*

The difference between subspace-hereditary and locally characterized properties is that for subspace-hereditary properties, there is no bound on the size of a witness for non-satisfiability. Even this issue can be tackled to some extent currently. One can define a notion of "complexity" for subspace-hereditary properties, based on the notion of Cauchy-Schwarz complexity for linear forms defined by Green and Tao [GT10b] in their celebrated work on arithmetic sequences in primes. (For the precise definition, see [BFL13] or [BFH+13]; we forego the details here.) The work [BFH+13] shows that the proof of Theorem 4.5 extends to bounded complexity subspace-hereditary properties.

Theorem 4.7 ([BFH+13])**.** *Any subspace-hereditary, affine-invariant property of bounded complexity is obliviously one-sided testable.*

For instance, odd-cycle-freeness has complexity 1, even though there is no bound on the size of witnesses needed to prove non-membership; so its testability is also explained by this theorem (although with a *much* worse query complexity than in [BGRS12]). In order to resolve Question 4.6, we therefore need to settle the testability of subspace-hereditary properties with unbounded complexity. It is not clear if there exists any such natural property.

Question 4.8. *Does there exist a natural subspace-hereditary property of unbounded complexity?*

Finally, for one-sided testable properties, the analog of Theorem 4.3 is also known. If a property $\mathcal{P}$ is obliviously one-sided testable, then it is "semi-subspace-hereditary" [BGS10]. For details, consult the paper, but roughly speaking, a semi-subspace-hereditary property $\mathcal{P}$ is close to a subspace-hereditary property $\mathcal{P}'$, in the sense that $\mathcal{P}'$ contains $\mathcal{P}$ and any function $f : \mathbb{F}^n \to [R]$ that's ε-far from $\mathcal{P}$ is not in $\mathcal{P}'$ (for large enough n in terms of ε). Oblivious one-sided testability of subspace-hereditary properties implies oblivious one-sided testability of semi-subspace-hereditary properties also. So, an affirmative answer to Question 4.6 would provide a characterization of the natural, one-sided testable properties.

For two-sided testability, the current state of affairs is not as advanced. The only relevant work is the very recent paper by Hatami and Lovett [HL13], where they prove the following.

Theorem 4.9 ([HL13])**.** *If an affine-invariant property $\mathcal{P}$ is testable (not necessarily one-sided), then for every $\varepsilon, \delta > 0$, one can distinguish with probability at least $2/3$ between functions that are δ-close to $\mathcal{P}$ and functions that are $(\delta + \varepsilon)$-far from $\mathcal{P}$ using only a constant number of queries (that may depend on ε and δ).*

Thus, one can estimate the distance to the property within any fixed additive error in constant time. The constant-time algorithm itself is also simple to state. [HL13] shows that there exists an integer $k_{\mathcal{P},\varepsilon,\delta}$ such that if a random subspace H of dimension $k_{\mathcal{P},\varepsilon,\delta}$ is chosen, then with high probability, the distance of the restriction $f|_H$ to $\mathcal{P}$ is within δ of the distance of f to $\mathcal{P}$.

Although this result does not directly have any consequences for the characterization question, historical precedent suggests ultimately it should. For graph properties, a similar result that testability implies distance estimability [FN07] was a key component behind the characterization by Alon et al. [AFNS06] of graph property testability.

5 Testability of Locally Characterized Properties

In this section, we aim to sketch the main components in the proof of Theorem 4.5 (which also form most of the proof of Theorem 4.7).

[9] In [BGS10], a positive answer to this question was conjectured. We currently feel unsure about the truth of this, and so we pose the issue as a question.

5.1 A brief history

The proof has an interesting history. Initially, attention was focused on the case when $R = 2$ and the locally characterized property $\mathcal{P}$ is *monotone*. Monotone means that if a function $f : \mathbb{F}^n \to \{0,1\}$ is in $\mathcal{P}$, then changing the value of f from 1 to 0 at some point still keeps f in $\mathcal{P}$. For instance, the property that there does not exist a subspace H of dimension k such that $f|_H \equiv 1$ is a monotone k-locally characterized property. Green [Gre05] had already shown testability of a special case of this problem, namely properties satisfied by functions f that do not contain in its support ($f^{-1}(1)$) solutions to a fixed linear equation. A particular case of interest is *Boolean triangle-freeness*, which is satisfied by functions $f : \mathbb{F}_2^n \to \{0,1\}$ for which there are no three elements $x, y, z \in f^{-1}(1)$ such that $x + y + z = 0$. He left open the question of whether testability holds when the support is forbidden to contain solutions to a fixed system of linear equations (rather than a single linear equation). [BCSX11] showed that Green's Fourier-analytic proof holds whenever a system of linear equations has Cauchy-Schwarz complexity 1. [KSV09] independently proved a similar statement and importantly, also showed that Green's result could be reproved by a reduction to graph properties. This latter result gave hope that perhaps Theorem 4.5 could be established by reducing testing of affine-invariant properties to testing hypergraph properties. Shapira [Sha09] and Král, Serra and Vena [KSV12] independently used this approach successfully to show testability for all monotone locally characterized affine-invariant properties.

For non-monotone locally characterized properties (such as all the examples in the previous section), however, the more analytic approach by Green [Gre05] and by Bhattacharyya et al. [BCSX11] seemed more promising. This was borne out by [BGS10] where every affine-invariant property of complexity 1 was showed to be one-sided testable, again through Fourier analysis. The barrier of complexity 1 was crossed in [BFL13] by using higher-order Fourier analysis, although this work restricted itself to properties over finite fields of large characteristic. Finally, [BFH+13] proved Theorem 4.5 over all fields, even of characteristic 2, by relying on a recent proof of the Gowers inverse theorem over low characteristic fields by Tao and Ziegler [TZ11].

This story is similar in spirit to the route taken to prove one-sided testability of hereditary k-uniform hypergraph properties[10] [RS09]. First, Alon et al. [AFKS00] showed testability of subgraph-freeness in graphs. Then, Alon and Shapira [AS08b] showed that every monotone graph property is testable. Subsequently, they [AS08a] extended it to hereditary graph properties. Finally, Rödl and Schacht [RS09] extended the result from graphs to hypergraphs, using previously established generalizations of graph-theoretic machinery to k-uniform hypergraphs [RS04, NRS06]. Graphs seem to be analogous to affine-invariant properties of complexity 1 and hypergraphs of bounded uniformity analogous to affine-invariant properties of bounded complexity. Indeed, a reader with some knowledge about the proof techniques in graph property testing will find that the overall structure of the proof of Theorem 4.5 in [BFH+13] very familiar. However, the actual arguments which are needed to simulate graph-theoretic arguments for the affine-invariant case are very different. Indeed, while the (hyper)graph-theoretic machinery is purely combinatorial, the analogous analytic machinery does not currently have constructive proofs. We return to these issues later, but first, let us delve into the proof of Theorem 4.5.

5.2 A sketch of the proof

We illustrate the proof with an example. Let $\mathcal{P}$ be a property of functions $f : \mathbb{F}^n \to \{0,1\}$, where f is in $\mathcal{P}$ exactly when there are no $x, y, z \in \mathbb{F}^n$ satisfying $f(x) = f(x + y + z) = 1$ and $f(x + y) = f(x + z) = 0$. It is easy to check that $\mathcal{P}$ is affine-invariant and 2-locally characterized. We analyze the tester T which picks random x, y, z uniformly from $\mathbb{F}^n$ and rejects iff $f(x) = f(x + y + z) = 1$ and $f(x + y) = f(x + z) = 0$. Clearly, if $f \in \mathcal{P}$, T accepts with probability 1. Now, we need to lower bound its rejection probability when f is a function that is ε-far from $\mathcal{P}$.

Let us try to do this analysis when f is of a "structured" form. We will then try to use the intuition obtained to treat arbitrary ε-far f.

To begin with, consider functions f that only depend on the first r coordinates of its input, where r is a constant. So, if $n > r$ and $w \in \mathbb{F}^r$, then $f(w \circ z)$ is the same for all values of $z \in \mathbb{F}^{n-r}$. In other words, there

[10] A k-uniform hypergraph property $\mathcal{P}$ is hereditary if it is closed under removal of vertices (and all incident edges).

are p^r equal sized cells $C_w = \{w \circ z : z \in \mathbb{F}^{n-r}\}$ on each of which f is constant. Now, suppose f does not satisfy $\mathcal{P}$, meaning there exist $\bar{x}, \bar{y}, \bar{z}$ with $f(\bar{x}) = f(\bar{x} + \bar{y} + \bar{z}) = 1$ and $f(\bar{x} + \bar{y}) = f(\bar{x} + \bar{z}) = 0$. Suppose $\bar{x} \in C_{w_1}$, $\bar{x} + \bar{y} \in C_{w_2}$, $\bar{x} + \bar{z} \in C_{w_3}$ and $\bar{x} + \bar{y} + \bar{z} \in C_{w_4}$. It follows that f is 1 on all of C_{w_1} and C_{w_4} and 0 on all of C_{w_2} and C_{w_3}. But then observe that for any $a \in C_{w_1}, b \in C_{w_2}, c \in C_{w_3}$, it's true that $a + b + c \in C_{w_4}$, and any such a, b, c produces x, y, z such that $f(x) = f(x + y + z) = 1$ and $f(x + y) = f(x + z) = 0$. So, there are at least p^{3n-3r} many such violating x, y, z, so that the rejection probability of T is at least p^{-3r}, a constant.

Next, consider functions f of the form $f(x) = \Gamma(\ell_1(x), \ell_2(x), \ldots, \ell_r(x))$ where $\ell_1, \ldots, \ell_r : \mathbb{F}^n \to \mathbb{F}$ are linear functions and $\Gamma : \mathbb{F}^r \to \{0, 1\}$ is arbitrary (the previous paragraph is a special case where the linear functions are all projections). We can assume $\ell_1, \ldots, \ell_r$ are linearly independent without loss of generality (otherwise, r can be made smaller). Now, observe that we can compose f with a bijective affine transformation A such that $f \circ A$ is now a function of r projections. Moreover, $f \circ A \notin \mathcal{P}$ because of affine invariance. We can thus use the previous paragraph to conclude that the rejection probability is at least p^{-3r} in this case also.

This motivates considering functions f of the form $f(x) = \Gamma(P_1(x), P_2(x), \ldots, P_r(x))$, where each $P_i : \mathbb{F}^n \to \mathbb{F}$ is a polynomial of degree at most d and $\Gamma : \mathbb{F}^r \to \{0, 1\}$ is arbitrary. Again, for such a function f, there are p^r many cells $C_w = \{x : P_1(x) = w(1), P_2(x) = w(2), \ldots, P_r(x) = w(r)\}$, one for each $w \in \mathbb{F}^r$, such that f is constant on each cell. Assuming $f \notin \mathcal{P}$ implies four cells $C_{w_1}, C_{w_2}, C_{w_3}, C_{w_4}$ such that there exist $\bar{x}, \bar{y}, \bar{z}$ with $\bar{x} \in C_{w_1}$, $\bar{x} + \bar{y} \in C_{w_2}$, $\bar{x} + \bar{z} \in C_{w_3}$ and $\bar{x} + \bar{y} + \bar{z} \in C_{w_4}$, and moreover, f is 1 on all of C_{w_1} and C_{w_4} and 0 on all of C_{w_2} and C_{w_3}. We would like to argue that there are many x, y, z with $x \in C_{w_1}$, $x + y \in C_{w_2}$, $x + z \in C_{w_3}$ and $x + y + z \in C_{w_4}$. In other words, we want to lower bound:

$$\mathop{\mathbf{P}}_{x,y,z}[(P_i(x)) = w_1 \wedge (P_i(x + y)) = w_2 \wedge (P_i(x + z)) = w_3 \wedge (P_i(x + y + z)) = w_4] \tag{1}$$

where $(P_i(x))$ is denoting the tuple $(P_1(x), P_2(x), \ldots, P_r(x))$. How do we analyze this quantity? In fact, how do we even lower bound $\mathbf{P}_x[(P_i(x)) = w_1]$? A priori, this could be very small. Indeed, we don't have any control over w_1. So, we somehow need to make sure that $\mathbf{P}_x[(P_i(x)) = w_1]$ is large for every value of w_1.

The key step here is to recognize that $\mathbf{P}_x[(P_i(x)) = w_1] \geqslant 0.99p^{-r}$ if the *rank* of $P_1, \ldots, P_r$ is sufficiently large, where rank of $P_1, \ldots, P_r$ is defined as the minimum over all $(\lambda_1, \ldots, \lambda_r) \neq (0, \ldots, 0)$ of the d-rank of $\lambda_1 P_1 + \cdots + \lambda_r P_r$, where $d = \deg(\lambda_1 P_1 + \cdots + \lambda_r P_r)$ and d-rank as defined in Section 3. This is due to the following theorem.

Theorem 5.1 (Theorem 4 of [KL08]). *For any $\varepsilon > 0$ and integer $d > 0$, there exists $r = r_{5.1}(d, \varepsilon)$ such that the following is true. If $P : \mathbb{F}^n \to \mathbb{F}$ is a degree-d polynomial with d-rank greater than r, then*[11] $\mathsf{bias}(P) \stackrel{\text{def}}{=} |\mathbf{E}_x[\mathsf{e}(P(x))]| < \varepsilon$.

Once we have that the bias of every nonzero linear combination of $P_1, \ldots, P_r$ is small, it is an easy matter to show that $(P_i(x)) = (P_1(x), P_2(x), \ldots, P_r(x))$ is nearly equidistributed.

We have to somehow make sure that the rank of $P_1, \ldots, P_r$ is large, but let's assume this and skip ahead and try lower bounding (1) itself. Here potentially lies danger! Observe that if say P_1 is linear (and so is also automatically high rank), we have the identity $P_1(x) - P_1(x + y) - P_1(x + z) + P_1(x + y + z) = 0$. So, unless $w_1(1) - w_2(1) - w_3(1) + w_4(1) = 0$, the quantity in (1) equals 0. This, by itself, is not so bad because we already know that for $\bar{x}, \bar{y}, \bar{z}$, we have $P_1(\bar{x}) = w_1(1), P_1(\bar{x} + \bar{y}) = w_2(1), P_1(\bar{x} + \bar{z}) = w_3(1)$ and $P_1(\bar{x} + \bar{y} + \bar{z}) = w_4(1)$, so that it must be the case $w_1(1) - w_2(1) - w_3(1) + w_4(1) = 0$. But this example gives rise to the fear that perhaps $(P_1(x), P_1(x + y), P_1(x + z), P_1(x + y + z))$ could be heavily biased towards some small subset T of tuples without always being in T. The following result, which is actually the technical centerpiece of [BFL13] and [BFH+13], ensures that this fear is unjustified if the rank of $P_1, \ldots, P_r$ is large enough.

[11] $\mathsf{e}(x) \stackrel{\text{def}}{=} e^{2\pi i x/p}$.

Theorem 5.2 (Near orthogonality dichotomy). *Given $\varepsilon > 0$, suppose $P_1, \dots, P_r$ is a collection of polynomials of degree $d > 0$ and rank $> r_{5.2}(d, \varepsilon)$ and Λ is a tuple of integers $(a_1, \dots, a_r, b_1, \dots, b_r, c_1, \dots, c_r, d_1, \dots, d_r)$.*

$$P_\Lambda(x, y, z) = \sum_{i \in [r]} a_i P_i(x) + \sum_{i \in [r]} b_i P_i(x + y) + \sum_{i \in [r]} c_i P_i(x + z) + \sum_{i \in [r]} d_i P_i(x + y + z).$$

Then, one of the two statements below is true.

- *For every $i \in [r]$, it holds that $a_i Q_i(x) + b_i Q_i(x + y) + c_i Q_i(x + z) + d_i Q_i(x + y + z) \equiv 0$ for all polynomials Q_i with the same degree as P_i. Clearly, $P_\Lambda \equiv 0$ in this case.*
- *P_Λ is non-constant. Moreover, $|\mathbf{E}_{x,y,z}[\mathsf{e}(P_\Lambda(x, y, z))]| < \varepsilon$.*

From Theorem 5.2, it is not hard to lower bound (1) by using Fourier-type arguments.

Let us now return to the deferred question of how we can make sure that $P_1, \dots, P_r$ have high rank. The answer is by repeated refinement. If it is the case that $P_1, \dots, P_r$ have low rank, meaning that some nonzero linear combination of $P_1, \dots, P_r$ is a function of a small number of strictly lower degree polynomials, we can replace the highest degree polynomial in the support of the linear combination with the lower degree polynomials. The induction stops at some point because the degree sequence of the polynomial collection decreases according to the lexicographic order. In this way, for any monotone growth function $R : \mathbb{Z}^+ \to \mathbb{Z}^+$, starting from a collection of polynomials $\{P_1, \dots, P_C\}$, we can produce a new collection $\{P'_1, \dots, P'_d\}$, where d is bounded as a function of R and c, the partition of $\mathbb{F}^n$ produced by $\{P'_1, \dots, P'_d\}$ is a refinement of the partition produced by $\{P_1, \dots, P_c\}$, and the rank of the collection $P'_1, \dots, P'_d$ is at least $R(d)$.

The above more or less completes the argument for when f is structured of the form $\Gamma(P_1(x), P_2(x), \dots, P_r(x))$. Now, let us look at the other end of the "structured-versus-random" dichotomy. It's clear that if one of $f_1, f_2, f_3, f_4 : \mathbb{F}_2^n \to [-1, 1]$ is a random function, then $\mathbf{E}_{x,y,z}[f_1(x)f_2(x + y)f_3(x + z)f_4(x + y + z)]$ is very small with high probability. Can we "derandomize" this fact? Luckily, such averages have been extensively studied before in additive combinatorics and analytic number theory. In particular, we have the *Gowers norm* $\| \cdot \|_{U^{d+1}}$ [Gow01], which allows us to say the following: If $\|f_1\|_{U^{d+1}} < \varepsilon$, $f_2, \dots, f_m$ are arbitrary functions that are bounded inside $[-1, 1]$, and $L_1, \dots, L_m$ are linear forms of complexity at most d, then

$$\left| \mathop{\mathbf{E}}_{x_1, \dots, x_\ell \in \mathbb{F}^n} \left[\prod_{i=1}^{m} f_i(L_i(x_1, \dots, x_\ell)) \right] \right| \leqslant \varepsilon. \tag{2}$$

In our case, the set of linear forms $\{x, x + y, x + z, x + y + z\}$ has complexity 1, so that making the U^2 norm of f_1 be small will be enough to make $\mathbf{E}_{x,y,z}[f_1(x)f_2(x + y)f_3(x + z)f_4(x + y + z)]$ be small[12].

What is the Gowers norm? The Gowers norm of order d for a function f is given by the expected (multiplicative) derivative of f in d random directions at a random point.

Definition 5.3 (Gowers norm). *Given a function $f : \mathbb{F}^n \to \mathbb{C}$ and an integer $d \geqslant 1$, the* Gowers norm of order d *for f is:*

$$\|f\|_{U^d} = \left| \mathop{\mathbf{E}}_{h_1, \dots, h_d, x \in \mathbb{F}^n} [(\Delta_{h_1} \Delta_{h_2} \cdots \Delta_{h_d} f)(x)] \right|^{1/2^d}$$

where $\Delta_h f(x) = f(x + h)\overline{f(x)}$.

Now, the question is how to merge the arguments for structured and random functions so as to handle arbitrary functions $f : \mathbb{F}^n \to \{0, 1\}$ that are ε-far from $\mathcal{P}$. Here, we can make use of *decomposition theorems* pioneered by Green and Tao [Gre07], which are analogous to the Szemerédi regularity lemma for graphs [Sze78].

[12] The U^2 norm is given by the Fourier coefficients, so this whole example can be handled more elementarily (as was done in [BGS10]). But let us ignore this so that the argument also works for the general case of affine forms with higher complexity.

Theorem 5.4 (Strong Decomposition Theorem; Theorem 4.4 of [BFL13])**.** *Suppose $\delta > 0$ and $d \geqslant 1$ are integers. Let $\eta : \mathbb{N} \to \mathbb{R}^+$ be an arbitrary non-increasing function and $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary non-decreasing function. Then there exist $N = N_{5.4}(\delta, \eta, r, d)$ and $C = C_{5.4}(\delta, \eta, r, d)$ such that the following holds.*

Given $f : \mathbb{F}^n \to \{0,1\}$ where $n > N$, there exist three functions $f_1, f_2, f_3 : \mathbb{F}^n \to \mathbb{R}$ and a collection of polynomials $P_1, \ldots, P_C$ of degree at most d such that the following conditions hold:

(i) $f = f_1 + f_2 + f_3$.

(ii) For any $x \in \mathbb{F}^n$, if $(P_1(x), P_2(x), \ldots, P_C(x)) = w$, then $f_1(x) = \mathbf{E}_{y:(P_1(x),P_2(x),\ldots,P_C(x))=w}[f(y)]$.

(iii) $\|f_2\|_{U^{d+1}} \leqslant 1/\eta(C)$.

(iv) $\|f_3\|_2 \leqslant \delta$.

(v) f_1 and $f_1 + f_3$ have range $[0,1]$; f_2 and f_3 have range $[-1,1]$.

(vi) Rank of $P_1, \ldots, P_C$ is at least $r(C)$.

In the above decomposition, f_1 is constant on each cell of the partition formed by $P_1, \ldots, P_C$, so that f_1 is some function of $P_1, \ldots, P_C$. And, f_2 has small U^{d+1}-norm and f_3 is small in L^2. With the decomposition theorem in hand, we can now try to lower bound the rejection probability of the tester:

$$\mathop{\mathbf{E}}_{x,y,z}[f(x)g(x+y)g(x+z)f(x+y+z)] \tag{3}$$

where $g(\cdot) = 1 - f(\cdot)$. We apply Theorem 5.4 to each of f and g (we can make sure we use the same set of polynomials to define f_1 and g_1). Expanding the multiplication, we find we have a sum of 81 expectations. 65 of them have some occurrence of f_2 or g_2, and so, each of these terms we can bound using (2). The other 16 come from

$$\mathop{\mathbf{E}}_{x,y,z}[(f_1+f_3)(x)\cdot(g_1+g_3)(x+y)\cdot(g_1+g_3)(x+z)\cdot(f_1+f_3)(x+y+z)]. \tag{4}$$

Now, $f_1 + f_3$ is in $[0,1]$, so it is enough to lower-bound this expectation when $x, x+y, x+z, x+y+z$ come from four specific cells from the partition formed by $P_1, \ldots, P_C$. In particular, because f is not in $\mathcal{P}$, we know there are four cells $C_1 \ni \bar{x}, C_2 \ni \bar{x}+\bar{y}, C_3 \ni \bar{x}+\bar{z}, C_4 \ni \bar{x}+\bar{y}+\bar{z}$ for some $\bar{x}, \bar{y}, \bar{z}$. Moreover, using the fact that f is ε-far from $\mathcal{P}$, we can ensure some other properties of the cells, such as f_1 is constant and close to 1 on all of C_1 and C_3 and g_1 is constant and close to 1 on all of C_2 and C_4. The argument from the discussion about "structured" functions shows that there is a constant density of x, y, z such that $x \in C_1, x+y \in C_2, x+z \in C_3, x+y+z \in C_4$. Thus, we have a lower bound on $\mathbf{E}_{x,y,z:x\in C_1, x+y\in C_2, x+z\in C_3, x+y+z\in C_4}[f_1(x)g_1(x+y)g_1(x+z)f_1(x+y+z)]$.

As for the terms which involve f_3 and g_3, the situation is a bit complicated, because δ, the upper bound on $\|f_3\|_2$ and $\|g_3\|_2$, does not depend on C. Fortunately, we can simulate a technique from [AFKS00] for graph property testing to resolve this situation. Another refinement of the partition is made such that inside each cell of the original partition, there is a subcell of the new partition so that the L^2 norms of f_3 and g_3 restricted to those subcells are decreasing as a function of C. An additional equidistribution result about high rank polynomials is also needed (somewhat similar in spirit to some claims in [GT10a]), but we omit more details here.

This more or less completes the proof sketch of Theorem 4.5…except for a crucial detail! Let's go back to Theorem 5.4 and ask how it can be proved. Given an arbitrary function f, if $\|f\|_{U^{d+1}}$ is small, then we are already done. Otherwise, we repeatedly appeal to the *Gowers inverse theorem* to find a finite collection of polynomials $P_1, \ldots, P_C$ of degree $\leqslant d$ such that $f = \Gamma(P_1, \ldots, P_C) + h$, where $\|h\|_{U^{d+1}}$ is small and Γ is some function. When $d < p$, the Gowers inverse theorem is the following statement:

Theorem 5.5 (Gowers Inverse Theorem [BTZ10])**.** *Given positive integers $d < p$, for every $\delta > 0$, there exists $\varepsilon = \varepsilon_{5.5}(\delta, p)$ such that if $f : \mathbb{F}^n \to \mathbb{R}$ satisfies $\|f\|_\infty \leqslant 1$ and $\|f\|_{U^{d+1}} \geqslant \delta$, then there exists a polynomial $P : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree at most d so that $|\mathbf{E}_x[f(x) \cdot \mathsf{e}(P(x))]| \geqslant \varepsilon$.*

But if $d < p$ (e.g., p is a fixed small constant), this theorem may no longer hold. Indeed, [GT09, LMS08] proved that if f equals the symmetric degree-4 polynomial over $\mathbb{F}_2$ and $d = 3$, we have an explicit counterexample to such a claim. Fortunately, Bergelson, Tao and Ziegler [BTZ10, TZ10, TZ11] showed that it is possible to salvage the decomposition theorem by replacing classical $\mathbb{F}$-valued polynomials with *non-classical polynomials*. For an integer $k > 0$, let $\mathbb{U}_k$ denote $\frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$. Note that $\mathbb{U}_1$ can be identified with $\mathbb{F}$.

Definition 5.6 (Non-classical polynomials)**.** *For integers $d, k > 0$, a* non-classical polynomial of degree $\leqslant d$ and depth $\leqslant k$ *is a function $P : \mathbb{F}^n \to \mathbb{U}_{k+1}$ such that for all $h_1, \ldots, h_{d+1}, x \in \mathbb{F}^n$, it holds that*

$$(D_{h_1} \cdots D_{h_{d+1}} P)(x) = 0. \tag{5}$$

where $D_h P(x) = P(x + h) - P(x)$.

Classical polynomials have depth 0. For a polynomial with non-zero depth, consider the function $P : \mathbb{F}_2 \to \mathbb{U}_2$ with $P(0) = 0$ and $P(1) = 1/4$. Now, we have $D_1P(0) = 1/4$ and $D_1P(1) = 3/4$, and $D_1D_1P(0) = 1/2$ and $D_1D_1P(1) = 1/2$; so $D_1D_1D_1P \equiv 0$. Thus, P is a non-classical polynomial of degree 2 and depth 1. Note that there are no classical polynomials over $\mathbb{F}_2$ of degree 2.

It is clear that a function $f : \mathbb{F}^n \to \mathbb{C}$ with $\|f\|_\infty \leqslant 1$ satisfies $\|f\|_{U^{d+1}} = 1$ if and only if $f = \mathsf{e}(P)$ for a non-classical polynomial P of degree $\leqslant d$. Tao and Ziegler's result shows that this correspondence is robust.

Theorem 5.7 (Theorem 1.11 of [TZ11])**.** *Suppose $\delta > 0$ and $d \geqslant 1$ is an integer. There exists an $\varepsilon = \varepsilon_{5.7}(\delta, d)$ such that the following holds. For every function $f : \mathbb{F}^n \to \mathbb{R}$ with $\|f\|_\infty \leqslant 1$ and $\|f\|_{U^{d+1}} \geqslant \delta$, there exists a non-classical polynomial P of degree $\leqslant d$ so that $|\mathbf{E}_{x \in \mathbb{F}^n} f(x)\mathsf{e}(-P(x))| \geqslant \varepsilon$.*

A (stronger) generalization of Theorem 5.1 to non-classical polynomials also holds:

Theorem 5.8 (Theorem 1.20 of [TZ11])**.** *For any $\varepsilon > 0$ and integer $d > 0$, there exists an integer $r = r_{5.8}(d, \varepsilon)$ such that the following is true. For any non-classical polynomial P of degree $\leqslant d$, if $\mathsf{rank}(P) > r$, then $\|\mathsf{e}(P)\|_{U^d} \leqslant \varepsilon$.*

Note that $\mathsf{bias}(P) = \|\mathsf{e}(P)\|_{U^1} \leqslant \|\mathsf{e}(P)\|_{U^d}$, so that Theorem 5.1 also follows. The goal is now to use Theorem 5.7 and Theorem 5.8 to generalize all of the previous discussion from classical polynomials to non-classical ones. This is what [BFH+13] accomplishes. Non-classical polynomials introduce subtleties in a lot of places. For instance, the definition of rank needs to be changed:

Definition 5.9 (Rank)**.** *A sequence of non-classical polynomials $P_1, \ldots, P_C$ with respective depths $k_1, \ldots, k_C$ is said to have* rank r *if r is the least integer for which there exist $(\lambda_1, \ldots, \lambda_C) \in \mathbb{Z}^C$ so that $(\lambda_1 \mod p^{k_1+1}, \ldots, \lambda_C \mod p^{k_C+1}) \neq (0, \ldots, 0)$ and the polynomial $Q = \sum_{i=1}^{C} \lambda_i P_i$ satisfies d-$\mathsf{rank}(Q) \leqslant r$ where $d = \max_i \deg(\lambda_i P_i)$.*

Note that since λ can be a multiple of p, rank measured with respect to $\deg(\lambda P)$ is not the same as rank measured with respect to $\deg(P)$. So, for instance, if P is a non-classical polynomial of degree d and depth k, then

$$\mathsf{rank}(\{P\}) = \min\{d\text{-}\mathsf{rank}(P), (d - (p-1))\text{-}\mathsf{rank}(pP), \cdots, (d - k(p-1))\text{-}\mathsf{rank}(p^k P)\}.$$

This makes refinement of a set of polynomials into a high rank set of polynomials trickier. Also, the proof of Theorem 5.2 cannot use the monomial structure of polynomials (as was done in a predecessor paper by Hatami and Lovett [HL11] for the high characteristic case) and can only use the fact that the $(d+1)$'th order derivative of degree-d polynomials vanish. [BFH+13] gives such a proof of Theorem 5.2 which uses Theorem 5.8.

6 Open Directions

Beyond the questions asked in the text, here are some of our personal favorites:

– Tao and Ziegler's proofs for Theorem 5.7 and Theorem 5.8 are completely non-explicit. The function $\varepsilon_{5.7}$ is known to exist and be positive, but there is no explicit nonzero lower bound. Similarly, for the quantity $r_{5.8}$, it is known to be finite but we do not have any explicit upper-bound. Thus, the testability algorithms described here do not have any explicit performance guarantees.

  If we consider the barriers faced to make the proof be constructive, one problem particularly stands out. Suppose we have a collection $\mathcal{F}$ of polynomials $P_1, \ldots, P_m$. As discussed in the last section, it's possible to form a new collection $\mathcal{F}' = \{Q_1, \ldots, Q_M\}$ such that the partition produced by $\mathcal{F}'$ is a refinement of the partition produced by $\mathcal{F}$ and additionally, $\mathcal{F}'$ has high rank, hence enjoying pseudorandomness properties. But we do not have an explicit way to construct $\mathcal{F}'$. Is there one?

  In joint work with Hatami and Tulsiani [BHT13], we show that for a different but related notion of pseudorandomness, it is possible to explicitly (in fact, in polynomial time) construct a pseudorandom refinement. More specifically, for a function $\gamma : \mathbb{Z}^+ \to (0,1)$, a collection of C polynomials is said to be γ-uniform if every nonzero linear combination of the polynomials has Gowers norm smaller than $\gamma(C)$. We show that given any collection $\mathcal{F}$ of polynomials of degree $d < p$ and a function $\gamma : \mathbb{Z}^+ \to (0,1)$, there exists a randomized algorithm running in time $O(n^d)$ that with high probability outputs a collection $\mathcal{F}'$ that is γ-uniform.

  This result has the following application. Suppose P is a polynomial of degree d for which it is known that there exists a polynomial Q of degree k such that $\mathsf{dist}(P, Q) < 1 - \frac{1}{p} - \varepsilon$ for some $k \leqslant d < p$. This implies that $\|\mathsf{e}(P')\|_{U^{k+1}} > \varepsilon$ for some multiple P' of P. Now the Gowers inverse theorem (Theorem 5.5) gives a way to reverse this implication. Starting from $\|\mathsf{e}(P')\|_{U^{k+1}} > \varepsilon$, it implies the existence of some polynomial Q' of degree k such that $\mathsf{dist}(P, Q') \leqslant 1 - \frac{1}{p} - \eta$ for some $\eta > 0$ depending on ε. Using the above result for constructing γ-uniform refinements, we can make the last statement algorithmic. So we have that if a polynomial P of degree d is within relative Hamming distance $1 - \frac{1}{p} - \varepsilon$ of some unknown polynomial Q of degree k over a prime field $\mathbb{F}$ (for $k < d < p$), then there is an $O(n^d)$ time algorithm for finding a degree-k polynomial Q' which is within distance $1 - \frac{1}{p} - \eta$ of P, for some η depending on ε.

  This is remarkable because $1 - \frac{1}{p}$ is well beyond the list decoding radius of $1 - \frac{k}{p}$ for Reed-Muller codes of order $k < p$ [Gop10]. Thus, for specially structured received words (those which are themselves Reed-Muller codes of a higher degree d), it is possible to discover a nearby codeword in polynomial time even when it is much farther away than the list decoding radius[13]. This immediately suggests the following question:

  Question 6.1. *Given a function f over $\mathbb{F}^n$, if there exists Q of degree k such that $\mathrm{dist}(f, Q) \leqslant 1 - \frac{1}{p} - \varepsilon$, can one find a Q' of degree k (in time polynomial in n) such that $\mathrm{dist}(f, Q') \leqslant 1 - \frac{1}{p} - \eta$ for some η depending on ε?*

  The only such result currently known is by Tulsiani and Wolf [TW11] who give a decoding algorithm that works for $k = 2$ over $\mathbb{F}_2$. For larger k, the question is open. A positive answer would make the Gowers inverse theorem algorithmic.

– As the above discussion indicates, the current proof does not give any useful bound for the rejection probability in Theorem 4.5. So, it happens that there are natural concrete problems for which we do not know any explicit bounds for the rejection probability of the natural test! For example, consider the property $\mathcal{P}$ that is satisfied by polynomials which can be written as a product of two degree 2 polynomials. [BFH+13] shows that $\mathcal{P}$ is testable, but since that is the only known proof, there is no explicit bound on the performance of the test.

  For starters, one can consider very simple affine-invariant properties, such as Boolean triangle-freeness. For this property, there is an explicit analysis [Gre05] showing that the canonical test which samples

[13] In fact, it's almost as far away as possible: the distance between a random function and polynomials of degree k is $1 - \frac{1}{k} - O(1)$.

random x, y and checks whether the input function f satisfies $f(x) = f(y) = f(x+y) = 1$ rejects ε-far functions with probability at least $W(\varepsilon)$, where $W(\varepsilon) = 1/2^{2^{2^{\cdots}}}$ with the height of the tower of 2's is polynomial in $1/\varepsilon$. How far is this bound from the truth?

Question 6.2. *Does the canonical test reject functions ε-far from Boolean triangle-freeness with probability at most $\exp(-\Omega(1/\varepsilon))$?*

The best bound known so far for this question is polynomial in ε. With Xie, we [BX10] showed a bound of $\varepsilon^{4.847}$. In a recent work, Fu and Kleinberg [FK13] improved this to $\varepsilon^{6.619}$ by using a construction from Coppersmith and Winograd's famous matrix multiplication algorithm.

– While Theorem 4.5 shows that any locally characterized affine-invariant properties is testable, it is not clear how this helps to show any particular property is testable. Proving a property is locally characterized often seems a big challenge in itself. For instance, consider the splittability property mentioned in Section 3. To show that it's locally characterized, one has to prove that if every hyperplane restriction of a function is splittable, then the function is itself splittable. [BFH+13] proves this statement, not only for splittability, but for all *degree-structural* properties. Informally speaking, a degree-structural property is a property $\mathcal{P}$ which is satisfied by a function only when it can be written as a certain combination of low-degree polynomials. See [BFH+13] for the formal definition. The property of whether a function is of the form $q_1q_2 + q_3q_4$ where q_1, q_2, q_3, q_4 are quadratics is another example of a degree-structural property. The proof uses the Gowers inverse theorem, near orthogonality of high rank non-classical polynomials, and in fact, nearly all the machinery needed to prove Theorem 4.5. We wonder whether there is a more elementary proof of local characterization for degree-structural properties (or even, for specific properties of interest such as splittability).

There still remain some natural affine-properties which we cannot prove are locally characterized and hence cannot prove one-sided testability.

Question 6.3. *Is the property of having spectral norm $\leqslant k$ locally characterized?*

Wimmer and Yoshida [WY13] have shown this property to be two-sided testable.

– The characterization of testable graph properties by Borgs et al. [BCL+06] relied on the study of limit objects of graph sequences, called *graphons*. What is the analogous limit object for functions on $\mathbb{F}^n$? To be more precise, consider a function $f : \mathbb{F}^n \to \{0,1\}$. Restricting the function to a random k-dimensional subspace gives a probability distribution $\mu_k(f)$ on $\{\mathbb{F}^k \to \{0,1\}\}$. A sequence of functions $\{f_i : \mathbb{F}^{n_i} \to \{0,1\}\}_{i\in\mathbb{Z}^+}$ is said to converge if for every k, the sequence $\{\mu_k(f_i)\}_{i\in\mathbb{Z}^+}$ converges. We want to find a limit object for such convergent sequences of functions. Very recently, Hatami, Hatami and Hirst [HHH13] characterized the limit objects as certain measurable functions, using the equidistribution results developed in [BFH+13]. However, certain basic questions remain unresolved, as indicated in [HHH13]:

Question 6.4. *For a set $\mathcal{L} = \{L_1, L_2, \ldots, L_m\}$ of linear forms, each on ℓ variables, and a function $f : \mathbb{F}^n \to \mathbb{C}$, define:*

$$t_{\mathcal{L}}(f) = \mathop{\mathbf{E}}_{x_1,\ldots,x_\ell}\left[\prod_{L\in\mathcal{L}} f(L(x_1,\ldots,x_\ell))\right]$$

Now, suppose that two limit objects Γ_1, Γ_2 satisfy $t_{\mathcal{L}}(\Gamma_1) = t_{\mathcal{L}}(\Gamma_2)$ for every system of linear forms $\mathcal{L}$. Then, what is the relationship between Γ_1 and Γ_2?

The goal is to develop this theory in order to fully solve the characterization question for affine-invariant properties.

Acknowledgements

First of all, thanks to Lane Hemaspaandra for his infinite patience during the writing of this survey, to Ronitt Rubinfeld for recommending me to write it, and to Shachar Lovett for comments on this writeup. I want to thank all my coauthors of papers in this area: Victor Chen, Madhu Sudan, Ning Xie, Jakob Nordstrom, Elena Grigorescu, Asaf Shapira, Eldar Fischer, Shachar Lovett, Hamed Hatami, Pooya Hatami, and Madhur Tulsiani. Also, I am grateful to Noga Alon, Ronitt Rubinfeld, and Alex Samorodnitsky for many helpful suggestions along the way.

References

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

[AFNS06] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. In *Proc. 38th Annual ACM Symposium on the Theory of Computing*, pages 251–260, 2006.

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

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

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

[BCL+06] Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós, Balázs Szegedy, and Katalin Vesztergombi. Graph limits and parameter testing. In *Proc. 38th Annual ACM Symposium on the Theory of Computing*, pages 261–270, 2006.

[BCSX11] Arnab Bhattacharyya, Victor Chen, Madhu Sudan, and Ning Xie. Testing linear-invariant non-linear properties. *Theory Comput.*, 7(1):75–99, 2011.

[BFH+13] Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every locally characterized affine-invariant property is testable. In *Proc. 45th Annual ACM Symposium on the Theory of Computing*, pages 429–436, 2013.

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

[BFL13] Arnab Bhattacharyya, Eldar Fischer, and Shachar Lovett. Testing low complexity affine-invariant properties. In *Proc. 24th ACM-SIAM Symposium on Discrete Algorithms*, pages 1337–1355, 2013.

[BFLS91] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *Proc. 23rd Annual ACM Symposium on the Theory of Computing*, pages 21–32, New York, 1991. ACM Press.

[BGH+12] Boaz Barak, Parikshit Gopalan, Johan Hastad, Raghu Meka, Prasad Raghavendra, and David Steurer. Making the long code shorter. In *Proc. 53rd Annual IEEE Symposium on Foundations of Computer Science*, pages 370–379, 2012.

[BGRS12] Arnab Bhattacharyya, Elena Grigorescu, Prasad Raghavendra, and Asaf Shapira. Testing odd-cycle-freeness in boolean functions. *Combin. Probab. Comput.*, 21:835–855, 11 2012.

[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, 2010.

[BHT13] Arnab Bhattacharyya, Pooya Hatami, and Madhur Tulsiani. Algorithmic regularity for polynomials and applications. Manuscript, October 2013.

[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.

[BSMSS11] Eli Ben-Sasson, Ghid Maatouk, Amir Shpilka, and Madhu Sudan. Symmetric LDPC codes are not necessarily locally testable. In *Annual IEEE Conference on Computational Complexity*, pages 55–65. IEEE Computer Society, 2011.

[BTZ10] Vitaly Bergelson, Terence Tao, and Tamar Ziegler. An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}^\omega$. *Geom. Funct. Anal.*, 19(6):1539–1596, 2010.

[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, 2010.

[FGL+96] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996.

[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.

[FK13] Hu Fu and Robert Kleinberg. Improved lower bounds for testing triangle-freeness in boolean functions via fast matrix multiplication. Technical report, August 2013. http://arxiv.org/abs/1308.1643.

[FN07] Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM J. Comput.*, 37(2):482–501, 2007.

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

[GK11] Oded Goldreich and Tali Kaufman. Proximity oblivious testing and the role of invariances. In *Approximation, randomization, and combinatorial optimization*, volume 6845 of *Lecture Notes in Comput. Sci.*, pages 579–592. Springer, Heidelberg, 2011.

[GL89] Oded Goldreich and Leonid Levin. A hard-core predicate for all one-way functions. In *Proc. 21st Annual ACM Symposium on the Theory of Computing*, pages 25–32, 1989.

[Gop10] Parikshit Gopalan. A Fourier-analytic approach to Reed-Muller decoding. In *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science*, pages 685–694, 2010.

[GOS+09] Parikshit Gopalan, Ryan O'Donnell, Rocco A. Servedio, Amir Shpilka, and Karl Wimmer. Testing Fourier dimensionality and sparsity. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming (ICALP '09)*, pages 500–512, 2009.

[Gow01] William T. Gowers. A new proof of Szeméredi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[GR11] Oded Goldreich and Dana Ron. On proximity-oblivious testing. *SIAM J. Comput.*, 40(2):534–566, 2011.

[Gre05] Ben Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376, 2005.

[Gre07] Ben Green. Montréal notes on quadratic Fourier analysis. Technical report, April 2007. http://arxiv.org/abs/math/0604089.

[GS12] Oded Goldreich and Igor Shinkar. Two-sided proximity oblivious testing. In *APPROX-RANDOM*, pages 565–578, 2012.

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

[GT09] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contrib. Discrete Math.*, 4(2):1–36, 2009.

[GT10a] Ben Green and Terence Tao. *An Irregular Mind: Szemerédi is 70*, volume 21 of *Bolyai Society Mathematical Studies*, chapter An arithmetic regularity lemma, associated counting lemma, and applications, pages 261–334. Springer, 2010.

[GT10b] Ben Green and Terence Tao. Linear equations in primes. *Ann. of Math*, 171:1753–1850, 2010.

[HHH13] Hamed Hatami, Pooya Hatami, and James Hirst. Limits of boolean functions on $\mathbb{F}_p^n$. Technical report, August 2013. http://arxiv.org/abs/1308.4108.

[HL11] Hamed Hatami and Shachar Lovett. Correlation testing for affine invariant properties on $\mathbb{F}_p^n$ in the high error regime. In *Proc. 43rd Annual ACM Symposium on the Theory of Computing*, pages 187–194, 2011.

[HL13] Hamed Hatami and Shachar Lovett. Estimating the distance from testable affine-invariant properties. Technical report, June 2013. http://arxiv.org/abs/1306.0649v1, to appear in FOCS '13.

[KL08] Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. In *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 166–175, 2008.

[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, 2008.

[KSV09] Daniel Král', Oriol Serra, and Lluís Vena. A combinatorial proof of the removal lemma for groups. *J. Combin. Theory Ser. A*, 116(4):971–978, May 2009.

[KSV12] Daniel Král, Oriol Serra, and Lluís Vena. A removal lemma for systems of linear equations over finite fields. *Israel J. Math*, pages 1–15, 2012.

[Lip89] Richard Lipton. New directions in testing. In *Proc. DIMACS workshop on Distributed computing and Cryptography*, 1989.

[LMS08] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 547–556, New York, NY, USA, 2008. ACM.

[NRS06] Brendan Nagle, Vojtěch Rödl, and Mathias Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures Algorithms*, 28(2):113–179, 2006.

[Ron09] Dana Ron. Algorithmic and analysis techniques in property testing. *Foundations and Trends in Theoretical Computer Science*, 5(2):73–205, 2009.

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

[RS04] Vojtěch Rödl and Jozef Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures Algorithms*, 25(1):1–42, 2004.

[RS09] Vojtěch Rödl and Mathias Schacht. Generalizations of the removal lemma. *Combinatorica*, 29(4):467–501, 2009.

[RS11] Ronitt Rubinfeld and Asaf Shapira. Sublinear time algorithms. *SIAM J. Disc. Math.*, pages 1562–1588, 2011.

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

[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.

[STV01] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the XOR Lemma. *J. Comp. Sys. Sci.*, 62(2):236–266, 2001.

[Sud10] Madhu Sudan. Invariance in property testing. Technical Report 10-051, Electronic Colloquium in Computational Complexity, March 2010.

[Sud11] Madhu Sudan. Testing linear properties: Some general themes. *SIGACT News, Complexity Theory Column*, 42(1):59–80, March 2011.

[Sze78] Endre Szemerédi. Regular partitions of graphs. In J.C. Bremond, J.C. Fournier, M. Las Vergnas, and D. Sotteau, editors, *Proc. Colloque Internationaux CNRS 260 – Problèmes Combinatoires et Théorie des Graphes*, pages 399–401, 1978.

[TSZ03] Amnon Ta-Shma, Shmuel Safra, and David Zuckerman. Extractors from Reed-Muller codes. In *Proc. 44th Annual IEEE Symposium on Foundations of Computer Science*, page 126, 2003.

[TW11] Madhur Tulsiani and Julia Wolf. Quadratic Goldreich-Levin theorems. In *Proc. 52nd Annual IEEE Symposium on Foundations of Computer Science*, pages 619–628. IEEE, 2011.

[TZ10] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Analysis & PDE*, 3(1):1–20, 2010.

[TZ11] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *Ann. Comb.*, (to appear), 2011. http://arxiv.org/abs/1101.1469.

[WY13] Karl Wimmer and Yuichi Yoshida. Testing linear-invariant function isomorphism. In *Proc. 40th Annual International Conference on Automata, Languages, and Programming*, pages 840–850, 2013.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de