# Estimating the distance from testable affine-invariant properties

Hamed Hatami

McGill University

hatami@cs.mcgill.ca

Shachar Lovett

UC San Diego

slovett@cs.ucsd.edu

June 3, 2013

## Abstract

Let $\mathcal{P}$ be an affine invariant property of functions $\mathbb{F}_p^n \to [R]$ for fixed $p$ and $R$. We show that if $\mathcal{P}$ is locally testable with a constant number of queries, then one can estimate the distance of a function $f$ from $\mathcal{P}$ with a constant number of queries. This was previously unknown even for simple properties such as cubic polynomials over $\mathbb{F}_2$.

Our test is simple: take a restriction of $f$ to a constant dimensional affine subspace, and measure its distance from $\mathcal{P}$. We show that by choosing the dimension large enough, this approximates with high probability the global distance of $f$ from $\mathcal{P}$. The analysis combines the approach of Fischer and Newman [SIAM J. Comp 2007] who established a similar result for graph properties, with recently developed tools in higher order Fourier analysis, in particular those developed in Bhattacharyya et al. [STOC 2013].

## 1   Introduction

Blum, Luby, and Rubinfeld [BLR93] observed that given a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$, it is possible to inquire the value of $f$ on a few random points, and accordingly probabilistically distinguish between the case that $f$ is a linear function and the case that $f$ has to be modified on at least $\varepsilon > 0$ fraction of points to become a linear function. Inspired by this observation, Rubinfeld and Sudan [RS93] defined the concept of property testing which is now a major area of research in theoretical computer science. Roughly speaking, to test a function for a property means to examine the value of the function on a few random points, and accordingly (probabilistically) distinguish between the case that the function has the property and the case that it is not too close to any function with that property.

The focus of our work is on testing properties of multivariate functions over finite fields. Fix a prime $p \geqslant 2$ and an integer $R \geqslant 2$ throughout. Let $\mathbb{F} = \mathbb{F}_p$ be a prime field and $[R] = \{0, \ldots, R-1\}$. We consider properties of functions $f : \mathbb{F}^n \to [R]$. We are interested in testing the *distance* of a function $f : \mathbb{F}^n \to [R]$ to a property. Here the distance corresponds to the minimum fraction of the points on which the function can be modified in order to satisfy the property. Fischer and Newman [FN07] showed that it is possible to estimate the distance from a graph to any given *testable* graph property. In this article we extend this result to the algebraic setting of affine-invariant properties on functions $f : \mathbb{F}^n \to [R]$. Furthermore we show that the Fischer-Newman test can be replaced by a more natural one: pick a sufficiently large subgraph $H$ randomly and estimate the distance of $H$ to the property. Analogously, in our setting, we pick a sufficiently large

affine subspace of $\mathbb{F}^n$ randomly, and measure the distance of the restriction of the function to this subspace from the property.

## 1.1  Testability

Given a property $\mathcal{P}$ of functions in $\{\mathbb{F}^n \to [R] \mid n \in \mathbb{N}\}$, we say that $f : \mathbb{F}^n \to [R]$ is $\varepsilon$-far from $\mathcal{P}$ if

$$\min_{g \in \mathcal{P}} \mathbf{Pr}_{x \in \mathbb{F}^n} [f(x) \neq g(x)] > \varepsilon,$$

and we say that it is $\varepsilon$-close otherwise.

**Definition 1.1** (Testability). *A property $\mathcal{P}$ is said to be* testable *(with two-sided error) if there is a function $q : (0, 1) \to \mathbb{N}$ 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 $2/3$ if $f \in \mathcal{P}$ and rejects with probability at least $2/3$ if $f$ is $\varepsilon$-far from $\mathcal{P}$.*

Note that if we do not require any restrictions on $\mathcal{P}$, then the algebraic structure of $\mathbb{F}^n$ becomes irrelevant, and $\mathbb{F}^n$ would be treated as a generic set of size $|\mathbb{F}|^n$. To take the algebraic structure into account, we have to require certain "invariance" conditions.

We say that a property $\mathcal{P} \subseteq \{\mathbb{F}^n \to [R] \mid n \in \mathbb{N}\}$ is *affine-invariant* if for any $f \in \mathcal{P}$ and any affine transformation $A : \mathbb{F}^n \to \mathbb{F}^n$, we have $Af := f \circ A \in \mathcal{P}$ (an affine transformation $A$ is of the form $L + c$ where $L$ is linear and $c$ is a constant vector in $\mathbb{F}^n$). Some well-studied examples of affine-invariant properties include Reed-Muller codes (in other words, bounded degree polynomials) [BFL91, BFLS91, FGL+96, RS96, AKK+05] and Fourier sparsity [GOS+09]. 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]). In particular, they asked for necessary and sufficient conditions for the testability of affine-invariant properties. This question initiated an active line of research, which have led to a near complete characterization of testable affine invariant properties, at least in the regime of one-sided error [BCSX11, KSV12, Sha09, BGS10, BFL12, BFH+13].

It is not difficult to see that for affine-invariant properties testability has an equivalent "non-algorithmic" definition through the distribution of restrictions to affine subspaces. We will describe a restriction of $\mathbb{F}^n$ to an affine subspace of dimension $k$ by an affine embedding $A : \mathbb{F}^k \to \mathbb{F}^n$ (an affine embedding is an injective affine transformation). The restriction of $f : \mathbb{F}^n \to [R]$ to the subspace is then given by $Af : \mathbb{F}^k \to [R]$.

**Proposition 1.2.** *An affine-invariant property $\mathcal{P}$ is testable if and only if for every $\varepsilon > 0$, there exist a constant $k$ and a set $\mathcal{H} \subseteq \{\mathbb{F}^k \to [R]\}$, such that for a function $f : \mathbb{F}^n \to [R]$ and a random affine embedding $A : \mathbb{F}^k \to \mathbb{F}^n$ the following holds. If $f \in \mathcal{P}$, then*

$$\mathbf{Pr}[Af \in \mathcal{H}] > 2/3,$$

*and if $f$ is $\varepsilon$-far from $\mathcal{P}$, then*

$$\mathbf{Pr}[Af \notin \mathcal{H}] > 2/3.$$

## 1.2 Our contribution

For a property $\mathcal{P}$ and a positive real $\delta$, let $\mathcal{P}_\delta$ denote the set of all functions that are $\delta$-close to the property. Our main result is the following theorem.

**Theorem 1.3.** *For every testable affine-invariant property $\mathcal{P}$ and every $\delta > 0$, the property $\mathcal{P}_\delta$ is testable.*

Theorem 1.3 says that for every $\varepsilon, \delta > 0$ one can probabilistically distinguish between functions that are $\delta$-close to the property and the functions that are $(\delta + \varepsilon)$-far from the property using only a constant number of queries. In fact the test is very natural. We show that there exists a constant $k_{\varepsilon, \delta, \mathcal{P}}$ such that for a random affine embedding $A : \mathbb{F}^k \to \mathbb{F}^n$, with probability at least $2/3$, $\operatorname{dist}(Af, \mathcal{P})$ provides a sufficiently accurate estimate of $\operatorname{dist}(f, \mathcal{P})$. Hence our test will be the following: Pick a random affine embedding $A : \mathbb{F}^k \to \mathbb{F}^n$. If $\operatorname{dist}(Af, \mathcal{P}) < \delta + \frac{\varepsilon}{2}$ accept, otherwise reject. This corresponds to taking $\mathcal{H} = \left\{ h : \mathbb{F}^k \to [R] \mid \operatorname{dist}(h, \mathcal{P}) \leqslant \delta + \frac{\varepsilon}{2} \right\}$ in Proposition 1.2.

We note that previously it was unknown if one can test distance to even simple properties, such as cubic polynomials over $\mathbb{F}_2$. The reason was that one specific natural test (the Gowers norm, or derivatives test) was shown not to perform well for such properties. Our work shows that a natural test indeed works, albeit the number of queries have to grow as a function of $\varepsilon$. We do not know if this is necessary for simple properties, such as cubic polynomials over $\mathbb{F}_2$, and leave this as an open problem.

On a technical level, our work combines two technologies developed in previous works. The first is the work of Fischer and Newman [FN07] which obtained similar results for graph properties. The second is higher order Fourier analysis, in particular a recent strong equidistribution theorem established in Bhattacharyya et al. [BFH+13]. From a high level, the approach for the graph case and the affine-invariant case are similar. One applies a regularization process, which allows to represent a graph (or a function) by a small structure. Then, one argues that a large enough random sample of the graph or function should have a similar small structure representing it. Hence, properties of the main object can be approximated by properties of a large enough sample of it. Fischer and Newman [FN07] implemented this idea in the graph case. We follow a similar approach in the algebraic case, which inevitably introduces some new challenges. One may see this result as an outcome of the large body of work on higher-order Fourier analysis developed in recent years. Once the machinery was developed, we can now apply it in various frameworks which were not accessible previously.

## 1.3 Proof overview

Let $R = 2$ for the simplicity of exposition, e.g. we consider functions $f : \mathbb{F}^n \to \{0, 1\}$. Let $\mathcal{P}$ be an affine invariant property of functions $\{\mathbb{F}^n \to \{0, 1\} : n \in \mathbb{N}\}$ which is locally testable, and fix $\varepsilon, \delta > 0$. We want to show that there exists an $m$ (which depends only on $\mathcal{P}, \varepsilon, \delta$) such that the following holds. Let $f : \mathbb{F}^n \to \{0, 1\}$ be a function, and let $\tilde{f}$ be the restriction of the function to a random $m$-dimensional affine subspace of $\mathbb{F}^n$. Then

- Completeness: If $f$ is $\delta$-close to $\mathcal{P}$ then, with high probability, $\tilde{f}$ is $(\delta + \varepsilon/2)$-close to $\mathcal{P}$.

- Soundness: If $f$ is $(\delta + \varepsilon)$-far from $\mathcal{P}$ then, with high probability, $\tilde{f}$ is $(\delta + \varepsilon/2)$-far from $\mathcal{P}$.

Once we show that we are done, as the local test computes the distance of $\tilde{f}$ from $\mathcal{P}$. If it is below $\delta + \varepsilon/2$ we declare that $f$ is $\delta$-close to $\mathcal{P}$; otherwise we declare it is $(\delta + \varepsilon)$-far from $\mathcal{P}$. The test

3

correctness follows immediately from the completeness and soundness. We next argue why these hold.

Let us first fix notations. Let $A : \mathbb{F}^m \to \mathbb{F}^n$ be a random full rank affine transformation. Then, a restriction of $f$ to a random $m$-dimensional affine subspace can be equivalently described by $\tilde{f} = Af$. The proof of the completeness is simple. If $f$ is $\delta$-close to a function $g : \mathbb{F}^n \to \{0,1\}$ which is in $\mathcal{P}$, then with high probability over a random restriction, the distance of $Af$ and $Ag$ is also at most $\delta + o_m(1)$. This is true because a random affine subspace is pairwise independent with regards to whether an element is contained in it. This, combined with Chebyshev's inequality implies the result. Then, by choosing $m$ large enough we get the error term down to $\varepsilon/2$.

The main work (as in nearly all works in property testing) is to establish soundness. That is, we wish to show that if a function $f$ is far from $\mathcal{P}$ then, with high probability, a random restriction of it is also from from the property. The main idea is to show that if for a typical restriction $Af$ is $\delta$-close to a function $h : \mathbb{F}^m \to \{0,1\}$ which is in $\mathcal{P}$, then $h$ can be "pulled back" to a function $g : \mathbb{F}^n \to \{0,1\}$ which is both roughly $\delta$-close to $f$ and also very close to $\mathcal{P}$. This will contradict our initial assumption that $f$ is $(\delta + \varepsilon)$-far from $\mathcal{P}$. In order to do so we apply the machinery of higher order Fourier analysis. The first description will hide various "cheats" but will present the correct general outline. We then note which steps need to be fixed to make this argument actually work.

First, we apply the assumption that $\mathcal{P}$ is locally testable to derive there exist a constant dimension $k = k(\mathcal{P}, \varepsilon)$ so that a random restriction to a $k$-dimensional subspace can distinguish functions in $\mathcal{P}$ from functions which are $\varepsilon/4$-far from $\mathcal{P}$. We want to decompose $f$ to "structured" parts which we will study, and "pseudo-random" parts which do not affect the distribution of restrictions to $k$-dimensional subspaces. In order to do so, for a function $f : \mathbb{F}^n \to \{0,1\}$ define by $\mu_{f,k}$ the distribution of its restriction to $k$-dimensional subspaces. That is, for $v : \mathbb{F}^k \to \{0,1\}$ let

$$\mu_{f,k}[v] = \mathbf{Pr}_A[Af = v].$$

We need to slightly generalize this definition to functions where the output $f(x)$ can be random. In our context, a randomized function is a function $f : \mathbb{F}^n \to [0,1]$, which describes a distribution over functions $f' : \mathbb{F}^n \to \{0,1\}$, where for all $x$ independently $\mathbf{Pr}[f'(x) = 1] = f(x)$. We extend the definition of $\mu_{f,k}$ to randomized functions by $\mu_{f,k}[v] = \mathbf{E}_{A,f'} \mu_{f',k}[v]$. By our definition, if two functions $f, g : \mathbb{F}^n \to [0,1]$ have distributions $\mu_{f,k}$ and $\mu_{g,k}$ close in statistical distance, then random restrictions to $k$-dimensional affine subspaces cannot distinguish $f$ from $g$. This will be useful in the analysis of the soundness.

We next decompose our function $f$ based on the above intuition. The formal notion of pseudo-randomness we use is that of Gowers uniformity. Informally, the $d$-th Gowers uniformity measures correlation with polynomials of degree less than $d$. However, it turns to capture much more than that. For example, one can show that by choosing $d$ large enough ($d = p^k$ suffices) then for any functions $f, g : \mathbb{F}^n \to [0,1]$, if $\|f - g\|_{U^d}$ is small enough then $\mu_{f,k}$ and $\mu_{g,k}$ are close in statistical distance. Thus, it makes sense to approximate $f$ as

$$f = f_1 + f_2$$

where $f_1$ is structured (to be explained soon) and $\|f_2\|_{U^d}$ is small enough. This will allow us to replace $f$ with $f_1$ for the purposes of analyzing its restrictions to $k$-dimensional subspaces. The structure of $f_1$ is as follows: it is a function of a constant number $C = C(\mathcal{P}, \varepsilon)$ of polynomials of

4

degree less than $d$. That is,
$$f_1(x) = \Gamma(P_1(x), \ldots, P_C(x)),$$
where $P_1, \ldots, P_C$ are polynomials and $\Gamma : \mathbb{F}^C \to \{0, 1\}$ is some function (not necessarily a low degree polynomial). The benefit of this decomposition is that $f_1$ is "dimension-less" in the sense that $\Gamma$ does not depend on $n$; however, the polynomials $P_1, \ldots, P_C$ do depend on $n$. One can however "regularize" these polynomials in order to obtain "random-looking" (or high rank) polynomials. It can be shown that all properties of high rank polynomials are governed just by their degree (which is at most $d$), hence essentially the entire description of $f_1$ does not depend on $n$.

The next step is to show that the same type of decomposition can be applied to the restriction $Af$ of $f$. Clearly, $Af = Af_1 + Af_2$. We show that with high probability over the choice of $A$,

- $Af_1 = \Gamma(Q_1(x), \ldots, Q_C(x))$ where $Q_i = AP_i$ are the restrictions of $P_1, \ldots, P_C$; and $Q_1, \ldots, Q_C$ are still of "high enough rank" to behave like random polynomials.

- $\|Af_2\|_{U^d} \approx \|f_2\|_{U^d}$ so we can still approximate $Af \approx Af_1$ with respect to the distribution of their restrictions to random $k$-dimensional subspaces.

We next apply the same decomposition process to $h$, which we recall is the assumed function (in $m$ variables) which is $(\delta + \varepsilon/2)$-close to $Af$. By choosing the conditions of regularity of $h$ slightly weaker than those of $f$ (but still strong enough), we get that we can decompose
$$h = h_1 + h_2$$
where
$$h_1(x) = \Gamma'(Q_1(x), \ldots, Q_{C'}(x))$$
for some $C' > C$ and $\|h_2\|_{U^d}$ is very small. The important aspect here is that, we can approximate $h$ by the structured function $h_1$, and moreover that the polynomials $Q_1, \ldots, Q_C$ which compose $Af_1$ are part of the description of $h_1$. That is, both $Af_1$ and $h_1$ can be defined in terms of the same basic building blocks (high rank polynomials $Q_1, \ldots, Q_C$).

The next step is to "pull back" $h$ to a function defined on $\mathbb{F}^n$. An easy first step is to pull back $h_1$. We need to define for $C < i \leqslant C'$ pullback polynomials $P_i : \mathbb{F}^n \to \{0, 1\}$ of $Q_i : \mathbb{F}^n \to \{0, 1\}$ such that both $Q_i = AP_i$; and such that $P_1, \ldots, P_{C'}$ are of high rank. This can be done for example by letting $P_i = DQ_i$ for any affine map $D : \mathbb{F}^n \to \mathbb{F}^m$ for which $AD$ is the identity map on $\mathbb{F}^m$. This provides a pull-back $\phi$ of the "coarse" description of $f_1$ of $h_1$, but does not in general generate a function close to $f$ (it makes sense, since we still haven't used the finer "pseudo-random" structure of $f$). Formally, we set $\phi(x) = \Gamma'(P_1(x), \ldots, P_{C'}(x))$. However, we can already show something about $\phi$: it is very close to $\mathcal{P}$. More concretely, its distribution over restrictions to $d$-dimensional subspaces is very close to that of $h$. Hence, the tester which distinguishes function in $\mathcal{P}$ from those $(\varepsilon/4)$-far from $\mathcal{P}$ cannot distinguish $\phi$ from functions in $\mathcal{P}$, hence $\phi$ must be $(\varepsilon/4)$-close to $\mathcal{P}$.

The next step is to define a more refined pull-back of $f$. Define an atom as a subset $\{x \in \mathbb{F}^n : P_1(x) = a_1, \ldots, P_{C'}(x) = a_{C'}\}$ for values $a_1, \ldots, a_{C'} \in \mathbb{F}$. Note that the functions $f_1, h_1$ are constant over atoms. We next define $\psi : \mathbb{F}^n \to [0, 1]$ by redefining $\phi$ inside each atom, so that the average over the atoms of $\phi, \psi$ is the same, but such that $\psi$ is as close as possible to $f$ given this constraint. For example, if in an atom the average of $f$ is higher than the value $\phi$ assigns to this atom (and so it needs to be reduced to match $\phi$), we set for all $x$ in this atom $\psi(x) = 0$ if $\phi(x) = 0$ and $\psi(x) = \alpha$ if $f(x) = 1$, where $\alpha$ is appropriately chosen so that the averages match. We then show that $\psi$ is a proper pull-back of $h$ in the sense that

5

- The distance between $f, \psi$ is very close to the distance between $Af, h$, which we recall is at most $\delta + \varepsilon/2$.

- $\psi$ is nearly $\varepsilon + 4$ close to $\mathcal{P}$ in the distributional sense.

To finalize, we show that sampling a function $g : \mathbb{F}^n \to \{0, 1\}$ based on $\psi$ has the same properties, which shows that $f$ is not $(\delta + \varepsilon)$-far from $\mathcal{P}$.

Let us remark on a few technical points overlooked in the above description. First, there are the exact notions of "high rank polynomials". It turns that in order to make this entire argument work, one needs to consider more general objects, called non-classical polynomials. We rely on a series of results on the distributional properties of high-rank non-classical polynomials, in particular these recently established in [BFH$^+$13]. Also, the decomposition theorems are actually to three parts,

$$f = f_1 + f_2 + f_3,$$

where $f_1$ is structured as before, $\|f_2\|_2$ is somewhat small (but not very small) and $\|f_3\|_{U^d}$ is very small. This requires a somewhat more refined analysis to make the argument work, but does not create any significant change in the proof outline as described above.

## 1.4 Comparison with graph property testing

The main outline of our proof follows closely that of Fischer and Newman [FN07]. They study graph properties, where decompositions are given by the Szemerédi regularity lemma. Their test, in the notation above, can be described as measuring the distance between $\Gamma$ and all potential $\Gamma'$ which can be achieved from graphs that have the property. Our argument (when applied to graph properties instead of affine invariant properties) shows that a much more natural test achieves the same behaviour: choose a random small subgraph and measure its distance from the property. In quantitative terms it is hard to compare the two results, as both get outrageous bounds coming from the bounds in the regularity lemma. So, we view this part of our work as having contribution in the simplicity of the test, and not in terms of the simplicity of the proof or the quantitative bounds (which are both very similar).

The more challenging aspect of our work is to take this approach and carry it out in the affine invariant settings. The main reason is that in the affine invariant setup the structural parts have more structure in them than in the graph setting. In the graph setup, the structure of a graph can be represented by a constant size graph with weighted edges. In the affine invariant case, the structured part is a constant size function applied to polynomials. However there will be *no constant bound* on the number of variables, and they can grow as $n$ grows. So, at first glance, these "compact descriptions" have sizes which grow with the input size; this is very different from the graph case. The reason these compact descriptions are useful is because, as long as the polynomials participating in them are "random enough", then their exact definitions do not matter, just a few simple properties of them (their degree, and "depth" for non-classical polynomials). This is fueled by the recent advances on higher-order Fourier analysis. In essence, the state of the art has reached a stage where these tools are powerful enough to simulate the counterpart arguments which were initially developed in the context of graph properties. Still, the affine invariant case is more complex, and there are some problems which we do not know yet how to handle. For example,

- A complete classification of one-sided testable properties (e.g. can properties of "infinite complexity" be locally testable?); See [BFH$^+$13]

- A complete classification of two-sided testable properties.

- Properties where any non-trivial distance from them can be witnessed by a constant number of queries (also called correlation testing [HL11]). For example, can one test correlation to cubics over $\mathbb{F}_2$ using a constant number of queries?

## 2 Background

We need to recall some definitions and results about higher order Fourier analysis. Most of the material in this section is directly quoted from the full version of [BFH$^+$13].

**Notation** We shorthand $\mathbb{F} = \mathbb{F}_p$ for a prime finite field. For $f : \mathbb{F}^n \to \mathbb{C}$ we denote $\|f\|_1 = \mathbf{E}[|f(x)|]$, $\|f\|_2^2 = \mathbf{E}[|f(x)|^2]$ where $x \in \mathbb{F}^n$ is chosen uniformly and $\|f\|_\infty = \max|f(x)|$. Note that $\|f\|_1 \leqslant \|f\|_2 \leqslant \|f\|_\infty$. The expression $o_m(1)$ denotes quantities which approach zero as $m$ grows. We shorthand $x \pm \varepsilon$ for any quantity in $[x - \varepsilon, x + \varepsilon]$.

### 2.1 Uniformity norms and non-classical polynomials

**Definition 2.1** (Multiplicative Derivative)**.** *Given a function $f : \mathbb{F}^n \to \mathbb{C}$ and an element $h \in \mathbb{F}^n$, define the* multiplicative derivative *in direction $h$ of $f$ to be the function $\Delta_h f : \mathbb{F}^n \to \mathbb{C}$ satisfying $\Delta_h f(x) = f(x+h)\overline{f(x)}$ for all $x \in \mathbb{F}^n$.*

The *Gowers norm* of order $d$ for a function $f : \mathbb{F}^n \to \mathbb{C}$ is the expected multiplicative derivative of $f$ in $d$ random directions at a random point.

**Definition 2.2** (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 given by*

$$\|f\|_{U^d} = \left| \mathop{\mathbf{E}}_{y_1,\ldots,y_d,x \in \mathbb{F}^n} [(\Delta_{y_1}\Delta_{y_2}\cdots\Delta_{y_d}f)(x)] \right|^{1/2^d}.$$

Note that as $\|f\|_{U^1} = |\mathbf{E}[f]|$ the Gowers norm of order 1 is only a semi-norm. However for $d > 1$, it is not difficult to show that $\|\cdot\|_{U^d}$ is indeed a norm.

If $f = e^{2\pi i P/p}$ where $P : \mathbb{F}^n \to \mathbb{F}$ is a polynomial of degree $< d$, then $\|f\|_{U^d} = 1$. If $d < p$ and $\|f\|_\infty \leqslant 1$, then in fact, the converse holds, meaning that any function $f : \mathbb{F}^n \to \mathbb{C}$ satisfying $\|f\|_\infty \leqslant 1$ and $\|f\|_{U^d} = 1$ is of this form. But when $d \geqslant p$, the converse is no longer true. In order to characterize functions $f : \mathbb{F}^n \to \mathbb{C}$ with $\|f\|_\infty \leqslant 1$ and $\|f\|_{U^d} = 1$, we define the notion of *non-classical polynomials*.

Non-classical polynomials might not be necessarily $\mathbb{F}$-valued. We need to introduce some notation. Let $\mathbb{T}$ denote the circle group $\mathbb{R}/\mathbb{Z}$. This is an abelian group with group operation denoted $+$. For an integer $k \geqslant 0$, let $\mathbb{U}_k$ denote $\frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$, a subgroup of $\mathbb{T}$. Let $\iota : \mathbb{F} \to \mathbb{U}_1$ be the injection $x \mapsto \frac{|x|}{p} \mod 1$, where $|x|$ is the standard map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\}$. Let $\mathsf{e} : \mathbb{T} \to \mathbb{C}$ denote the character $\mathsf{e}(x) = e^{2\pi i x}$.

**Definition 2.3** (Additive Derivative). *Given a function[1] $P : \mathbb{F}^n \to \mathbb{T}$ and an element $h \in \mathbb{F}^n$, define the* additive derivative *in direction $h$ of $f$ to be the function $D_h P : \mathbb{F}^n \to \mathbb{T}$ satisfying $D_h P(x) = P(x + h) - P(x)$ for all $x \in \mathbb{F}^n$.*

**Definition 2.4** (Non-classical polynomials). *For an integer $d \geqslant 0$, a function $P : \mathbb{F}^n \to \mathbb{T}$ is said to be a* non-classical polynomial of degree $\leqslant d$ *(or simply a* polynomial of degree $\leqslant d$*) if for all $y_1, \ldots, y_{d+1}, x \in \mathbb{F}^n$, it holds that*

$$(D_{y_1} \cdots D_{y_{d+1}} P)(x) = 0. \tag{1}$$

*The* degree *of $P$ is the smallest $d$ for which the above holds. A function $P : \mathbb{F}^n \to \mathbb{T}$ is said to be a* classical polynomial of degree $\leqslant d$ *if it is a non-classical polynomial of degree $\leqslant d$ whose image is contained in $\iota(\mathbb{F})$.*

It is a direct consequence 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 $P : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$.

**Lemma 2.5** (Lemma 1.7 in [TZ11]). *A function $P : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leqslant d$ if and only if $P$ can be represented as*

$$P(x_1, \ldots, x_n) = \alpha + \sum_{\substack{0 \leqslant d_1, \ldots, d_n < p; k \geqslant 0: \\ 0 < \sum_i d_i \leqslant d - k(p-1)}} \frac{c_{d_1, \ldots, d_n, k} |x_1|^{d_1} \cdots |x_n|^{d_n}}{p^{k+1}} \mod 1,$$

*for a unique choice of $c_{d_1, \ldots, d_n, k} \in \{0, 1, \ldots, p - 1\}$ and $\alpha \in \mathbb{T}$. The element $\alpha$ is called the* shift *of $P$, and the largest integer $k$ such that there exist $d_1, \ldots, d_n$ for which $c_{d_1, \ldots, d_n, k} \neq 0$ is called the* depth *of $P$. Classical polynomials correspond to polynomials with 0 shift and 0 depth.*

Also, for convenience of exposition, we will assume throughout this paper that the shifts of all polynomials are zero. This can be done without affecting any of the results in this work. Hence, all polynomials of depth $k$ take values in $\mathbb{U}_{k+1}$.

## 2.2 Uniformity over linear forms

A linear form in $k$ variables is $L = (\ell_1, \ldots, \ell_k) \in \mathbb{F}^k$. We interpret it as a linear operator $L : (\mathbb{F}^n)^k \to \mathbb{F}^n$ given by $L(x_1, \ldots, x_k) = \sum_{i=1}^k \ell_i x_i$.

**Definition 2.6** (Cauchy-Schwarz complexity, [GT10]). *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a set of linear forms. The* Cauchy-Schwarz complexity *of $\mathcal{L}$ is the minimal $s$ such that the following holds. For every $i \in [m]$, we can partition $\{L_j\}_{j \in [m] \setminus \{i\}}$ into $s + 1$ subsets such that $L_i$ does not belong to the linear span of each subset.*

Following is a lemma due to Green and Tao [GT10] based on repeated applications of the Cauchy-Schwarz inequality.

---

[1]We try to adhere to the following convention: upper-case letters (e.g. $F$ and $P$) to denote functions mapping from $\mathbb{F}^n$ to $\mathbb{T}$ or to $\mathbb{F}$, lower-case letters (e.g. $f$ and $g$) to denote functions mapping from $\mathbb{F}^n$ to $\mathbb{C}$, and upper-case Greek letters (e.g. $\Gamma$ and $\Sigma$) to denote functions mapping $\mathbb{T}^C$ to $\mathbb{T}$. By abuse of notation, we sometimes conflate $\mathbb{F}$ and $\iota(\mathbb{F})$.

**Lemma 2.7.** *Let $f_1, \ldots, f_m : \mathbb{F}^n \to [-1, 1]$. Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of $m$ linear forms in $k$ variables of Cauchy-Schwarz complexity $s$. Then:*

$$\left| \mathop{\mathbf{E}}_{x_1, \ldots, x_k \in \mathbb{F}^n} \left[ \prod_{i=1}^m f_i(L_i(x_1, \ldots, x_\ell)) \right] \right| \leqslant \min_{i \in [m]} \|f_i\|_{U^{s+1}}$$

We would need to apply Lemma 2.7 in this paper in the special case when $\mathcal{L}$ corresponds to all $p^k$ linear forms describing all points in an affine subspace of dimension $k$. We would care only about some upper bound on the Cauchy-Schwarz complexity of the system. The following claim follows immediately from the definitions.

**Claim 2.8.** *Let $\mathcal{L} = \{(1, a_1, \ldots, a_k) : a_1, \ldots, a_k \in \mathbb{F}\}$. Then the Cauchy-Schwarz complexity of $\mathcal{L}$ is at most $p^k$.*

## 2.3 Polynomial factors and rank

**Definition 2.9** (Factors). *If $X$ is a finite set then by a* factor *$\mathcal{B}$ we mean simply a partition of $X$ into finitely many pieces called* atoms.

A function $f : X \to \mathbb{C}$ is called *$\mathcal{B}$-measurable* if it is constant on atoms of $\mathcal{B}$. For any function $f : X \to \mathbb{C}$, we may define the conditional expectation

$$\mathbf{E}[f|\mathcal{B}](x) = \mathop{\mathbf{E}}_{y \in \mathcal{B}(x)} [f(y)],$$

where $\mathcal{B}(x)$ is the unique atom in $\mathcal{B}$ that contains $x$. Note that $\mathbf{E}[f|\mathcal{B}]$ is $\mathcal{B}$-measurable.

A finite collection of functions $\phi_1, \ldots, \phi_C$ from $X$ to some other finite space $Y$ naturally define a factor $\mathcal{B} = \mathcal{B}_{\phi_1, \ldots, \phi_C}$ whose atoms are sets of the form $\{x : (\phi_1(x), \ldots, \phi_C(x)) = (y_1, \ldots, y_C)\}$ for some $(y_1, \ldots, y_C) \in Y^C$. By an abuse of notation we also use $\mathcal{B}$ to denote the map $x \mapsto (\phi_1(x), \ldots, \phi_C(x))$, thus also identifying the atom containing $x$ with $(\phi_1(x), \ldots, \phi_C(x))$.

**Definition 2.10** (Polynomial factors). *If $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ is a sequence of polynomials, then the factor $\mathcal{B}_{P_1, \ldots, P_C}$ is called a* polynomial factor.

The *complexity* of $\mathcal{B}$, denoted $|\mathcal{B}|$, is the number of defining polynomials $C$. The *degree* of $\mathcal{B}$ is the maximum degree among its defining polynomials $P_1, \ldots, P_C$. If $P_1, \ldots, P_C$ are of depths $k_1, \ldots, k_C$, respectively, then $\|\mathcal{B}\| = \prod_{i=1}^C p^{k_i+1}$ is called the *order* of $\mathcal{B}$.

Notice that the number of atoms of $\mathcal{B}$ is bounded by $\|\mathcal{B}\|$.

Next we need to define the notion of the *rank* of a polynomial or a polynomial factor.

**Definition 2.11** (Rank of a polynomial). *Given a polynomial $P : \mathbb{F}^n \to \mathbb{T}$ and an integer $d > 1$, the $d$-rank of $P$, denoted $\mathsf{rank}_d(P)$, is defined to be the smallest integer $r$ such that there exist polynomials $Q_1, \ldots, Q_r : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d - 1$ and a function $\Gamma : \mathbb{T}^r \to \mathbb{T}$ satisfying $P(x) = \Gamma(Q_1(x), \ldots, Q_r(x))$. If $d = 1$, then $1$-rank is defined to be $\infty$ if $P$ is non-constant and $0$ otherwise.*

*The* rank *of a polynomial $P : \mathbb{F}^n \to \mathbb{T}$ is its $\deg(P)$-rank.*

A high-rank polynomial of degree $d$ is, intuitively, a "generic" degree-$d$ polynomial. There are no unexpected ways to decompose it into lower degree polynomials. The following theorem shows that a high rank polynomial is distributed close to uniform.

9

**Theorem 2.12** (Theorem 4 of [KL08])**.** *For any $\varepsilon > 0$ and integer $d > 0$, there exists $r = r_{2.12}(d, \varepsilon)$ such that the following is true. If $P : \mathbb{F}^n \to \mathbb{T}$ is a degree-$d$ polynomial with rank greater than $r$, then $|\mathbf{E}_x[\mathsf{e}\,(P(x))]| < \varepsilon$.*

Next, we will formalize the notion of a generic collection of polynomials. Intuitively, it should mean that there are no unexpected algebraic dependencies among the polynomials.

**Definition 2.13** (Rank and Regularity)**.** *A polynomial factor $\mathcal{B}$ defined by a sequence of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective depths $k_1, \ldots, k_C$ is said to have $\mathrm{rank}$ $r$ 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 $\mathsf{rank}_d(Q) \leqslant r$ where $d = \max_i \deg(\lambda_i P_i)$.*

*Given a polynomial factor $\mathcal{B}$ and a function $r : \mathbb{N} \to \mathbb{N}$, we say $\mathcal{B}$ is $r$-regular if $\mathcal{B}$ is of rank larger than $r(|\mathcal{B}|)$.*

Note that since $\lambda$ 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 $\mathcal{B}$ is the factor defined by a single polynomial $P$ of degree $d$ and depth $k$, then

$$\mathsf{rank}(\mathcal{B}) = \min\left\{ \mathsf{rank}_d(P), \mathsf{rank}_{d-(p-1)}(pP), \cdots, \mathsf{rank}_{d-k(p-1)}(p^k P) \right\}.$$

Regular factors indeed do behave like generic collections of polynomials, and thus, given any factor $\mathcal{B}$ that is not regular, it will often be useful to *regularize* $\mathcal{B}$, that is, find a refinement $\mathcal{B}'$ of $\mathcal{B}$ that is regular up to our desires. We distinguish between two kinds of refinements.

**Definition 2.14** (Semantic and syntactic refinements)**.** *A polynoial factor $\mathcal{B}'$ is called a* syntactic refinement *of $\mathcal{B}$, and denoted $\mathcal{B}' \succeq_{syn} \mathcal{B}$, if the sequence of polynomials defining $\mathcal{B}'$ extends that of $\mathcal{B}$. It is called a* semantic refinement*, and denoted $\mathcal{B}' \succeq_{sem} \mathcal{B}$ if the induced partition is a combinatorial refinement of the partition induced by $\mathcal{B}$. In other words, if for every $x, y \in \mathbb{F}^n$, $\mathcal{B}'(x) = \mathcal{B}'(y)$ implies $\mathcal{B}(x) = \mathcal{B}(y)$.*

The following lemma shows that every polynomial factor can be refined to be arbitrarily regular without increasing its complexity by more than a constant.

**Lemma 2.15** (Polynomial Regularity Lemma, Lemma 2.19 of [BFH$^+$13])**.** *Let $r : \mathbb{N} \to \mathbb{N}$ be a non-decreasing function and $d > 0$ be an integer. Then, there is a function $C_{2.15}^{(r,d)} : \mathbb{N} \to \mathbb{N}$ such that the following is true. Suppose $\mathcal{B}$ is a factor defined by polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ of degree at most $d$. Then, there is an $r$-regular factor $\mathcal{B}'$ consisting of polynomials $Q_1, \ldots, Q_{C'} : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$ such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$ and $C' \leqslant C_{2.15}^{(r,d)}(C)$.*

The first step towards showing that regular factors behave like generic collections of polynomials is to show that they are almost equipartitions.

**Lemma 2.16** (Size of atoms, Lemma 3.2 of [BFH$^+$13])**.** *Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of degree $d > 0$, complexity $C$, and rank $r_{2.12}(d, \varepsilon)$, defined by a tuple of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ having respective depths $k_1, \ldots, k_C$. Suppose $b = (b_1, \ldots, b_C) \in \mathbb{U}_{k_1+1} \times \cdots \times \mathbb{U}_{k_C+1}$. Then*

$$\mathbf{Pr}_x[\mathcal{B}(x) = b] = \frac{1}{\|\mathcal{B}\|} \pm \varepsilon.$$

*In particular, for $\varepsilon < \frac{1}{\|\mathcal{B}\|}$, $\mathcal{B}(x)$ attains every possible value in its range and thus has $\|\mathcal{B}\|$ atoms.*

Finally we state the regularity lemma, the basis of the higher order Fourier analysis.

**Theorem 2.17** (Theorem 4.4 of [BFL12]). *Suppose $\zeta > 0$ is a real and $d, k \geqslant 1$ are integers. Let $\eta : \mathbb{N} \to \mathbb{R}^+$ be an arbitrary non-increasing function, and let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary non-decreasing function. Let $\mathcal{B}_0$ be a polynomial factor of degree $d$ and complexity $C_0$. Then, there exist $C = C_{2.17}(\eta, r, \zeta, C_0, d)$ such that the following holds.*

*Every function $f : \mathbb{F}^n \to \{0, 1\}$ has a decomposition $f = f_1 + f_2 + f_3$ such that the following is true:*

- *$f_1 = \mathbf{E}[f|\mathcal{B}_1]$ for a polynomial factor $\mathcal{B}_1 \succeq_{sem} \mathcal{B}_0$ of degree $d$ and complexity $C_1 \leqslant C$.*
- *$\|f_2\|_2 < \zeta$ and $\|f_3\|_{U^{d+1}} < \eta(|\mathcal{B}|)$.*
- *The functions $f_1$ and $f_1 + f_3$ have range $[0, 1]$; $f_2$ and $f_3$ have range $[-1, 1]$.*
- *$\mathcal{B}_1$ is $r$-regular.*

*Furthermore if $\mathsf{rank}(\mathcal{B}_0) \geqslant r_{2.17}(\eta, r, \zeta, C_0, d)$, then one can assume that $\mathcal{B}_1 \succeq_{syn} \mathcal{B}_0$.*

## 2.4 Useful claims

We prove in this subsection a few useful claims relating to polynomial factors.

Sometimes one needs to refine the decomposition given by the factor $\mathcal{B}_1$ in Theorem 2.17. The following simple lemma is useful in such situations.

**Lemma 2.18.** *Let $\mathcal{B}$ be a polynomial factor of complexity $C$ and degree $d$, and let $f : \mathbb{F}^n \to \{0, 1\}$ be decomposed into $f = \mathbf{E}[f|\mathcal{B}] + f_2 + f_3$ for $f_2, f_3 : \mathbb{F}^n \to [-1, 1]$. If $\mathcal{B}' \succeq_{sem} \mathcal{B}$ is a polynomial factor of degree $d$ and complexity $C'$, then*

$$\left\| \mathbf{E}\left[f|\mathcal{B}\right] - \mathbf{E}\left[f|\mathcal{B}'\right]\right\|_1 \leqslant \|f_2\|_2 + p^{dC'}\|f_3\|_{U^{d+1}}.$$

*Proof.* We have

$$\left\| \mathbf{E}\left[f|\mathcal{B}\right] - \mathbf{E}\left[f|\mathcal{B}'\right]\right\|_1 = \left\| \mathbf{E}\left[f_2|\mathcal{B}'\right] + \mathbf{E}\left[f_3|\mathcal{B}'\right]\right\|_1 \leqslant \|f_2\|_1 + \left\| \mathbf{E}\left[f_3|\mathcal{B}'\right]\right\|_1.$$

The claim follows since $\|f_2\|_1 \leqslant \|f_2\|_2$ and Claim 2.20 below shows that $\|\mathbf{E}\left[f_3|\mathcal{B}'\right]\|_1 \leqslant p^{dC'}\|f_3\|_{U^{d+1}}$. $\square$

**Claim 2.19.** *Let $f : \mathbb{F}^n \to [-1, 1]$, and let $\mathcal{B}$ be a polynomial factor of degree $d$ and complexity $C$. Then for any atom $b$ of $\mathcal{B}$,*

$$\|f(x)1_{\mathcal{B}(x)=b}\|_{U^{d+1}} \leqslant \|f\|_{U^{d+1}}.$$

*Proof.* Let $\mathcal{B}$ be defined by polynomials $P_1, \ldots, P_C$ of depths $k_1, \ldots, k_C$, respectively. An atom $b \in \mathcal{B}$ is defined by $b = \{x \in \mathbb{F}^n : P_i(x) = b_i\}$. So

$$
\begin{aligned}
\|f(x)1_{\mathcal{B}(x)=b}\|_{U^{d+1}} &= \left\| f(x) \prod_{i=1}^{C} \frac{1}{p^{k_i+1}} \sum_{\lambda_i=0}^{p^{k_i+1}-1} \mathsf{e}\left(\lambda_i(P_i(x) - b_i)\right) \right\|_{U^{d+1}} \\
&\leqslant \prod_{i=1}^{C} p^{-(k_i+1)} \cdot \sum_{\substack{(\lambda_1, \ldots, \lambda_C) \\ \in \prod_i [0, p^{k_i+1}-1]}} \left\| f(x)\mathsf{e}\left(\sum_i \lambda_i(P_i(x) - b_i)\right) \right\|_{U^{d+1}} \\
&= \left(\prod_{i=1}^{C} p^{-(k_i+1)}\right) \cdot \left(\prod_{i=1}^{C} p^{k_i+1}\|f\|_{U^{d+1}}\right) = \|f\|_{U^{d+1}}.
\end{aligned}
$$

Let us $\square$

11

**Claim 2.20.** *Let $f : \mathbb{F}^n \to [-1, 1]$, and let $\mathcal{B}$ be a polynomial factor of degree $d$ and complexity $C$. Then $\|\mathbf{E}\left[f|\mathcal{B}\right]\|_1 \leqslant p^{dC}\|f\|_{U^{d+1}}$.*

*Proof.* By the monotonicity of the Gowers norms and Claim 2.19, for every atom $b \in \mathcal{B}$

$$\left|\mathbf{E}\left[f(x)1_{\mathcal{B}(x)=b}\right]\right| \leqslant \left\|f(x)1_{\mathcal{B}(x)=b}\right\|_{U^{d+1}} \leqslant \|f\|_{U^{d+1}}.$$

Hence

$$\|\mathbf{E}\left[f|\mathcal{B}\right]\|_1 = \sum_{b \in \mathcal{B}} \left|\mathbf{E}\left[f(x)1_{\mathcal{B}(x)=b}\right]\right| \leqslant p^{dC}\|f\|_{U^{d+1}}.$$

$\square$

We need a simple bound on Gowers uniformity norms in terms of $L_1$ norm.

**Claim 2.21.** *Let $f : \mathbb{F}^n \to [-1, 1]$. For any $d \geqslant 1$,*

$$\|f\|_{U^d} \leqslant \|f\|_1^{1/2^d}.$$

*Proof.* By the definition of the $U^d$ norm and the boundedness of $f$,

$$\|f\|_{U^d}^{2^d} \leqslant \mathop{\mathbf{E}}_{x,y_1,\ldots,y_d} \prod_{I \subseteq [d]} |f(x + \sum_{i \in I} y_i)| \leqslant \mathbf{E}|f(x)| = \|f\|_1.$$

$\square$

We will also need the following lemma about restrictions of high rank polynomials to affine subspaces.

**Lemma 2.22.** *For $\varepsilon > 0$ and positive $d, e, r$, if $m \geqslant m_{2.22}(\varepsilon, d, e, r)$ and $n \geqslant m$ then the following holds. For every polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree $d$, depth $e$ and rank $\geqslant r$, a random affine embedding $A : \mathbb{F}^m \to \mathbb{F}^n$ satisfies that*

$$\mathbf{Pr}[\deg(P) < d \text{ or } (\mathsf{rank}_d(BP) < r) \text{ or } \mathrm{depth}(P) < e] < \varepsilon.$$

*Proof.* Let $\mathcal{P}'$ denote the property of "bad" restrictions of $P$. That is, $\mathcal{P}'$ is the property of functions which are either polynomials of degree less than $d$; or polynomials of degree $d$ and rank less than $r$; or polynomials of degree $d$ and depth less than $e$. By assumption $P \notin \mathcal{P}'$. Theorem 1.7 in [BFH$^+$13] shows that any degree-structural property, and in particular $\mathcal{P}'$, is locally defined. Theorem 1.2 in [BFH$^+$13] shows that any such property is locally testable. Furthermore, as all elements in $\mathcal{P}'$ are polynomials of degree $d$, then $P$ is $\eta$-far from $\mathcal{P}'$, where $\eta \geqslant p^{-\lceil d/(p-1)\rceil}$ is the minimal distance of polynomials of degree $d$ (it is not hard to see that the minimal distance for non-classical polynomials of a given degree is achieved by a classical polynomial). Hence, by Proposition 1.2 there exists $m = m(\eta, d, e, r)$ such that for a random affine embedding $A : \mathbb{F}^m \to \mathbb{F}^n$, $\mathbf{Pr}[AP \in \mathcal{P}'] < \varepsilon$. $\square$

# 3    Some remarks on testability

Let us discuss some results related to testablity of affine-invariant properties. To simplify the presentation we focus on the special case of $R = 2$. The proof easily generalizes to $R > 2$, by decomposing every function $f : \mathbb{F}^n \to [R]$ as $f = f^{(1)} + \ldots f^{(R)}$ where $f^{(i)}$ is the indicator function of the set $\{x : f(x) = i\}$.

Consider a function $f : \mathbb{F}^n \to \{0, 1\}$, and a positive integer $k \leqslant n$. Let $A : \mathbb{F}^k \to \mathbb{F}^n$ be a random affine embedding. Let $\mu_{f,k}$ denote the distribution of $Af : \mathbb{F}^k \to \{0, 1\}$. So via this sampling, every function $f : \mathbb{F}^n \to \{0, 1\}$ defines a probability measure $\mu_{f,k}$ on the set of functions $\{\mathbb{F}^k \to \{0, 1\}\}$. We denote by $\mu_{f,k}[v]$ the probability that $\mu_{f,k}$ assigns to $v : \mathbb{F}^k \to \{0, 1\}$.

This can be generalized to functions $f : \mathbb{F}^n \to [0, 1]$. We view such functions as distribution over functions $f' : \mathbb{F}^n \to \{0, 1\}$, where $\mathbf{Pr}[f'(x) = 1] = f(x)$ independently for all $x \in \mathbb{F}^n$. Let again $A : \mathbb{F}^k \to \mathbb{F}^n$ be a random affine embedding, and we denote by $\mu_{f,k}$ the distribution of $Af' : \mathbb{F}^k \to \{0, 1\}$. This is a generalization of the former case as a function $f : \mathbb{F}^n \to \{0, 1\}$ can be identified with the function that maps every $x \in \mathbb{F}^n$ to the point-mass probability distribution over $\{0, 1\}$ which is concentrated on $f(x)$.

The following simple corollary follows easily from the definition of testability.

**Corollary 3.1.** *If an affine-invariant property $\mathcal{P}$ is testable, then for every $\varepsilon > 0$, there exist $k \geqslant 1$ so that the following holds. For any function $f : \mathbb{F}^n \to \{0, 1\}$ with $\mathrm{dist}(f, \mathcal{P}) \geqslant \varepsilon$, and any function $g : \mathbb{F}^n \to \{0, 1\}$ in $\mathcal{P}$, the statistical distance between $\mu_{f,k}$ and $\mu_{g,k}$ is at least $1/3$.*

*Proof.* From the definition of testability in Proposition 1.2, there exist $k \geqslant 1$ and a family $\mathcal{H} \subseteq \{\mathbb{F}^k \to \{0, 1\}\}$, such that for a random affine embedding $A : \mathbb{F}^k \to \mathbb{F}^n$, $\mathbf{Pr}[Ag \in \mathcal{H}] > 2/3$ and $\mathbf{Pr}[Af \in \mathcal{H}] \leqslant 1/3$. Hence, the statistical distance between $\mu_{f,k}$ and $\mu_{g,k}$ is at least $1/3$. $\square$

We can deduce the following useful corollary. If $f : \mathbb{F}^n \to \{0, 1\}$ has a distribution $\mu_{f,k}$ which is very close to $\mu_{g,k}$ for a function $g \in \mathcal{P}$, then $f$ must be close to $\mathcal{P}$. In fact, the same holds for $f : \mathbb{F}^n \to [0, 1]$, except now the results holds with high probability over $f' : \mathbb{F}^n \to \{0, 1\}$ sampled from $f$.

**Corollary 3.2.** *For $\varepsilon > 0$ let $k \geqslant 1$ be given by Corollary 3.1, and assume that $n \geqslant n_{3.2}(k, \varepsilon)$. Let $f : \mathbb{F}^n \to [0, 1]$ and $g : \mathbb{F}^n \to \{0, 1\}$ so that $g \in \mathcal{P}$ and the statistical distance between $\mu_{f,k}, \mu_{g,k}$ is at most $1/4$. Let $f' : \mathbb{F}^n \to \{0, 1\}$ be sampled by taking $f'(x) = 1$ with probability $f(x)$ independently for all $x \in \mathbb{F}^n$. Then with probability at least $99\%$ over the choice of $f'$,*

$$\mathrm{dist}(f', \mathcal{P}) \leqslant \varepsilon.$$

*Proof.* We will show that by choosing $n$ large enough, the distribution $\mu_{f,k}$ and $\mu_{f',k}$ are very close in statistical distance (say, distance $\leqslant 1/12$) with high probability (say, 99%). The corollary then follows from Corollary 3.1 applied to $f'$ and $g$. Let $v : \mathbb{F}^k \to \{0, 1\}$ be a function. By definition

$$\mu_{f,k}[v] = \underset{A,f'}{\mathbf{E}} \, \mathbf{Pr}[Af' = v] = \underset{f'}{\mathbf{E}} \, \mu_{f',k}[v].$$

Moreover, for two affine embeddings $A_1, A_2 : \mathbb{F}^k \to \mathbb{F}^n$, if their images are disjoint then $A_1 f'$ and $A_2 f'$ are independent. Since the probability over a random choice of $A_1, A_2$ that their images intersect is at most $p^{2k-n}$, we get that

$$\mathbf{Var}[\mu_{f',k}[v]] \leqslant p^{2k-n} = o_n(1).$$

This means that $\mu_{f',k}[v] = \mu_{f,k}[v] + o_n(1)$ with probability $1 - o_n(1)$. The result now follows from applying the union bound over all possible values of $v$. □

We next argue that for two functions $f, g : \mathbb{F}^n \to [0, 1]$, for any $k \geqslant 1$, there exists a $d \geqslant 1$ such that, if $\|f - g\|_{U^d}$ is small enough, then the statistical distance of $\mu_{f,k}, \mu_{g,k}$ is arbitrarily small. This is useful, since it shows that in this case if $g \in \mathcal{P}$ then $f$ must be close to $\mathcal{P}$ provided that $k$ is large enough.

**Lemma 3.3.** *For every $\varepsilon > 0, k \geqslant 1$, there exists $\rho > 0, d \geqslant 1$ such that the following holds. If $f, g : \mathbb{F}^n \to [0, 1]$ are functions such that $\|f - g\|_{U^d} \leqslant \rho$, then the statistical distance between $\mu_{f,k}$ and $\mu_{g,k}$ is at most $\varepsilon$.*

*Proof.* Let $A : \mathbb{F}^k \to \mathbb{F}^n$ be a random affine embedding. Consider $v : \mathbb{F}^k \to \{0, 1\}$. For $y \in \mathbb{F}^k$, define $f_y(x) = 1 - v(y) - (-1)^{v(y)} f(x)$. The probability that $\mu_{f,k}$ samples $v$ is given by

$$\mu_{f,k}[v] = \mathop{\mathbf{E}}_{A,f'} \mathbf{Pr}[Af' = v] = \mathop{\mathbf{E}}_{A} \prod_{y \in \mathbb{F}^k} f_y(Ay).$$

Similarly define $g_y(x) = 1 - v(y) - (-1)^{v(y)} g(x)$ to obtain

$$\mu_{g,k}[v] = \mathop{\mathbf{E}}_{A} \prod_{y \in \mathbb{F}^k} g_y(Ay).$$

Let $<$ define an arbitrary ordering on $\mathbb{F}^k$. We rewrite $\mu_{f,k}[v] - \mu_{g,k}[v]$ as a telescopic sum

$$\mu_{f,k}[v] - \mu_{g,k}[v] = \sum_{z \in \mathbb{F}^k} \mathop{\mathbf{E}}_{A} \left( \prod_{y<z} f_y(Ay) \right) \cdot (f_z(Az) - g_z(Az)) \cdot \left( \prod_{y>z} g_y(Ay) \right).$$

We bound each term in the sum. To do so, we will apply Lemma 2.7. Note that the set of linear forms $\{Ay : y \in \mathbb{F}^k\}$ is exactly that given in Claim 2.8 and its Cauchy-Schwarz complexity is at most $p^k$. Note that $\|f_y\|_\infty \leqslant 1$. Hence for $d = p^k + 1$ we get that

$$\left| \mathop{\mathbf{E}}_{A} \left( \prod_{y<z} f_y(Ay) \right) \cdot (f_z(Az) - g_z(Az)) \cdot \left( \prod_{y>z} g_y(Ay) \right) \right| \leqslant \|f_z - g_z\|_{U^d} = \|f - g\|_{U^d}.$$

We conclude that $|\mu_{f,k}[v] - \mu_{g,k}[v]| \leqslant p^k \|f - g\|_{U^d}$ and hence the statistical distance between $\mu_{f,k}$ and $\mu_{g,k}$ is bounded by $2^{p^k} p^k \|f - g\|_{U^d}$. The lemma follows for $\rho = 2^{-p^k} p^{-k} \varepsilon$. □

The following corollary is immediate.

**Corollary 3.4.** *For every $\varepsilon > 0$ there exist $d \geqslant 1, \rho > 0$ such that the following holds. Let $f, g : \mathbb{F}^n \to \{0, 1\}$ be functions and assume that $g \in \mathcal{P}$. If $\|f - g\|_{U^d} \leqslant \rho$ then $f$ is $\varepsilon$-close to $\mathcal{P}$.*

Structured parts obtained from the decomposition theorems are of the form $f(x) = \Gamma(P_1(x), \ldots, P_C(x))$ where $P_1, \ldots, P_C$ are polynomials. We would argue that if they have large enough rank, then $\mu_{f,k}$ essentially depends just on $\Gamma$ and the degrees and depths of $P_1, \ldots, P_C$, and not on the specific polynomials.

14

**Lemma 3.5.** *For any $\varepsilon > 0$ and $k, d \geqslant 1$, there exists $r = r_{3.5}(k, d, \varepsilon) : \mathbb{N} \to \mathbb{N}$ such the following holds. Let $P_1, \ldots, P_C$ be an $r$-regular factor over $\mathbb{F}^n$ of degree at most $d$. Let $Q_1, \ldots, Q_C$ be an $r$-regular factor over $\mathbb{F}^m$ of degree at most $d$. Assume that both $P_i, Q_i$ have degree $d_i \leqslant d$ and depth $k_i$, for all $i \leqslant C$. Let $\Gamma : \prod_{i=1}^{C} \mathbb{U}_{k_i+1} \to [0, 1]$ be a function. Let $f : \mathbb{F}^n \to [0, 1]$ be defined as $f(x) = \Gamma(P_1(x), \ldots, P_C(x))$ and $g : \mathbb{F}^m \to [0, 1]$ be defined as $g(x) = \Gamma(Q_1(x), \ldots, Q_C(x))$. Then $\mu_{f,k}$ and $\mu_{g,k}$ have statistical distance at most $\varepsilon$.*

*Proof.* Let $A : \mathbb{F}^k \to \mathbb{F}^n$ be a random affine embedding. For $y \in \mathbb{F}^k$ define $\Gamma_y : \prod_{i=1}^{C} \mathbb{U}_{k_i+1} \to [0, 1]$ as

$$\Gamma_y(z_1, \ldots, z_C) = 1 - v(y) - (-1)^{v(y)} \Gamma(z_1, \ldots, z_C).$$

The probability that $\mu_{f,k}$ samples $v : \mathbb{F}^k \to \{0, 1\}$ is

$$\mu_{f,k}[v] = \mathbf{E}_{A} \prod_{y \in \mathbb{F}^k} \Gamma_y(P_1(Ay), \ldots, P_C(Ay)).$$

Expanding each $\Gamma_y$ in the Fourier basis gives

$$\Gamma_y(z_1, \ldots, z_C) = \sum_{\alpha \in \prod_{i=1}^{C} \mathbb{U}_{k_i+1}} \hat{\Gamma}_y(\alpha) \mathrm{e}\left( \sum_{j=1}^{C} \alpha_j z_j \right).$$

Note that $\|\widehat{\Gamma_y \Gamma_y}\|_\infty \leqslant 1$. Plugging this into the equation for $\mu_{f,k}[v]$ and expanding gives

$$\mu_{f,k}[v] = \sum_{\alpha : \mathbb{F}^k \to \prod_{i=1}^{C} \mathbb{U}_{k_i+1}} c_\alpha \, \mathbf{E}_{A} \, \mathrm{e}\left( \sum_{y \in \mathbb{F}^k} \sum_{j=1}^{C} \alpha(y)_j P_j(Ay) \right),$$

where $c_\alpha := \prod_{y \in \mathbb{F}^k} \hat{\Gamma}_y(\alpha(y))$. Note that $|c_\alpha| \leqslant 1$ and that it depends only on $\Gamma$ and the depths of the polynomials, and not on the specific polynomials. We will apply the same expansion to $g$ and obtain

$$\mu_{g,k}[v] = \sum_{\alpha : \mathbb{F}^k \to \prod_{i=1}^{C} \mathbb{U}_{k_i+1}} c_\alpha \, \mathbf{E}_{A} \, \mathrm{e}\left( \sum_{y \in \mathbb{F}^k} \sum_{j=1}^{C} \alpha(y)_j Q_j(Ay) \right).$$

We next apply Theorem 3.3 in [BFH+13]. It states that linear combination of systems of high rank polynomials evaluated over affine linear forms, such as $\sum_{y \in \mathbb{F}^k} \sum_{j=1}^{C} \alpha(y)_j P_j(Ay)$, are either identically zero or very close to uniformly distributed. Concretely, the theorem states that for any parameter $\gamma(C) > 0$, if we choose $r(C)$ large enough, then for any polynomials $P_1, \ldots, P_C$ of degrees $d_1, \ldots, d_C$ and depths $k_1, \ldots, k_C$ and rank at least $r(C)$, either

$$\sum_{y \in \mathbb{F}^k} \sum_{j=1}^{C} \alpha(y)_j P_j(Ay) \equiv 0$$

or

$$\left| \mathbf{E}_{A} \, \mathrm{e}\left( \sum_{y \in \mathbb{F}^k} \sum_{j=1}^{C} \alpha(y)_j P_j(Ay) \right) \right| \leqslant \gamma(C).$$

15

Note that we can apply this theorem both to $P_1, \ldots, P_C$ and to $Q_1, \ldots, Q_C$, obtaining the same results. Hence, we conclude that

$$|\mu_{f,k}[v] - \mu_{g,k}[v]| \leqslant \left(\prod_{i=1}^{C} p^{k_i+1}\right) \gamma(C) \leqslant p^{dC} \gamma(C).$$

Choosing $\gamma(C) := p^{-dC} 2^{-p^k} \varepsilon$ we obtain that $\mu_{f,k}$ and $\mu_{g,k}$ have statistical distance at most $\varepsilon$. $\quad \square$

## 4 Proof of Theorem 1.3

To simplify the presentation we prove the theorem for the special case of $R = 2$. The proof easily generalizes to $R > 2$, by decomposing every function $f : \mathbb{F}^n \to [R]$ as $f = f^{(1)} + \ldots f^{(R)}$ where $f^{(i)}$ is the indicator function of the set $\{x : f(x) = i\}$.

Let $\mathcal{P} \subseteq \{f : \mathbb{F}^n \to \{0,1\} : n \in \mathbb{N}\}$ be a testable affine-invariant property, and let $\delta, \varepsilon > 0$ be the parameters given in Theorem 1.3. We will show that a local test can distinguish between functions which are $\delta$-close to $\mathcal{P}$ to functions which are $\delta + \varepsilon$ far from $\mathcal{P}$.

Let $m = m(\mathcal{P}, \delta, \varepsilon)$ be a sufficiently large integer to be determined later, and let $A : \mathbb{F}^m \to \mathbb{F}^n$ be a random affine embedding. We will establish the following two statements.

**(i)** If $f : \mathbb{F}^n \to \{0,1\}$ is $\delta$-close to $\mathcal{P}$, then

$$\mathbf{Pr}[\mathrm{dist}(Af, \mathcal{P}) < \delta + (\varepsilon/2)] > 2/3.$$

**(ii)** If $f : \mathbb{F}^n \to \{0,1\}$ is $(\delta + \varepsilon)$-far from $\mathcal{P}$, then

$$\mathbf{Pr}[\mathrm{dist}(Af, \mathcal{P}) > \delta + (\varepsilon/2)] > 2/3.$$

### 4.1 Proof of (i)

Since $f$ is $\delta$-close to $\mathcal{P}$, there exists a function $g \in \mathcal{P}$ such that $\alpha := \|f - g\|_1 \leqslant \delta$. Note that

$$\mathbf{E}_A[\|Af - Ag\|_1] = \mathbf{E}_{x \in \mathbb{F}^m, A}[|Af(x) - Ag(x)|] = \mathbf{E}_{x \in \mathbb{F}^n}[|f(x) - g(x)|] = \alpha.$$

Furthermore

$$\mathbf{E}_A\left[\|Af - Ag\|_1^2\right] = \mathbf{E}_{x,y \in \mathbb{F}^m, A}[|Af(x) - Ag(x)||Af(y) - Ag(y)|] \leqslant \alpha^2 + \mathbf{Pr}[x = y] \leqslant \alpha^2 + \frac{1}{p^m}.$$

Hence $\mathbf{Var}[\mathrm{dist}(Af, \mathcal{P})] \leqslant \frac{1}{p^m}$, and thus by Chebyshev's inequality

$$\mathbf{Pr}[\mathrm{dist}(Af, \mathcal{P}) \geqslant \delta + (\varepsilon/2)] < \frac{4}{p^m \varepsilon^2} < 1/3,$$

provided that $m$ is sufficiently large.

## 4.2 Proof of (ii)

We apply Corollary 3.2 and Corollary 3.4 with parameter $\varepsilon/8$ to obtain $k, d \geqslant 1$ and $\rho > 0$, so that the following two statements hold.

- If $f : \mathbb{F}^n \to \{0,1\}$, $g : \mathbb{F}^m \to \{0,1\}$ are functions, $g \in \mathcal{P}$, and $\mu_{f,k}$ and $\mu_{g,k}$ have statistical distance at most $1/4$, then $f$ is $(\varepsilon/8)$-close to $\mathcal{P}$.

- If $f, g : \mathbb{F}^n \to \{0,1\}$ are functions and $\|f - g\|_{U^d} < \rho$, then $\mu_{f,k}$ and $\mu_{g,k}$ have statistical distance at most $1/100$.

Let $f : \mathbb{F}^n \to \{0,1\}$ be a function which is $(\delta + \varepsilon)$-far from $\mathcal{P}$. We start by decomposing $f$ to a structured part and a pseudo-random part. Our decompositions will use a number of parameters. We already fixed $\varepsilon, \delta$ and $d$. Let $\gamma > 0$, $\eta_0, \eta_1 : \mathbb{N} \to \mathbb{R}^+$ and $r_0, r_1 : \mathbb{N} \to \mathbb{N}$ be parameters to be determined later. For the reader who wishes to verify that these definitions are not cyclical, we note that $\gamma$ will depend just on $\varepsilon, \delta, d$; that $\eta_1, r_1$ will depend just on $\varepsilon, \delta, d, \gamma$; and that $\eta_0, r_0$ will depend on $\varepsilon, \delta, d, \gamma, \eta_1, r_1$.

We apply Theorem 2.17 to $f$ with parameters $d, \gamma, r_0, \eta_0$ and a trivial initial factor, and obtain an $r_0$-regular polynomial factor $\mathcal{B}_0$ of degree less than $d$, and a decomposition

$$f = f_1 + f_2 + f_3,$$

where $f_1 = \mathbf{E}[f|\mathcal{B}_0], \|f_2\|_2 < \gamma, \|f_3\|_{U^d} < \eta_0(|\mathcal{B}_0|)$. Next, we project this decomposition to $Af$. Suppose that $\mathcal{B}_0$ is defined by polynomials $P_1, \ldots, P_C$. Denote by $Q_i := AP_i$ for $i = 1, \ldots, C$, and let $\tilde{\mathcal{B}}_0$ be the polynomial factor over $\mathbb{F}^m$ defined by the $Q_i$'s. We decompose

$$Af = Af_1 + Af_2 + Af_3.$$

The following claim argues that $Af_1, Af_2, Af_3$ have similar properties to $f_1, f_2, f_3$ with high probability, assuming that $m$ is chosen large enough and that $r_0$ is chosen to grow fast enough.

**Claim 4.1.** *Assume that $r_0(C) \geqslant r_{2.12}(d, 1/(2p^{dC}))$. If $m \geqslant m_{4.1}(d, \gamma, r_0, \eta_0)$, then the following events hold with probability at least 99%.*

$(\mathbf{E}_1)$ *The polynomials $Q_1, \ldots, Q_C$ have the same degrees and depths as $P_1, \ldots, P_C$, respectively, and $\tilde{\mathcal{B}}_0$ is $r_0$-regular.*

$(\mathbf{E}_2)$ *We have $\|Af_2\|_2 \leqslant 2\gamma$ and $\|Af_3\|_{U^d} \leqslant 2\eta_0(|\mathcal{B}_0|)$.*

$(\mathbf{E}_3)$ $\left\| \mathbf{E}[Af|\tilde{\mathcal{B}}_0] - A\,\mathbf{E}[f|\mathcal{B}_0] \right\|_\infty \leqslant \gamma.$

*Proof.* Lemma 2.22 shows that the probability that $(\mathbf{E}_1)$ does not hold can be made arbitrarily small by setting the parameters properly. To prove that $(\mathbf{E}_2)$ holds with high probability, we will show that $\mathbf{Pr}\left[\|Af_3\|_{U^d} > \|f_3\|_{U^d} + o_m(1)\right] = o_m(1)$ and $\mathbf{Pr}\left[\|Af_2\|_2 > \|f\|_2 + o_m(1)\right] = o_m(1)$. We only prove the former, as the proof of the latter is easy and similar. To do so, We will establish that $\mathbf{E}_A \|Af_3\|_{U^d}^{2^d} = \|f_3\|_{U^d}^{2^d} + o_m(1)$ and that $\mathbf{Var}(\|Af_3\|_{U^d}^{2^d}) = o_m(1)$, and apply Chebyshev's inequality.

We first establish the first moment calculation. Let $y_1, \ldots, y_d, x$ be uniform random variables taking values in $\mathbb{F}^m$. Note that if $y_1, \ldots, y_d, x$ are linearly independent then $Ay_1, \ldots, Ay_d, Ax$ are

17

linearly independent uniform random variables taking values in $\mathbb{F}^n$. The probability that they are not linearly independent is at most $p^{d+1-m} = o_m(1)$. Combining this with the fact that $\|f_3\|_\infty \leqslant 1$ we get

$$\mathbf{E}_A \|Af_3\|_{U^d}^{2^d} = \mathbf{E}\left[A\Delta_{y_1,\ldots,y_d}f_3(x)\right] = \mathbf{E}\left[\Delta_{Ay_1,\ldots,Ay_d}f_3(Ax)\right] = \|f_3\|_{U^d}^{2^d} \pm p^{d+1-m}.$$

We proceed to the second moment calculation. Let $y_1',\ldots,y_d',x'$ be independent uniform random variables taking values in $\mathbb{F}^m$ uniformly and independently of $y_1,\ldots,y_d,x$. Similarly, if $y_1,\ldots,y_d,x,y_1',\ldots,y_d',x'$ are linearly independent then $Ay_1,\ldots,Ay_d,Ax,Ay_1',\ldots,Ay_d',Ax'$ are linearly independent uniform random variables taking values in $\mathbb{F}^n$. Hence same as before,

$$\mathbf{E}_A \|Af_3\|_{U^d}^{2^{d+1}} = \mathbf{E}\left[(A\Delta_{y_1,\ldots,y_d}f_3(x))\left(A\Delta_{y_1',\ldots,y_d'}f_3(x')\right)\right] = \|f_3\|_{U^d}^{2^{d+1}} \pm p^{2d+2-m}.$$

Thus, $\mathbf{Var}(\|Af_3\|_{U^d}^{2^d}) = o_m(1)$ and by Chebyshev's inequality ($\mathbf{E}_2$) holds with high probability assuming $m$ is chosen large enough.

We next establish that ($\mathbf{E}_3$) holds with high probability. Similarly to the previous calculation, this will also be shown by performing a first and second moment calculation and applying Chebyshev's inequality. Consider an atom $b_0 \in \mathbb{T}^C$ of $\mathcal{B}_0$. Since $\tilde{\mathcal{B}}_0$ is defined by $AP_1,\ldots,AP_C$, we have that $\mathbf{E}[Af(y)|\tilde{\mathcal{B}}_0(y) = b_0] = \mathbf{E}[Af(y)|\mathcal{B}_0(Ay) = b_0]$, hence

$$\mathbf{E}_A[\mathbf{E}[Af(y)|\mathcal{B}_0(Ay) = b_0]] = \mathbf{E}_{x\in\mathbb{F}^n}[f(x)|\mathcal{B}_0(x) = b_0],$$

and

$$
\begin{aligned}
\mathbf{E}_A\left[\mathbf{E}[Af(y)|\mathcal{B}_0(Ay) = b_0]^2\right] &= \mathbf{E}_{y_1,y_2\in\mathbb{F}^m,A}[f(Ay_1)f(Ay_2)|\mathcal{B}_0(Ay_1) = \mathcal{B}_0(Ay_2) = b_0] \\
&= \mathbf{E}[f(x)|\mathcal{B}_0(x) = b_0]^2 \pm \mathbf{Pr}[y_1 = y_2|\mathcal{B}_0(Ay_1) = \mathcal{B}_0(Ay_2) = b_0] \\
&= \mathbf{E}[f(x)|\mathcal{B}_0(x) = b_0]^2 \pm p^{-m}/(|\{x \in \mathbb{F}^m : \mathcal{B}_0(x) = b_0\}|p^{-n})^2 \\
&= \mathbf{E}[f(x)|\mathcal{B}_0(x) = b_0]^2 \pm 4|\mathcal{B}_0|^2 p^{-m},
\end{aligned}
$$

where in the last step we applied Theorem 2.12 and the assumption on the rank of $\mathcal{B}_0$ to lower bound the size of the atom defined by $b_0$. So by Chebyshev's inequality $\mathbf{E}[Af(y)|\tilde{\mathcal{B}}_0(y) = b_0]$ is concentrated around $\mathbf{E}_x[f(x)|\mathcal{B}_0(x) = b_0]$. Since the number of atoms is bounded by $\|\mathcal{B}_0\|$, we obtain that with probability $1 - o_m(1)$ this holds of every atom. $\qquad\square$

Claim 4.1 shows that 99% of the affine embeddings $A$ satisfy ($\mathbf{E}_1$), ($\mathbf{E}_2$), ($\mathbf{E}_3$). Let us assume towards contradiction that $\mathbf{Pr}_A[\text{dist}(Af, \mathcal{P}) > \delta + (\varepsilon/2)] \leqslant 1/3$. So, we can fix an embedding $A$ so that ($\mathbf{E}_1$), ($\mathbf{E}_2$), ($\mathbf{E}_3$) hold, and find a function $h : \mathbb{F}^m \to \{0,1\}$ in $\mathcal{P}$ for which $\text{dist}(Af, h) \leqslant \delta + (\varepsilon/2)$. We fix $A$ and $h$ for the reminder of the proof.

The next step is to decompose $h$. However, we wish to maintain the regular factor $\tilde{\mathcal{B}}_0$ we obtained for $Af$. So, we apply Theorem 2.17 to $h$ with parameters $d, \gamma, r_1, \eta_1$ and initial factor $\tilde{\mathcal{B}}_0$, and obtain an $r_1$-regular polynomial factor $\tilde{\mathcal{B}}_1$ of degree $d$, and a decomposition

$$h = h_1 + h_2 + h_3,$$

where $h_1 = \mathbf{E}[h|\tilde{\mathcal{B}}_1], \|h_2\|_2 < \gamma, \|h_3\|_{U^d} < \eta_1(|\mathcal{B}_1|)$. Furthermore, we will assume that $r_0$ is much larger than $r_1$, so that by Theorem 2.17 we get that $\tilde{\mathcal{B}}_1$ is a syntactic refinement of $\tilde{\mathcal{B}}_0$ (which we recall that by Claim 4.1 is $r_0$-regular). Concretely, this will require us to assume that $r_0(C) \geqslant$

$r_{2.17}(\eta_1, r_1, \gamma, C, d)$. So, $\tilde{\mathcal{B}}_1$ is defined by polynomials $Q_1, \ldots, Q_{C'}$ for a constant $C' > C$, where we recall that $\tilde{\mathcal{B}}_0$ was defined by $Q_1, \ldots, Q_C$. Our construction so far guarantees that for $i \leqslant C$ we have that $Q_i = AP_i$ and that $P_i, Q_i$ have the same depth and degree. We would like to guarantee this also for $i > C$. That is, we would like to find polynomials $P_i$ for $C < i \leqslant C'$ defined over $\mathbb{F}^n$ for which $Q_i = AP_i$ as well, and such that $P_1, \ldots, P_{C'}$ is of high rank.

Let $A' : \mathbb{F}^n \to \mathbb{F}^m$ be any affine transformation satisfying $AA' = I_m$. For $i = C+1, \ldots, C'$ set $P_i := A'Q_i$. Note that applying an affine transformation cannot increase degree, depth, or rank. Hence, since $P_i = A'Q_i$ and $Q_i = AP_i$ we must have that $P_i, Q_i$ have the same degree and depth for $C < i \leqslant C'$. Moreover, by $(\mathbf{E}_1)$ we know that this also holds for $i \leqslant C$, hence it holds for all $1 \leqslant i \leqslant C'$. Furthermore, since by construction $Q_1, \ldots, Q_{C'}$ are $r_1$-regular than so are $P_1, \ldots, P_{C'}$. We denote by $\mathcal{B}_1$ the polynomial factor defined by $P_1, \ldots, P_{C'}$. The following claim summarizes the notation and the facts we established so far.

**Claim 4.2.** *The factor $\mathcal{B}_0$ is an $r_0$-regular factor over $\mathbb{F}^n$ defined by $P_1, \ldots, P_C$. The factor $\tilde{\mathcal{B}}_0$ is an $r_0$-regular factor over $\mathbb{F}^m$ defined by $Q_1, \ldots, Q_C$. The factor $\mathcal{B}_1$ is an $r_1$-regular factor over $\mathbb{F}^n$ defined by $P_1, \ldots, P_{C'}$. The factor $\tilde{\mathcal{B}}_1$ is an $r_1$-regular factor over $\mathbb{F}^m$ defined by $Q_1, \ldots, Q_{C'}$. We further have:*

- $Q_i = AP_i$.

- $P_i, Q_i$ *have the same degree and depth.*

- $f = f_1 + f_2 + f_3$ *where* $f_1 = \mathbf{E}[f|\mathcal{B}_0], \|f_2\|_2 \leqslant \gamma, \|f_3\|_{U^d} \leqslant \eta_0(C)$.

- $h : \mathbb{F}^m \to \{0, 1\}$ *is a function in $\mathcal{P}$ to $Af$ for which* $\mathrm{dist}(Af, h) \leqslant \delta + \varepsilon/2$.

- $h = h_1 + h_2 + h_3$ *where* $h_1 = \mathbf{E}[h|\tilde{\mathcal{B}}_1], \|h_2\|_2 \leqslant \gamma, \|h_3\|_{U^d} \leqslant \eta_1(C')$.

We already know that the property that $h$ is in $\mathcal{P}$ does not actually depend on $h_3$. We will shortly show that for a small enough choice of $\gamma$, it also does not depend on $h_2$, hence all the information is essentially just in $h_1$. We would like to lift $h_1$ from $\mathbb{F}^m$ to $\mathbb{F}^n$ in order to get a function $g : \mathbb{F}^n \to \{0, 1\}$ which is close to $\mathcal{P}$. Moreover, we would like to do so in a way so that $\|f - g\|_1 \approx \mathrm{dist}(Af, h)$, hence this lifting must be done carefully. We start by lifting $h_1 : \mathbb{F}^m \to [0, 1]$ to a function $\phi : \mathbb{F}^n \to [0, 1]$ which has a similar structure. We would later use $\phi$ to define the required function $g$ as discussed above.

The two factors $\mathcal{B}_1$ and $\tilde{\mathcal{B}}_1$ are both of large rank (at least $r_1(C')$) and their defining polynomials are in a degree and depth preserving one-to-one correspondence. This naturally defines an operator $\mathcal{T}$ that maps functions $\phi : \mathbb{F}^n \to \mathbb{C}$ measurable with respect to $\mathcal{B}_1$ to functions $\mathcal{T}\phi : \mathbb{F}^m \to \mathbb{C}$ measurable with respect to $\tilde{\mathcal{B}}_1$. More precisely, $\mathcal{T} : \Gamma(P_1, \ldots, P_{C'}) \mapsto \Gamma(Q_1, \ldots, Q_{C'})$ for every function $\Gamma$. Note that $\mathcal{T}$ is invertible. So, recalling that $h_1 = \mathbb{F}^m \to [0, 1]$ is measurable with respect to $\tilde{\mathcal{B}}_1$, it is natural to define $\phi : \mathbb{F}^n \to [0, 1]$ by

$$\phi := \mathcal{T}^{-1} h_1.$$

We first argue that $\phi$ is close to $\mathbf{E}[f|\mathcal{B}_1]$ assuming the underlying parameters are properly chosen.

**Claim 4.3.** *Assume that* $\eta_0(C) \leqslant p^{-dC_{2.17}(\eta_1, r_1, \gamma, C, d)}\gamma$ *and* $r_1(C) \geqslant r_{2.12}(d, p^{-dC}\gamma)$. *Then*

$$\|\mathbf{E}[f|\mathcal{B}_1] - \phi\|_1 \leqslant \delta + \varepsilon/2 + 9\gamma.$$

*Proof.* By Lemma 2.16 and the condition on $r_1$, every function $\phi : \mathbb{F}^n \to [-1,1]$ satisfies

$$\left|\|\mathcal{T}\phi\|_1 - \|\phi\|_1\right| \leqslant \gamma. \tag{2}$$

We can thus write

$$
\begin{aligned}
\|\mathbf{E}[f|\mathcal{B}_1] - \mathcal{T}^{-1}\mathbf{E}[h|\tilde{\mathcal{B}}_1]\|_1 \quad &\leqslant \quad \|\mathcal{T}\mathbf{E}[f|\mathcal{B}_1] - \mathbf{E}[h|\tilde{\mathcal{B}}_1]\|_1 + \gamma \\
&\leqslant \quad \|\mathcal{T}\mathbf{E}[f|\mathcal{B}_1] - \mathcal{T}\mathbf{E}[f|\mathcal{B}_0]\|_1 + \|\mathcal{T}\mathbf{E}[f|\mathcal{B}_0] - \mathbf{E}[Af|\tilde{\mathcal{B}}_0]\|_1 + \\
&\quad + \|\mathbf{E}[Af|\tilde{\mathcal{B}}_0] - \mathbf{E}[Af|\tilde{\mathcal{B}}_1]\|_1 + \|\mathbf{E}[Af|\tilde{\mathcal{B}}_1] - \mathbf{E}[h|\tilde{\mathcal{B}}_1]\|_1 + \gamma. \tag{3}
\end{aligned}
$$

We will show that each of the terms is bounded by $O(\gamma)$, except for the fourth term for which

$$\|\mathbf{E}[Af|\tilde{\mathcal{B}}_1] - \mathbf{E}[h|\tilde{\mathcal{B}}_1]\|_1 \leqslant \|Af - h\|_1 \leqslant \delta + \varepsilon/2.$$

To bound the first term, we apply (2) and Lemma 2.18,

$$\|\mathcal{T}\mathbf{E}[f|\mathcal{B}_1] - \mathcal{T}\mathbf{E}[f|\mathcal{B}_0]\|_1 \leqslant \|\mathbf{E}[f|\mathcal{B}_1] - \mathbf{E}[f|\mathcal{B}_0]\|_1 + \gamma \leqslant \|f_2\|_2 + p^{dC'}\|f_3\|_{U^d} + \gamma \leqslant 3\gamma.$$

To bound the second term, by $(\mathbf{E}_3)$

$$\|\mathcal{T}\mathbf{E}[f|\mathcal{B}_0] - \mathbf{E}[Af|\tilde{\mathcal{B}}_0]\|_1 = \|A\mathbf{E}[f|\mathcal{B}_0] - \mathbf{E}[Af|\tilde{\mathcal{B}}_0]\|_1 \leqslant \|A\mathbf{E}[f|\mathcal{B}_0] - \mathbf{E}[Af|\tilde{\mathcal{B}}_0]\|_\infty \leqslant \gamma.$$

To bound the third term, by Lemma 2.18 and $(\mathbf{E}_2)$,

$$\|\mathbf{E}[Af|\tilde{\mathcal{B}}_0] - \mathbf{E}[Af|\tilde{\mathcal{B}}_1]\|_1 \leqslant \|Af_2\|_2 + p^{dC'}\|Af_3\|_{U^d} \leqslant 2\gamma + 2\gamma \leqslant 4\gamma.$$

$\square$

The function $\phi$ is defined over $\mathbb{F}^n$ and has the same structure as $h_1$, however we only guaranteed that it is close to $\mathbf{E}[f|\mathcal{B}_1]$. This cannot be avoided since $\phi$ is $\mathcal{B}_1$-measurable, and thus it cannot approximate $f$ inside the atoms. The next step is to define a function whose average on atoms is $\phi$, and that simultaneously has a small distance from $f$. It will be obtained by perturbing $f$ inside each atom in order to obtain the right average while making the changes minimal. It will be convenient to first define such a function $\psi$ mapping $\mathbb{F}^n$ to $[0,1]$. Later we will use it to sample $g : \mathbb{F}^n \to \{0,1\}$ with the required properties.

Define a function $\psi : \mathbb{F}^n \to [0,1]$ as follows. Fix $x \in \mathbb{F}^n$ and let $b_1 = \mathcal{B}_1(x)$ be its corresponding atom in $\mathcal{B}_1$. Let $\alpha = \mathbf{E}[f|\mathcal{B}_1](b_1)$ be the average of $f$ over the atom and $\beta = \phi(b_1)$ be the value $\phi$ attains on the atom. We set

$$
\psi(x) = \left\{
\begin{array}{ll}
\frac{\beta - \alpha}{1 - \alpha} & \text{If } \alpha \leqslant \beta \text{ and } f(x) = 0 \\
1 & \text{If } \alpha \leqslant \beta \text{ and } f(x) = 1 \\
0 & \text{If } \alpha > \beta \text{ and } f(x) = 0 \\
\frac{\beta}{\alpha} & \text{If } \alpha > \beta \text{ and } f(x) = 1
\end{array}
\right.
$$

The following claim summarizes the properties of $\psi$, assuming the underlying parameters are properly chosen.

**Claim 4.4.** *Assume that $\eta_0(C) \leqslant p^{-dC_{2.17}(\eta_1, r_1, \gamma, C, d)}\gamma$. Then*

(i) $\mathbf{E}[\psi|\mathcal{B}_1] = \phi$.

*(ii)* $\|f - \psi\|_1 = \|\mathbf{E}[f|\mathcal{B}_1] - \phi\|_1$.

*(iii)* $\|\psi - \phi\|_{U^d} \leqslant \gamma + 3\gamma^{1/2^d}$.

*Proof.* Items *(i)*,*(ii)* follow immediately from the definition of $\psi$. To establish *(iii)*, decompose $\psi - \phi$ to the atoms of $\mathcal{B}_1$,

$$\psi(x) - \phi(x) = \sum_{b_1 \in \mathcal{B}_1} (\psi(x) - \phi(b_1)) \, 1_{\mathcal{B}_1(x) = b_1}.$$

Fix an atom $b_1$ of $\mathcal{B}_1$ and set $\alpha = \mathbf{E}[f|\mathcal{B}_1](b_1)$ and $\beta = \phi(b_1)$. If $\alpha \leqslant \beta$, then

$$(\psi(x) - \beta) \, 1_{\mathcal{B}_1(x) = b_1} = \left( f(x) + \frac{\beta - \alpha}{1 - \alpha}(1 - f(x)) - \beta \right) 1_{\mathcal{B}_1(x) = b_1} = \frac{1 - \beta}{1 - \alpha}(f(x) - \alpha) 1_{\mathcal{B}_1(x) = b_1}.$$

If $\alpha > \beta$ then

$$(\psi(x) - \beta) \, 1_{\mathcal{B}_1(x) = b_1} = \left( \frac{\beta}{\alpha} f(x) - \beta \right) 1_{\mathcal{B}_1(x) = b_1} = \frac{\beta}{\alpha}(f(x) - \alpha) 1_{\mathcal{B}_1(x) = b_1}.$$

Recall that $f = f_1 + f_2 + f_3$ with $f_1 = \mathbf{E}[f|\mathcal{B}_0], \|f\|_2 \leqslant \gamma, \|f_3\|_{U^d} \leqslant \eta_0(|\mathcal{B}_0|)$. Define $f_i' = f_i - \mathbf{E}[f_i|\mathcal{B}_1]$. Note that $f(x) - \mathbf{E}[f|\mathcal{B}_1](x) = f_1'(x) + f_2'(x) + f_3'(x)$. Furthermore, $f_1' = 0$ since $\mathcal{B}_1$ is a refinement of $\mathcal{B}_0$. Define $\xi(x) = (1 - \beta)/(1 - \alpha)$ if $x$ belongs to an atom $b_1$ with $\alpha \leqslant \beta$, and $\xi(x) = \beta/\alpha$ otherwise. We thus get that

$$\psi(x) - \phi(x) = \sum_{b \in \mathcal{B}_1} \xi(x)(f_2'(x) + f_3'(x)) 1_{\mathcal{B}_1(x) = b_1} = \xi(x)(f_2'(x) + f_3'(x)).$$

We now bound $\|\psi - \phi\|_{U^d}$. Now, by Claim 2.21 and the fact that $\|\xi\|_\infty \leqslant 1$,

$$\|\xi \cdot f_2'\|_{U^d}^{2^d} \leqslant \|\xi \cdot f_2'\|_1 \leqslant \|f_2\|_1 + \|\mathbf{E}[f_2|\mathcal{B}_1]\|_1 \leqslant 2\|f_2\|_1 \leqslant 2\|f_2\|_2 \leqslant 2\gamma.$$

By Claim 2.20 , we have

$$\|\xi \cdot \mathbf{E}[f_3|\mathcal{B}_1]\|_{U^d}^{2^d} \leqslant \|\xi \cdot \mathbf{E}[f_3|\mathcal{B}_1]\|_1 \leqslant \|\mathbf{E}[f_3|\mathcal{B}_1]\|_1 \leqslant p^{d|\mathcal{B}_1|} \|f_3\|_{U^d} \leqslant p^{d|\mathcal{B}_1|} \eta_0(|\mathcal{B}_0|) \leqslant \gamma.$$

Thus recalling that $\|\mathcal{B}_1\|$ is an upper-bound on the number of atoms of $\mathcal{B}_1$, by Claim 2.19 we have

$$
\begin{aligned}
\|\xi \cdot f_3'\|_{U^d} &\leqslant \|\xi \cdot E[f_3|\mathcal{B}_1]\|_{U^d} + \|\xi \cdot f_3\|_{U^d} \leqslant \gamma^{1/2^d} + \sum_{b_1 \in \mathcal{B}_1} \|f_3(x) 1_{\mathcal{B}_1(x) = b_1}\|_{U^d} \\
&\leqslant \gamma^{1/2^d} + \|\mathcal{B}_1\| \|f_3\|_{U^d} \leqslant \gamma^{1/2^d} + \|\mathcal{B}_1\| \eta_0(|\mathcal{B}_0|) \leqslant \gamma^{1/2^d} + \gamma.
\end{aligned}
$$

$\square$

We are ready to conclude the proof. Let $g : \mathbb{F}^n \to \{0, 1\}$ be sampled with $\mathbf{Pr}[g(x) = 1] = \psi(x)$ independently for all $x \in \mathbb{F}^n$. The following simple claim states that with high probability, $g$ behaves like $\psi$.

**Claim 4.5.** *If $n$ is large enough then with probability at least 99% over the choice of $g$,*

- $\|f - g\|_1 \leqslant \|f - \psi\|_1 + \gamma$.

- $\|g - \psi\|_{U^d} \leqslant \gamma$.

*Proof.* Both items are simple. The first item holds since $\mathbf{E}_g \|f - g\|_1 = \|f - \psi\|_1$, and $\mathbf{Var}\|f - g\|_1 \leqslant p^{-n}$. Hence by Chebyshev's inequality, $\|f - g\|_1 = \|f - \psi\|_1 + \gamma$ with probability $1 - o_n(1)$. For the second item,

$$\mathbf{E}_g \|g - \psi\|_{U^d}^{2^d} = \mathbf{E}_{x, y_1, \ldots, y_d \in \mathbb{F}^n} \prod_{I \subseteq [d]} \mathbf{E}_g [g(x + \sum_{i \in I} y_i) - \psi(x + \sum_{i \in I} y_i)].$$

The inner expectation is zero if all the sums $x + \sum_{i \in I} y_i$ are distinct. In particular this holds if $y_1, \ldots, y_d$ are linearly independent, which happens with probability at least $1 - p^{d-n}$. Hence

$$\mathbf{E}_g \|g - \psi\|_{U^d}^{2^d} \leqslant p^{d-n} = o_n(1),$$

and so by choosing $n$ large enough we get that with high probability $\|g - \psi\|_{U^d} \leqslant \gamma$. $\qquad \square$

Fix $g : \mathbb{F}^n \to \{0, 1\}$ satisfying Claim 4.5. Let us summarize the facts that we know so far. By Claim 4.3, Claim 4.4 and Claim 4.5 we know that $f$ and $g$ are close. By choosing $\gamma \leqslant \varepsilon/40$ we get that

$$\|f - g\|_1 \leqslant \delta + \varepsilon/2 + 10\gamma \leqslant \delta + 3\varepsilon/4.$$

We next argue that $g$ is very close to $\mathcal{P}$. Recall that $h \in \mathcal{P}$ and $h_1 = \mathbf{E}[h | \tilde{\mathcal{B}}_1]$. We have that

$$\|h - h_1\|_{U^d} \leqslant \|h_2\|_{U^d} + \|h_3\|_{U^d} \leqslant \|h_2\|_2^{1/2^d} + \|h_3\|_{U^d} \leqslant \gamma^{1/2^d} + \eta_1(C'),$$

where we applied Claim 2.21 and the assumptions on $h_2, h_3$. By choosing $\gamma \leqslant (\rho/2)^{2^d}$ and $\eta_1(C') \leqslant \rho/2$, we get that $\|h - h_1\|_{U^d} \leqslant \rho$. Hence by Lemma 3.3 the statistical distance between $\mu_{h,k}$ and $\mu_{h_1,k}$ is at most $1/100$. Next, by definition $h_1 = \Gamma(Q_1, \ldots, Q_{C'})$ and $\phi = \Gamma(P_1, \ldots, P_{C'})$, where both $Q_1, \ldots, Q_C$ and $P_1, \ldots, P_C$ are $r_1$-regular. Making sure that $r_1(C') \geqslant r_{3.5}(C)$ guarantees that conditions of Lemma 3.5 are satisfied, and hence $\mu_{h_1,k}$ and $\mu_{\phi,k}$ also have statistical distance at most $1/100$. Next, by Claim 4.4 and Claim 4.5 we have

$$\|g - \phi\|_{U^d} \leqslant 2\gamma + 3\gamma^{1/2^d}.$$

Choosing $\gamma$ so that $2\gamma + 3\gamma^{1/2^d} \leqslant \rho$, we get that also the statistical distance between $\mu_{g,k}$ and $\mu_{\phi,k}$ is at most $1/100$. Hence, combining all these together gives that

The statistical distance between $\mu_{g,k}$ and $\mu_{h,k}$ is at most $3/100$.

Applying Corollary 3.4 we establish that

$$\text{dist}(g, \mathcal{P}) \leqslant \varepsilon/8.$$

Hence we reached a contradiction, since $f$ is $(\delta + 3\varepsilon/4)$-close to a function $g$ which is $(\varepsilon/8)$-close to $\mathcal{P}$, that is $f$ is $(\delta + 7\varepsilon/8)$-close to $\mathcal{P}$, contradicting our initial assumption that $f$ is $(\delta + \varepsilon)$-far from $\mathcal{P}$.

# References

[AKK⁺05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Transactions on Information Theory*, 51(11):4032–4039, 2005. 2

[BCSX11] Arnab Bhattacharyya, Victor Chen, Madhu Sudan, and Ning Xie. Testing linear-invariant non-linear properties. *Theory of Computing*, 7(1):75–99, 2011. 2

[BFH⁺13] Arnab Bhattacharyya, Eldar Fische, Hamed Hatami, , Pooya Hatami, and Shachar Lovett. Every locally characterized afine-invariant property is testabl. In *Proc. 45th Annual ACM Symposium on the Theory of Computing*, page to appear, 2013. Full version at arXiv:1212.3849. 2, 3, 6, 7, 10, 12, 15

[BFL91] László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991. 2

[BFL12] Arnab Bhattacharyya, Eldar Fischer, and Shachar Lovett. Testing low complexity affine-invariant properties. Technical report, October 2012. http://arxiv.org/abs/1201.0330v2 (to appear in SODA '13). 2, 11

[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. 2

[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. 2

[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. 1

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

[FN07] Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM J. Comput.*, 37(2):482–501 (electronic), 2007. 1, 3, 6

[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. 2

[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. 2

[GT10]    Ben Green and Terence Tao.   Linear equations in primes.   *Annals of Mathematics*, 171:1753–1850, 2010. 8

[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. 7

[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. 10

[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. 2

[KSV12]   Daniel Král, Oriol Serra, and Lluís Vena.  A removal lemma for systems of linear equations over finite fields. *Israel Journal of Mathematics*, pages 1–15, 2012. Preprint available at http://arxiv.org/abs/0809.1846. 2

[RS93]    Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials and their applications to program testing. Technical report, Ithaca, NY, USA, 1993. 1

[RS96]    Ronitt Rubinfeld and Madhu Sudan.  Robust characterizations of polynomials with applications to program testing. *SIAM J. on Comput.*, 25:252–271, 1996. 2

[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. 2

[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. 8