Electronic Colloquium on Computational Complexity, Report No. 184 (2012)

Every locally characterized affine-invariant property is testable

Arnab Bhattacharyya* Eldar Fischer† Hamed Hatami‡ Pooya Hatami§
Shachar Lovett¶

December 26, 2012

Abstract

Let $\mathbb{F} = \mathbb{F}_p$ for any fixed prime $p \geqslant 2$. An affine-invariant property is a property of functions on $\mathbb{F}^n$ that is closed under taking affine transformations of the domain. We prove that all affine-invariant properties that have local characterizations are testable. In fact, we give a proximity-oblivious test for any such property $\mathcal{P}$, meaning that given an input function f, we make a constant number of queries to f, always accept if f satisfies $\mathcal{P}$, and otherwise reject with probability larger than a positive number that depends only on the distance between f and $\mathcal{P}$. More generally, we show that any affine-invariant property that is closed under taking restrictions to subspaces and has bounded complexity is testable.

We also prove that any property that can be described as the property of being decomposable into a known structure of low-degree polynomials is locally characterized and is, hence, testable. For example, whether a function is a product of two degree-d polynomials, whether a function splits into a product of d linear polynomials, and whether a function has low rank are all examples of degree-structural properties and are therefore locally characterized.

Our results use a new Gowers inverse theorem by Tao and Ziegler for low characteristic fields that decomposes any polynomial with large Gowers norm into a function of a small number of low-degree *non-classical polynomials*. We establish a new equidistribution result for high rank non-classical polynomials that drives the proofs of both the testability results and the local characterization of degree-structural properties.

1 Introduction

The field of property testing, as initiated by [BLR93, BFL91] and defined formally by [RS96, GGR98], is the study of algorithms that query their input a very small number of times and with high probability decide correctly whether their input satisfies a given property or is "far" from satisfying that property. A property is called *testable*, or sometimes *strongly testable* or *locally testable*, if the number of queries can be made independent of the size of the object without affecting the correctness probability. Perhaps surprisingly, it has been found that a large number of natural properties satisfy this strong requirement; see e.g. the surveys [Fis04, Rub06, Ron09, Sud10] for a general overview.

*DIMACS & Rutgers University. Email: arnabb@dimacs.rutgers.edu.
†Technion. Email: eldar@cs.technion.ac.il.
‡McGill University. Email: hatami@cs.mcgill.ca.
§University of Chicago. Email: pooya@cs.uchicago.edu.
¶Institute for Advanced Study and University of California, San Diego. Email: slovett@math.ias.edu.

ISSN 1433-8092

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 $\mathbb{F} = \mathbb{F}_p$. We consider properties of functions $f : \mathbb{F}^n \to \{1, \ldots, R\}$. Our main result shows that any such property that is invariant with respect to affine transformations on $\mathbb{F}^n$ and that is locally characterized is testable. Furthermore, we show that a large class of natural algebraic properties whose query complexity had not been previously studied are locally characterized affine-invariant properties and are, hence, testable. Our results constitute an exact characterization of proximity-obliviously testable properties, the most common notion of testability considered for algebraic properties. In the rest of this section, we motivate and describe our results in more detail.

1.1 Testability and Invariances

Let $[R]$ denote the set $\{1, \ldots, R\}$. Given a property $\mathcal{P}$ of functions in $\{\mathbb{F}^n \to [R] \mid n \in \mathbb{Z}_{\geqslant 0}\}$, we say that $f : \mathbb{F}^n \to [R]$ is ε-*far* from $\mathcal{P}$ if

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

and we say that it is ε-*close* otherwise.

Definition 1.1 (Testability)**.** *A property $\mathcal{P}$ is said to be* testable *(with one-sided error) if there are functions $q : (0,1) \to \mathbb{Z}_{>0}$, $\delta : (0,1) \to (0,1)$, 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, always accepts if $f \in \mathcal{P}$ and rejects with probability at least $\delta(\varepsilon)$ if f is ε-far from $\mathcal{P}$. If, furthermore, q is a constant function, then $\mathcal{P}$ is said to be* proximity-obliviously testable (PO testable)*.*

As an example of a testable (in fact, PO testable) property, let us recall the famous result by Blum, Luby and Rubinfeld [BLR93] which initiated this whole line of research. 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 if f is linear and rejects with probability $\Omega(\varepsilon)$ if f is ε-far from linear.

Linearity, in addition to being testable, is also an example of a *linear-invariant* property. We say that a property $\mathcal{P} \subseteq \{\mathbb{F}^n \to [R]\}$ is linear-invariant if it is the case that for any $f \in \mathcal{P}$ and for any linear transformation $L : \mathbb{F}^n \to \mathbb{F}^n$, it holds that $f \circ L \in \mathcal{P}$. Similarly, an *affine-invariant* property is closed under composition with affine transformations $A : \mathbb{F}^n \to \mathbb{F}^n$ (an affine transformation A is of the form $L + c$ where L is linear and c is a constant). The property of a function $f : \mathbb{F}^n \to \mathbb{F}$ being affine is testable by a simple reduction to [BLR93], and is itself affine-invariant. Other well-studied examples of affine-invariant (and hence, linear-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. In particular, they asked for necessary and sufficient conditions for the testability of affine-invariant properties.

1.2 Locally Characterized Properties

The result summarized in the title of this paper gives a necessary and sufficient condition for affine-invariant properties to be PO testable. Let us first see why "local characterization" is a necessary condition for PO testability.

For a PO testable property $\mathcal{P}$, if a function f does not satisfy $\mathcal{P}$, then by Definition 1.1, the tester rejects f with positive probability. Since the test always accepts functions with the property, there must be q points $x_1, \ldots, x_q \in \mathbb{F}^n$ that form a witness for non-membership in $\mathcal{P}$. These are the queries that cause the tester to reject. Thus, denoting $\sigma = (f(x_1), \ldots, f(x_q)) \in [R]^q$, we say that $\mathcal{C} = (x_1, x_2, \ldots, x_q; \sigma)$ forms a *q-local constraint* for $\mathcal{P}$. This means that whenever the constraint is violated by a function g, i.e., $(g(x_1), \ldots, g(x_q)) = \sigma$, we know that g is not in $\mathcal{P}$. A property $\mathcal{P}$ is *q-locally characterized* if there exists a collection of q-local constraints $\mathcal{C}_1, \ldots, \mathcal{C}_m$ such that $g \in \mathcal{P}$ if and only if none of the constraints $\mathcal{C}_1, \ldots, \mathcal{C}_m$ are violated. It follows from the above discussion that if $\mathcal{P}$ is PO testable with q queries, then $\mathcal{P}$ is q-locally characterized. We say $\mathcal{P}$ is *locally characterized* if it is q-locally characterized for some constant q.

We now give some examples of locally characterized affine-invariant properties. Consider the property of being affine. It is 4-locally characterized because a function f is affine if and only if $f(x) - f(x+y) - f(x+z) + f(x+y+z) = 0$ for every $x, y, z \in \mathbb{F}^n$. Note that this characterization automatically suggests a 4-query test: pick random $x, y, z \in \mathbb{F}^n$ and check whether the identity holds or not for that choice of x, y, z. More generally, consider the property of being a polynomial of degree at most d, for some fixed integer $d > 0$. The property is known to be PO testable due to independent work of [KR06, JPRZ04], and their test is based upon a $p^{\lceil \frac{d+1}{p-1} \rceil}$-local characterization. Again, the test is simply to pick a random constraint and check if it is violated.

Indeed, for any q-locally characterized property $\mathcal{P}$ defined by constraints $\mathcal{C}_1, \ldots, \mathcal{C}_m$, one can design the following q-query test: choose a constraint $\mathcal{C}_i$ uniformly at random and reject only if the input function violates $\mathcal{C}_i$. Clearly, if the input function f is in $\mathcal{P}$, the test always accepts. The question is the probability with which a function ε-far from $\mathcal{P}$ is rejected. We show that for affine-invariant properties, this test always rejects with probability bounded away from zero for every constant $\varepsilon > 0$.

Theorem 1.2. *Every q-locally characterized affine-invariant property is proximity-obliviously testable with q queries.*

1.3 Subspace Hereditary Properties

Just as a necessary condition for PO testability is local characterization, one can formulate a natural condition that is (almost) necessary for testability in general. In the context of affine-invariant properties, the condition can be succinctly stated as follows:

Definition 1.3 (Subspace hereditary properties)**.** *An affine-invariant property $\mathcal{P}$ is said to be* (affine) subspace hereditary *if for any $f : \mathbb{F}^n \to [R]$ satisfying $\mathcal{P}$, the restriction of f to any affine subspace of $\mathbb{F}^n$ also satisfies $\mathcal{P}$.*

In [BGS10], it is shown that every affine-invariant property testable by a "natural" tester is very "close" to a subspace hereditary property[1]. Thus, if we gloss over some technicalities, subspace hereditariness is a necessary condition for testability. In the opposite direction, [BGS10] conjectures the following:

[1]We omit the technical definitions of "natural" and "close", since they are unimportant here. Informally, the behavior of a "natural" tester is independent of the size of the domain and "close" means that the property deviates from an actual affine subspace hereditary property on functions over a finite domain. See [BGS10] for details, or [AS08a] for the analogous definitions in a graph-theoretic context.

Conjecture 1.4 ([BGS10])**.** *Every subspace hereditary property is testable.*

Resolving Conjecture 1.4 would yield a combinatorial *characterization* of the (natural) one-sided testable affine-invariant properties, similar to the characterization for testable dense graph properties [AS08a]. Although we are not yet able to confirm or refute the full Conjecture 1.4, we can show testability if we make an additional assumption of "bounded complexity", defined formally in Section 1.5.2.

Theorem 1.5 (Informal)**.** *Every subspace hereditary property of "bounded complexity" is testable.*

We will formally define the notion of complexity later on in Section 1.5.2, but for now, it suffices to know that it is an integer that we will associate with each property (independent of n). Also, q-locally characterized properties are of complexity at most q. All natural affine-invariant properties that we know of have bounded complexity (in fact, most are locally characterized). So, the subspace hereditary properties not covered by Theorem 1.5 seem to be mainly of theoretical interest.

1.4 Degree-structural Properties

The conditions required in Theorem 1.2 and Theorem 1.5 are very general, and so, we expect that they are satisfied by many interesting algebraic properties. This, in fact, turns out to be the case. We show that a class of properties that we call *degree-structural* are all locally characterized and are, hence, testable by Theorem 1.2. We give the definition below in Definition 1.6. First let us list some examples of degree-structural properties. Let d be a fixed positive integer. Each of the following definitions defines a degree-structural property.

Degree $\leqslant d$: The degree of the function $F : \mathbb{F}^n \to \mathbb{F}$ as a polynomial is at most d;

Splitting: A function $F : \mathbb{F}^n \to \mathbb{F}$ *splits* if it can be written as a product of at most d linear functions;

Factorization: A function $F : \mathbb{F}^n \to \mathbb{F}$ *factors* if $F = GH$ for polynomials $G, H : \mathbb{F}^n \to \mathbb{F}$ such that $\deg(G) \leqslant d-1$ and $\deg(H) \leqslant d-1$;

Sum of two products: A function $F : \mathbb{F}^n \to \mathbb{F}$ is a *sum of two products* if there are polynomials G_1, G_2, G_3, G_4 such that $F = G_1G_2+G_3G_4$ and $\deg(G_i) \leqslant d-1$ for $i \in \{1, 2, 3, 4\}$;

Having square root: A function $F : \mathbb{F}^n \to \mathbb{F}$ *has a square root* if $F = G^2$ for a polynomial G with $\deg(G) \leqslant d/2$;

Low d-rank: for a fixed integer $r > 0$, a function $F : \mathbb{F}^n \to \mathbb{F}$ *has d-rank at most r* if there exist polynomials $G_1, \ldots, G_r : \mathbb{F}^n \to \mathbb{F}$ of degree $\leqslant d-1$ and a function $\Gamma : \mathbb{F}^r \to \mathbb{F}$ such that $F = \Gamma(G_1, \ldots, G_r)$.

In fact, roughly speaking, any property that can be described as the property of decomposing into a known structure of low-degree polynomials is degree-structural.

Definition 1.6 (Degree-structural property)**.** *Given an integer $c > 0$, a vector of non-negative integers $\mathbf{d} = (d_1, \ldots, d_c) \in \mathbb{Z}^c_{\geqslant 0}$, and a function $\Gamma : \mathbb{F}^c \to \mathbb{F}$, define the* $(c, \mathbf{d}, \Gamma)$-structured property *to be the collection of functions $F : \mathbb{F}^n \to \mathbb{F}$ for which there exist polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{F}$ satisfying $F(x) = \Gamma(P_1(x), \ldots, P_c(x))$ for all $x \in \mathbb{F}^n$ and* $\deg(P_i) \leqslant d_i$ *for all $i \in [c]$.*

We say a property $\mathcal{P}$ is degree-structural *if there exist integers $\sigma, \Delta > 0$ and a set of tuples $S \subset \{(c, \mathbf{d}, \Gamma) \mid c \in [\sigma], \mathbf{d} \in [0, \Delta]^c, \Gamma : \mathbb{F}^c \to \mathbb{F}\}$, such that a function $F : \mathbb{F}^n \to \mathbb{F}$ satisfies $\mathcal{P}$ if and only if F is $(c, \mathbf{d}, \Gamma)$-structured for some $(c, \mathbf{d}, \Gamma) \in S$. We call σ the* scope *and Δ the* max-degree *of the degree-structural property $\mathcal{P}$.*

It is straightforward to see that the examples above satisfy this definition. Our main result for degree-structural properties is the following:

Theorem 1.7. *Every degree-structural property with bounded scope and max-degree is a locally characterized affine-invariant property.*

Combining Theorem 1.7 with Theorem 1.2 implies PO testability for all degree-structural properties.

1.5 Formal version of the Main Result

In this section, we describe our main result, Theorem 1.5, rigorously. Theorem 1.2 follows as a corollary. We first need to set up some notions. Just as a locally characterized property can be described by a list of constraints, subspace hereditary properties can also be described similarly, but here, the size of the list can be infinite. For affine-invariant properties, we can represent the constraints in a very special form, as "induced affine constraints". We first describe these, then define the notion of complexity, and finally state the theorem.

1.5.1 Affine constraints

A *linear form on k variables* is a vector $L = (w_1, w_2, \ldots, w_k) \in \mathbb{F}^k$ that is interpreted as a function from $(\mathbb{F}^n)^k$ to $\mathbb{F}^n$ via the map $(x_1, \ldots, x_k) \mapsto w_1x_1 + w_2x_2 + \cdots + w_kx_k$. A linear form $L = (w_1, w_2, \ldots, w_k)$ is said to be *affine* if $w_1 = 1$. From now, linear forms will always be assumed to be affine.

We specify a partial order $\preceq$ among affine forms. We say $(w_1, \ldots, w_k) \preceq (w'_1, \ldots, w'_k)$ if $|w_i| \leqslant |w'_i|$ for all $i \in [k]$, where $|\cdot|$ is the obvious map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\}$. An affine constraint is a collection of affine forms, with the added technical restriction of being downward-closed with respect to $\preceq$. For future references we state this as the following definition.

Definition 1.8 (Affine constraints)**.** *An* affine constraint of size m on k variables *is a tuple $A = (L_1, \ldots, L_m)$ of m affine forms $L_1, \ldots, L_m$ over $\mathbb{F}$ on k variables, where:*

(i) $L_1(x_1, \ldots, x_k) = x_1$*;*
(ii) If a belongs to A and $a' \preceq a$, then a' also belongs to A.

Any subspace hereditary property can be described using affine constraints and forbidden patterns, in the following way.

Definition 1.9 (Properties defined by induced affine constraints)**.**

> *An* induced affine constraint of size m on ℓ variables *is a pair (A, σ) where A is an affine constraint of size m on ℓ variables and $\sigma \in [R]^m$.*
>
> *Given such an induced affine constraint (A, σ), a function $f : \mathbb{F}^n \to [R]$ is said to be (A, σ)-*free *if there exist no $x_1, \ldots, x_\ell \in \mathbb{F}^n$ such that $(f(L_1(x_1, \ldots, x_\ell)), \ldots, f(L_m(x_1, \ldots, x_\ell))) = \sigma$. On the other hand, if such $x_1, \ldots, x_\ell$ exist, we say that f* induces (A, σ) at $x_1, \ldots, x_\ell$*.*

Given a (possibly infinite) collection $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\}$ of induced affine constraints, a function $f : \mathbb{F}^n \to [R]$ is said to be $\mathcal{A}$-free *if it is (A^i, σ^i)-free for every $i \geqslant 1$.*

As an example consider the property of having degree at most 1 as a polynomial, for function $F : \mathbb{F}^n \to \mathbb{F}$. It is easy to see that F satisfies this property if and only if $F(x_1) - F(x_1+x_2) - F(x_1+x_3) + F(x_1+x_2+x_3) = 0$ for all $x_1, x_2, x_3 \in \mathbb{F}$. Consequently the property can be defined by the set of induced affine constraints that forbid any values for $F(x_1), F(x_1+x_2), F(x_1+x_3), F(x_1+x_2+x_3)$ that do not satisfy the identity $F(x_1) - F(x_1+x_2) - F(x_1+x_3) + F(x_1+x_2+x_3) = 0$.

The connection between affine subspace hereditariness and affine constraints is given by the following simple observation.

Observation 1.10. *An affine-invariant property $\mathcal{P}$ is subspace hereditary if and only if it is equivalent to the property of $\mathcal{A}$-freeness for some fixed collection $\mathcal{A}$ of induced affine constraints.*

Proof. Given an affine invariant property $\mathcal{P}$, a simple (though inefficient) way to obtain the set $\mathcal{A}$ is to let it be the following: For every n and a function $f : \mathbb{F}^n \to [R]$ that is not in $\mathcal{P}$, we include in $\mathcal{A}$ the constraint (A_f, σ_f), where A_f is indexed by members of $\mathbb{F}^n$ and contains $\{L_z(X_1, \ldots, X_{n+1}) = X_1 + \sum_{i=1}^n z_i X_{i+1} : z = (z_1, \ldots, z_n) \in \mathbb{F}^n\}$, and σ_f is just set to f.

Setting $X_1 = 0$ and X_i to the ith standard vector e_i for every $i \in [n]$ shows that f is not (A_f, σ_f)-free. Hence the property defined by $\mathcal{A}$ is contained in $\mathcal{P}$. The containment in the other direction follows from $\mathcal{P}$ being affine-invariant and hereditary.

The other direction of the observation is trivial. $\square$

1.5.2 Complexity of linear forms

Green and Tao, in their seminal work on arithmetic progressions in primes, introduced the following notion of complexity of linear forms.

Definition 1.11 (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 d such that the following holds. For every $i \in [m]$, we can partition $\{L_j\}_{j\in[m]\setminus\{i\}}$ into $d+1$ subsets such that L_i does not belong to the linear span of any subset.*

If $\mathcal{L} = \{L_1, \ldots, L_m\}$ contains two linear forms that are multiples of each other (that is $L_i = \lambda L_j$ for $i \neq j$ and $\lambda \in \mathbb{F}$), then the complexity of $\mathcal{L}$ is infinity. Otherwise its complexity is at most $|\mathcal{L}|-2$. Note that sets of affine linear forms are always of finite complexity.

Given a collection $\mathcal{A} = \big\{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i)\big\}$ of induced affine constraints, we say that $\mathcal{A}$ is of *complexity* $\leqslant d$ if for each i, the collection of affine forms A^i is of complexity $\leqslant d$ according to Definition 1.11.

1.5.3 Statement of the main result

Theorem 1.12 (Main theorem)**.** *For any integer $d > 0$ and (possibly infinite) fixed collection $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\}$ of induced affine constraints, each of complexity $\leqslant d$, there are functions $q_{\mathcal{A}} : (0,1) \to \mathbb{Z}^+$, $\delta_{\mathcal{A}} : (0,1) \to (0,1)$ and a tester T which, for every $\varepsilon > 0$, makes $q_{\mathcal{A}}(\varepsilon)$ queries, accepts $\mathcal{A}$-free functions and rejects functions ε-far from $\mathcal{A}$-free with probability at least $\delta_{\mathcal{A}}(\varepsilon)$. Moreover, $q_{\mathcal{A}}$ is a constant if $\mathcal{A}$ is of finite size.*

We do not have any explicit bounds on $\delta_{\mathcal{A}}$ because the analysis depends on previous work based on ergodic theory. It would of course be interesting to have explicit bounds for some of the properties described in 1.2.

Let us finally note that Theorem 1.2 is quite nontrivial even if $\mathcal{A}$ consists only of a single induced affine constraint of complexity greater than 1. Indeed, previously it was not known how to show testability in this case. A more detailed account of previous work is given in Section 1.7.

1.6 Overview of Proofs

1.6.1 Testability

Let us now give an overview of our proof of Theorem 1.12. For simplicity of exposition, assume for now that $\mathcal{A}$ consists only of a single induced affine constraint (A, σ) where A is the tuple of affine linear forms $(L_1, \dots, L_m)$, each over ℓ variables, and $\sigma \in [R]^m$. Let d be the complexity of the constraint. For $i \in [R]$, let $f^{(i)} : \mathbb{F}^n \to \{0,1\}$ be the indicator function for the set $f^{-1}(\{i\})$. Our goal will be to show that, when f is ε-far from (A, σ)-free, then:

$$\mathop{\mathbf{E}}_{x_1,\dots,x_\ell}\left[f^{(\sigma_1)}(L_1(x_1,\dots,x_\ell)) \cdot f^{(\sigma_2)}(L_2(x_1,\dots,x_\ell)) \cdot \dots \cdot f^{(\sigma_m)}(L_m(x_1,\dots,x_\ell))\right] \geqslant \delta(\varepsilon), \qquad (1)$$

for some function $\delta : (0,1) \to (0,1)$. If Eq. (1) is true, then a valid test would be to simply pick ℓ points uniformly at random and reject only if $f(L_1(x_1,\dots,x_\ell)) = \sigma_1, \dots, f(L_m(x_1,\dots,x_\ell)) = \sigma_m$.

Studying averages of products, as in (1), has been crucial to a wide range of problems in additive combinatorics and analytic number theory. Szemerédi's theorem about the density of arithmetic progressions in subsets of the integers is a classic example. Szemerédi's work [Sze75] arguably initiated such questions in additive combinatorics, but the major development which led to a more systematic understanding of these averages was Gowers' definition of a new notion of uniformity in a Fourier-analytic proof for Szemerédi's theorem [Gow01]. In particular, Gowers introduced the *Gowers norm* $\|\cdot\|_{U^{d+1}}$, which allows us to say the following about (1): 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

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

This observation leads to the study of *decomposition theorems*, that express an arbitrary function f as a sum of two functions g and h, where g is "structured" in a sense we describe soon and h has low $(d+1)$-th order Gowers norm. Decomposing each $f^{(\sigma_i)}$ in this way into $g^{(\sigma_i)}$ and $h^{(\sigma_i)}$, substituting into Eq. (1) and expanding, we get inside the expectation a sum of 2^m terms. All these terms, except one, contain some $h^{(\sigma_i)}$ in the product and can be bounded by the above mentioned property of the Gowers norm. In fact, we can make the Gowers norm small enough that we can effectively discard all these terms inside the expectation. The term remaining is the product of the "structured" functions,

$$\mathop{\mathbf{E}}_{x_1,\dots,x_\ell}\left[g^{(\sigma_1)}(L_1(x_1,\dots,x_\ell))g^{(\sigma_2)}(L_2(x_1,\dots,x_\ell))\dots g^{(\sigma_m)}(L_m(x_1,\dots,x_\ell))\right], \qquad (2)$$

and the goal is to lower-bound this expectation.

To describe the structure of g, let us go over how the decomposition into g and h is obtained. Given an arbitrary function f, if $\|f\|_{U^{d+1}}$ is small, then we are already done. Otherwise, we repeatedly apply 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. But there is a catch in this nice-looking structural theorem! If $p > d$, $P_1, \ldots, P_C$ are indeed "classical" degree-d $\mathbb{F}$-valued polynomials over $\mathbb{F}^n$. However, in our setting, where p is a fixed small constant, such a decomposition may no longer hold. Indeed, [GT09, LMS08] proved that if f equals the symmetric degree-4 polynomial 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 by "non-classical" polynomials. These polynomials may take values in $\mathbb{Z}_{p^k}$ for some integer k. More precisely, a non-classical polynomial of degree d is a function P from $\mathbb{F}^n$ to $\mathbb{Z}_{p^k}$ such that the $(d+1)$-th order derivative of P is zero. The integer $k-1$ is called the "depth" of P. Classical polynomials have depth 0.

We use the result of [TZ11] to obtain non-classical polynomials $P_1, \ldots, P_C$ of degree $\leqslant d$ such that each $g^{(\sigma_i)} = \Gamma_i(P_1, \ldots, P_C)$ for some function Γ_i. We return now to the goal of lower-bounding Eq. (2). By a sequence of steps already introduced in [BGS10] and [BFL12] (inspired by similar techniques on graph property testing in [AFKS00, AS08b, AS08a]), we reduce to the problem of lower-bounding the probability

$$\mathop{\mathbf{Pr}}_{x_1,\ldots,x_\ell}\left[\bigwedge_{i\in[C],j\in[m]} P_i(L_j(x_1,\ldots,x_\ell)) = b_{i,j}\right]$$

where each $b_{i,j}$ is an arbitrary fixed element in the range of P_i. That is, we want to show that the polynomials $\{P_i \circ L_j \mid i \in [C], j \in [m]\}$ behave like independent random variables distributed nearly uniformly on their range. Of course, this cannot be completely true. For example, if P_i is linear, $P_i(x_1+x_2+x_3)-P_i(x_1+x_2)-P_i(x_1+x_3)+P_i(x_1)$ is identically zero and so, $\{P_i(x_1+x_2+x_3), P_i(x_1+x_2), P_i(x_1+x_3), P_i(x_1)\}$ are correlated. Moreover, because the polynomials are non-classical, pP is a non-constant polynomial of lower degree than P and satisfies other identities not satisfied by P itself. What we show is that if the collection of polynomials $P_1, \ldots, P_C$ is of high *rank*, then besides correlations which are forced by the degree and depth of the polynomials, there are no other dependencies. This equidistribution result for high rank non-classical polynomials is the technical crux of our work. Our proof technique is very different from the similar equidistribution claim in [HL11a, HL11b] for classical polynomials, since that proof uses the monomial structure of classical polynomials.

Let us briefly describe what we mean by a high rank collection of non-classical polynomials $P_1, \ldots, P_C$. We say that the rank of the collection is r if there exist integers $\lambda_1, \ldots, \lambda_C$ such that $\lambda_1 P_1, \ldots, \lambda_C P_C$ are not all identically zero but $\sum_{i=1}^{C} \lambda_i P_i = \Gamma(Q_1, \ldots, Q_r)$ for some r polynomials $Q_1, \ldots, Q_r$ each of degree $< \max_i \deg(\lambda_i P_i)$ and some function Γ. So, if the rank of a collection of polynomials is high, that means that no linear combination of the polynomials, unless it is trivially zero, has an explanation in terms of a small number of lower degree polynomials. Intuitively, a high rank collection of degree d polynomials is like a random or generic collection of degree d polynomials. It does not have unexpected low-degree correlations, and it is robust to common operations such as taking projections or multiplying by constants or taking derivatives.

This finishes the high-level overview of the proof, although there are some additional issues that we have swept under the rug. One problem is that the decomposition theorem actually decomposes a given function f to a sum of three functions f_1, f_2, f_3, not into two functions g and h as in the description above. The functions f_1 and f_2 correspond to g and h, respectively, and f_3 is an additional function that has low L^2-norm. Now, the closeness to equidistribution of the non-classical polynomials $P_1, \ldots, P_C$ describing f_1 and the smallness of the Gowers norm for f_2 can be made arbitrarily small as a function of C and are thus, essentially negligible for the purposes of the proof. On the other hand, the bound on the L^2-norm for f_3 is only moderately small and cannot be made to decrease as a function of the complexity of the decomposition. To get around this issue, we use a sequence of two decompositions, and make the norm of f_3 decrease as a function of the size of the first decomposition. We hope that these iterated decomposition theorems (proved in a prequel [BFL12] to this paper) are of independent interest.

1.6.2 Degree-Structural Properties

Next, we give an overview of our proof of Theorem 1.7. For the sake of concreteness, let us focus on a particular degree-structural property, say, the property $\mathcal{P}$ of having a square root, as defined in Section 1.4. To show that $\mathcal{P}$ is locally characterized, we find a constant $K = K(\mathcal{P})$ such that if a function $F : \mathbb{F}^n \to \mathbb{F}$ does not have a square root, then there must exist a subspace H of dimension K such that F restricted to H also does not have a square root.

So, suppose we are given a function $F : \mathbb{F}^n \to \mathbb{F}$ such that $n \geqslant K$ and every hyperplane restriction has a square root of degree $\leqslant d/2$. For large enough constant K, this automatically implies $\deg(F) \leqslant d$. We first *regularize* f, meaning that we find polynomials $P_1, \ldots, P_C$ of degree $\leqslant d$ such that $P_1, \ldots, P_C$ are of high rank and $F = \Gamma(P_1, \ldots, P_C)$ for some function Γ. Note that here, just as in the proof of the testability result, we need to allow $P_1, \ldots, P_C$ to be non-classical polynomials. Now, for some i such that $F|_{x_i=0}$ has a square root, let $P'_1, \ldots, P'_C$ be the restrictions of $P_1, \ldots, P_C$ to $x_i = 0$. So, $\Gamma(P'_1, \ldots, P'_C) = G^2$ for some polynomial G. The polynomials $P'_1, \ldots, P'_C$ can be shown to be of high rank also. This implies that we can extend the collection of polynomials $P'_1, \ldots, P'_C$ to $P'_1, \ldots, P'_C, Q_1, \ldots Q_D$ such that the new collection is also of high rank and $G = \Delta(P'_1, \ldots, P'_C, Q_1, \ldots, Q_D)$ for some function Δ. Hence

$$\Gamma(P'_1, \ldots, P'_C) = (\Delta(P'_1, \ldots, P'_C, Q_1, \ldots, Q_D))^2.$$

Because of the high rank of the collection $\{P'_1, \ldots, P'_C, Q_1, \ldots, Q_D\}$, the equidistribution result described in the last section allows us to conclude that in fact:

$$\Gamma(x_1, \ldots, x_C) = (\Delta(x_1, \ldots, x_C, y_1, \ldots y_D))^2$$

for all $x_1, \ldots, x_C, y_1, \ldots, y_D$ in the ranges of $P'_1, \ldots, P'_C, Q_1, \ldots, Q_D$, respectively. Therefore, if we set $\tilde{G} = \Delta(P_1, \ldots, P_C, 0, \ldots, 0)$, then $F = \tilde{G}^2$. It is immediate[2] that $\deg(\tilde{G}) \leqslant d/2$, and so, F has a square root.

It is curious that our proof of Theorem 1.7, which is entirely about classical polynomials, requires the use of non-classical polynomials. Also, as we mentioned earlier, there are no effective bounds on $K(\mathcal{P})$ that arise from our argument. It would be interesting to obtain better bounds (both upper and lower) for the locality of degree-structural properties.

1.7 Comparison with Previous Work

This work is part of, and culminates a sequence of works investigating the relationship between affine-invariance and testability. As described, Kaufman and Sudan [KS08] initiated the pro-

[2]For other degree-structural properties, the degree bound may not be immediate at this last step, and we need to argue it separately, again using the equidistribution result for high-rank non-classical polynomials.

gram. Subsequently, Bhattacharyya, Chen, Sudan, and Xie [BCSX11] investigated *monotone* linear-invariant properties of functions $f : \mathbb{F}_2^n \to \{0,1\}$, where a property $\mathcal{P}$ is monotone if it satisfies the condition that for any function $g \in \mathcal{P}$, modifying g by changing some outputs from 1 to 0 does not make it violate $\mathcal{P}$. Král, Serra and Vena [KSV12] and, independently, Shapira [Sha09] showed testability for any monotone linear-invariant property characterized by a finite number of linear constraints (of arbitrary complexity). For general non-monotone properties, Bhattacharyya, Grigorescu, and Shapira proved in [BGS10] that affine-invariant properties of functions in $\{\mathbb{F}_2^n \to \{0,1\}\}$ are testable if the complexity of the property is 1. Earlier this year, Bhattacharrya, Fischer and Lovett in [BFL12] generalized [BGS10] to show that affine-invariant properties of complexity $< p$ are testable. In this paper, we only have to restrict the complexity to be bounded, but the bound can be independent of p.

In terms of techniques, the general framework of the proof for testability here is very much the same as in [BGS10] or [BFL12]. However, the main difference here is that we work with collections of non-classical polynomials, rather than classical ones. Because the degrees of non-classical polynomials can change when multiplied by constants, the notions of rank and regularity are much more subtle. We need to establish a new version of a "polynomial regularity lemma" which allows us to decompose a given polynomial collection into a high rank collection of non-classical polynomials. Also, as discussed earlier, we establish a new equidistribution theorem for non-classical polynomials. We expect that these results will be of independent interest.

At a high level, the argument to prove our main theorem mirrors ideas used in a sequence of works [AFKS00, AS08b, AS08a, FN07, AFNS06, BCL+06] to characterize the testable graph properties. In particular, the technique of simultaneously decomposing the domain into a coarse partition and a fine partition with very strong regularity properties is due to [AFKS00], and the compactness argument used to handle infinitely many constraints is due to [AS08b].

1.8 Organization

In Section 2, we assemble all the technical components and establish some basic notions such as non-classical polynomials, rank and regularity. In Section 3, we show the equidistribution result for non-classical polynomials. In Section 4, we use the results established thus far to prove Theorem 1.7. Section 5 is devoted to proving Theorem 1.12.

2 Preparation

2.1 Notation

For integers a, b, we let $[a]$ denote the set $\{1, 2, \ldots, a\}$ and $[a, b]$ denote the set $\{a, a+1, \ldots, b\}$.

Fix a prime field $\mathbb{F} = \mathbb{F}_p$ for a prime $p \geqslant 2$. As we defined earlier $|\cdot|$ denotes the standard map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\} \subset \mathbb{Z}$.

We use the shorthand $x = a \pm \varepsilon$ to mean $a - \varepsilon \leqslant x \leqslant a + \varepsilon$.

2.2 Locality

In the context of affine-invariant properties, we can define the notion of local characterization in a more algebraic way than we did in the introduction. Recall that a *hyperplane* is an affine subspace of codimension 1.

Definition 2.1 (Locally characterized properties)**.** *An affine-invariant property $\mathcal{P} \subset \{\mathbb{F}^n \to [R] : n \geqslant 0\}$ is said to be* locally characterized *if both of the following hold:*

For every function $f : \mathbb{F}^n \to [R]$ in $\mathcal{P}$ and every hyperplane $A \leqslant \mathbb{F}^n$, $f|_A \in \mathcal{P}$.

There exists a constant $K \geqslant 1$ such that if $f : \mathbb{F}^n \to [R]$ does not belong to $\mathcal{P}$ and $n > K$, then there exists a hyperplane $B \leqslant \mathbb{F}^n$ such that $f|_B \notin \mathcal{P}$.

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

The following observation shows that an affine-invariant property is locally characterized if and only if it can be described using a bounded number of induced affine constraints from the previous section, and hence, is locally characterized in the sense of the introduction.

Lemma 2.2. *If $\mathcal{P} \subseteq \{\mathbb{F}^n \to [R] : n \geqslant 0\}$ is a locally characterized affine-invariant property with locality K, then $\mathcal{P}$ is equivalent to $\mathcal{A}$-freeness, where $\mathcal{A}$ is a finite collection of induced affine constraints, with each constraint of size p^K on $K+1$ variables. On the other hand, if $\mathcal{P}$ is equivalent to $\mathcal{A}$-freeness, where $\mathcal{A}$ is a collection of induced affine constraints with each constraint on $\leqslant K+1$ variables, then $\mathcal{P}$ has locality at most K.*

Finally, we also make formal note of the observation in the introduction that if a property is testable, then it must be locally characterized.

Remark 2.3. *If K is a fixed integer and $\mathcal{P} \subset \{\mathbb{F}^n \to [R]\}$ is an affine-invariant property that is testable with K queries, then $\mathcal{P}$ is a locally characterized property with locality K.*

So, we can view our main result as a converse statement.

2.3 Derivatives and Polynomials

Definition 2.4 (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 is the expected multiplicative derivative of f in d random directions at a random point.

Definition 2.5 (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*

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

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 2.6 (Additive Derivative)**.** *Given a function*[3] $P : \mathbb{F}^n \to \mathbb{T}$ *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 2.7 (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* $h_1, \ldots, h_{d+1}, x \in \mathbb{F}^n$, *it holds that*

$$(D_{h_1} \cdots D_{h_{d+1}} P)(x) = 0. \tag{3}$$

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 $\leqslant d$.

The following lemma of Tao and Ziegler shows that a classical polynomial P of degree d must always be of the form $\iota \circ Q$, where $Q : \mathbb{F}^n \to \mathbb{F}$ is a polynomial (in the usual sense) of degree d. It also characterizes the structure of non-classical polynomials.

Lemma 2.8 (Part of Lemma 1.7 in [TZ11])**.** *Let* $d \geqslant 1$ *be an integer.*

(i) A function $P : \mathbb{F}^n \to \mathbb{T}$ *is a polynomial of degree* $\leqslant d+1$ *if and only if* $D_h P$ *is a polynomial of degree* $\leqslant d$ *for all* $h \in \mathbb{F}^n$.

(ii) A function $P : \mathbb{F}^n \to \mathbb{T}$ *is a classical polynomial of degree* $\leqslant d$ *if* $P = \iota \circ Q$, *where* $Q : \mathbb{F}^n \to \mathbb{F}$ *has a representation of the form*

$$Q(x_1, \ldots, x_n) = \sum_{\substack{0 \leqslant d_1, \ldots, d_n < p: \\ \sum_i d_i \leqslant d}} c_{d_1, \ldots, d_n} x_1^{d_1} \cdots x_n^{d_n},$$

for a unique choice of coefficients $c_{d_1, \ldots, d_n} \in \mathbb{F}$.

(iii) 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*

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

[3] 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. Γ and Σ) to denote functions mapping $\mathbb{T}^C$ to $\mathbb{T}$. By abuse of notation, we sometimes conflate $\mathbb{F}$ and $\iota(\mathbb{F})$.

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

(iv) If $P : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of depth k, then it takes values in a coset of the subgroup $\mathbb{U}_{k+1}$. In particular, a polynomial of degree $\leqslant d$ takes on at most $p^{\lfloor \frac{d-1}{p-1} \rfloor + 1}$ distinct values.

Note that Lemma 2.8 (iii) immediately implies the following important observation[4] :

Remark 2.9. *If $Q : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree d and depth k, then pQ is a polynomial of degree $\max(d-p+1,0)$ and depth $k-1$. In other words, if Q is classical, then pQ vanishes, and otherwise, its degree decreases by $p-1$ and its depth by 1. Also, if $\lambda \in [1,p-1]$ is an integer, then $\deg(\lambda Q) = d$ and $\mathrm{depth}(\lambda Q) = k$.*

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 $\mathbb{U}_{k+1}$.

2.4 Inverse Theorem

There is a tight connection between polynomials and Gowers norms. In one direction, it is a straightforward consequence of the monotonicity of the Gowers norm ($\|f\|_{U^d} \leqslant \|f\|_{U^{d+1}}$) and invariance of the Gowers norm with respect to modulation by lower degree polynomials ($\|f\|_{U^{d+1}} = \|f \cdot \mathsf{e}(P)\|_{U^{d+1}}$ for polynomials P of degree $\leqslant d$) that if f is δ-*correlated* with a polynomial P of degree $\leqslant d$, meaning

$$|\mathop{\mathbf{E}}_{x} f(x)\mathsf{e}(-P(x))| \geqslant \delta$$

for some $\delta > 0$, then

$$\|f\|_{U^{d+1}} \geqslant \delta.$$

In the other direction, we have the following "Inverse theorem for the Gowers norm".

Theorem 2.10 (Theorem 1.11 of [TZ11])**.** *Suppose $\delta > 0$ and $d \geqslant 1$ is an integer. There exists an $\varepsilon = \varepsilon_{2.10}(\delta, d)$ such that the following holds. For every function $f : \mathbb{F}^n \to \mathbb{C}$ with $\|f\|_\infty \leqslant 1$ and $\|f\|_{U^{d+1}} \geqslant \delta$, there exists a polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$ that is ε-correlated with f, meaning*

$$\left|\mathop{\mathbf{E}}_{x \in \mathbb{F}^n} f(x)\mathsf{e}(-P(x))\right| \geqslant \varepsilon.$$

2.5 Rank

We will often need to study Gowers norms of exponentials of polynomials. As we describe below if this analytic quantity is non-negligible, then there is an algebraic explanation for this: it is possible to decompose the polynomial as a function of a constant number of low-degree polynomials. To state this rigorously, let us define the notion of *rank* of a polynomial.

[4]Recall that $\mathbb{T}$ is an additive group. If $n \in \mathbb{Z}$ and $x \in \mathbb{T}$, then nx is shorthand for $\underbrace{x + \cdots + x}_{n \text{ terms}}$ if $n \geqslant 0$ and $\underbrace{-x - \cdots - x}_{-n \text{ terms}}$ otherwise.

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 ∞ 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.*

Note that for integer $\lambda \in [1, p-1]$, $\mathrm{rank}(P) = \mathrm{rank}(\lambda P)$. The following theorem of Tao and Ziegler shows that high rank polynomials have small Gowers norms.

Theorem 2.12 (Theorem 1.20 of [TZ11])**.** *For any $\varepsilon > 0$ and integer $d > 0$, there exists an integer $r = r_{2.12}(d, \varepsilon)$ such that the following is true. For any polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$, if $\|\mathsf{e}(P)\|_{U^d} \geqslant \varepsilon$, then $\mathsf{rank}_d(P) \leqslant r$.*

For future use, we also record here a simple lemma stating that restrictions of high rank polynomials to hyperplanes generally preserve degree and high rank.

Lemma 2.13. *Suppose $P : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree d and rank $\geqslant r$, where $r > p+1$. Let A be a hyperplane in $\mathbb{F}^n$, and denote by P' the restriction of P to A. Then, P' is a polynomial of degree d and rank $\geqslant r - p$, unless $d = 1$ and P is constant on A.*

Proof. For the case $d = 1$, we can check directly that either P' is constant or else, P' is a non-constant degree-1 polynomial and so has rank infinity.

So, assume $d > 1$. By making an affine transformation, we can assume without loss of generality that A is the hyperplane $\{x_1 = 0\}$. Let $\pi : \mathbb{F}^n \to \mathbb{F}^{n-1}$ be the projection to A. Let $P'' = P - P' \circ \pi$. Clearly, P'' is zero on A. For $x \in \mathbb{F} \setminus \{0\}$, let $h_x = (x, 0, \ldots, 0) \in \mathbb{F}^n$. Note that $D_{h_x} P''$ is of degree $\leqslant d-1$ and that $(D_{h_x} P'')(y) = P''(y + h_x)$ for all $y \in A$. Hence, for every $x \in \mathbb{F} \setminus \{0\}$, P'' on $h_x + A$ agrees with a polynomial Q_x of degree $\leqslant d-1$. So, for a function $\Gamma : \mathbb{T}^{p+1} \to \mathbb{T}$, we can write $P = \Gamma(\iota(x_1), P', Q_1, Q_2, \ldots, Q_{p-1})$, where $\iota(x_1), Q_1, \ldots, Q_{p-1}$ are of degree $\leqslant d-1$.

Now, if P' itself is of degree $d-1$, then P is of rank $\leqslant p+1 < r$, a contradiction. If P' is of rank $< r-p$, then again P is of rank $< r - p + p = r$, a contradiction. □

2.6 Polynomial factors

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, and the property of high rank is robust against various operations such as restrictions to hyperplanes, taking derivatives, multiplying by integers, etc. 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.

A collection of polynomials automatically defines a partition of the domain. Here, we set up some notation.

Definition 2.14. *A* polynomial factor *I is a sequence of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$. We also identify it with the function $\mathcal{B} : \mathbb{F}^n \to \mathbb{T}^C$ mapping x to $(P_1(x), \ldots, P_C(x))$. An* atom *of $\mathcal{B}$ is a nonempty preimage $\mathcal{B}^{-1}(y)$ for some $y \in \mathbb{T}^C$. When there is no ambiguity, by an abuse of notation we identify an atom of $\mathcal{B}$ with the common value $\mathcal{B}(x)$ of all x in the atom.*

The partition induced by $\mathcal{B}$ *is the partition of $\mathbb{F}^n$ given by $\{\mathcal{B}^{-1}(y) : y \in \mathbb{T}^C\}$. 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 $\|\mathcal{B}\|$. The rank of a factor can now be defined as follows.

Definition 2.15 (Rank and Regularity)**.** *A polynomial factor $\mathcal{B}$ defined by a sequence of polynomials $P_1, \dots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective depths $k_1, \dots, k_C$ is said to have* rank r *if r is the least integer for which there exist $(\lambda_1, \dots, \lambda_C) \in \mathbb{Z}^C$ so that $(\lambda_1 \mod p^{k_1+1}, \dots, \lambda_C \mod p^{k_C+1}) \neq (0, \dots, 0)$ 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 $r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$, we say $\mathcal{B}$ is r-regular *if $\mathcal{B}$ is of rank larger than $r(|\mathcal{B}|)$.*

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 $\mathcal{B}$ 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 a generic collection of polynomials, as we shall establish in a precise sense in Section 3. 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 2.16 (Semantic and syntactic refinements)**.** *$\mathcal{B}'$ 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 relation $\succeq$ (without subscripts) is a synonym for $\succeq_{syn}$.*

Remark 2.17. *Clearly, being a syntactic refinement is stronger than being a semantic refinement. But observe that if $\mathcal{B}'$ is a semantic refinement of $\mathcal{B}$, then there exists a syntactic refinement $\mathcal{B}''$ of $\mathcal{B}$ that induces the same partition of $\mathbb{F}^n$, and for which $|\mathcal{B}''| \leqslant |\mathcal{B}'| + |\mathcal{B}|$, because we can define $\mathcal{B}''$ by just adding the defining polynomials of $\mathcal{B}$ to those of $\mathcal{B}'$.*

The following lemma is the workhorse that allows us to construct regular refinements.

Lemma 2.18 (Polynomial Regularity Lemma)**.** *Let $r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a non-decreasing function and $d > 0$ be an integer. Then, there is a function $C_{2.18}^{(r,d)} : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ such that the following is true. Suppose $\mathcal{B}$ is a factor defined by polynomials $P_1, \dots, P_C : \mathbb{F}^n \to \mathbb{T}$ of degree at most d. Then, there is an r-regular factor $\mathcal{B}'$ consisting of polynomials $Q_1, \dots, Q_{C'} : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$ such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$ and $C' \leqslant C_{2.18}^{(r,d)}(C)$.*

Moreover, if $\mathcal{B}$ is itself a refinement of some $\hat{\mathcal{B}}$ that has rank $> (r(C') + C')$ and consists of polynomials, then additionally $\mathcal{B}'$ will be a syntactic refinement of $\hat{\mathcal{B}}$.

Proof. We can prove our lemma starting from Lemma 9.6 of [TZ11]. To explain, let us define the notion of an *extended* factor. We say a polynomial factor $\mathcal{B}$ is extended if for any polynomial $Q \in \mathcal{B}$ that is not classical, $pQ \in \mathcal{B}$ also. Note that an extended factor defined by polynomials $P_1, \dots, P_C$ is of high rank if for all tuples $(\lambda_1, \dots, \lambda_C) \in [0, p-1]^C$, unless all the λ_i's are zero, $\sum_i \lambda_i P_i$ is of high $(\max_i \deg(\lambda_i P_i))$-rank. Tao and Ziegler proved the following:

Lemma 2.19 (Lemma 9.6 of [TZ11])**.** *Let $r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a non-decreasing function and $d > 0$ be an integer. Then, there are functions $\bar{C}^{(r,d)} : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ and $I^{(d)} : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ such that the following is true. Suppose $\mathcal{B}$ is an extended polynomial factor defined by polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$. Then, there is a subspace $V \leqslant \mathbb{F}^n$ and an r-regular extended factor $\bar{\mathcal{B}}$ consisting of polynomials $Q_1, \ldots, Q_{\bar{C}} : V \to \mathbb{T}$ such that $2 \leqslant \deg(Q_i) \leqslant d$ for each i, $\bar{\mathcal{B}}$ semantically refines the factor defined by $P_1|_V, \ldots, P_C|_V$, $\bar{C} \leqslant \bar{C}^{(r,d)}(C)$, and $\dim(V) \geqslant n - I^{(d)}(\bar{C})$.*

Let $\mathcal{B}_1$ be the extended factor defined by $\{p^k P_i \mid 0 \leqslant k \leqslant \mathrm{depth}(P_i), i \in [C]\}$. Apply Lemma 2.19 to $\mathcal{B}_1$ in order to obtain a bounded index subspace V_1 and an extended R_1-regular factor $\bar{\mathcal{B}}_1$ defined by polynomials $Q_1, \ldots, Q_{\bar{C}} : V_1 \to \mathbb{T}$, where R_1 is a growth function (growing even faster than r) we specify later on in the proof and $\bar{C} \leqslant \bar{C}^{(R_1,d)}(|\mathcal{B}_1|)$. For $a \in \mathbb{F}^n/V_1$ and $P \in \mathcal{B}_1$, define $P^a : V_1 \to \mathbb{F}^n$ to be $P^a(x) = D_a P(x) = P(a+x) - P(x)$. Each P^a is of degree $\leqslant d-1$. Also, since V_1 is the intersection of $I \leqslant I^{(d)}(\bar{C})$ hyperplanes, we can decide which coset in $\mathbb{F}^n/V_1$ an element $x \in \mathbb{F}^n$ belongs to as a function of $I \leqslant I^{(d)}(\bar{C})$ (classical) linear functions $\pi_1, \ldots, \pi_I$. Let $\mathcal{B}_2$ be the extended factor obtained by adding to $\bar{\mathcal{B}}_1$ all the polynomials $\{P^a \mid P \in \mathcal{B}_1, a \in \mathbb{F}^n/V_1\}$ and $\pi_1, \ldots, \pi_I$. Consider $x \in \mathbb{F}^n$ and let $x = a + y$ where $y \in V_1$, and $a \in \mathbb{F}^n/V_1$. Since $P(x) = P(y) - P^{-a}(x)$, each polynomial in $\mathcal{B}_1$ is a function of the polynomials in $\mathcal{B}_2$ over all of $\mathbb{F}^n$, and so $\mathcal{B}_2$ is a semantic refinement of $\mathcal{B}_1$ (and a syntactic refinement of $\bar{\mathcal{B}}_1$). Note that $|\mathcal{B}_2| \leqslant \bar{C} + dCI^{(d)}(\bar{C}) + I^{(d)}(\bar{C}) < \bar{C} + 2d\bar{C}I^{(d)}(\bar{C})$.

Now, suppose we repeat the steps in the previous paragraph with $\mathcal{B}_2$ taking the place of $\mathcal{B}_1$ and a different function R_2 taking the place of R_1. We specify R_2 later, but we will choose it so that it grows faster than r. The new application of Lemma 2.19 to $\mathcal{B}_2$ produces an extended factor $\bar{\mathcal{B}}_2$ that is R_2-regular and a bounded index subspace V_2 such that the polynomials in $\mathcal{B}_2$ restricted to V_2 are measurable with respect to $\bar{\mathcal{B}}_2$. We argue that $\bar{\mathcal{B}}_2$ differs from $\mathcal{B}_2$ only by polynomials of degree $\leqslant d-1$. Suppose $\mathcal{B}_2$ is not R_2-regular to start off with. The function R_1 is chosen so that $\bar{\mathcal{B}}_1$'s rank, $R_1(|\bar{\mathcal{B}}_1|) > R_2(|\mathcal{B}_2|) + |\mathcal{B}_2|$. This means that if a linear combination of polynomials in $\mathcal{B}_2$, $\sum_{S \in \mathcal{B}_2} \lambda_S S$, has rank $\leqslant R_2(|\mathcal{B}_2|)$ and $d' = \max_{S: \lambda_S \neq 0} \deg(S)$, then there must be an $S \notin \bar{\mathcal{B}}_1$ with $\lambda_S \neq 0$ and degree d', since otherwise the rank condition of $\bar{\mathcal{B}}_1$ would be violated. Since all the polynomials in $\mathcal{B}_2$ which are not in $\bar{\mathcal{B}}_1$ have degree $\leqslant d-1$, we conclude that $d' \leqslant d-1$. Inspecting the proof of Lemma 2.19 in [TZ11] shows that this means $\bar{\mathcal{B}}_2$ consists of the polynomials of $\bar{\mathcal{B}}_1$ along with other polynomials of degree $\leqslant d-1$. In the same way as in the previous paragraph, we obtain an extended factor $\mathcal{B}_3 \succeq_{syn} \bar{\mathcal{B}}_2$, so that $\mathcal{B}_3$ is a semantic refinement of $\mathcal{B}_2$ over all of $\mathbb{F}^n$. Note that since all the polynomials of $\bar{\mathcal{B}}_1$ are already in $\bar{\mathcal{B}}_2$, we only need to add $\{P^a \mid P \in \mathcal{B}_2 \setminus \bar{\mathcal{B}}_1, a \in \mathbb{F}^n/V_2\}$, together with some linear functions. All these polynomials have degree at most $\leqslant d-2$.

We keep repeating this process to obtain a sequence of extended factors $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \ldots$ and $\bar{\mathcal{B}}_1, \bar{\mathcal{B}}_2, \bar{\mathcal{B}}_3, \ldots$. Each $\mathcal{B}_{i+1}$ semantically refines $\mathcal{B}_i$ and syntactically refines $\bar{\mathcal{B}}_i$. The process stops at step i if $\mathcal{B}_i$ becomes R_i-regular, where the sequence of growth functions R_i satisfies $R_i(m) > R_{i+1}(m + 2dmI^{(d)}(m)) + m + 2dmI^{(d)}(m)$ and $R_d(m) = r(m)$. The functions R_i are chosen so that $R_i(|\bar{\mathcal{B}}_i|) > R_{i+1}(|\mathcal{B}_{i+1}|) + |\mathcal{B}_{i+1}|$, and therefore, by the above argument, $\mathcal{B}_{i+1}$ differs from $\mathcal{B}_i$ by polynomials of degree $\leqslant d-i$. So, we must stop after obtaining $\mathcal{B}_d$ in the sequence. Also, since each R_i grows faster than r, note that R_i-regularity for any $i \in [d]$ implies r-regularity. So, it must be that some $\mathcal{B}_i$ for $i \leqslant d$ already becomes r-regular.

Given an extended factor $\mathcal{B}''$ of rank $> r$, we can get a (standard) factor $\mathcal{B}'$ of rank $> r$ by letting $\mathcal{B}'$ be defined by the smallest subset of polynomials S such that $\{p^i P \mid P \in S, i \in \mathbb{Z}_{\geqslant 0}\} \supseteq \mathcal{B}''$. The last statement of the lemma follows from the same considerations as used above to argue that $\bar{\mathcal{B}}_i$ syntactically refines $\mathcal{B}_i$. $\square$

3 Equidistribution of Regular Factors

In this section, we make precise the intuition that a high-rank collection of polynomials often behaves like a collection of independent random variables. The key technical tool is the connection between the combinatorial notion of rank and the analytic notion of bias, given in Theorem 2.12. A weaker statement, that was established earlier by Kaufman and Lovett[5] and used by Tao and Ziegler in their proof of Theorem 2.12, is the following.

Theorem 3.1 (Theorem 4 of [KL08])**.** *For any $\varepsilon > 0$ and integer $d > 0$, there exists $r = r_{3.1}(d, \varepsilon)$ 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 $|\mathbf{E}_x[\mathsf{e}(P(x))]| < \varepsilon$.*

Proof. Given Theorem 2.12, this follows directly from easy fact that $|\mathbf{E}[f]| \leqslant \|f\|_{U^d}$ for every $d \geqslant 2$, and every $f : \mathbb{F}^n \to \mathbb{C}$. ☐

Using a standard observation that relates the bias of a function to its distribution on its range, we can conclude the following.

Lemma 3.2 (Size of atoms)**.** *Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of degree $d > 0$, complexity C, and rank $r_{3.1}(d, \varepsilon)$, 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*

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

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

Proof.

$$\begin{aligned}
\Pr_x[\mathcal{B}(x) = b] &= \mathbf{E}_x\left[\prod_i \frac{1}{p^{k_i+1}} \sum_{\lambda_i=0}^{p^{k_i+1}-1} \mathsf{e}\left(\lambda_i(P_i(x) - b_i)\right)\right] \\
&= \prod_i p^{-(k_i+1)} \cdot \sum_{\substack{(\lambda_1,\ldots,\lambda_C) \\ \in \prod_i [0, p^{k_i+1}-1]}} \mathbf{E}_x\left[\mathsf{e}\left(\sum_i \lambda_i(P_i(x) - b_i)\right)\right] \\
&= \prod_i p^{-(k_i+1)} \cdot \left(1 \pm \varepsilon \prod_i p^{k_i+1}\right) = \frac{1}{\|\mathcal{B}\|} \pm \varepsilon.
\end{aligned}$$

The first equality uses the fact that $P_i(x) - b_i$ is in $\mathbb{U}_{k_i+1}$ and that for any nonzero $x \in \mathbb{U}_{k_i+1}$, $\sum_{\lambda=0}^{p^{k+1}-1} \mathsf{e}(\lambda x) = 0$. The third equality uses Theorem 3.1 and the fact that unless every $\lambda_i = 0$, the polynomial $\sum_i \lambda_i(P_i(x) - b_i)$ has rank at least $r_{3.1}(d, \varepsilon)$. ☐

For our applications, we need to not only understand the distribution of $\mathcal{B}(x) = (P_i(x))$ but also, more generally, $(P_i(L_j(x)))$ for a given sequence of linear forms $L_1, \ldots, L_m : (\mathbb{F}^n)^\ell \to \mathbb{F}^n$. To this end, we first show the following dichotomy theorem.

[5] Kaufman and Lovett proved Theorem 3.1 for classical polynomials. But their proof also works for non-classical ones without modification.

Theorem 3.3 (Near orthogonality)**.** *Given $\varepsilon > 0$, suppose $\mathcal{B} = (P_1, \ldots, P_C)$ is a polynomial factor of degree $d > 0$ and rank $> r_{2.12}(d, \varepsilon)$, $A = (L_1, \ldots, L_m)$ is an affine constraint on ℓ variables, and Λ is a tuple of integers $(\lambda_{i,j})_{i \in [C], j \in [m]}$. Define*

$$P_{A,\mathcal{B},\Lambda}(x_1, \ldots, x_\ell) = \sum_{i \in [C], j \in [m]} \lambda_{i,j} P_i(L_j(x_1, \ldots, x_\ell)).$$

Then, one of the two statements below is true.

For every $i \in [C]$, it holds that $\sum_{j \in [m]} \lambda_{i,j} Q_i(L_j(\cdot)) \equiv 0$ for all polynomials $Q_i : \mathbb{F}^n \to \mathbb{T}$ with the same degree and depth as P_i. Clearly, $P_{A,\mathcal{B},\Lambda} \equiv 0$ in this case.

$P_{A,\mathcal{B},\Lambda}$ is non-constant. Moreover, $|\mathbf{E}_{x_1,\ldots,x_\ell}[\mathsf{e}\left(P_{A,\mathcal{B},\Lambda}(x_1, \ldots, x_\ell)\right)]| < \varepsilon$.

Proof. If $\lambda_{i,j} \neq 0$, then $\lambda_{i,j} P_i$ can be assumed to be non-constant, since otherwise, we can set $\lambda_{i,j}$ to 0. Let the depths of $P_1, \ldots, P_C$ be $k_1, \ldots, k_C$ respectively. For each $j \in [m]$, we let $(w_{j,1}, \ldots, w_{j,\ell})$ denote the vector corresponding to the affine form L_j; recall that $w_{j,1} = 1$. For any affine form L_j, let $|L_j|$, its *weight*, denote the sum $\sum_{t=2}^{\ell} |w_{j,t}|$.

For each i, perform the following step independently. If there exists a j such that $|L_j| > \deg(\lambda_{i,j} P_i)$ and $\lambda_{i,j} \neq 0$, then use Eq. (3) to replace $\lambda_{i,j} P_i(L_j(\cdot))$ by a linear combination over $\mathbb{Z}$ of $P_i(L_{j'}(\cdot))$ with $L_{j'} \preceq L_j$, and repeat until no such j exists. Here we use the assumption in part (ii) of Definition 1.8 that such $L_{j'} \in A$. At the end of this process, we obtain a new tuple of coefficients $\Lambda' = (\lambda'_{i,j})$; we can assume that each $\lambda'_{i,j} \in [0, p^{k_i+1} - 1]$ after quotienting with $p^{k_i+1}\mathbb{Z}$.

If all the $\lambda'_{i,j}$ are zero, then for the original coefficients $(\lambda_{i,j})$ also, $\sum_{j \in [m]} \lambda_{i,j} P_i(L_j(\cdot))$ is identically zero for every i individually. Indeed, $\sum_{j \in [m]} \lambda_{i,j} Q_i(L_j(\cdot))$ is zero for any Q_i with the same degree and depth as P_i because the above transformation from Λ to Λ' only depended upon the degree and depth of P_i.

Otherwise, Λ' does not consist of all zeroes, and for every nonzero $\lambda'_{i,j}$, we have $|L_j| \leqslant \deg(\lambda'_{i,j} P_i)$. In this case we show that $|\mathbf{E}[\mathsf{e}\left(P_{A,\mathcal{B},\Lambda'}(x_1, \ldots, x_\ell)\right)]| < \varepsilon$. At a high level, our goal is to express the bias of $P_{A,\mathcal{B},\Lambda'}$ in terms of the Gowers norm of a linear combination of P_i's and then use Theorem 2.12.

Suppose without loss of generality that the form L_1 satisfies:

(i) $\lambda'_{i,1} \neq 0$ for some $i \in [C]$.

(ii) L_1 is maximal in the sense that for every $j \neq 1$, either $\lambda'_{i,j} = 0$ for all $i \in [C]$ or it is the case that $|w_{j,t}| < |w_{1,t}|$ for some $t \in [\ell]$.

We want to "derive" $P_{A,\mathcal{B},\Lambda'}$ until we kill all $P_i(L_j(\cdot))$ terms for $j > 1$. Given a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_\ell) \in \mathbb{F}^\ell$, an element $y \in \mathbb{F}^n$, and a function $P : (\mathbb{F}^n)^\ell \to \mathbb{T}$, let us define

$$\bar{D}_{\boldsymbol{\alpha},y} P(x_1, \ldots, x_\ell) = P(x_1 + \alpha_1 y, \ldots, x_\ell + \alpha_\ell y) - P(x_1, \ldots, x_\ell).$$

Note that

$$\begin{aligned}\bar{D}_{\boldsymbol{\alpha},y}(P_i \circ L_j)(x_1, \ldots, x_\ell) &= P_i(L_j(x_1, \ldots, x_\ell) + L_j(\boldsymbol{\alpha})y) - P_i(L_j(x_1, \ldots, x_\ell)) \\ &= (D_{L_j(\boldsymbol{\alpha})y} P_i)(L_j(x_1, \ldots, x_\ell)).\end{aligned}$$

Thus, if $L_j(\boldsymbol{\alpha}) = \langle L_j, \boldsymbol{\alpha} \rangle = 0$, then $\bar{D}_{\boldsymbol{\alpha},y} P_i \circ L_j \equiv 0$ for all choices of y.

Set $\Delta = |L_1| = \sum_{i=2}^{\ell} w_{1,i}$, and let $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_\Delta \in \mathbb{F}^\ell$ be the set of all vectors of the form $(-w, 0, \ldots, 0, 1, 0, \ldots, 0)$ where 1 is in the ith coordinate for $i \in [2, \ell]$ and $0 \leqslant w \leqslant |w_{1,i}| - 1$ is an integer. Note that that $\langle L_1, \boldsymbol{\alpha}_k \rangle \neq 0$ for all $k \in [\Delta]$, but for any $j > 1$, by maximality of L_1, there exists some $k \in [\Delta]$ such that $\langle L_j, \boldsymbol{\alpha}_k \rangle = 0$. Consequently,

$$\begin{aligned}
\bar{D}_{\boldsymbol{\alpha}_\Delta, y_\Delta} \cdots \bar{D}_{\boldsymbol{\alpha}_1, y_1} P_{A,\mathcal{B},\Lambda'}(x_1, \ldots, x_\ell) &= \left(\bar{D}_{\boldsymbol{\alpha}_\Delta, y_\Delta} \cdots \bar{D}_{\boldsymbol{\alpha}_1, y_1} \sum_{i=1}^{C} \lambda'_{i,1} P_i \circ L_1 \right)(x_1, \ldots, x_\ell) \\
&= (D_{\langle L_1, \boldsymbol{\alpha}_\Delta \rangle y_\Delta} \cdots D_{\langle L_1, \boldsymbol{\alpha}_1 \rangle y_1} \sum_{i=1}^{C} \lambda'_{i,1} P_i)(L_1(x_1, \ldots, x_\ell)).
\end{aligned}$$

Therefore

$$\mathop{\mathbf{E}}_{\substack{y_1, \ldots, y_\Delta, \\ x_1, \ldots, x_\ell}} [\mathsf{e}\left((\bar{D}_{\boldsymbol{\alpha}_\Delta, y_\Delta} \cdots \bar{D}_{\boldsymbol{\alpha}_1, y_1} P_{A,\mathcal{B},\Lambda'})(x_1, \ldots, x_\ell) \right)] = \left\| \sum_{i=1}^{C} \lambda'_{i,1} P_i \right\|_{U^\Delta}^{2^\Delta}. \tag{4}$$

On the other hand we have the following claim.

Claim 3.4.

$$\mathop{\mathbf{E}}_{\substack{y_1, \ldots, y_\Delta, \\ x_1, \ldots, x_\ell}} [\mathsf{e}\left((\bar{D}_{\boldsymbol{\alpha}_\Delta, y_\Delta} \cdots \bar{D}_{\boldsymbol{\alpha}_1, y_1} P_{A,\mathcal{B},\Lambda'})(x_1, \ldots, x_\ell) \right)] \geqslant \left(\left| \mathop{\mathbf{E}}_{x_1, \ldots, x_\ell} \mathsf{e}\left(P_{A,\mathcal{B},\Lambda'}(x_1, \ldots, x_\ell) \right) \right| \right)^{2^\Delta}.$$

Proof. It suffices to show that for any function $P(x_1, \ldots, x_\ell)$ and nonzero $\boldsymbol{\alpha} \in \mathbb{F}^\ell$,

$$\left| \mathop{\mathbf{E}}_{y, x_1, \ldots, x_\ell} [\mathsf{e}\left((\bar{D}_{\boldsymbol{\alpha}, y} P)(x_1, \ldots, x_\ell) \right)] \right| \geqslant \left| \mathop{\mathbf{E}}_{x_1, \ldots, x_\ell} [\mathsf{e}\left(P(x_1, \ldots, x_\ell) \right)] \right|^2.$$

Recall that $(\bar{D}_{\boldsymbol{\alpha}, y} P)(x_1, \ldots, x_\ell) = P(x_1 + \alpha_1 y, \ldots, x_\ell + \alpha_\ell y) - P(x_1, \ldots, x_\ell)$. Without loss of generality, suppose $\alpha_1 \neq 0$. We make a change of coordinates so that $\boldsymbol{\alpha}$ can be assumed to be $(1, 0, \ldots, 0)$. More precisely, define $P' : (\mathbb{F}^n)^\ell \to \mathbb{T}$ as

$$P'(x_1, \ldots, x_\ell) = P\left(x_1, \frac{x_2 + \alpha_2 x_1}{\alpha_1}, \frac{x_3 + \alpha_3 x_1}{\alpha_1}, \ldots, \frac{x_\ell + \alpha_\ell x_1}{\alpha_1} \right),$$

so that $P(x_1, \ldots, x_\ell) = P'(x_1, \alpha_1 x_2 - \alpha_2 x_1, \alpha_1 x_3 - \alpha_3 x_1, \ldots, \alpha_1 x_\ell - \alpha_\ell x_1)$, and thus $(\bar{D}_{\boldsymbol{\alpha}, y} P)(x_1, \ldots, x_\ell) = P'(x_1 + \alpha_1 y, \alpha_1 x_2 - \alpha_2 x_1, \ldots, \alpha_1 x_\ell - \alpha_\ell x_1) - P'(x_1, \alpha_1 x_2 - \alpha_2 x_1, \ldots, \alpha_1 x_\ell - \alpha_\ell x_1)$. Therefore

$$\begin{aligned}
&\left| \mathop{\mathbf{E}}_{y, x_1, \ldots, x_\ell} [\mathsf{e}\left((\bar{D}_{\boldsymbol{\alpha}, y} P)(x_1, \ldots, x_\ell) \right)] \right| \\
&= \left| \mathop{\mathbf{E}}_{y, x_1, \ldots, x_\ell} [\mathsf{e}\left(P'(x_1 + \alpha_1 y, \alpha_1 x_2 - \alpha_2 x_1, \ldots, \alpha_1 x_\ell - \alpha_\ell x_1) - P'(x_1, \alpha_1 x_2 - \alpha_2 x_1, \ldots, \alpha_1 x_\ell - \alpha_\ell x_1) \right)] \right| \\
&= \left| \mathop{\mathbf{E}}_{y, x_1, \ldots, x_\ell} [\mathsf{e}\left(P'(x_1 + \alpha_1 y, x_2, \ldots, x_\ell) - P'(x_1, x_2, \ldots, x_\ell) \right)] \right| = \mathop{\mathbf{E}}_{x_2, \ldots, x_\ell} \left| \mathop{\mathbf{E}}_{x_1} [\mathsf{e}\left(P'(x_1, x_2, \ldots, x_\ell) \right)] \right|^2 \\
&\geqslant \left| \mathop{\mathbf{E}}_{x_1, x_2, \ldots, x_\ell} [\mathsf{e}\left(P'(x_1, x_2, \ldots, x_\ell) \right)] \right|^2 = \left| \mathop{\mathbf{E}}_{x_1, x_2, \ldots, x_\ell} [\mathsf{e}\left(P(x_1, x_2, \ldots, x_\ell) \right)] \right|^2.
\end{aligned}$$

□

Therefore, combining Eq. (4) with Claim 3.4, we get:

$$\left\| \mathsf{e}\left(\sum_{i\in[C]} \lambda'_{i,1} P_i(x)\right)\right\|_{U^\Delta} \geqslant \left| \mathop{\mathbf{E}}_{x_1,\dots,x_\ell} \mathsf{e}\left(P_{A,\mathcal{B},\Lambda'}(x_1,\dots,x_\ell)\right)\right|.$$

Suppose $|\mathbf{E}_{x_1,\dots,x_\ell}\, \mathsf{e}\left(P_{A,\mathcal{B},\Lambda'}(x_1,\dots,x_\ell)\right)| \geqslant \varepsilon$. Then, by the above inequality and Theorem 2.12, we get that $\sum_{P_i\in\mathcal{B}} \lambda'_{i,1} P_i(x)$ is a function of $r = r_{2.12}(d,\varepsilon)$ polynomials of degree $\Delta - 1$. But recall that if $\lambda'_{i,1} \neq 0$, then $\deg(\lambda'_{i,1}P_i) \geqslant |L_1| = \Delta$. Also, there exists a nonzero $\lambda'_{i,1}$. This is a contradiction to our assumption that the factor $\mathcal{B}$ is of rank $> r_{2.12}(d,\varepsilon)$. $\square$

Remark 3.5. *The proof of Theorem 3.3 also shows the following. Suppose, in the setting of Theorem 3.3, that for every $P_i \in \mathcal{B}$ and $L_j \in A$, either $|L_j| \leqslant \deg(\lambda_{i,j}P_i)$ or $\lambda_{i,j} = 0$. Then, unless every $\lambda_{i,j} = 0 \pmod{p^{k_i+1}}$, we have that $P_{A,\mathcal{B},\Lambda}$ is non-constant and $|\mathbf{E}[\mathsf{e}\left(P_{A,\mathcal{B},\Lambda}(x_1,\dots,x_\ell)\right)]| < \varepsilon$. The only modification needed to the above proof is that the transformation from Λ to Λ' can be omitted.*

To show equidistribution of $(P_i(L_j(x_1,\dots,x_\ell)))$, we can use Theorem 3.3 in the same manner we used Theorem 3.1 to show the equidistribution of $(P_i(x))$ in Lemma 3.2. Before we do so, however, let us give a name to those Λ for which the first case of Theorem 3.3 holds.

Definition 3.6. *Given an affine constraint $A = (L_1,\dots,L_m)$ on ℓ variables and integers $d,k > 0$ such that $d > k(p-1)$, the (d,k)*-dependency set of *A is the set of tuples $(\lambda_1,\dots,\lambda_m) \in [0, p^{k+1}-1]$ such that $\sum_{i=1}^m \lambda_i P(L_i(x_1,\dots,x_\ell)) \equiv 0$ for every polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree d and depth k.*

Theorem 3.3 says that if $\mathcal{B}$ is a regular factor, $P_{A,\mathcal{B},\Lambda} \equiv 0$ exactly when the first condition holds. In other words:

Corollary 3.7. *Fix an integer $C > 0$, tuples $(d_1,\dots,d_C) \in \mathbb{Z}^C_{>0}$ and $(k_1,\dots,k_C) \in \mathbb{Z}^C_{\geqslant 0}$, and an affine constraint $(L_1,\dots,L_m)$ on ℓ variables. For $i \in [C]$, let Λ_i be the (d_i,k_i)-dependency set of A.*

Then, for any polynomial factor $\mathcal{B} = (P_1,\dots,P_C)$, where each P_i has degree d_i and depth k_i, and $\mathcal{B}$ has rank $> r_{2.12}\left(\max_i d_i, \frac{1}{2}\right)$, it is the case that a tuple $(\lambda_{i,j})_{i\in[C],j\in[m]}$ satisfies

$$\sum_{i=1}^{C}\sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(x_1,\dots,x_\ell)) \equiv 0$$

if and only if for every $i \in [C]$, $(\lambda_{i,1} \pmod{p^{k_i+1}}, \dots, \lambda_{i,m} \pmod{p^{k_i+1}}) \in \Lambda_i$.

Proof. The "if" direction is obvious. For the "only if" direction, we use Theorem 3.3 to conclude that if $\sum_{i,j} \lambda_{i,j} P_i(L_j(\cdot)) \equiv 0$, it must be that for every $i \in [C]$, $\sum_j \lambda_{i,j} Q_i(L_j(\cdot)) \equiv 0$ for any polynomial Q_i with degree d_i and depth k_i. This is equivalent to saying $(\lambda_{i,1} \pmod{p^{k_i+1}}, \dots, \lambda_{i,m} \pmod{p^{k_i+1}}) \in \Lambda_i$. $\square$

Remark 3.8. *For large characteristic fields, Hatami and Lovett [HL11a] showed that the analog of Corollary 3.7 is true even without the rank condition.*

The distribution of $(P_i(L_j(x_1,\dots,x_\ell)))$ is only going to be supported on atoms which respect the constraints imposed by dependency sets. This is obvious: if P is a polynomial of degree d and depth k, $(\lambda_1,\dots,\lambda_m)$ are in the (d,k)-dependency set of $(L_1,\dots,L_m)$, and $P(L_j(x_1,\dots,x_\ell)) = b_j$, then $\sum_j \lambda_j b_j = 0$. We call atoms which respect this constraint for all P_i in a factor *consistent*. Formally:

Definition 3.9 (Consistency). *Let A be an affine constraint of size m. A sequence of elements $b_1,\dots,b_m \in \mathbb{T}$ are said to be* (d,k)-consistent with A *if $b_1,\dots,b_m \in \mathbb{U}_{k+1}$ and for every tuple $(\lambda_1,\dots,\lambda_m)$ in the (d,k)-dependency set of A, it holds that $\sum_{i=1}^m \lambda_i b_i = 0$.*

Given a polynomial factor $\mathcal{B}$ defined by polynomials $P_1,\dots P_C : \mathbb{F}^n \to \mathbb{T}$ of degrees $d_1,\dots,d_C$ and depths $k_1,\dots,k_C$ respectively, a sequence of atoms $b_1,\dots,b_m \in \mathbb{T}^C$ are said to be consistent with A *if for every $i \in [C]$, the elements $b_{1,i},\dots,b_{m,i}$ are (d_i,k_i)-consistent with A.*

Now, the proof of equidistribution of $(P_i(L_j(x_1,\dots,x_\ell))$ is straightforward.

Theorem 3.10. *Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of degree $d > 0$, complexity C, and rank $r_{3.1}(d,\varepsilon)$, that is defined by a tuple of polynomials $P_1,\dots,P_C : \mathbb{F}^n \to \mathbb{T}$ having respective degrees $d_1,\dots,d_C$ and respective depths $k_1,\dots,k_C$. Let $A = (L_1,\dots,L_m)$ be an affine constraint on ℓ variables.*

Suppose $b_1,\dots,b_m \in \mathbb{T}^C$ are atoms of $\mathcal{B}$ that are consistent with A. Then

$$\Pr_{x_1,\dots,x_\ell}[\mathcal{B}(L_j(x_1,\dots,x_\ell)) = b_j \ \forall j \in [m]] = \frac{\prod_{i=1}^C |\Lambda_i|}{\|\mathcal{B}\|^m} \pm \varepsilon$$

where Λ_i is the (d_i,k_i)-dependency set of A.

Proof. The proof is similar to that of Lemma 3.2.

$$\begin{aligned}
&\Pr_{x_1,\dots,x_\ell}[P_i(L_j(x_1,\dots,x_\ell)) = b_{i,j} \ \forall i \in [C], \forall j \in [m]] \\
&= \mathop{\mathbf{E}}_{x_1,\dots,x_\ell}\left[\prod_{i,j} \frac{1}{p^{k_i+1}} \sum_{\lambda_{i,j}=0}^{p^{k_i+1}-1} \mathsf{e}\left(\lambda_{i,j}(P_i(L_j(x_1,\dots,x_\ell)) - b_{i,j})\right)\right] \\
&= \left(\prod_i p^{-(k_i+1)}\right)^m \sum_{\substack{(\lambda_{i,j}) \\ \in \prod_{i,j}[0,p^{k_i+1}-1]}} \mathsf{e}\left(-\sum_{i,j} \lambda_{i,j} b_{i,j}\right) \mathbf{E}\left[\mathsf{e}\left(\sum_{i,j} \lambda_{i,j} P_i(L_j(x_1,\dots,x_\ell)\right)\right] \\
&= p^{-m\sum_{i=1}^C (k_i+1)} \cdot \left(\prod_{i=1}^C |\Lambda_i| \ \pm \ \varepsilon p^{m\sum_{i=1}^C (k_i+1)}\right),
\end{aligned}$$

where the last line follows because by Corollary 3.7, $\sum_{i,j} \lambda_{i,j} P_i(L_j(\cdot))$ is identically zero for $\prod_i |\Lambda_i|$ many tuples $(\lambda_{i,j})$ and, in that case, $\sum_{i,j} \lambda_{i,j} b_{i,j} = 0$ because of the consistency requirement. For any other tuple $(\lambda_{i,j})$, the expectation in the third line is bounded by ε in absolute value. $\square$

4 Degree-structural Properties

In this section, we prove Theorem 1.7 in the introduction stating that if $\mathcal{P}$ is degree-structural (recall Definition 1.6), then $\mathcal{P}$ is locally characterized. The proof uses many of the tools established in Section 3.

Theorem 1.7 (restated). *Every degree-structural property with bounded scope and max-degree is a locally characterized affine-invariant property.*

Proof. Let $\mathcal{P}$ be a degree-structural property with scope σ and max-degree Δ. Denote by S the set of tuples $(c, \mathbf{d}, \Gamma)$ such that $c \leqslant \sigma$ and $\mathcal{P}$ is the union over all $(c, \mathbf{d}, \Gamma) \in S$ of $(c, \mathbf{d}, \Gamma)$-structured functions. It is clear that $\mathcal{P}$ is affine-invariant, as having degree bounded by a constant is an affine-invariant property. It is also immediate that $\mathcal{P}$ is closed under taking restrictions to subspaces, since if F is $(c, \mathbf{d}, \Gamma)$-structured, then F restricted to any hyperplane is also $(c, \mathbf{d}, \Gamma)$-structured. The non-trivial part of the theorem is to show that the locality is bounded. In other words we need to show that there is a constant K such that for $n \geqslant K$, if $F : \mathbb{F}^n \to \mathbb{T}$ is a function with $F|_A \in \mathcal{P}$ for every hyperplane $A \leqslant \mathbb{F}^n$, then $F \in \mathcal{P}$.

First, let us bound the degree of F. We know that $F|_A \in \mathcal{P}$ for every hyperplane A. Therefore, $\deg(F|_A) \leqslant p\sigma\Delta$ for every A, as $F|_A$ is a function of at most σ polynomials each of degree at most Δ over a field of characteristic p. It follows that F itself is of degree $\leqslant p\sigma\Delta$.

Let $r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a function to be fixed later. Define $r_2 : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ so that $r_2(m) > r(C_{2.18}^{(r,p\sigma\Delta)}(m+\sigma)) + C_{2.18}^{(r,p\sigma\Delta)}(m+\sigma) + p$.

We apply Lemma 2.18to $\{F\}$ to find an r_2-regular polynomial factor $\mathcal{B}$ of degree $\leqslant p\sigma\Delta$, defined by polynomials $R_1, \dots, R_C : \mathbb{F}^n \to \mathbb{T}$, where $C \leqslant C_{2.18}^{r_2,d}(1)$. Since F is measurable with respect to $\mathcal{B}$, there exists a function $\Sigma : \mathbb{T}^C \to \mathbb{F}$, such that $F(x) = \Sigma(R_1(x), \dots, R_C(x))$.

From each R_i pick a monomial with degree equal to $\deg(R_i)$ and a monomial (possibly the same one) with depth equal to $\mathrm{depth}(R_i)$. By taking K to be sufficiently large, we can gaurantee the existence of an $i_0 \in [n]$ such that x_{i_0} is not involved in any of these monomials. Consequently $\deg(R'_i) = \deg(R_i)$ and $\mathrm{depth}(R'_i) = \mathrm{depth}(R_i)$ for all $i \in [C]$, where $R'_1, \dots, R'_C$ are the restrictions of $R_1, \dots, R_C$, respectively, to the hyperplane $\{x_{i_0} = 0\}$. Also by Lemma 2.13, $R'_1, \dots, R'_C$ have rank $> r_2(C) - p$. Since $F|_{x_{i_0}=0} \in \mathcal{P}$, by definition of $\mathcal{P}$, there must exist $(c, \mathbf{d}, \Gamma) \in S$ with $c \leqslant \sigma$ such that

$$\Sigma(R'_1, \dots, R'_C) = \Gamma(P_1, \dots, P_c),$$

where $\deg(P_i) \leqslant d_i$ for all $i \in [c]$.

Now, apply Lemma 2.18 to find an r-regular refinement of the factor defined by the tuple of polynomials $(R'_1, \dots, R'_C, P_1, \dots, P_c)$. Because of our choice of r_2 and the last part of Lemma 2.18, we obtain a syntactic refinement of $\{R'_1, \dots, R'_C\}$. That is, we obtain a tuple $\mathcal{B}'$ of polynomials $R'_1, \dots, R'_C, S_1, \dots, S_D : \mathbb{F}^n \to \mathbb{T}$ such that it has degree $\leqslant p\sigma\Delta$, its rank $> r(C+D)$, and $C + D \leqslant C_{2.18}^{(r)}(C+\sigma)$, and for each $i \in [c]$, $P_i = \Gamma_i(R'_1, \dots, R'_C, S_1, \dots, S_D)$ for some function $\Gamma_i : \mathbb{T}^{C+D} \to \mathbb{T}$. So for all $x \in \mathbb{F}^n$,

$$\begin{aligned}\Sigma(R'_1(x), \dots, R'_C(x)) = \\ \Gamma(\Gamma_1(R'_1(x), \dots, R'_C(x), S_1(x), \dots, S_D(x)), \dots, \Gamma_c(R'_1(x), \dots, R'_C(x), S_1(x), \dots, S_D(x))).\end{aligned}$$

Applying Lemma 3.2, we see that if the rank of $\mathcal{B}'$ is $> r_{3.2}(p\sigma\Delta, \varepsilon)$ where $\varepsilon > 0$ is sufficiently small (say $\varepsilon = \|\mathcal{B}'\|/2$), then $(R'_1(x), \dots, R'_C(x), S_1(x), \dots, S_D(x))$ acquires every value in its range. Thus, we have the identity

$$\Sigma(a_1, \dots, a_c) = \Gamma(\Gamma_1(a_1, \dots, a_C, b_1, \dots, b_D), \dots, \Gamma_c(a_1, \dots, a_C, b_1, \dots, b_D)),$$

for every $a_i \in \mathbb{U}_{\mathrm{depth}(R'_i)+1}$ and $b_i \in \mathbb{U}_{\mathrm{depth}(S_i)+1}$. Thus, we can substitute R_i for R'_i and 0 for S_i in the above equation and still retain the identity

$$\begin{aligned}F(x) &= \Sigma(R_1(x), \dots, R_C(x))\\ &= \Gamma(\Gamma_1(R_1(x), \dots, R_C(x), 0, \dots, 0), \dots, \Gamma_c(R_1(x), \dots, R_C(x), 0, \dots, 0))\\ &= \Gamma(Q_1(x), \dots, Q_c(x))\end{aligned}$$

where $Q_i : \mathbb{F}^n \to \mathbb{T}$ are defined as $Q_i(x) = \Gamma_i(R_1(x), \dots, R_C(x), 0, \dots, 0)$. Since for every i, $\deg(R_i) = \deg(R'_i)$ and $\mathrm{depth}(R_i) = \mathrm{depth}(R'_i)$, we can apply Theorem 4.1 below to conclude that $\deg(Q_i) \leqslant \deg(P_i) \leqslant d_i$ for every $i \in [c]$, as long as the rank of $\mathcal{B}'$ is $> r_{4.1}(p\sigma\Delta)$. Finally, we show that $Q_1, \dots, Q_c$ map to $\mathbb{U}_1 = \iota(F)$ and, so, are classical. Indeed, since $P_1, \dots, P_c$ are classical, $\Gamma_1, \dots, \Gamma_c$ must map to $\iota(F)$ on all of $\prod_{i=1}^{C} \mathbb{U}_{\mathrm{depth}(R'_i)+1} \times \prod_{i=1}^{D} \mathbb{U}_{\mathrm{depth}(S_i)+1} \supseteq \prod_{i=1}^{C} \mathbb{U}_{\mathrm{depth}(R_i)+1} \times \{0\}^D$. Hence, $F \in \mathcal{P}$. □

The following theorem, used in the proof above, shows that a function of a high rank collection of polynomials has the degree one would expect. Thus, it displays yet another way in which high-rank polynomials behave "generically". The proof is via another application of the near-orthogonality result in Theorem 3.3.

Theorem 4.1. *For an integer $d > 0$, let $P_1, \dots, P_C : \mathbb{F}^n \to \mathbb{T}$ be polynomials of degree $\leqslant d$ and rank $> r_{4.1}(d)$, and let $\Gamma : \mathbb{T}^C \to \mathbb{T}$ be an arbitrary function. Define the polynomial $F : \mathbb{F}^n \to \mathbb{T}$ by $F(x) = \Gamma(P_1(x), \dots, P_C(x))$. Then, for every collection of polynomials $Q_1, \dots Q_C : \mathbb{F}^n \to \mathbb{T}$ with $\deg(Q_i) \leqslant \deg(P_i)$ and $\mathrm{depth}(Q_i) \leqslant \mathrm{depth}(P_i)$ for all $i \in [C]$, if $G : \mathbb{F}^n \to \mathbb{T}$ is the polynomial $G(x) = \Gamma(Q_1(x), \dots, Q_C(x))$, it holds that $\deg(G) \leqslant \deg(F)$.*

Proof. Let $f(x) = \mathsf{e}(F(x))$ and $\gamma(x_1, \dots, x_C) = \mathsf{e}(\Gamma(x_1, \dots, x_C))$. Let $D = \deg(F)$. Then, for every $x, y_1, \dots, y_{D+1} \in \mathbb{F}^n$,

$$\Delta_{y_{D+1}} \cdots \Delta_{y_1} f(x) = 1.$$

We need to show that $g(x) = \mathsf{e}(G(x))$ also satisfies $\Delta_{y_{D+1}} \cdots \Delta_{y_1} g(x) = 1$.

Let $k_1, \dots, k_C$ be the depths of $P_1, \dots, P_C$, respectively. Then, each P_i takes values in $\mathbb{U}_{k_i+1}$. Let Σ denote the group $\mathbb{Z}_{p^{k_1+1}} \times \cdots \times \mathbb{Z}_{p^{k_C+1}}$. Considering the Fourier transform of γ, we have

$$f(x) = \gamma(P_1(x), \dots, P_C(x)) = \sum_{\beta \in \Sigma} \hat{\gamma}(\beta) \mathsf{e}\left(\sum_{i=1}^{C} \beta_i P_i(x)\right).$$

Next, we look at the derivative.

$$\begin{aligned}\Delta_{y_{D+1}} \cdots \Delta_{y_1} f(x) &= \Delta_{y_{D+1}} \cdots \Delta_{y_1} \left(\sum_{\beta \in \Sigma} \hat{\gamma}(\beta) \mathsf{e}\left(\sum_{i=1}^{C} \beta_i P_i(x)\right)\right)\\ &= \sum_{\alpha_J \in \Sigma : J \subseteq [D+1]} \prod_{J \subseteq [D+1]} \hat{\gamma}(\alpha_J) \mathsf{e}\left((-1)^{|J|+1} \sum_{i=1}^{C} \alpha_{J,i} P_i\left(x + \sum_{j \in J} y_j\right)\right).\end{aligned}$$

Denoting $\delta(\alpha) = \prod_{J \subseteq [D+1]} \hat{\gamma}(\alpha_J)$ for $\alpha = (\alpha_J)_{J \subseteq [D+1]} \in \Sigma^{\mathcal{P}([D+1])}$, we have

$$\Delta_{y_{D+1}} \cdots \Delta_{y_1} f(x) = \sum_{\alpha \in \Sigma^{\mathcal{P}([D+1])}} \delta(\alpha) \mathsf{e}\left(\sum_{i=1}^{C} \sum_{J \subseteq [D+1]} (-1)^{|J|+1} \alpha_{J,i} P_i\left(x + \sum_{j \in J} y_j\right)\right). \tag{5}$$

For any i, if there is a J such that $|J|> \deg(\alpha_{J,i}P_i)$, we can use Eq. (3) to rewrite $\alpha_{J,i}P_i\left(x+\sum_{j\in J}y_j\right)$ as a linear combination (over $\mathbb{Z}$) of $\left\{P_i\left(x+\sum_{j\in J'}y_j\right):|J'|<|J|\right\}$. We repeat this process until for every i and J, either $\alpha_{J,i}=0$ or $|J|\leqslant \deg(\alpha_{J,i}P_i)$. Denoting by $\mathcal{A}$ the set of $\alpha\in\Sigma^{\mathcal{P}([D+1])}$ that satisfy this condition, we have obtained a new set of coefficients $\delta'(\alpha)$ such that

$$\Delta_{y_{D+1}}\cdots\Delta_{y_1}f(x)=\sum_{\alpha\in\mathcal{A}}\delta'(\alpha)\mathsf{e}\left(\sum_{i=1}^{C}\sum_{J\subseteq[D+1]}\alpha_{J,i}P_i\left(x+\sum_{j\in J}y_j\right)\right).$$

Now, the crucial observation is that if instead of $P_1,\ldots,P_C$, we had $Q_1,\ldots,Q_C$, the same decomposition applies.

$$\Delta_{y_{D+1}}\cdots\Delta_{y_1}g(x)=\sum_{\alpha\in\mathcal{A}}\delta'(A)\mathsf{e}\left(\sum_{i=1}^{C}\sum_{J\subseteq[D+1]}\alpha_{J,i}Q_i\left(x+\sum_{j\in J}y_j\right)\right). \tag{6}$$

The reason is that Eq. (5) remains valid as is if f is replaced by g and the P_i's are replaced by Q_i's, and furthermore since $\deg(P_i)\geqslant\deg(Q_i)$ and $\operatorname{depth}(P_i)\geqslant\operatorname{depth}(Q_i)$, the applications of Eq. (3) remain valid also. Therefore, Eq. (6) is also valid [6].

But now, we argue that $\delta'(\cdot)$ are uniquely determined. Let $k=\max_i k_i\leqslant d/(p-1)$.

Claim 4.2. *If $P_1,\ldots,P_C$ are of rank $> r_{2.12}(d,1/|\mathcal{A}|)+1$, the functions*

$$\left\{\mathsf{e}\left(\sum_{i=1}^{C}\sum_{J\subseteq[D+1]}\alpha_{J,i}P_i\left(x+\sum_{j\in J}y_j\right)\right)\ :\ \alpha\in\mathcal{A}\right\}$$

are linearly independent over $\mathbb{C}$.

Proof. Note that all these functions have L^2-norm equal to 1. Hence it suffices to show that their pairwise inner products are all bounded in absolute value by $1/|\mathcal{A}|$. To prove this consider $\alpha,\beta\in\mathcal{A}$, and note that by Theorem 3.3 and, in particular, Remark 3.5, unless all the $\alpha_{J,i}-\beta_{J,i}$ are zero,

$$\left|\mathbf{E}\left[\mathsf{e}\left(\sum_{i=1}^{C}\sum_{J\subseteq[D+1]}(\alpha_{J,i}-\beta_{J,i})P_i\left(x+\sum_{j\in J}y_j\right)\right)\right]\right|<\frac{1}{|\mathcal{A}|}.$$

□

Therefore, since $\Delta_{y_{D+1}}\cdots\Delta_{y_1}f(x)=1$, we must have $\delta'(\alpha)=1$ when α is the all-zero tuple, and $\delta'(\alpha)=0$ for every nonzero α. Plugging into Eq. (6), we get $\Delta_{y_{D+1}}\cdots\Delta_{y_1}g(x)=1$. □

[6] Note that in Eq. (6), one could haven onzero $\alpha_{J,i}$ and $|J|>\deg(\alpha_{J,i}Q_i)$, for $A=(\alpha_J)_{J\subseteq[D+1]}$ with $\delta'(A)\neq 0$.

5 Property Testing

5.1 Decomposition Theorems

We have already seen a decomposition theorem in Theorem 2.12: if a polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$ satisfies $\|\mathsf{e}(P)\|_{U^d} \geqslant \varepsilon$, then there exists a factor $\mathcal{B}$ of complexity $\leqslant r_{2.12}(d, \varepsilon)$ such that P is a function of the polynomials defining $\mathcal{B}$. In this section, we establish decomposition theorems that can be applied to arbitrary functions and not just low-degree polynomials with large Gowers norms.

To describe such decompositions, we need to consider conditional expectations over polynomial factors. Note that a polynomial factor defines a partition of $\mathbb{F}^n$, and thus one can consider the conditional expectation of functions $f : F^n \to \mathbb{C}$ with respect to this partition:

Definition 5.1 (Expectation over polynomial factor)**.** *Given a factor $\mathcal{B}$ and a function $f : \mathbb{F}^n \to \{0,1\}$, the* expectation *of f over an atom $y \in \mathbb{T}^{|\mathcal{B}|}$, denoted by $\mathbf{E}[f|y]$, is the average $f(x)$ over $\{x : \mathcal{B}(x) = y\}$. The* conditional expectation *of f over $\mathcal{B}$, is the real-valued function over $\mathbb{F}^n$ given by $\mathbf{E}[f|\mathcal{B}](x) = \mathbf{E}[f|\mathcal{B}(x)]$. In particular, it is constant on every atom of the polynomial factor and hence it is a function of the polynomials defining $\mathcal{B}$.*

The Strong Decomposition Theorem below shows that any Boolean function can be decomposed into the sum of a conditional expectation over a high rank factor, a function with small Gowers norm, and a function with small L^2-norm.

Theorem 5.2 (Strong Decomposition Theorem; Theorem 4.4 of [BFL12])**.** *Suppose $\delta > 0$ and $d, C_0 \geqslant 1$ 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 $N = N_{5.2}(\delta, \eta, r, d, p)$ and $C = C_{5.2}(\delta, \eta, r, d, C_0)$ such that the following holds.*

Given $f : \mathbb{F}^n \to \{0,1\}$ where $n > N$ and a polynomial factor $\mathcal{B}_0$ of degree at most d and complexity at most C_0, there exist three functions $f_1, f_2, f_3 : \mathbb{F}^n \to \mathbb{R}$ and a polynomial factor $\mathcal{B} \succeq_{sem} \mathcal{B}_0$ of degree at most d and complexity at most C such that the following hold:

$f = f_1 + f_2 + f_3$
$f_1 = \mathbf{E}[f|\mathcal{B}]$
$\|f_2\|_{U^{d+1}} \leqslant 1/\eta(|\mathcal{B}|)$
$\|f_3\|_2 \leqslant \delta$
f_1 and $f_1 + f_3$ have range $[0,1]$; f_2 and f_3 have range $[-1,1]$.
$\mathcal{B}$ is r-regular

It turns out though that this Strong Decomposition Theorem is not quite sufficient for our needs. The issue is that the bound on f_3 above is a constant δ. Ideally, we would want δ to decrease as a function of the complexity of the polynomial factor, but it is not possible to achieve this. We resolve the issue by using the non-negativity of f_1 and $f_1 + f_3$ to *localize* our analysis to certain atoms c' so that there is a guarantee that the L^2-norm of f_3 conditioned inside c' (i.e., $\mathbf{E}_{x \in c'}[|f_3(x)|^2]$) is small. Inspired by [AFKS00], we choose these atoms c' to be atoms from a polynomial factor $\mathcal{B}'$, with each atom c' contained inside an atom c of a coarser factor $\mathcal{B}$. To make the localization argument work, it is necessary that the function $f : \mathbb{F}^n \to \{0,1\}$have roughly the same density on c and c'. To this end, we make the following definition.

Definition 5.3 (Polynomial factor represents another factor)**.** *Given a function $f : \mathbb{F}^n \to \{0,1\}$, a polynomial factor $\mathcal{B}'$ that refines another factor $\mathcal{B}$ and a real $\zeta \in (0,1)$, we say $\mathcal{B}'$ ζ-*represents $\mathcal{B}$ with respect to f *if for at most ζ fraction of atoms c of $\mathcal{B}$, more than ζ fraction of the atoms c' lying inside c satisfy $|\mathbf{E}[f|c] - \mathbf{E}[f|c']| > \zeta$.*

We now describe a Super Decomposition Theorem and a corollary, which produces two factors $\mathcal{B}' \succeq_{syn} \mathcal{B}$ such that f_1 is a conditional expectation over the finer factor $\mathcal{B}'$ while f_3 has L^2-norm that decreases with the complexity of the coarser factor $\mathcal{B}$.

Theorem 5.4 (Super Decomposition Theorem; Theorem 4.9 of [BFL12])**.** *Suppose $\zeta > 0$ and $d, C_0 \geqslant 1$ are integers. Let $\eta : \mathbb{N} \to \mathbb{R}^+$ and $\delta : \mathbb{N} \to \mathbb{R}^+$ be arbitrary non-increasing functions, and $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary non-decreasing function. Then there exist $N = N_{5.4}(\delta, \eta, r, d, \zeta, C_0)$ and $C = C_{5.4}(\delta, \eta, r, d, \zeta, C_0)$ such that the following holds.*

Given $f : \mathbb{F}^n \to \{0,1\}$ where $n > N$ and a polynomial factor $\mathcal{B}_0$ of degree at most d and complexity at most C_0, there exist functions $f_1, f_2, f_3 : \mathbb{F}^n \to \mathbb{R}$, a semantic refinement $\mathcal{B}$ of $\mathcal{B}_0$ of degree at most d and a syntactic refinement $\mathcal{B}'$ of $\mathcal{B}$ of degree at most d and of complexity at most C, such that the following hold:

$f = f_1 + f_2 + f_3$
$f_1 = \mathbf{E}[f|\mathcal{B}']$
$\|f_2\|_{U^{d+1}} \leqslant \eta(|\mathcal{B}'|)$
$\|f_3\|_2 \leqslant \delta(|\mathcal{B}|)$
f_1 and $f_1 + f_3$ have range $[0,1]$; f_2 and f_3 have range $[-1,1]$.
$\mathcal{B}$ and $\mathcal{B}'$ are both r-regular.
$\mathcal{B}'$ ζ-represents $\mathcal{B}$ with respect to f.

Although the above Super Decomposition Theorem may be useful by itself for other applications, we will need a particular variant. We make a careful choice of one atom of $\mathcal{B}'$ for every atom of $\mathcal{B}$, such that the following conditions are satisfied:

The choice of atoms will be made in a "uniform" manner. This part is helped by $\mathcal{B}'$ being a syntactic refinement. We will in fact set the "subatom ID" (the values of the polynomials appearing in $\mathcal{B}'$ and not in $\mathcal{B}$) to be the same for all atoms of $\mathcal{B}$.

The L^2-norm of f_3 will be small on every chosen subatom of $\mathcal{B}'$.

Most subatoms will "well-represent" their corresponding atoms from $\mathcal{B}$, in terms of the corresponding conditional expectation of f.

Before stating this formally, let us also take this opportunity to remark that it is possible to decompose several functions $f^{(1)}, \ldots, f^{(R)} : \mathbb{F}^n \to \{0,1\}$ simultaneously with a single polynomial factor, instead of a single function $f : \mathbb{F}^n \to \{0,1\}$ as we have done so far. Alternatively, this could be thought of as decomposing a single "vector" function $f : \mathbb{F}^n \to \{0,1\}^R$. The proofs of the above theorems can be adapted in a straightforward way to establish this.

Theorem 5.5 (Subatom Selection; Theorem 4.12 of [BFL12])**.** *Suppose $\zeta > 0$ and $d \geqslant 1$ is an integer. Let $\eta, \delta : \mathbb{N} \to \mathbb{R}^+$ be arbitrary non-increasing functions, and let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary non-decreasing function. Then, there exist $C = C_{5.5}(\delta, \eta, r, \zeta, R)$ such that the following holds.*

Given $f^{(1)}, \ldots, f^{(R)} : \mathbb{F}^n \to \{0,1\}$, there exist functions $f_1^{(i)}, f_2^{(i)}, f_3^{(i)} : \mathbb{F}^n \to \mathbb{R}$ for all $i \in [R]$, a polynomial factor $\mathcal{B}$ of degree d with atoms denoted by elements of $\mathbb{T}^{|\mathcal{B}|}$, a syntactic refinement $\mathcal{B}' \succeq_{syn} \mathcal{B}$ of degree d with complexity at most C and atoms denoted by elements of $\mathbb{T}^{|\mathcal{B}|} \times \mathbb{T}^{|\mathcal{B}'|-|\mathcal{B}|}$, and an element $s \in \mathbb{T}^{|\mathcal{B}'|-|\mathcal{B}|}$ such that the following is true:

$f^{(i)} = f_1^{(i)} + f_2^{(i)} + f_3^{(i)}$ for every $i \in [R]$.
$f_1^{(i)} = \mathbf{E}[f^{(i)}|\mathcal{B}']$ for every $i \in [R]$.
$\|f_2^{(i)}\|_{U^{d+1}} < \eta(|\mathcal{B}'|)$ for every $i \in [R]$.
For every $i \in [R]$, $f_1^{(i)}$ and $f_1^{(i)} + f_3^{(i)}$ have range $[0,1]$, and $f_2^{(i)}$ and $f_3^{(i)}$ have range $[-1,1]$.
$\mathcal{B}$ and $\mathcal{B}'$ are both r-regular.
for every atom $c \in \mathbb{T}^{|\mathcal{B}|}$ of $\mathcal{B}$, the subatom $c' = (c,s) \in \mathbb{T}^{|\mathcal{B}'|}$ has the property that

$$\mathop{\mathbf{E}}_x[(f_3^{(i)}(x))^2 \mid \mathcal{B}'(x) = (c,s)] < (\delta(|\mathcal{B}|))^2$$

for every $i \in [R]$.

$$\mathop{\mathbf{Pr}}_{\text{atom } c \text{ of } \mathcal{B}}[\exists_{i\in[R]}|\mathbf{E}[f^{(i)}|c] - \mathbf{E}[f^{(i)}|(c,s)]| > \zeta] < \zeta$$

where we denote $\mathbf{E}[f|c] = \mathbf{E}[f(x)|\mathcal{B}(x) = c]$ and $\mathbf{E}[f|(c,s)] = \mathbf{E}[f(x)|\mathcal{B}'(x) = (c,s)]$.

5.2 Big Picture Functions

Suppose we have a function $f : \mathbb{F}^n \to [R]$, and we want to find out whether it induces a particular affine constraint (A, σ), where $A = (L_1, \ldots, L_m)$ is a sequence of affine forms on ℓ variables and $\sigma \in [R]^m$. Now, suppose $\mathbb{F}^n$ is partitioned by a polynomial factor $\mathcal{B}$ defined by polynomials $P_1, \ldots, P_C$ of degrees $d_1, \ldots, d_C$ and depths $k_1, \ldots, k_C$. Then, observe that if $b_1, \ldots, b_m \in \mathbb{T}^C$ denote the atoms of $\mathcal{B}$ containing $L_1(x_1, \ldots, x_\ell), \ldots, L_m(x_1, \ldots, x_\ell)$ respectively, it must be the case that $b_1, \ldots, b_m$ must be consistent with A (as defined in Definition 3.9). Thus, to locate where f might induce (A, σ), we should restrict our search to sequences of atoms consistent with A.

It will be convenient for us to "blur" a given function so as to retain only atom-level information about it.

Definition 5.6. *Given a function $f : \mathbb{F}^n \to [R]$ and a polynomial factor $\mathcal{B}$, the* big picture function *of f is the function $f_{\mathcal{B}} : \mathbb{T}^{|\mathcal{B}|} \to 2^{[R]}$, where $2^{[R]}$ denotes the power set of R, defined by $f_{\mathcal{B}}(y) = \{f(x) : \mathcal{B}(x) = y\}$. In other words, $f_{\mathcal{B}}(y)$ is the set of all values that f takes within the corresponding atom of $\mathcal{B}$.*

On the other hand, given any function $g : \mathbb{T}^C \to 2^{[R]}$, and a vector of degrees $\mathbf{d} = (d_1, \ldots, d_C)$ and depths $\mathbf{k} = (k_1, \ldots, k_C)$ (which we think of as corresponding to the degrees and depths of some future polynomial factor of complexity C), we will define what it means for such a function to "induce" a copy of a given constraint.

Definition 5.7 (Partially induce)**.** *Suppose we are given vectors $\mathbf{d} = (d_1, \ldots, d_C) \in \mathbb{Z}_{>0}^C$ and $\mathbf{k} = (k_1, \ldots, k_C) \in \mathbb{Z}_{\geqslant 0}^C$, a function $g : \prod_{i\in[C]} \mathbb{U}_{k_i+1} \to 2^{[R]}$, and an induced affine constraint (A, σ) of size m over ℓ variables. We say that g* partially $(\mathbf{d}, \mathbf{k})$-induces *(A, σ) if there exist $\{b_j = (b_{j,1}, \ldots, b_{j,C}) \in \mathbb{T}^C : j \in [m]\}$ such that the following hold.*

For every $i \in [C]$, $(b_{1,i}, b_{2,i}, \ldots, b_{m,i}) \in \mathbb{T}^m$ is (d_i, k_i)-consistent with A

$\sigma_j \in g(b_j)$ for each $j \in [m]$.

The big picture function defined above extracts a finitary description of a function $f : \mathbb{F}^n \to [R]$ in relation to some $\mathcal{B}$, which we will later obtain through a decomposition theorem. Regardless of how we obtained $\mathcal{B}$, moving from an induced constraint of f to a partially induced constraint of the big picture function $f_{\mathcal{B}}$ is always guaranteed.

Observation 5.8. *If $f : \mathbb{F}^n \to [R]$ induces a constraint (A, σ), then for a factor $\mathcal{B}$ defined by polynomials of respective degrees $(d_1, \ldots, d_{|\mathcal{B}|}) = \mathbf{d}$ and respective depths $(k_1, \ldots, k_{|\mathcal{B}|}) = \mathbf{k}$, the big picture function $f_{\mathcal{B}}$ partially $(\mathbf{d}, \mathbf{k})$-induces (A, σ).*

To handle a possibly infinite collection $\mathcal{A}$ of affine constraints, we will employ a compactness argument, analogous to one used in [AS08b] to bound the size of the constraint partially induced by the big picture function. Let us make the following definition:

Definition 5.9 (The compactness function)**.** *Suppose we are given a positive integer C and a possibly infinite collection of induced affine constraints $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots\}$, where each affine constraint (A^i, σ^i) is of size m_i and of complexity at most d. For fixed $\mathbf{d} = (d_1, \ldots, d_C) \in [d]^C$ and $\mathbf{k} = (k_1, \ldots, k_C) \in \left[0, \left\lfloor \frac{d-1}{p-1} \right\rfloor\right]^C$, denote by $\mathcal{G}(\mathbf{d}, \mathbf{k})$ to be the set of functions $g : \prod_{i=1}^{C} \mathbb{U}_{k_i+1} \to 2^{[R]}$ that partially $(\mathbf{d}, \mathbf{k})$-induce some $(A^i, \sigma^i) \in \mathcal{A}$. Now, we define the following function:*

$$\Psi_{\mathcal{A}}(C) = \max_{\mathbf{d},\mathbf{k}} \max_{g \in \mathcal{G}(\mathbf{d},\mathbf{k})} \min_{\substack{(A^i,\sigma^i) \text{ partially} \\ (\mathbf{d},\mathbf{k})\text{-induced by } g}} m_i$$

where the outer max *is over vectors $\mathbf{d} = (d_1, \ldots, d_C) \in [d]^C$ and $\mathbf{k} = (k_1, \ldots, k_C) \in \left[0, \left\lfloor \frac{d-1}{p-1} \right\rfloor\right]^C$. Whenever $\mathcal{G}(\mathbf{d}, \mathbf{k})$ is empty, we set the corresponding maximum to 0.*

Note that the above is indeed finite, as both the number of possible degree and depth sequences (bounded by d^{2C}) and the size of $\mathcal{G}(d_1, \ldots, d_C)$ (bounded by $2^{Rp^{dC}}$) are finite. The compactness function allows to bound an induced constraint in advance, at least (for now) in the realm of big picture functions:

Observation 5.10. *Let $\mathbf{d} \in \mathbb{Z}_{>0}^C, \mathbf{k} \in \mathbb{Z}_{\geqslant 0}^C$ be sequences, for which a function $g : \mathbb{T}^C \to 2^{[R]}$ partially $(\mathbf{d}, \mathbf{k})$-induces some constraint from $\mathcal{A}$. Then g will necessarily partially induce some $(A^i, \sigma^i) \in \mathcal{A}$ whose size is at most $\Psi_{\mathcal{A}}(|\mathcal{B}|)$.*

Proof. This is immediate, as a g satisfying the above in particular belongs to $\mathcal{G}(\mathbf{d}, \mathbf{k})$. □

5.3 Counting Lemma

A function $f : \mathbb{F}^n \to [-1, 1]$ has low Gowers norm, then in many ways, it behaves like a "pseudo-random" function mapping to $[-1, 1]$. For instance, it means that $\mathbf{E}_x[f(x)] = \|f\|_{U^1}$ is small, so that the sign of the function is positive and negative roughly equal number of times. Using the Cauchy-Schwarz inequality repeatedly leads to the following stronger lemma. Refer to **??** for the term "complexity".

Lemma 5.11 (Counting Lemma)**.** *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 ℓ variables of complexity s. Then:*

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

5.4 Proof of Testability

We prove the main result, Theorem 1.12, in this section. In fact, we will show the following.

Theorem 5.12. *Suppose we are given a possibly infinite collection of labeled affine constraints $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\}$ where each (A^i, σ^i) is an affine constraint of size m_i on ℓ_i variables. Then, there are functions $\ell_{\mathcal{A}}(\cdot)$ and $\delta_{\mathcal{A}}(\cdot)$ such that the following is true for any $\varepsilon \in (0,1)$. If a function $f : \mathbb{F}^n \to [R]$ is ε-far from being $\mathcal{A}$-free, then f induces $\delta \cdot p^{n\ell_i}$ many copies of (A^i, σ^i) where $\ell_i < \ell_{\mathcal{A}}(\varepsilon)$ and $\delta > \delta_{\mathcal{A}}(\varepsilon)$.*

Moreover, if $\mathcal{A}$ is locally characterized, then $\ell_A(\varepsilon)$ is a constant independent of ε.

Theorem 1.12 immediately follows. Consider the following test: choose uniformly at random $x_1, \ldots, x_{\ell_{\mathcal{A}}(\varepsilon)} \in \mathbb{F}^n$, let H denote the affine space $\{x_1 + \sum_{j=2}^{\ell_{\mathcal{A}}(\varepsilon)} c_j x_j : c_j \in \mathbb{F}\}$, and check whether f restricted to H is $\mathcal{A}$-free or not, thus making $\leqslant p^{\ell_{\mathcal{A}}(\varepsilon)}$ queries. By Theorem 5.12, if f is ε-far from $\mathcal{A}$-freeness, this test rejects with probability at least $\delta_{\mathcal{A}}(\varepsilon)$.

Proof of Theorem 5.12:

Preliminaries. Fix a function $f : \mathbb{F}^n \to [R]$ that is ε-far from being $\mathcal{A}$-free. Let d be the maximum complexity of an affine constraint A^i appearing in $\mathcal{A}$. For $i \in [R]$, define $f^{(i)} : \mathbb{F}^n \to \{0,1\}$ so that $f^{(i)}(x)$ equals 1 when $f(x) = i$ and equals 0 otherwise. Additionally, set the following parameters, where $\Psi_{\mathcal{A}} : \mathbb{Z}^+ \to \mathbb{Z}^+$ is the compactness function from Definition 5.9:

$$\begin{aligned}
\alpha(C) &= p^{-2dC\Psi_A(C)} \\
\rho(C) &= r_{2.12}(d, \alpha(C)) \\
\Delta(C) &= \frac{1}{16}\left(\frac{\varepsilon}{8R}\right)^{\Psi_{\mathcal{A}}(C)} \\
\eta(C) &= \frac{1}{8p^{dC\Psi_{\mathcal{A}}(C)}}\left(\frac{\varepsilon}{24R}\right)^{\Psi_{\mathcal{A}}(C)} \\
\zeta &= \frac{\varepsilon}{8R}
\end{aligned}$$

Decomposing by regular factors. Next, apply Theorem 5.5 to the functions $f^{(1)}, f^{(2)}, \ldots, f^{(R)}$ in order to get polynomial factors $\mathcal{B}' \succeq_{syn} \mathcal{B}$ of complexity at most $C_{5.5}(\Delta, d, \rho, \zeta, \eta)$, an element $s \in \mathbb{T}^{|\mathcal{B}'|-|\mathcal{B}|}$, and functions $f_1^{(i)}, f_2^{(i)}, f_3^{(i)} : \mathbb{F}^n \to \mathbb{R}$ for each $i \in [R]$. The sequence of polynomials generating $\mathcal{B}'$ will be denoted by $P_1, \ldots, P_{|\mathcal{B}'|}$. Since $\mathcal{B}'$ is a syntactic refinement, we can assume $\mathcal{B}$ is generated by the polynomials $P_1, \ldots, P_{|\mathcal{B}|}$. Let $C = |\mathcal{B}|$ and $C' = |\mathcal{B}'|$. Note that the number of atoms $\|\mathcal{B}\| < p^{(k_{\max}+1)C} \leqslant p^{dC}$, where $k_{\max} \leqslant \lfloor (d-1)/(p-1) \rfloor$ is the maximum depth of a polynomial in $\mathcal{B}$. Denote the degree of P_i by d_i and the depth of P_i by k_i.

Cleanup. Based on $\mathcal{B}'$ and $\mathcal{B}$, we construct a function $F : \mathbb{F}^n \to [R]$ that is ε-close to f and hence, still violates $\mathcal{A}$-freeness. The "cleaner" structure of F will help us locate the induced constraint violated by f. F is constructed by executing the following steps in order:

1. For every $z \in \mathbb{F}^n$, let $F(z) = f(z)$.

2. For every atom c of $\mathcal{B}$ for which there exists $i \in [R]$ such that $|\mathbf{Pr}[F(x) = i \mid \mathcal{B}(x) = c] - \mathbf{Pr}[F(x) = i \mid \mathcal{B}'(x) = (c,s)]| > \varepsilon/(8R)$, do the following. For every $z \in \mathcal{B}^{-1}(c)$, set $F(z) = \arg\max_{j \in [R]} \mathbf{Pr}[F(x) = j \mid \mathcal{B}'(x) = (c,s)]$, the most popular value inside the subatom (c,s).

3. For every atom c of $\mathcal{B}$, for every $i \in [R]$ such that $0 < \mathbf{Pr}[F(x) = i \mid \mathcal{B}'(x) = (c,s)] < \varepsilon/(8R)$, set $F(z) = \arg\max_{j \in [R]} \mathbf{Pr}_{x \in (c,s)}[F(x) = j]$ for any $z \in F^{-1}(i) \cap \mathcal{B}^{-1}(c)$.

Lemma 5.13. *F is $\varepsilon/2$-close to f, and therefore, F is not $\mathcal{A}$-free.*

Proof. Observe that the second step changes the value of F on at most $\frac{\varepsilon}{8R}\|\mathcal{B}\|$ atoms, since $\mathcal{B}'$ $\varepsilon/(8R)$-represents $\mathcal{B}$ with respect to each $f^{(1)}, \ldots, f^{(R)}$. By Lemma 3.2, each atom occupies at most $\frac{1}{\|\mathcal{B}\|} + \alpha(C)$ fraction of the entire domain. So, the fraction of points whose values changed in the second step is at most $\frac{\varepsilon}{8R}\|\mathcal{B}\| \cdot (\frac{1}{\|\mathcal{B}\|} + \alpha(C)) < \frac{\varepsilon}{4R}$.

The third step doesn't apply to any atom of $\mathcal{B}$ affected by the second step. Therefore, in the third case, if $\mathbf{Pr}[F(x) = i \mid \mathcal{B}'(x) = (c,s)] < \varepsilon/(8R)$, then $\mathbf{Pr}[f(x) = i \mid \mathcal{B}(x) = c] < \varepsilon/(4R)$. Hence, the fraction of the domain modified in the third case is at most $\varepsilon/4$.

The distance of F from f is bounded by $\varepsilon/(4R) + \varepsilon/4 < \varepsilon/2$. ☐

Locating a violated constraint. We now want to use F to "find" the affine constraint induced in f. Letting $\mathbf{d}$ denote the vector $(d_1, \ldots, d_C)$ and $\mathbf{k}$ denote the vector $(k_1, \ldots, k_C)$, we have by Observation 5.8 that the big picture function $F_{\mathcal{B}}$ of F will partially $(\mathbf{d}, \mathbf{k})$-induce some constraint from $\mathcal{A}$, and hence by Observation 5.10, it will partially $(\mathbf{d}, \mathbf{k})$-induce some (A^i, σ^i) for which $m_i \leqslant \Psi_{\mathcal{A}}(|\mathcal{B}|)$. This will be the constraint of which we will find many copies in the original f.

Let $A \stackrel{\text{def}}{=} A^i$, $\sigma \stackrel{\text{def}}{=} \sigma^i$, $m \stackrel{\text{def}}{=} m_i$, and let $\ell \stackrel{\text{def}}{=} \ell_i$. Denote the affine forms in A by $(L_1, \ldots, L_m)$ and the vector σ by $(\sigma_1, \ldots, \sigma_m)$. Since we can assume $\ell_i \leqslant m_i$ (without loss of generality by making a change of variables), we can now define

$$\ell_{\mathcal{A}}(\varepsilon) = \Psi_{\mathcal{A}}(C_{5.5}(\Delta, \eta, \rho, \zeta, R)). \tag{7}$$

Let $b_1 = (b_{1,1}, \ldots, b_{1,C})$, $\ldots$, $b_m = (b_{1,m}, \ldots, b_{m,C}) \in \prod_{i=1}^{C} \mathbb{U}_{k_i+1}$ index the atoms of $\mathcal{B}$ where (A^i, σ^i) is partially $(\mathbf{d}, \mathbf{k})$-induced by $F_{\mathcal{B}}$, the big picture function of the cleanup function F, i.e., $b_1, \ldots, b_m$ are consistent with A, and $\sigma_i \in F_{\mathcal{B}}(b_i)$ for every $j \in [m]$. Also, let $b'_1, \ldots, b'_m \in \prod_{i=1}^{C'} \mathbb{U}_{k_i+1}$ index the associated subatoms of $\mathcal{B}'$, obtained by letting $b'_j = (b_j, s)$ for every $j \in [m]$. Observe that:

Lemma 5.14. *The subatoms $b'_1, \ldots, b'_m$ are consistent with A.*

Proof. Since $b_1, \ldots, b_m$ are already consistent with A, we only need to show that for every $i \in [C+1, C']$, the sequence $(b'_{1,i}, \ldots, b'_{m,i}) = (s_{i-C}, s_{i-C}, \ldots, s_{i-C})$ is (d_i, k_i)-consistent. This holds because a constant function is of degree $\leqslant d_i$. ☐

The main analysis. Our goal is to lower bound:

$$\begin{aligned} &\mathop{\mathbf{Pr}}_{x_1, \ldots, x_\ell \in \mathbb{F}^n} [f(L_1(x_1, \ldots, x_\ell)) = \sigma_1 \wedge \cdots \wedge f(L_m(x_1, \ldots, x_\ell)) = \sigma_m] \\ &= \mathop{\mathbf{E}}_{x_1, \ldots, x_\ell \in \mathbb{F}^n} \left[f^{(\sigma_1)}(L_1(x_1, \ldots, x_\ell)) \cdots f^{(\sigma_m)}(L_m(x_1, \ldots, x_\ell)) \right] \end{aligned} \tag{8}$$

The theorem obviously follows if the above expectation is more than the respective $\delta_{\mathcal{A}}(\varepsilon)$. We rewrite the expectation as:

$$\mathop{\mathbf{E}}_{x_1,\ldots,x_\ell\in\mathbb{F}^n}\left[(f_1^{(\sigma_1)}+f_2^{(\sigma_1)}+f_3^{(\sigma_1)})(L_1(x_1,\ldots,x_\ell))\cdots(f_1^{(\sigma_m)}+f_2^{(\sigma_m)}+f_3^{(\sigma_m)})(L_m(x_1,\ldots,x_\ell))\right]. \tag{9}$$

We can expand the expression inside the expectation as a sum of 3^m terms. The expectation of any term which is a multiple of $f_2^{(\sigma_j)}$ for any $j\in[m]$ has an absolute value upper bound of $\|f_2^{(\sigma_j)}\|_{U^{d+1}}\leqslant\eta(|\mathcal{B}'|)$, because of Lemma 5.11 and the fact that the complexity of A^i is bounded by d. Hence, the expression (9) is at least:

$$\mathop{\mathbf{E}}_{x_1,\ldots,x_\ell}\left[(f_1^{(\sigma_1)}+f_3^{(\sigma_1)})(L_1(x_1,\ldots,x_\ell))\cdots(f_1^{(\sigma_m)}+f_3^{(\sigma_m)})(L_m(x_1,\ldots,x_\ell))\right]-3^m\eta(|\mathcal{B}'|). \tag{10}$$

Before we continue, to ease notation, for the rest of the proof we will now define an indicator function. $\mathcal{I}_{(L_1,\ldots,L_m)}^{(b_1',\ldots,b_m')}(x_1,\ldots,x_\ell)$ will be set to 1 if $\mathcal{B}'(L_j(x_1,\ldots,x_\ell))=b_j'$ for every $j\in[m]$, and it will be set to 0 otherwise.

Now, because of the non-negativity of $f_1^{(\sigma_j)}+f_3^{(\sigma_j)}$ for every $j\in[m]$, the expectation in (10) is at least:

$$\mathop{\mathbf{E}}_{x_1,\ldots,x_\ell}\left[\left(f_1^{(\sigma_1)}+f_3^{(\sigma_1)}\right)(L_1(x_1,\ldots,x_\ell))\cdots\left(f_1^{(\sigma_m)}+f_3^{(\sigma_m)}\right)(L_m(x_1,\ldots,x_\ell))\cdot\mathcal{I}_{(L_1,\ldots,L_m)}^{(b_1',\ldots,b_m')}(x_1,\ldots,x_\ell)\right].$$

In other words, what we are doing now is counting only patterns that arise from the selected subatoms $b_1',\ldots,b_m'$. We next expand the product inside the expectation into 2^m terms. The main contribution will come from:

$$\mathop{\mathbf{E}}_{x_1,\ldots,x_\ell}\left[f_1^{(\sigma_1)}(L_1(x_1,\ldots,x_\ell))\cdots f_1^{(\sigma_m)}(L_m(x_1,\ldots,x_\ell))\cdot\mathcal{I}_{(L_1,\ldots,L_m)}^{(b_1',\ldots,b_m')}(x_1,\ldots,x_\ell)\right]. \tag{11}$$

But first, let us show that the contribution from each of the other 2^m-1 terms is small. Consider a term that contains $f_3^{(\sigma_k)}$ for some $k\in[m]$. Letting g denote an arbitrary function with $\|g\|_\infty\leqslant 1$, such a term is of the form:

$$\mathop{\mathbf{E}}_{x_1,\ldots,x_\ell}\left[f_3^{(\sigma_k)}(L_k(x_1,\ldots,x_\ell))g(x_1,\ldots,x_\ell)\cdot\mathcal{I}_{(L_1,\ldots,L_m)}^{(b_1',\ldots,b_m')}(x_1,\ldots,x_\ell)\right]. \tag{12}$$

By our definition of affine constraints, $L_k(x_1,\ldots,x_\ell)$ is of the form $x_1+\sum_{i\in[\ell]}\alpha_i x_i$ for some $\alpha_i\in\mathbb{F}$. We now change the summation variables of the expectation by replacing x_1 with $z=x_1+\sum_{i\in[2,\ell]}\alpha_i x_i$,. Letting $L_1',\ldots,L_m'$ denote the linear forms as they appear after the change, we first note that $a_k'(z,x_2,\ldots,x_\ell)$ will equal z. We can now bound the square of (12) using Cauchy-Schwarz as:

$$\begin{aligned}
&\left(\mathop{\mathbf{E}}_{x_1,\ldots,x_\ell}\left[f_3^{(\sigma_k)}(L_k(x_1,\ldots,x_\ell))g(x_1,\ldots,x_\ell)\cdot\mathcal{I}_{(L_1,\ldots,L_m)}^{(b_1',\ldots,b_m')}(x_1,\ldots,x_\ell)\right]\right)^2\\
&\leqslant\left(\mathop{\mathbf{E}}_{z,x_2,\ldots,x_\ell}\left[\left|f_3^{(\sigma_k)}(z)\right|\cdot\mathcal{I}_{(a_1',\ldots,a_m')}^{(b_1',\ldots,b_m')}(z,x_2,\ldots,x_\ell)\right]\right)^2\\
&\leqslant\mathop{\mathbf{E}}_{z}\left[|f_3^{(\sigma_k)}(z)|^2\cdot\mathcal{I}_{(\mathrm{id})}^{(b_k')}(z)\right]\cdot\mathop{\mathbf{E}}_{z}\left[\left(\mathop{\mathbf{E}}_{x_2,\ldots,x_\ell}\left[\mathcal{I}_{(a_1',\ldots,a_m')}^{(b_1',\ldots,b_m')}(z,x_2,\ldots,x_\ell)\right]\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \Delta^2(C)\cdot \Pr_z[\mathcal{B}'(z)=b'_k]\cdot \mathop{\mathbf{E}}_z\left[\left(\mathop{\mathbf{E}}_{x_2,\ldots,x_\ell}\left[\mathcal{I}_{(a'_1,\ldots,a'_m)}^{(b'_1,\ldots,b'_m)}(z,x_2,\ldots,x_\ell)\right]\right)^2\right]\\
&\leqslant \Delta^2(C)\cdot\left(\frac{1}{\|\mathcal{B}'\|}+\alpha(C')\right)\cdot \mathop{\mathbf{E}}_z\left[\left(\mathop{\mathbf{E}}_{x_2,\ldots,x_\ell}\prod_{\substack{i\in[C']\\ j\in[m]}}\frac{1}{p^{k_i+1}}\sum_{\lambda_{i,j}=0}^{p^{k_i+1}-1}\mathsf{e}\left(\lambda_{i,j}\cdot(P_i(L'_j(z,x_2,\ldots,x_\ell))-b'_{i,j})\right)\right)^2\right]\\
&\leqslant \frac{2\Delta^2(C)}{\|\mathcal{B}'\|^{2m+1}}\mathop{\mathbf{E}}_z\left[\left(\sum_{\substack{(\lambda_{i,j})\\ \in\prod_{i,j}[0,p^{k_i+1}-1]}}\mathsf{e}\left(-\sum_{\substack{i\in[C']\\ j\in[m]}}\lambda_{i,j}b'_{i,j}\right)\mathop{\mathbf{E}}_{x_2,\ldots,x_\ell}\mathsf{e}\left(\sum_{\substack{i\in[C']\\ j\in[m]}}\lambda_{i,j}P_i(a'_j(z,x_2,\ldots,x_\ell))\right)\right)^2\right]\\
&\leqslant \frac{2\Delta^2(C)}{\|\mathcal{B}'\|^{2m+1}}\sum_{\substack{(\lambda_{i,j}),(\tau_{i,j})\\ \in\prod_{i,j}[0,p^{k_i+1}-1]}}\mathsf{e}\left(-\sum_{\substack{i\in[C']\\ j\in[m]}}\lambda_{i,j}b'_{i,j}\right)\mathsf{e}\left(\sum_{\substack{i\in[C']\\ j\in[m]}}\tau_{i,j}b'_{i,j}\right)\cdot\\
&\qquad\mathop{\mathbf{E}}_{\substack{z,x_2,\ldots,x_\ell\\ y_2,\ldots,y_\ell}}\left[\mathsf{e}\left(\sum_{\substack{i\in[C']\\ j\in[m]}}\lambda_{i,j}P_i(a'_j(z,x_2,\ldots,x_\ell))\right)\mathsf{e}\left(-\sum_{\substack{i\in[C']\\ j\in[m]}}\tau_{i,j}P_i(a'_j(z,y_2,\ldots,y_\ell))\right)\right]\\
&\leqslant \frac{2\Delta^2(C)}{\|\mathcal{B}'\|^{2m+1}}\sum_{\substack{(\lambda_{i,j}),(\tau_{i,j})\\ \in\prod_{i,j}[0,p^{k_i+1}-1]}}\left|\mathop{\mathbf{E}}_{\substack{z,x_2,\ldots,x_\ell\\ y_2,\ldots,y_\ell}}\left[\mathsf{e}\left(\sum_{\substack{i\in[C']\\ j\in[m]}}\lambda_{i,j}P_i(a'_j(z,x_2,\ldots,x_\ell))-\sum_{\substack{i\in[C']\\ j\in[m]}}\tau_{i,j}P_i(a'_j(z,y_2,\ldots,y_\ell))\right)\right]\right|
\end{aligned}
\tag{13}
$$

We can bound the above using Theorem 3.3. Let A' denote the set of $2m$ linear forms: $\left\{L'_j(z,x_2,\ldots,x_\ell)\mid j\in[m]\right\}\cup\left\{L'_j(z,y_2,\ldots,y_\ell)\mid j\in[m]\right\}$. Let Λ_i and Λ'_i denote the (d_i,k_i)-dependency set of A and A' respectively.

Lemma 5.15. *For each i, $|\Lambda'_i|=|\Lambda_i|^2\cdot p^{k_i+1}$*

Proof. Firstly, it's clear that the dependency sets of $\{L_j : j\in[m]\}$ and of $\{L'_j : j\in[m]\}$ are the same. So, redefine Λ'_i to be the (d_i,k_i)-dependency set of $\{L_j(z,x_2,\ldots,x_\ell)\}\cup\{L_j(z,y_2,\ldots,y_\ell)\}$. Recall that $L_1(z,x_2,\ldots,x_\ell)=L_1(z,y_2,\ldots,y_\ell)=z$.

For any $\tau^1,\tau^2\in\Lambda_i$ and any $\alpha\in[0,p^{k+1}-1]$, note that $(\tau^1_1+\alpha \pmod{p^{k+1}},\tau^1_2,\ldots,\tau^1_m,\tau^2_1-\alpha \pmod{p^{k_i+1}},\tau^2_2,\ldots,\tau^2_m)\in\Lambda'_i$. Hence, $|\Lambda'_i|\geqslant|\Lambda_i|^2\cdot p^{k_i+1}$. To show $|\Lambda'_i|\leqslant|\Lambda_i|^2\cdot p^{k_i+1}$, we give a map from Λ'_i to $\Lambda_i\times\Lambda_i$ that is p^{k_i+1}-to-1. Suppose $\sum_{j=1}^m\tau^1_jQ(L_j(z,x_2,\ldots,x_\ell))+\sum_{j=1}^m\tau^2_jQ(L_j(z,y_2,\ldots,y_\ell))\equiv 0$ for every polynomial Q of degree d_i and depth k_i. Now, observe that $\sum_{j=2}^m\tau^1_jQ(L_j(z,x_2,\ldots,x_\ell))$ and $\sum_{j=2}^m\tau^2_jQ(L_j(z,y_2,\ldots,y_\ell))$ must both individually be multiples of $Q(z)$, since otherwise, their sum could not be a multiple of $Q(z)$ because of the presence of x_i's or y_i's. Hence, let $\sum_{j=2}^m\tau^1_jQ(L_j(z,x_2,\ldots,x_\ell))=\alpha_1Q(z)$ and $\sum_{j=2}^m\tau^2_jQ(L_j(z,y_2,\ldots,y_\ell))=\alpha_2Q(z)$, so that $\alpha_1+\tau^1_1+\alpha_2+\tau^2_2=0$. Then, the map $\tau^1\circ\tau^2$ to $((-\alpha_1 \pmod{p^{k_i+1}}),\tau^1_2,\ldots,\tau^1_m),(-\alpha_2 \pmod{p^{k_i+1}}),\tau^2_2,\ldots,\tau^2_m))$ is a map from Λ'_i to $\Lambda_i\times\Lambda_i$. It can be checked that this map is p^{k_i+1}-to-1. ☐

Applying Theorem 3.3 (just as in the proof of Theorem 3.10), we get that (13), and therefore the square of (12), is at most:

$$\frac{2\Delta^2(C)}{\|\mathcal{B}'\|^{2m+1}}\left(\prod_{i=1}^{C'}|\Lambda_i|^2 p^{k_i+1}+\|\mathcal{B}'\|^{2m}\alpha(C')\right)\leqslant 2\Delta^2(C)\left(\frac{\prod_{i=1}^{C'}|\Lambda_i|^2}{\|\mathcal{B}'\|^{2m}}+\alpha(C')\right) \tag{14}$$

Finally, we turn to the main term (11). We already know from Lemma 5.14 that the subatoms $b'_1,\dots,b'_m$ are consistent with A. We can now lower-bound (11) as follows:

$$\begin{aligned}
&\mathop{\mathbf{E}}_{x_1,\dots,x_\ell}\left[f_1^{(\sigma_1)}(L_1(x_1,\dots,x_\ell))\cdots f_1^{(\sigma_m)}(L_m(x_1,\dots,x_\ell))\cdot\mathcal{I}^{(b'_1,\dots,b'_m)}_{(L_1,\dots,L_m)}(x_1,\dots,x_\ell)\right]\\
&=\mathbf{Pr}[\mathcal{B}'(L_1(x_1,\dots,x_\ell))=b'_1\wedge\cdots\wedge\mathcal{B}'(L_m(x_1,\dots,x_\ell))=b'_m]\cdot\\
&\qquad\mathop{\mathbf{E}}_{x_1,\dots,x_\ell}\left[f_1^{(\sigma_1)}(L_1(x_1,\dots,x_\ell))\cdots f_1^{(\sigma_m)}(L_m(x_1,\dots,x_\ell))|\forall j\in[m],\ \mathcal{B}'(L_j(x_1,\dots,x_\ell))=b'_j\right]\\
&\geqslant\left(\frac{\prod_{i=1}^{C'}|\Lambda_i|}{\|\mathcal{B}'\|^m}-\alpha(C')\right)\cdot\left(\frac{\varepsilon}{8R}\right)^m
\end{aligned} \tag{15}$$

Let us justify the last line. The first term is due to the lower bound on the probability from Theorem 3.10. The second term in (15) is because each $f_1^{(\sigma_j)}$ is constant on the atoms of $\mathcal{B}'$, and because by construction, the big picture function $F_{\mathcal{B}}$ of the cleanup function F, on which (A,σ) was partially induced, supports a value inside an atom b of $\mathcal{B}$ only if the original function f acquires the value on at least an $\varepsilon/(8R)$ fraction of the subatom (c,s).

Setting $\beta=(\prod_{i=1}^{C'}|\Lambda_i|/\|\mathcal{B}'\|)^m$ and combining the bounds from (10), (14) and (15), we get that (8) is at least:

$$\begin{aligned}
&(\beta-\alpha(C'))\cdot\left(\frac{\varepsilon}{8R}\right)^m-\sqrt{2\Delta^2(C)\cdot(\beta^2+\alpha(C'))}-3^m\cdot\eta(C')\\
&>\frac{\beta}{2}\cdot\left(\frac{\varepsilon}{8R}\right)^{\Psi_{\mathcal{A}}(C)}-2\beta\cdot\Delta(C)-3^{\Psi_{\mathcal{A}}(C)}\cdot\eta(C')
\end{aligned}$$

Since $\|\mathcal{B}'\|\leqslant p^{dC'}$, $\frac{1}{\|\mathcal{B}'\|^{\Psi_{\mathcal{A}}(C)}}\leqslant\beta\leqslant 1$, $\Delta(C)=\frac{1}{16}(\frac{\varepsilon}{8R})^{\Psi_{\mathcal{A}}(C)}$, $\eta(C')<\frac{1}{8\|\mathcal{B}'\|^{\Psi_{\mathcal{A}}(C)}}\left(\frac{\varepsilon}{24R}\right)^{\Psi_{\mathcal{A}}(C)}$, and both C and C' are upper-bounded by $C_{5.5}(\Delta,\eta,\rho,\zeta,R)$, we can now define

$$\delta_{\mathcal{A}}(\varepsilon)=\frac{1}{4}p^{-d\Psi_{\mathcal{A}}(C_{5.5}(\Delta,\eta,\rho,\zeta,R))C_{5.5}(\Delta,\eta,\rho,\zeta,R)}\cdot\left(\frac{\varepsilon}{8R}\right)^{\Psi_{\mathcal{A}}(C_{5.5}(\Delta,\eta,\rho,\zeta,R))} \tag{16}$$

to conclude the proof.

□

References

[AFKS00] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000. 8, 10, 25

[AFNS06] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 251–260, 2006. 10

[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

[AS08a] Noga Alon and Asaf Shapira. A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. on Comput.*, 37(6):1703–1727, 2008. 3, 4, 8, 10

[AS08b] Noga Alon and Asaf Shapira. Every monotone graph property is testable. *SIAM J. on Comput.*, 38(2):505–522, 2008. 8, 10, 28

[BCL+06] Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós, Balázs Szegedy, and Katalin Vesztergombi. Graph limits and parameter testing. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 261–270, 2006. 10

[BCSX11] Arnab Bhattacharyya, Victor Chen, Madhu Sudan, and Ning Xie. Testing linear-invariant non-linear properties. *Theory of Computing*, 7(1):75–99, 2011. 10

[BFL91] László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991. 1, 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). 8, 9, 10, 25, 26

[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. 3, 4, 8, 10

[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, 2

[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. 8

[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

[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. 1

[FN07] Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM J. Comput.*, 37(2):482–501, 2007. 10

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998. 1

[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

[Gow01] William T. Gowers. A new proof of Szeméredi's theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001. 7

[GT09] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contributions to Discrete Mathematics*, 4(2):1–36, 2009. 8

[GT10] Ben Green and Terence Tao. Linear equations in primes. *Annals of Mathematics*, 171:1753–1850, 2010. 6

[HL11a] 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. 8, 20

[HL11b] Hamed Hatami and Shachar Lovett. Higher-order Fourier analysis of $\mathbb{F}_p^n$ and the complexity of systems of linear forms. *Geometric And Functional Analysis*, 21:1331–1357, 2011. 8

[JPRZ04] Charanjit S. Jutla, Anindya C. Patthak, Atri Rudra, and David Zuckerman. Testing low-degree polynomials over prime fields. In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 423–432, 2004. 3

[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. 17

[KR06] Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM J. on Comput.*, 36(3):779–802, 2006. 3

[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, 9

[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. 10

[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. 8

[Ron09] Dana Ron. Algorithmic and analysis techniques in property testing. *Foundations and Trends in Theoretical Computer Science*, 5(2):73–205, 2009. 1

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. on Comput.*, 25:252–271, 1996. 1, 2

[Rub06] Ronitt Rubinfeld. Sublinear time algorithms. In *Proceedings of International Congress of Mathematicians 2006*, volume 3, pages 1095–1110, 2006. 1

[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. 10

[Sud10] Madhu Sudan. Invariance in property testing. Technical Report 10-051, Electronic Colloquium in Computational Complexity, March 2010. 1

[Sze75] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. 7

[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. 8

[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, 12, 13, 14, 15, 16

ECCC ISSN 1433-8092

http://eccc.hpi-web.de