Electronic Colloquium on Computational Complexity, Report No. 1 (2012)

Testing Low Complexity Affine-Invariant Properties

Arnab Bhattacharyya*
Princeton University
arnabb@princeton.edu

Eldar Fischer†
Technion
eldar@cs.technion.ac.il

Shachar Lovett‡
Institute for Advanced Study
slovett@math.ias.edu

January 1, 2012

Abstract

Invariance with respect to linear or affine transformations of the domain is arguably the most common symmetry exhibited by natural algebraic properties. In this work, we show that any *low complexity* affine-invariant property of multivariate functions over finite fields is testable with a constant number of queries. This immediately reproves, for instance, that the Reed-Muller code over $\mathbb{F}_p$ of degree $d < p$ is testable, with an argument that uses no detailed algebraic information about polynomials, except that low degree is preserved by composition with affine maps.

The complexity of an affine-invariant property $\mathcal{P}$ refers to the maximum complexity, as defined by Green and Tao (Ann. Math. 2008), of the sets of linear forms used to characterize $\mathcal{P}$. A more precise statement of our main result is that for any fixed prime $p \geq 2$ and fixed integer $R \geq 2$, any affine invariant-property $\mathcal{P}$ of functions $f : \mathbb{F}_p^n \to [R]$ is testable, assuming the complexity of the property is less than p. Our proof involves developing analogs of graph-theoretic techniques in an algebraic setting, using tools from higher-order Fourier analysis.

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.

*Center for Computational Intractability. Supported by NSF Grants CCF-0832797, 0830673, and 0528414.
†Supported in part by an ERC-2007-StG grant number 202405.
‡Supported by NSF grant DMS-0835373.

ISSN 1433-8092

A fundamental problem in the area is then to find a combinatorial *characterization* of the testable properties. The characterization problem was explicitly raised even in the early work of [GGR98]. In this work, we make progress towards such a characterization for the class of affine-invariant properties of multivariate functions over finite fields. Before stating our results, let us define some useful notions that will be helpful to know throughout this paper.

1.1 Testability and Invariances

Fix a prime $p \geq 2$ and an integer $R \geq 2$ throughout. Given a property $\mathcal{P}$ of functions in $\{\mathbb{F}_p^n \to [R]\}$, we say that $f : \mathbb{F}_p^n \to [R]$ is ϵ-*far* from $\mathcal{P}$ if $\min_{g \in \mathcal{P}} \Pr_{x \in \mathbb{F}_p^n}[f(x) \neq g(x)] > \epsilon$, and we say that it is ϵ-*close* otherwise. $\mathcal{P}$ is said to be *testable* (with one-sided error) if there is a function $q : (0,1) \to \mathbb{Z}^+$ and an algorithm T that, given as input a parameter $\epsilon \in (0,1)$ and oracle access to a function $f : \mathbb{F}_p^n \to [R]$, makes at most $q(\epsilon)$ queries to the oracle for f, always accepts if $f \in \mathcal{P}$ and rejects with probability at least $2/3$ if f is ϵ-far from $\mathcal{P}$.

As an example of a testable property, let us recall the famous result by Blum, Luby and Rubinfeld [BLR93] which started off this whole line of research. They showed that for testing whether a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ is linear or whether it is ϵ-far from linear, it is enough to query the value of f at only $O(1/\epsilon)$ points of the domain.

Linearity, in addition to being testable, is also an example of an *affine-invariant* property. We say that a property $\mathcal{P} \subseteq \{\mathbb{F}_p^n \to [R]\}$ is affine-invariant if it is the case that for any $f \in \mathcal{P}$ and for any affine transformation $A : \mathbb{F}_p^n \to \mathbb{F}_p^n$, it holds that $f \circ A \in \mathcal{P}$. Similarly, a *linear-invariant* property is closed under composition with linear transformations $L : \mathbb{F}_p^n \to \mathbb{F}_p^n$. 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], homogenous polynomials of bounded degree [KS08], and subspace juntas [VX11].

In general, invariance under a large group of symmetries seems to be a common trait of mathematically natural properties, and in particular, affine invariance underlies most interesting properties that one would classify as "algebraic". Kaufman and Sudan in [KS08] made explicit note of this phenomenon and urged a study of the testability of properties with focus on their invariance. In their paper, Kaufman and Sudan showed that *linear* affine-invariant properties are automatically testable but left open the general question. Note that arbitrary affine-invariant properties are not testable; in fact, testing a random affine-invariant property requires querying nearly all of the domain. So, the question becomes: what is the minimal set of restrictions an affine-invariant property must satisfy in order to be testable? In order to state the conjectured answer to this question, as well as our progress here, we need to introduce some more notions.

1.2 Hereditariness and Induced Affine Constraints

We now introduce the subclass of affine-invariant properties which, we believe, captures every testable property.

Definition 1 (Affine subspace hereditary properties) *An affine-invariant property $\mathcal{P}$ is said to be* affine subspace hereditary *if for any $f : \mathbb{F}_p^n \to [R]$ satisfying $\mathcal{P}$, the restriction of f to any affine subspace of $\mathbb{F}_p^n$ also satisfies $\mathcal{P}$.*

Affine subspace hereditariness thus provides something like a uniformity condition, relating the definition of the property for different values of n. Specializing the conjecture in [BGS10] for linear-invariant properties to affine-invariant properties gives the following:

Conjecture 2 ([BGS10]) *Every affine subspace hereditary property is testable.*

Moreover, [BGS10] show that *every* affine-invariant property testable by a "natural" tester is very "close" to an affine subspace hereditary propery[1]. In fact, resolving Conjecture 2 would yield a combinatorial *characterization* of the (natural) one-sided testable affine-invariant properties, similar to the characterization for dense graph properties [AS08a].

Before proceeding, let us give some examples of affine subspace hereditary properties in order to build intuition about how to test them. Consider the property of linearity, by which we mean here that the function is a polynomial of degree at most 1. This is clearly an affine subspace hereditary property. As we remarked earlier, the property is known to be testable. Note that here, we could also have defined linearity as the condition of satisfying the identity $f(x) - f(x+y) - f(x+z) + f(x+y+z) = 0$ for every $x, y, z \in \mathbb{F}_p^n$. This is a "local" characterization of linearity in the sense that the functional equation does not depend on the value of n. Moreover, this characterization automatically suggests a linearity test: pick random $x, y, z \in \mathbb{F}_p^n$ 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 positive integer d. Again, the property is clearly affine subspace hereditary. It is also known to be testable [AKK+05] over finite fields. And just as in the case of linearity, the test arises out of a local characterization for degree d: the $(d+1)$th derivative in every $d+1$ directions at every point should be 0. The test is then to choose a random point and random $d+1$ directions and to check if the $(d+1)$th derivative in the chosen directions at the chosen point is 0 or not.

In fact, one can describe any affine subspace hereditary property using such local characterizations. To state this formally, let us make a useful definition.

Definition 3 (Affine constraints)

- *An* affine constraint of size m on ℓ variables *is a tuple* $A = (a_1, \ldots, a_m)$ *of* m *linear forms* $a_1, \ldots, a_m$ *over* $\mathbb{F}_p$ *on* ℓ *variables, where* $a_1(X_1, \ldots, X_\ell) = X_1$ *and for every* $i \geq 2$, $a_i(X_1, \ldots, X_\ell) = X_1 + \sum_{j=2}^{\ell} c_{i,j} X_j$ *where each* $c_{i,j} \in \mathbb{F}_p$.
- *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* $f : \mathbb{F}_p^n \to [R]$ *is said to be* (A, σ)-free *if there exist no* $x_1, \ldots, x_\ell \in \mathbb{F}_p^n$ *such that* $(f(a_1(x_1, \ldots, x_\ell)), \ldots, f(a_m(x_1, \ldots, x_\ell))) = \sigma$. *On the other hand, if such* $x_1, \ldots, x_\ell$ *exist, we say* f induces (A, σ) at $x_1, \ldots, x_\ell$.

[1] We omit the technical definitions of "natural" and "close" here, 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 essentially the same definitions in a graph-theoretic context.

- *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* $f : \mathbb{F}_p^n \to [R]$ *is said to be* $\mathcal{A}$-free *if it is* (A^i, σ^i)*-free for every* $i \geq 1$.

The connection between affine subspace hereditariness and affine constraints is given by the following proposition, whose (straightforward) proof we omit.

Proposition 4 *An affine-invariant property* $\mathcal{P}$ *is affine 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.*

Thus, resolving Conjecture 2 boils down to showing testability for all $\mathcal{A}$-freeness properties.

1.3 Main Result

We show that $\mathcal{A}$-freeness is testable as long as all affine constraints in $\mathcal{A}$ are of *complexity* less than p. We next define the complexity of an affine constraint, and more generally, of an arbitrary set of linear forms.

Definition 5 (Cauchy-Schwarz complexity, [GT10b]) *Let* $\mathcal{L} = \{L_1, \ldots, L_m\}$ *be a set of linear forms. The* (Cauchy-Schwarz) complexity of $\mathcal{L}$ *is the minimal* s *such that the following holds. For every* $i \in [m]$*, we can partition* $\{L_j\}_{j \in [m] \setminus \{i\}}$ *into* $s+1$ *subsets such that* L_i *does not belong to the linear span of each subset.*

We put our main finding into a theorem.

Theorem 6 (Main theorem) *For any* $\epsilon \in (0,1)$ *and for any (possibly infinite) fixed collection* $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\}$ *of induced affine constraints such that each* A^i *has complexity less than* p*, there is a function* $q_{\mathcal{A}} : (0,1) \to \mathbb{Z}^+$ *and a one-sided tester which determines whether a function* $f : \mathbb{F}_p^n \to [R]$ *is* $\mathcal{A}$*-free or* ϵ*-far from* $\mathcal{A}$*-free while making at most* $q_{\mathcal{A}}(\epsilon)$ *queries to* f.

The function $q_{\mathcal{A}}$ has rather horrible, Ackermann function-like, dependence on $1/\epsilon$. Our primary concern in this work though is to establish testability, and we make no effort in improving the growth of $q_{\mathcal{A}}$. We note though that recent work by Kalyanasundaram and Shapira [KS11] and by Conlon and Fox [CF11], building on previous work by Gowers [Gow97], suggests that very rapid growth of the query complexity function is inherent with our proof techniques.

Let us lastly note that Theorem 6 is quite nontrivial even when the collection $\mathcal{A}$ is finite. Indeed, even if $\mathcal{A}$ consists only of a single induced affine constraint of complexity > 1, it was not known previously how to show testability. We give more details about past work in Section 1.5.

1.4 Overview of the Proof

Let us give an overview of our proof of Theorem 6. 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 linear forms $(a_1, \ldots, a_m)$, each over ℓ variables, and $\sigma \in [R]^m$. For $i \in [R]$, let $f^{(i)} : \mathbb{F}_p^n \to \{0, 1\}$ be the indicator function for the set $f^{-1}(\{i\})$. Our goal will then be to show that, when f is ϵ-far from (A, σ)-free, then:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[f^{(\sigma_1)}(a_1(x_1, \ldots, x_\ell)) \cdot f^{(\sigma_2)}(a_2(x_1, \ldots, x_\ell)) \cdots f^{(\sigma_m)}(a_m(x_1, \ldots, x_\ell))\right] \tag{1}$$

is at least $\delta(\epsilon)$, where crucially, δ does not depend on n. If we could show this, then we would be done since a valid test would be to repeat the following procedure $O(1/\delta)$ times: uniformly pick $x_1, \ldots, x_\ell \in \mathbb{F}_p^n$ and reject if $(f(a_1(x_1, \ldots, x_\ell)), \ldots, f(a_m(x_1, \ldots, x_\ell))) = \sigma$.

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* $\|\cdot\|_{U^d}$ for a parameter $d \geq 1$, which allows us to say the following about (1). If, for a suitably large value of d,

$$\|f^{(\sigma_j)}\|_{U^d} < \epsilon$$

for some $j \in [m]$, then the entire expectation in (1) is less than ϵ.

This observation leads to the study of *decomposition theorems*, that express an arbitrary function as a linear combination of functions which have either small Gowers norm or are structured in some sense. This is an extension of classical Fourier analysis over $\mathbb{F}_p^n$ where a function is expressed as a linear combination of a small number of characters with high Fourier mass plus a small error term. To deal with Gowers norm, the "characters" need to be exponentials of not linear functions, as in classical Fourier analysis, but of higher degree polynomials. Approximate orthogonality among these "characters" was established by Green and Tao in [GT09] and by Kaufman and Lovett in [KL08]. At this stage, one might expect that results by Hatami and Lovett [HL11a, HL11b] can allow us to use orthogonality to approximate the expectation of the form in (1).

Unfortunately, the proof does not follow that easily from [HL11a]. There are two main reasons for this. The first is that the only information we have about the original function f is ϵ-farness from (A, σ)-freeness. Information about correlation, as was assumed in [HL11a], allows more straightforward application of the higher-order Fourier analytic tools. We use ideas from previous work on graph property testing, as in [AFKS00] and [AS08b], to locate regions of the domain in which we are guaranteed to find at least one induced occurrence of (A, σ). This leads to interesting formulations of the decomposition theorems which might be of independent interest.

The second problem we face is one which also arose in a work by Green and Tao on decomposition theorems (a.k.a., regularity lemmas) over the integers [GT10a]. Namely, the decomposition theorem we use decomposes an arbitrary function $f : \mathbb{F}_p^n \to \mathbb{R}$ as a sum of three functions f_1, f_2, f_3. f_1 consists of the approximate "characters" as mentioned above, f_2 has small Gowers norm, and f_3 has low L^2-norm. Now, the closeness to orthogonality for f_1 and the smallness of the Gowers norm for f_2 decreases as a function of the "complexity" of the decomposition, 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. The way we get around this is by making the norm decrease as a function of the complexity of a coarser decomposition, and we show that this is enough for our purposes.

1.5 Previous Work

This work is part of a sequence of works investigating the relationship between invariance and testability of properties. As described, Kaufman and Sudan [KS08] initiated the program. 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).

Progress has been significantly slower for the non-monotone properties. Bhattacharyya, Grigorescu, and Shapira proved in [BGS10] that linear-invariant properties of functions in $\{\mathbb{F}_2^n \to \{0,1\}\}$ are testable if the complexity of the property is 1. When restricted to affine-invariant properties, the result of [BGS10] is a special case of the main result here for $p = 2$. The previous works did not explicitly use higher-order Fourier analysis; [KSV12] and [Sha09] used variants of the hypergraph regularity lemma which are similar in spirit to higher-order Fourier analysis but are somewhat harder to manipulate due to the lack of analytic tools.

Higher-order Fourier analysis began with the work of Gowers [Gow98] and parallel ergodic-theoretic work by Host and Kra [HK05]. Applications to analytic number theory inspired much more study by Gowers, Green, Tao, Wolf, and Ziegler among others. A book in preparation by Tao [Tao11] surveys the current theory of higher-order Fourier analysis. Our work in this paper relies on decomposition theorems over finite fields of the type first explicitly described by Green in [Gre07].

At a high level, the argument to prove our main theorem mirrors ideas used in a sequence of works by Alon et al. [AFKS00, AS08b, AS08a, AFNS06] 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]. However, implementing these graph-theoretic techniques using higher-order Fourier analysis required several new ideas which, we hope, can be extended to prove Conjecture 2.

1.6 Further research

We study affine subspace hereditary properties, and show that if they are defined by affine constraints of low complexity then they are locally testable. There are several obvious possible generalizations to this work:

1. Remove the condition that the field size is larger than the complexity of the affine forms; this requires non-trivial generalizations of several technical lemmas to small fields.

2. Handle all linear invariant properties (and not just affine invariant properties).

A third generalization, which we suspect may be too strong to hold, is to remove the low complexity assumption on the linear forms. Indeed, in several analogs of this line of research in hypergraph testing, this requirement is analogous to requiring bounded uniformity from the hypergraphs, which is implicitly assumed in all previous works on hypergraph testing.

1.7 Organization

The proof of our main result requires some technical preparation. In Section 2, we describe useful arithmetic decomposition theorems and prove extensions that are helpful for proving testability. In Section 3, we show that we can accurately count the number of linear structures localized to particular cells of the decomposition. Finally, in Section 4, we complete the proof of Theorem 6.

2 Decomposition into Polynomial Factors

In this section, we build the arithmetic decomposition theorems which we will need later on. Defining the decompositions requires us to introduce the *Gowers norm* first.

2.1 The Gowers Norm

We define Gowers norms in the general setting of arbitrary finite Abelian groups.

Definition 7 (Gowers norm) *Let G be a finite abelian group and $f : G \to \mathbb{C}$. For an integer $k \geq 1$, the k'th Gowers norm of f, denoted $\|f\|_{U^k}$ is defined by:*

$$\|f\|_{U^k}^{2^k} = \mathop{\mathbb{E}}_{x,y_1,y_2,\ldots,y_k \in G} \left[\prod_{S \subseteq [k]} \mathcal{C}^{k-|S|} f\left(x + \sum_{i \in S} y_i\right) \right]$$

where $\mathcal{C}$ denotes the complex conjugation operator.

Two facts about the Gowers norm will be absolutely crucial in what follows. First is the Gowers Inverse theorem, established by [BTZ10, TZ10]. Throughout, we let $\mathsf{e}(x)$ denote the complex number $e^{2\pi i x/p}$ for $x \in \mathbb{F}_p$.

Theorem 8 (Gowers Inverse Theorem) *Given a positive integer $d < p$, for every $\delta > 0$, there exists $\epsilon = \epsilon_8(\delta)$ such that if $f : \mathbb{F}_p^n \to \mathbb{R}$ satisfies $\|f\|_\infty \leq 1$ and $\|f\|_{U^{d+1}} \geq \delta$, then there exists a polynomial $P : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree at most d so that $|\mathbb{E}_x[f(x) \cdot \mathsf{e}(P(x))]| \geq \epsilon$.*

The second is a lemma due to Green and Tao [GT10b] based on repeated applications of the Cauchy-Schwarz inequality. Refer to Definition 5 for the term "complexity".

Lemma 9 *Let $f_1, \ldots, f_m : \mathbb{F}_p^n \to [-1, 1]$. Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of m linear forms in ℓ variables of complexity s. Then:*

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

2.2 Decomposition Theorems

While partitioning the domain to affine linear subspaces would be the most intuitive for counting affine cubes, we in fact need higher degree algebraic partitions.

Definition 10 (Polynomial factor) *A* polynomial factor $\mathcal{B}$ *is a sequence of polynomials* $P_1, \ldots, P_C : \mathbb{F}_p^n \to \mathbb{F}_p$. *We also identify it with the function* $\mathcal{B} : \mathbb{F}_p^n \to \mathbb{F}_p^C$ *sending* x *to* $P_1(x), \ldots, P_C(x)$. *A* cell *of* $\mathcal{B}$ *is a preimage* $\mathcal{B}^{-1}(y)$ *for some* $y \in \mathbb{F}_p^C$. *The* partition induced by $\mathcal{B}$ *is the partition of* $\mathbb{F}_p^n$ *given by* $\left\{\mathcal{B}^{-1}(y) : y \in \mathbb{F}_p^C\right\}$. *The* complexity *of* $\mathcal{B}$ *is the number of defining polynomials* $|\mathcal{B}| = C$. *The* degree *of* $\mathcal{B}$ *is the maximum degree among its defining polynomials* $P_1, \ldots, P_C$.

Next, we define the notion of conditional expectation with respect to a given factor.

Definition 11 (Expectation over polynomial factor) *Given a factor* $\mathcal{B}$ *and a function* $f : \mathbb{F}_p^n \to \{0,1\}$, *the* expectation *of* f *over a cell* $y \in \mathbb{F}_p^{|\mathcal{B}|}$ *is the average* $\mathbb{E}[f|_{\mathcal{B}^{-1}(y)}]$, *which we denote by* $\mathbb{E}[f|y]$. *The* conditional expectation *of* f *over* $\mathcal{B}$, *is the real-valued function over* $\mathbb{F}_p^n$ *given by* $\mathbb{E}[f|\mathcal{B}](x) = \mathbb{E}[f|\mathcal{B}(x)]$. *In particular, it is constant on each atom of the polynomial factor.*

The decomposition theorems will iteratively partition the domain $\mathbb{F}_p^n$ into finer and finer partitions. We will need to be careful about distinguishing between two different types of refinements.

Definition 12 (Refinement of polynomial factor) $\mathcal{B}'$ *is called a* syntactic refinement *of* $\mathcal{B}$, *and denoted* $\mathcal{B}' \preceq_{syn} \mathcal{B}$, *if the sequence of polynomials defining* $\mathcal{B}'$ *extends that of* $\mathcal{B}$. *It is called a* semantic refinement, *and denoted* $\mathcal{B}' \preceq_{sem} \mathcal{B}$ *if the induced partition is a combinatorial refinement of the partition induced by* $\mathcal{B}$. *In other words, if for every* $x, y \in \mathbb{F}_2^n$, $\mathcal{B}'(x) = \mathcal{B}'(y)$ *implies* $\mathcal{B}(x) = \mathcal{B}(y)$. *The relation* $\preceq$ *(without subscripts) is a synonym for* $\preceq_{syn}$.

Clearly, being a syntactic refinement is stronger than being a semantic refinement. However in essence, these are almost the same thing.

Observation 13 *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* $\mathbb{F}_p^n$, *and for which* $|\mathcal{B}''| \leq |\mathcal{B}'| + |\mathcal{B}|$.

Proof: Just add the defining polynomials of $\mathcal{B}$ to those of $\mathcal{B}'$. ■

We can now describe our basic decomposition theorem[2].

[2]Most theorems in this section are implicit in previous work by Green and Tao (and explicit in [HL11a]). The exceptions are our remarks that pertain to the distinction between syntactic and semantic refinements and Theorem 22 and its Corollary. In any case, for completeness, we give the missing proofs for the decomposition theorems in the appendix.

Theorem 14 (Basic Decomposition Theorem) *Suppose $\delta > 0$ and $d, C_0 \geq 1$ are integers so that $d < p$. Let $\eta : \mathbb{N} \to \mathbb{R}^+$ be an arbitrary non-increasing function. Then there exist $N = N_{14}(\delta, \eta, d, p)$ and $C = C_{14}(\delta, \eta, d, C_0)$ such that the following holds.*

Given $f : \mathbb{F}_p^n \to \{0,1\}$ 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 $f_1, f_2, f_3 : \mathbb{F}_p^n \to \mathbb{R}$ and a polynomial factor $\mathcal{B} \preceq \mathcal{B}_0$ of degree at most d and complexity at most C such that the following hold:

- $f = f_1 + f_2 + f_3$
- $f_1 = \mathbb{E}[f|\mathcal{B}]$
- $\|f_2\|_{U^{d+1}} \leq \eta(|\mathcal{B}|)$
- $\|f_3\|_2 \leq \delta$
- *f_1 and $f_1 + f_3$ have range $[0,1]$; f_2 and f_3 have range $[-1,1]$.*

This basic decomposition theorem is not enough for our purposes though. Recall that our goal is to control expectations of the form[3]:

$$\mathop{\mathbb{E}}_{\mathbf{x}}\left[\prod_i f(L_i(\mathbf{x}))\right]$$

for some function $f : \mathbb{F}_p^n \to \{0,1\}$ and linear forms L_i over $\mathbb{F}_p$. The f in the above expectation will arise from an application of a decomposition theorem. If $f = f_2$ from Theorem 14 and so $\|f\|_{U^{d+1}}$ is small (for an appropriate d), then the expectation is already small by Lemma 9. The next decomposition theorem will allow us to control the expectation when $f = f_1$ from Theorem 14. It turns out that in order to do so, one needs the polynomials defining the factor to be independent in a strong sense. To this end, we define the *rank* of a polynomial factor [GT09].

Definition 15 (Rank of polynomial factors) *Suppose $\mathcal{B}$ is a polynomial factor defined by polynomials $P_1, \ldots, P_C : \mathbb{F}_p^n \to \mathbb{F}_p$. The* rank *of $\mathcal{B}$ is the largest integer r such that for every $(\alpha_1, \ldots, \alpha_C) \in \mathbb{F}_p^C \setminus \{0^C\}$, the polynomial $P_\alpha = \sum_{i=1}^C \alpha_i P_i$ cannot be expressed as a function of r polynomials of degree $d-1$, where $d = \max_{i \in [C]:\alpha_i \neq 0} \deg(P_i)$.*

The following result, proved by Kaufman and Lovett [KL08] for all p (extending previous work of Green and Tao [GT10b] over large characteristic fields), is crucial:

Theorem 16 *Suppose $\epsilon > 0$ and integer $d \geq 1$. Then, there exists $r = r_{16}(d, \epsilon)$ such that: If $P : \mathbb{F}_p^n \to \mathbb{F}_p$ is a degree-d polynomial with rank at least r, then $|\mathbb{E}_x[\mathsf{e}(P(x))]| < \epsilon$.*

As an example of how useful Theorem 16 is, consider the following simple lemma which states that every cell of a polynomial factor with large enough rank has approximately the same size.

[3] Actually, in the expectations we will analyze, each multiplicand in the product will involve a different "f" but this does not pose additional issues.

Lemma 17 *Given a polynomial factor $\mathcal{B}$ of degree d, complexity C, and rank at least $r_{16}(d,\epsilon)$ generated by the polynomials $P_1,\ldots,P_C:\mathbb{F}_p^n\to\mathbb{F}_p$, and an element $b\in\mathbb{F}_p^C$, we have that:*

$$\Pr_{x\in\mathbb{F}_p^n}[\mathcal{B}(x)=b]=p^{-C}\pm\epsilon$$

Proof: This is implicit in previous work, e.g. [Gre07]. For completeness, we repeat the argument:

$$\begin{aligned}\Pr_{x\in\mathbb{F}_p^n}[\mathcal{B}(x)=b]&=\mathop{\mathbb{E}}_x\left[\prod_{i\in[C]}\frac{1}{p}\sum_{\lambda_i\in\mathbb{F}_p}\mathsf{e}\left(\lambda_i\cdot(P_i(x)-b_i)\right)\right]\\&=p^{-C}\sum_{\lambda_i\in\mathbb{F}_p:i\in[C]}\mathop{\mathbb{E}}_x\left[\mathsf{e}\left(\sum_{i\in[C]}\lambda_i(P_i(x)-b_i)\right)\right]\\&=p^{-C}\left(1\pm p^C\epsilon\right)\end{aligned}$$

where the last line uses Theorem 16 whenever not all the λ_i equal 0. ■

Notice that for Lemma 17 to be nontrivial, we will want ϵ to be smaller than p^{-C}, and hence, the rank of the factor to be large as a function of C. The following theorem shows that one can indeed make the rank larger than an arbitrary function of the complexity, without making the complexity superconstant.

Lemma 18 ([GT09]) *Let $r:\mathbb{Z}^+\to\mathbb{Z}^+$ be a non-decreasing function. Then, there is a function $\tau_r:\mathbb{Z}^+\to\mathbb{Z}^+$ such that the following is true. If $\mathcal{B}$ is a polynomial factor with complexity at most C, then there is a semantic refinement $\mathcal{B}'\preceq_{sem}\mathcal{B}$ with complexity $C'\le\tau_r(C)$ and rank $>r(C')$.*

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

With Lemma 18 in hand, we can strengthen the Basic Decomposition Theorem so that the polynomials defining the factor are of arbitrarily high rank.

Theorem 19 (Strong Decomposition Theorem) *Suppose $\delta>0$ and $d,C_0\ge1$ are integers so that $d<p$. 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 $N=N_{19}(\delta,\eta,r,d,p)$ and $C=C_{19}(\delta,\eta,r,d,C_0)$ such that the following holds.*

Given $f:\mathbb{F}_p^n\to\{0,1\}$ 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 $f_1,f_2,f_3:\mathbb{F}_p^n\to\mathbb{R}$ and a polynomial factor $\mathcal{B}\preceq_{sem}\mathcal{B}_0$ of degree at most d and complexity at most C such that the following hold:

- $f=f_1+f_2+f_3$
- $f_1=\mathbb{E}[f|\mathcal{B}]$
- $\|f_2\|_{U^{d+1}}\le1/\eta(|\mathcal{B}|)$

- $\|f_3\|_2 \leq \delta$
- *f_1 and $f_1 + f_3$ have range $[0,1]$; f_2 and f_3 have range $[-1,1]$.*
- *$\mathcal{B}$ is of rank at least $r(|\mathcal{B}|)$*

Moreover, if $\mathcal{B}_0$ is a syntactic refinement of some $\hat{\mathcal{B}}$ of rank at least $r(C) + C$, then $\mathcal{B}$ will also be a syntactic refinement of $\hat{\mathcal{B}}$ (in particular, if $\mathcal{B}_0 = \hat{\mathcal{B}}$).

It turns out though that this Strong Decomposition Theorem is still not enough 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 we cannot achieve this. The way we resolve the issue is by using the non-negativity of f_1 and $f_1 + f_3$ to *localize* our analysis to certain cells c' so that there is a guarantee that the L^2-norm of f_3 conditioned inside c' (i.e., $\mathbb{E}_{x \in c'}[|f_3(x)|^2]$) is small. Inspired by [AFKS00], we choose these cells c' to be atoms from a polynomial factor $\mathcal{B}'$, with each cell c' contained inside a cell c of a coarser factor $\mathcal{B}$. To make the localization argument work, we will want that the function $f : \mathbb{F}_p^n \to \{0,1\}$ in consideration have roughly the same density on c and c'. To this end, we make the following definition:

Definition 20 (Polynomial factor represents another factor) *Given a function $f : \mathbb{F}_p^n \to \{0,1\}$, 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 cells c of $\mathcal{B}$, more than ζ fraction of the cells c' lying inside c satisfy $|\mathbb{E}[f|c] - \mathbb{E}[f|c']| > \zeta$.*

Using a standard defect version of Cauchy-Schwarz yields:

Observation 21 *If $f : \mathbb{F}_p^n \to \{0,1\}$ and $\mathcal{B}'$ is a refinement of $\mathcal{B}$ which is not ζ-representing with respect to f, then $\mathbb{E}[(\mathbb{E}[f|\mathcal{B}'])^2] \geq \mathbb{E}[(\mathbb{E}[f|\mathcal{B}])^2] + \gamma_{21}(\zeta)$.*

We now put forth a Super Decomposition Theorem and a corollary, which produces two factors $\mathcal{B}' \preceq_{syn} \mathcal{B}$ and picks out a representative subcell of $\mathcal{B}'$ inside each cell of $\mathcal{B}$ so that, as described above, the L^2-norm of f_3 inside each subcell is sufficiently small.

Theorem 22 (Super Decomposition Theorem) *Suppose $\zeta > 0$ and $d, C_0 \geq 1$ are integers so that $d < p$. 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 $N = N_{22}(\delta, \eta, r, d, \zeta, C_0)$ and $C = C_{22}(\delta, \eta, r, d, \zeta, C_0)$ such that the following holds.*

Given $f : \mathbb{F}_p^n \to \{0,1\}$ where $n > N$ and a polynomial factor $\mathcal{B}_0$ of degree at most d and complexity at most C_0, there exist functions $f_1, f_2, f_3 : \mathbb{F}_p^n \to \mathbb{R}$, 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:

- $f = f_1 + f_2 + f_3$
- $f_1 = \mathbb{E}[f|\mathcal{B}']$

- $\|f_2\|_{U^{d+1}} \leq \eta(|\mathcal{B}'|)$
- $\|f_3\|_2 \leq \delta(|\mathcal{B}|)$
- *f_1 and $f_1 + f_3$ have range $[0,1]$; f_2 and f_3 have range $[-1,1]$.*
- *$\mathcal{B}$ is of rank at least $r(|\mathcal{B}|)$.*
- *$\mathcal{B}'$ is of rank at least $r(|\mathcal{B}'|)$.*
- *$\mathcal{B}'$ ζ-represents $\mathcal{B}$ with respect to f.*

Proof: Set the following two parameters in order:

$$C'(m) = C_{19}(\delta(m), \eta, r, d, m)$$
$$R(m) = r(C'(m)) + C'(m)$$

Now, apply Theorem 19 to get a factor $\mathcal{B}_1 \preceq_{sem} \mathcal{B}_0$ of complexity at most $C = C_{19}(\delta(1), \eta, R, d, C_0)$. Apply Theorem 19 once again to get a factor $\mathcal{B}_2 \preceq_{sem} \mathcal{B}_1$ of complexity at most $C_{19}(\delta(C), \eta, r, d, C) = C'(C)$ and rank at least $r(C'(C))$. In fact, because of the last case of Theorem 19 and the rank lower bound for $\mathcal{B}_1$, $\mathcal{B}_2$ is actually a *syntactic* refinement of $\mathcal{B}_1$. If we let f_1, f_2, f_3 be the functions resulting from the last application of Theorem 19 and let $\mathcal{B} = \mathcal{B}_1$ and $\mathcal{B}' = \mathcal{B}_2$, then all the requirements of our Theorem are satisfied, except the last claim about ζ-representation.

To have $\mathcal{B}'$ ζ-represent $\mathcal{B}$, we repeat the above argument. With $\mathcal{B}_2$ taking the place of $\mathcal{B}_0$, we get $\mathcal{B}_4 \preceq_{syn} \mathcal{B}_3 \preceq_{sem} \mathcal{B}_2$. Continuing in this fashion, we get $\mathcal{B}_{i+1} \preceq_{syn} \mathcal{B}_i \preceq_{sem} \mathcal{B}_{i-1}$, for $i \in \{1, 3, 5, \dots\}$. Notice that $1 \geq \mathbb{E}[(\mathbb{E}[f|\mathcal{B}_{i+1}])^2] \geq \mathbb{E}[(\mathbb{E}[f|\mathcal{B}_i])^2] \geq 0$. By the pigeonhole principle, for some odd $i \leq 1/\gamma_{21}(\zeta)$, we must have $\mathbb{E}[(\mathbb{E}[f|\mathcal{B}_{i+1}])^2] - \mathbb{E}[(\mathbb{E}[f|\mathcal{B}_i])^2] \leq \gamma_{21}(\zeta)$, which means that $\mathcal{B}_{i+1}$ ζ-represents $\mathcal{B}_i$ with respect to f by Observation 21. Letting $\mathcal{B} = \mathcal{B}_i$, $\mathcal{B}' = \mathcal{B}_{i+1}$, and letting f_1, f_2, f_3 be the functions resulting from the $(i+1)$th application of Theorem 19 satisfies all our requirements. ■

Corollary 23 (Subatom Selection) *Suppose $\zeta > 0$ and $d \geq 1$ is an integer less than p. 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 $C = C_{23}(\delta, d, r, \zeta, \eta)$ such that the following holds.*

Given $f : \mathbb{F}_p^n \to \{0,1\}$, there exist functions $f_1, f_2, f_3 : \mathbb{F}_p^n \to \mathbb{R}$, a polynomial factor $\mathcal{B}$ with cells denoted by elements of $\mathbb{F}_p^{|\mathcal{B}|}$, a syntactic refinement $\mathcal{B}'$ of $\mathcal{B}$ with complexity at most C and cells denoted by elements of $\mathbb{F}_p^{|\mathcal{B}|} \times \mathbb{F}_p^{|\mathcal{B}'|-|\mathcal{B}|}$, and an element $s \in \mathbb{F}_p^{|\mathcal{B}'|-|\mathcal{B}|}$ such that the following is true:

- $f = f_1 + f_2 + f_3$
- $f_1 = \mathbb{E}[f|\mathcal{B}']$
- $\|f_2\|_{U^{d+1}} < \eta(|\mathcal{B}'|)$
- *f_1 and $f_1 + f_3$ have range $[0,1]$; f_2 and f_3 have range $[-1,1]$.*

- $\mathcal{B}$ *is of rank at least* $r(|\mathcal{B}|)$
- $\mathcal{B}'$ *is of rank at least* $r(|\mathcal{B}'|)$
- *For every* $c \in \mathbb{F}_p^{|\mathcal{B}|}$, *the subcell* $c' = (c, s) \in \mathbb{F}_2^{|\mathcal{B}'|}$ *has the property that* $\mathbb{E}_{x \in \mathcal{B}'^{-1}(c')}[(f_3(x))^2] < (\delta(|\mathcal{B}|))^2$.
- $\Pr_{c \in \mathbb{F}_p^{|\mathcal{B}|}}[|\mathbb{E}[f|c] - \mathbb{E}[f|(c,s)]| > \zeta] < \zeta$

Proof: Let $r'(C) = r_{16}(d, p^{-C}/10)$ do that by Theorem 16, if a polynomial factor $\mathcal{B}$ of degree d and rank at least $r'(|\mathcal{B}|)$, then for any $c \in \mathbb{F}_p^{|\mathcal{B}|}$,

$$0.9\ p^{-|\mathcal{B}|} \leq \Pr_{x \in \mathbb{F}_p^n}[\mathcal{B}(x) = c] \leq 1.1\ p^{-|\mathcal{B}|}.$$

Set $C_{23}(\delta, d, r, \zeta, \eta) = C_{22}(\Delta, \eta, r'', d, \zeta/4, 1)$, where $\Delta(m) = 0.1 \cdot \delta(m)/p^m$ and $r''(C) = \max(r(C), r'(C))$. Apply Theorem 22 with $\mathcal{B}_0$ being the trivial partitioning consisting of one cell. This yields a factor $\mathcal{B}$ with rank at least $r''(|\mathcal{B}|)$, and a syntactic refinement $\mathcal{B}'$ of $\mathcal{B}$ with rank at least $r''(|\mathcal{B}'|)$. Let s be a uniformly chosen random element from $\mathbb{F}_p^{|\mathcal{B}'|-|\mathcal{B}|}$.

Observe that for every cell $c \in \mathbb{F}_p^{|\mathcal{B}|}$ of $\mathcal{B}$, at most $0.1p^{-|\mathcal{B}|}$ fraction of the subcells $c' \in \{c\} \times \mathbb{F}_2^{|\mathcal{B}'|-|\mathcal{B}|}$ of $\mathcal{B}'$ have $\mathbb{E}_{x \in \mathcal{B}'^{-1}(c')}[(f_3(x))^2] > \delta(|\mathcal{B}|)^2$. It's so, since if for even one cell $c \in \mathbb{F}_p^{|\mathcal{B}|}$, this event does not occur, then $\|f_3\|_2^2 = \mathbb{E}_{x \in \mathbb{F}_p^n}[(f_3(x))^2] > \delta(|\mathcal{B}|)^2 \Pr_{x \in \mathbb{F}_p^n}[\mathcal{B}(x) = c] \geq 0.8\ \delta(|\mathcal{B}|)^2/p^{2|\mathcal{B}|} > \Delta(|\mathcal{B}|)^2$, a contradiction to the guarantee of Theorem 22. Hence, by the union bound, with probability at least $3/4$, for every $c \in \mathbb{F}_2^{|\mathcal{B}|}$, the subcell $c' = (c, s)$ has the property that $\mathbb{E}_{x \in \mathcal{B}'^{-1}(c')}[(f_3(x))^2] \leq \delta(|\mathcal{B}|)^2$.

Also, because $\mathcal{B}'$ $\zeta/4$-represents $\mathcal{B}$, the expected number of cells c for which $|\mathbb{E}[f|c] - \mathbb{E}[f|(c,s)]| > \zeta$ is less than $\zeta/4 \cdot p^{|\mathcal{B}|}$. So, with probability at least $3/4$,

$$\Pr_{c \in \mathbb{F}_p^{|\mathcal{B}|}}[|\mathbb{E}[f|c] - \mathbb{E}[f|(c,s)]| > \zeta] < \zeta$$

.

We conclude then that an s exists with both the desired properties. ■

The theorems so far referred to only a single function. Here, we actually require decomposition theorems which work for several functions simultaneously with a single polynomial factor. It is quite straightforward to adapt the previous proofs to get the following.

Theorem 24 (Subatom Selection - Multiple Functions) *Suppose* $\zeta > 0$ *and* $d \geq 1$ *is an integer less than* p. *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* $C = C_{24}(\delta, d, r, \zeta, \eta)$ *such that the following holds.*

Given $f^{(1)}, \ldots, f^{(R)} : \mathbb{F}_p^n \to \{0, 1\}$, *there exist functions* $f_1^{(i)}, f_2^{(i)}, f_3^{(i)} : \mathbb{F}_p^n \to \mathbb{R}$ *for all* $i \in [R]$, *a polynomial factor* $\mathcal{B}$ *with cells denoted by elements of* $\mathbb{F}_p^{|\mathcal{B}|}$, *a syntactic refinement* $\mathcal{B}'$ *of* $\mathcal{B}$ *with complexity at most* C *and cells denoted by elements of* $\mathbb{F}_p^{|\mathcal{B}|} \times \mathbb{F}_p^{|\mathcal{B}'|-|\mathcal{B}|}$, *and an element* $s \in \mathbb{F}_p^{|\mathcal{B}'|-|\mathcal{B}|}$ *such that the following is true:*

- $f^{(i)} = f_1^{(i)} + f_2^{(i)} + f_3^{(i)}$ *for each* $i \in [R]$.
- $f_1^{(i)} = \mathbb{E}[f^{(i)}|\mathcal{B}']$ *for each* $i \in [R]$.
- $\|f_2^{(i)}\|_{U^{d+1}} < \eta(|\mathcal{B}'|)$ *for each* $i \in [R]$.
- $\mathbb{E}_x[(f_3^{(i)}(x))^2 \mid \mathcal{B}'(x) = (c,s)] < (\delta(|\mathcal{B}|))^2$ *for each* $i \in [R]$, $c \in \mathbb{F}_p^{|\mathcal{B}|}$.
- *For each* $i \in [R]$, $f_1^{(i)}$ *and* $f_1^{(i)} + f_3^{(i)}$ *have range* $[0,1]$, *and* $f_2^{(i)}$ *and* $f_3^{(i)}$ *have range* $[-1,1]$.
- $\mathcal{B}'$ *is of rank at least* $r(|\mathcal{B}'|)$
- $\Pr_{c \in \mathbb{F}_p^{|\mathcal{B}|}}[|\mathbb{E}[f^{(i)}|c] - \mathbb{E}[f^{(i)}|(c,s)]| > \zeta] < \zeta$ *for each* $i \in [R]$

3 Counting Patterns inside Cells

Let $\mathcal{B}$ be a polynomial factor generated by the polynomials $P_1, \ldots, P_C : \mathbb{F}_p^n \to \mathbb{F}_p$, and let $b_1, \ldots, b_m \in \mathbb{F}_p^C$ denote the images of m cells of $\mathcal{B}$. We will want to estimate probabilities of the following form:

$$\Pr_{x_1,\ldots,x_\ell}[\mathcal{B}(a_1(x_1,\ldots,x_\ell)) = b_1 \wedge \mathcal{B}(a_2(x_1,\ldots,x_\ell)) = b_2 \wedge \cdots \wedge \mathcal{B}(a_m(x_1,\ldots,x_\ell)) = b_m] \tag{2}$$

where $(a_1, \ldots, a_m)$ is an affine constraint of size m on ℓ variables. In Lemma 17, we analyzed the expectation when $\ell = m = 1$ and $a_1(x_1) = x_1$. In order to deal with the more general form, let us reexpress (2) in the following way:

$$\begin{aligned}
&\Pr_{x_1,\ldots,x_\ell}[\mathcal{B}(L_1(x_1,\ldots,x_\ell)) = b_1 \wedge \cdots \wedge \mathcal{B}(L_m(x_1,\ldots,x_\ell)) = b_m] \\
&= \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in \mathbb{F}_p^n}\left[\prod_{i\in[C]}\prod_{j\in[m]}\frac{1}{p}\sum_{\lambda_{i,j}\in\mathbb{F}_p}\mathsf{e}\left(\lambda_{i,j}\cdot(P_i(a_j(x_1,\ldots,x_\ell) - b_{i,j}))\right)\right] \\
&= p^{-mC}\sum_{\substack{\lambda_{i,j}\in\mathbb{F}_p:\\ i\in[C],j\in[m]}}\mathsf{e}\left(-\sum_{i\in[C]}\sum_{j\in[m]}\lambda_{i,j}b_{i,j}\right)\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[\mathsf{e}\left(\sum_{i\in[C]}\sum_{j\in[m]}\lambda_{i,j}P_i(a_j(x_1,\ldots,x_\ell))\right)\right]
\end{aligned} \tag{3}$$

Hatami and Lovett in [HL11a, HL11b] studied expectations such as those in (3) and proved the following dichotomy.

Lemma 25 (Lemma 5.1 in [HL11b]) *Suppose we are given* $\epsilon \in (0,1)$, *positive integer* $d < p$ *and an affine constraint* $(a_1, \ldots, a_m)$ *of size* m *on* ℓ *variables. Let* $P_1, \ldots, P_C : \mathbb{F}_p^n \to \mathbb{F}_p$ *be a collection of polynomials of degree at most* d *such that the rank of the polynomial factor generated by* $P_1, \ldots, P_C$ *is at least* $r_{16}(d,\epsilon)$. *Then, for every set of coefficients* $\Lambda = \{\lambda_{i,j} \in \mathbb{F}_p : i \in [C], j \in [m]\}$, *if* $P_\Lambda : (\mathbb{F}_p^n)^\ell \to \mathbb{F}_p$ *is the polynomial defined by:*

$$P_\Lambda(X_1,\ldots,X_\ell) = \sum_{i=1}^{C}\sum_{j=1}^{m}\lambda_{i,j}P_i\left(a_j(X_1,\ldots,X_\ell)\right)$$

either P_Λ *is the zero polynomial or else,* $\left|\mathbb{E}_{x_1,\ldots,x_\ell\in\mathbb{F}_p^n}\mathsf{e}\left(P_\Lambda(x_1,\ldots,x_\ell)\right)\right| < \epsilon$.

Thus, to bound (3), we need to count the number of sets Λ such that $P_\Lambda \equiv 0$, in the language of Lemma 25. To this end, let us make the following definition, following the works of Gowers and Wolf [GW10b, GW10a].

Definition 26 (Dimension of linear forms) *For a positive integer d and linear form $L(X_1, \ldots, X_\ell) = \lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_\ell X_\ell$ where $\lambda_1, \ldots, \lambda_\ell \in \mathbb{F}_p$, let the d*th tensor power of *L denote:*

$$L^{\otimes d} \stackrel{\text{def}}{=} \left(\prod_{j=1}^{d} \lambda_{i_j} \ : \ i_1, \ldots, i_d \in [\ell] \right) \in \mathbb{F}_p^{\ell^d}$$

Given positive integers $d_1, \ldots, d_C$ and an affine constraint $A = (a_1, \ldots, a_m)$ of size m on ℓ variables, define the $(d_1, \ldots, d_C)$-dimension of *A to be:*

$$\sum_{i=1}^{C} \dim\left(\left\{ a_1^{\otimes d_i}, \ldots, a_m^{\otimes d_i} \right\} \right)$$

The following lemma shows the relevance of this definition:

Lemma 27 *Let the notation here be same as in Lemma 25. If $d_1, \ldots, d_C$ are the respective degrees of the polynomials $P_1, \ldots, P_C$ and if s is the $(d_1, \ldots, d_C)$-dimension of $(a_1, \ldots, a_m)$, then the number of sets Λ for which $P_\Lambda \equiv 0$ equals p^{mC-s}.*

Proof: Notice that we want to show that the number of sets Λ for which $P_\Lambda \equiv 0$ is dependent just on the degrees of the polynomials $P_1, \ldots, P_C$ and not on any other specifics. The reason we can claim this is the following lemma from [HL11b].

Lemma 28 (Lemma 5.2 in [HL11b]) *Suppose $\lambda_{i,j} \in \mathbb{F}_p$ for $i \in [C], j \in [m]$, and $d_1, \ldots, d_C \in [d]$, where $d < p$. Also, let let $A = \{a_1, \ldots, a_m\}$ be an affine constraint where each linear form a_j is on variables $X_1, \ldots, X_\ell$. Then, one of the following holds:*

- *For every collection of linearly independent polynomials $P_1, \ldots, P_C$ of degree $d_1, \ldots, d_C$ respectively:*
$$\sum_{i=1}^{C} \sum_{j=1}^{m} \lambda_{i,j} P_i\left(a_j(X_1, \ldots, X_\ell)\right) \equiv 0$$
- *For every collection of linearly independent polynomials $P_1, \ldots, P_C$ of degree $d_1, \ldots, d_C$ respectively:*
$$\sum_{i=1}^{C} \sum_{j=1}^{m} \lambda_{i,j} P_i\left(a_j(X_1, \ldots, X_\ell)\right) \not\equiv 0$$

Thus, for the purposes of our claim, instead of having the polynomials $P_1, \ldots, P_C$, we can substitute any collection of linearly independent polynomials of respective degrees $d_1, \ldots, d_C$. In particular, let

us define $P'_i(x) = x_i^{d_i}$ for each $i \in [C]$ (we assume $n > C$). Then, the polynomial $P'_\Lambda(X_1, \ldots, X_\ell) = \sum_{i=1}^{C} \sum_{j=1}^{m} \lambda_{i,j} P'_i\left(a_j(X_1, \ldots, X_\ell)\right)$ is identically zero exactly when for each $i \in [C]$,

$$\sum_{j=1}^{m} \lambda_{i,j} a_j^{\otimes d_i} = 0$$

Usual linear algebra and the definition of $(d_1, \ldots, d_C)$-dimension then shows that the set of Λ's for which $P'_\Lambda \equiv 0$ forms a linear subspace of codimension s. ■

At this point, we can quickly prove the main theorem of this section. For notational convenience, let us make the following definition.

Definition 29 *Given an affine constraint $A = (a_1, \ldots, a_m)$ and positive integers $d_1, \ldots, d_C$, we say that elements $b_1, \ldots, b_m$, where each $b_j = (b_{1,j}, \ldots, b_{C,j}) \in \mathbb{F}_p^C$, are* consistent with respect to *A and $d_1, \ldots, d_C$ if the following is true. For any set $\Lambda = \{\lambda_{i,j} \in \mathbb{F}_p : i \in [C], j \in [m]\}$ for which $\sum_{j \in [m]} \lambda_{i,j} (a_j(X_1, \ldots, X_\ell))^{\otimes d_i}$ equals 0 for all $i \in [C]$, it is the case that $\sum_{j \in [m]} \lambda_{i,j} b_{i,j} = 0$ as well for all $i \in [C]$.*

The Theorem shows that the expectation in (2) is nonzero if and only if $b_1, \ldots, b_m$ are consistent.

Theorem 30 *Let $\epsilon \in (0,1)$, let $A = (a_1, \ldots, a_m)$ be an affine constraint on ℓ variables, and let $\mathcal{B}$ be a polynomial factor of degree d, complexity C and rank at least $r_{16}(d, \epsilon)$ generated by the polynomials $P_1, \ldots, P_C : \mathbb{F}_p^n \to \mathbb{F}_p$. For each $i \in [C]$, let d_i be the degree of P_i. Let s denote the $(d_1, \ldots, d_C)$-dimension of A over $\mathbb{F}_p$. Finally, for each $j \in [m]$, fix the image of a cell in $\mathcal{B}$, indexed by $b_j = (b_{1,j}, \ldots, b_{C,j}) \in \mathbb{F}_p^C$.*

Suppose $b_1, \ldots, b_m$ are consistent with respect to A and $d_1, \ldots, d_C$. Then:

$$\Pr_{x_1, \ldots, x_\ell \in \mathbb{F}_p^n} \left[\mathcal{B}\left(a_j(x_1, \ldots, x_\ell)\right) = b_j \ \forall j \in [m]\right] = p^{-s} \pm \epsilon$$

If $b_1, \ldots, b_m$ are not consistent with respect to A and $d_1, \ldots, d_C$, then the above probability is 0.

Proof: Assume first that the supposition is true. Let us rewrite the probability in question as in (3):

$$p^{-mC} \sum_{\substack{\lambda_{i,j} \in \mathbb{F}_p: \\ i \in [C], j \in [m]}} \mathsf{e}\left(-\sum_{i \in [C]} \sum_{j \in [m]} \lambda_{i,j} b_{i,j}\right) \mathop{\mathbb{E}}_{x_1, \ldots, x_\ell} \left[\mathsf{e}\left(\sum_{i \in [C]} \sum_{j \in [m]} \lambda_{i,j} P_i(a_j(x_1, \ldots, x_\ell))\right)\right]$$

According to Lemma 25, the expectation in the above expression is at most ϵ in absolute value if $\sum_{i \in [C]} \sum_{j \in [m]} \lambda_{i,j} P_i(a_j(X_1, \ldots, X_\ell))$ is not the zero polynomial. On the other hand, by the argument of Lemma 27, if $\sum_{i \in [C], j \in [m]} \lambda_{i,j} P_i(a_j(X_1, \ldots, X_\ell)) \equiv 0$, then $\sum_{i \in [C], j \in [m]} \lambda_{i,j} a_j^{\otimes d_i}$ equals 0. Hence, in this case, by consistency, $\sum_{i \in [C]} \sum_{j \in [m]} \lambda_{i,j} b_{i,j} = 0$, and so, such a choice of $\{\lambda_{i,j}\}$ contributes 1 to the outermost summation. The number of such choices of $\{\lambda_{i,j}\}$ is p^{mC-s} by Lemma 27.

$$\Pr_{x_1, \ldots, x_\ell \in \mathbb{F}_p^n} \left[P_i\left(a_j(x_1, \ldots, x_\ell)\right) = b_{i,j} \ \forall i \in [C], \forall j \in [m]\right] = p^{-mC}(p^{mC-s} \pm p^{mC}\epsilon) = p^{-s} \pm \epsilon$$

The last part of the Theorem follows easily. Suppose the probability in question is nonzero and so, there exist $x_1, \ldots, x_\ell$ so that $\mathcal{B}(a_j(x_1, \ldots, x_\ell)) = b_{i,j}\ \forall i \in [C], \forall j \in [m]$. Then, $\sum_{i\in[C],j\in[m]} \lambda_{i,j} b_{i,j} = \sum_{i\in[C],j\in[m]} \lambda_{i,j} P_i(a_j(x_1, \ldots, x_\ell))$. But, by the argument of Lemma 27, $\sum_{j\in[m]} \lambda_{i,j} P_i(a_j(x_1, \ldots, x_\ell)) \equiv 0$ if $\sum_{j\in[m]} \lambda_{i,j}(a_j(X_1, \ldots, X_\ell))^{\otimes d_i} = 0$ for any $i \in [C]$, and so the supposition is true. ■

4 The Proof of Testability

We prove the main result, Theorem 6, in this section. In fact, we will show the following.

Theorem 31 *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 affine constraint* A^i *is of complexity less than* p *and consists of* m_i *linear forms on* ℓ_i *variables and and each pattern string* $\sigma^i \in [R]^{m_i}$*. Then, there are functions* $N_{\mathcal{A}}(\cdot)$, $\ell_{\mathcal{A}}(\cdot)$ *and* $\delta_{\mathcal{A}}(\cdot)$ *such that the following is true for any* $\epsilon \in (0,1)$*. If a function* $f : \mathbb{F}_p^n \to [R]$ *with* $n > N_{\mathcal{A}}(\epsilon)$ *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* $\ell_i < \ell_{\mathcal{A}}(\epsilon)$ *and* $\delta > \delta_{\mathcal{A}}(\epsilon)$*.*

Theorem 6 immediately follows. Consider the following test: choose uniformly at random $x_1, \ldots, x_{\ell_{\mathcal{A}}(\epsilon)} \in \mathbb{F}_p^n$, let H denote the affine space $\{x_1 + \sum_{j=2}^{\ell_{\mathcal{A}}(\epsilon)} c_j x_j : c_j \in \mathbb{F}_p\}$, and check whether f restricted to H is $\mathcal{A}$-free or not. By Theorem 31, if f is ϵ-far from $\mathcal{A}$-freeness, this test rejects with probability at least $\delta_{\mathcal{A}}(\epsilon)$. Repeating the test $O(1/\delta_{\mathcal{A}}(\epsilon))$ times then guarantees a constant rejection probability. And of course, if f is $\mathcal{A}$-free, the test always accepts.

Proof of Theorem 31: We begin with some preliminaries. Let d be the maximum complexity of an affine constraint A^i appearing in $\mathcal{A}$. By hypothesis, $d < p$. For $i \in [R]$, define $f^{(i)} : \mathbb{F}_p^n \to \{0,1\}$ 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 a fast-growing function that we define later:

$$\begin{aligned}
\alpha(C) &= p^{-2\Psi_A(C)C} \\
\rho(C) &= r_{16}(d, \alpha(C)) \\
\Delta(C) &= \frac{1}{8}\left(\frac{\epsilon}{8R}\right)^{\Psi_{\mathcal{A}}(C)} \\
\eta(C) &= \frac{1}{8(3p)^{C\Psi_{\mathcal{A}}(C)}}\left(\frac{\epsilon}{8R}\right)^{\Psi_{\mathcal{A}}(C)} \\
\zeta &= \frac{\epsilon}{8R}
\end{aligned}$$

Next, apply Theorem 24 to the functions $f^{(1)}, f^{(2)}, \ldots, f^{(R)}$ in order to get polynomial factors $\mathcal{B}' \preceq_{syn} \mathcal{B}$ of complexity at most $C_{24}(\Delta, d, \rho, \zeta, \eta)$, an element $s \in \mathbb{F}_p^{|\mathcal{B}'|-|\mathcal{B}|}$, and functions $f_1^{(i)}, f_2^{(i)}, f_3^{(i)} : \mathbb{F}_p^n \to \mathbb{R}$ 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, $\mathcal{B}$ is generated by the polynomials $P_1, \ldots, P_{|\mathcal{B}|}$.

Based on $\mathcal{B}'$ and $\mathcal{B}$, we construct a function $F : \mathbb{F}_p^n \to [R]$ that is ϵ-close to f and hence, still violates $\mathcal{A}$-freeness. The 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 $z \in \mathbb{F}_p^n$, let $F(z) = f(z)$.
2. For every cell c of $\mathcal{B}$ for which there exists $i \in [R]$ such that $|\Pr[f(x) = i \mid \mathcal{B}(x) = c] - \Pr[f(x) = i \mid \mathcal{B}'(x) = (c,s)]| > \epsilon/(8R)$, do the following. For every $z \in \mathcal{B}^{-1}(c)$, set $F(z) = \arg\max_{j \in [R]} \Pr[f(x) = j \mid \mathcal{B}'(x) = (c,s)]$, the most popular value inside the subatom (c,s).
3. For every cell c of $\mathcal{B}$, for every $i \in [R]$ such that $\Pr[f(x) = i \mid \mathcal{B}'(x) = (c,s)] < \epsilon/(8R)$, set $F(z) = \arg\max_{j \in [R]} \Pr_{x \in (c,s)}[f(x) = j]$ for any $z \in f^{-1}(i) \cap \mathcal{B}^{-1}(c)$.

Lemma 32 *F is $\epsilon/2$-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 $\epsilon/(8R)$ fraction of the cells, since $\mathcal{B}'$ $\epsilon/(8R)$-represents $\mathcal{B}$ with respect to each $f^{(1)}, \ldots, f^{(R)}$. By Lemma 17, each cell occupies at most $p^{-C} + \alpha(C)$ fraction of the entire domain. So, the fraction of points whose values changed in the second step is at most $\frac{\epsilon}{8R} p^C (p^{-C} + \alpha(C)) = \frac{\epsilon}{8R} + \frac{\alpha(C)\epsilon}{8R} p^C < \frac{\epsilon}{4R}$.

The third step doesn't apply to any cell of $\mathcal{B}$ affected by the second step. Therefore, in the third case, if $\Pr[f(x) = i \mid \mathcal{B}'(x) = (c,s)] < \epsilon/(8R)$, then $\Pr[f(x) = i \mid \mathcal{B}(x) = c] < \epsilon/(4R)$. Hence, the fraction of the domain modified in the third case is at most $\epsilon/4$. The distance of F from f is bounded by $\epsilon/(4R) + \epsilon/4 < \epsilon/2$. ■

We now want to use F to find the affine constraint induced in f. We extract a finitary description of F using its structure and then employ a compactness argument (analogous to one used in [AS08b]) to bound the size of the constraint induced by F. Let us make the following two definitions:

Definition 33 (Partially induce) *Suppose we are given positive integers $d_1, \ldots, d_C$, a function[4] $\mathcal{P} : \mathbb{F}_p^C \to 2^{[R]}$, and an induced affine constraint (A, σ) of size m on ℓ variables. We say that $\mathcal{P}$* partially $(d_1, \ldots, d_C)$-induces *(A, σ) if there exist $\{b_j = (b_{1,j}, \ldots, b_{C,j}) \in \mathbb{F}_p^C : j \in [m]\}$ making the following true.*

- *$b_1, \ldots, b_m$ are consistent with respect to A and $d_1, \ldots, d_C$.*
- *$\sigma_j \in \mathcal{P}(b_j)$ for each $j \in [m]$.*

Definition 34 (The function $\Psi_{\mathcal{A}}$) *Suppose we are given positive integer C and a possibly infinite collection of induced affine constraints $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\}$ where each affine constraint A^i is of size m_i and of complexity at most d. For fixed $d_1, \ldots, d_C \in [d]$, denote by $\mathcal{P}(d_1, \ldots, d_C)$ to be the set of functions $\mathcal{P} : \mathbb{F}_p^C \to 2^{[R]}$ that partially $(d_1, \ldots, d_C)$-induce some $(A^i, \sigma^i) \in \mathcal{A}$. Now, we define the following function:*

$$\Psi_{\mathcal{A}}(C) = \max_{d_1, \ldots, d_C \in [d]} \max_{\mathcal{P} \in \mathcal{P}(d_1, \ldots, d_C)} \min_{\substack{(A^i, \sigma^i) \text{ partially} \\ \text{induced by } \mathcal{P}}} m_i$$

($\Psi_{\mathcal{A}}$ is the same function appearing in the parameter settings at the start of the proof.)

Define the function $\mathcal{P} : \mathbb{F}_p^{|\mathcal{B}|} \to 2^{[R]}$ by letting $\mathcal{P}(c) = \{F(x) : x \in \mathcal{B}^{-1}(c)\}$ for each $c \in \mathbb{F}_p^{|\mathcal{B}|}$.

[4] $2^{[R]}$ denotes the set of all subsets of $[R]$.

Lemma 35 *If $d_1, \ldots, d_{|\mathcal{B}|}$ are the respective degrees of the polynomials $P_1, \ldots, P_{|\mathcal{B}|}$ generating $\mathcal{B}$, then $\mathcal{P}$ partially $(d_1, \ldots, d_{|\mathcal{B}|})$-induces some $(A^i, \sigma^i) \in \mathcal{A}$ of size at most $\Psi_{\mathcal{A}}(|\mathcal{B}|)$.*

Proof: Let $(A^i, \sigma^i) \in \mathcal{A}$ be a constraint of minimum size induced by F. The claim is that (A^i, σ^i) is also $(d_1, \ldots, d_{|\mathcal{B}|})$-partially induced by $\mathcal{P}$. To see this, suppose F induces (A^i, σ^i) at $x_1, \ldots, x_{\ell_i}$, and let $c_1, \ldots, c_{m_i} \in \mathbb{F}_p^{|\mathcal{B}|}$ be the images of m_i cells in $\mathcal{B}$ defined by $c_1 = \mathcal{B}(a_1(x_1, \ldots, x_{\ell_i})), c_2 = \mathcal{B}(a_2(x_1, \ldots, x_{\ell_i}))$, $\ldots$, $c_{m_i} = \mathcal{B}(a_{m_i}(x_1, \ldots, x_{\ell_i}))$ where $A^i = (a_1, \ldots, a_{m_i})$. Then, because of the last condition in Theorem 30, it must be the case that $c_1, \ldots, c_m$ are consistent with respect to A^i and $d_1, \ldots, d_{|\mathcal{B}|}$. This fulfills the first condition of Definition 33, and the second condition is true by the definition of $\mathcal{P}$. The bound of $\Psi_{\mathcal{A}}(|\mathcal{B}|)$ on m_i now follows from Definition 34 and the fact that each of $d_1, \ldots, d_{|\mathcal{B}|}$ is at most d. ■

Fix (A^i, σ^i) partially $(d_1, \ldots, d_{|\mathcal{B}|})$-induced by $\mathcal{P}$. Let $m \stackrel{\text{def}}{=} m_i \leq \Psi_{\mathcal{A}}(|\mathcal{B}|)$. Also, let $\ell \stackrel{\text{def}}{=} \ell_i$, and let $\sigma_1, \ldots, \sigma_m$ denote $\sigma_1^i, \ldots, \sigma_m^i$ respectively. As in the above proof of Lemma 35, denote the linear forms in A^i by $a_1, \ldots, a_m$, so that there exist $x_1, \ldots, x_\ell \in \mathbb{F}_p^n$ satisfying $F(a_j(x_1, \ldots, x_\ell)) = \sigma_j$ for each $j \in [m]$. Let $c_1 = (c_{1,1}, \ldots, c_{|\mathcal{B}|,1}), \ldots, c_m = (c_{1,m}, \ldots, c_{|\mathcal{B}|,m}) \in \mathbb{F}_p^{|\mathcal{B}|}$ index the cells of $\mathcal{B}$ where (A^i, σ^i) is induced, i.e., $c_j = \mathcal{B}(a_j(x_1, \ldots, x_\ell))$ for each $j \in [m]$. Also, let $c'_1, \ldots, c'_m \in \mathbb{F}_p^{|\mathcal{B}'|}$ index the associated subcells of $\mathcal{B}'$, obtained by letting $c'_j = c_j \circ s$ for each $j \in [m]$.

Our goal will now be to lower bound:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in \mathbb{F}_p^n} \left[f^{(\sigma_1)}(a_1(x_1, \ldots, x_\ell)) \cdots f^{(\sigma_m)}(a_m(x_1, \ldots, x_\ell)) \right] \tag{4}$$

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

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in \mathbb{F}_p^n} \left[(f_1^{(\sigma_1)} + f_2^{(\sigma_1)} + f_3^{(\sigma_1)})(a_1(x_1, \ldots, x_\ell)) \cdots (f_1^{(\sigma_m)} + f_2^{(\sigma_m)} + f_3^{(\sigma_m)})(a_m(x_1, \ldots, x_\ell)) \right] \tag{5}$$

We can expand the expression inside the expectation as a sum of 3^m terms. The expectation of any term which involves $f_2^{(\sigma_j)}$ for any $j \in [m]$ can be upperbounded in absolute value by $\|f_2^{(\sigma_j)}\|_{U^{d+1}} \leq \eta(|\mathcal{B}'|)$, because of Lemma 9 and the fact that the complexity of A^i is bounded by d. Hence, the expression (5) is at least:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell} \left[(f_1^{(\sigma_1)} + f_3^{(\sigma_1)})(a_1(x_1, \ldots, x_\ell)) \cdots (f_1^{(\sigma_m)} + f_3^{(\sigma_m)})(a_m(x_1, \ldots, x_\ell)) \right] - 3^m \eta(|\mathcal{B}'|) \tag{6}$$

Because of the non-negativity of $f_1^{(\sigma_j)} + f_3^{(\sigma_j)}$ for each $j \in [m]$, the expectation in (6) is at least:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell} \left[\begin{array}{l} \left(f_1^{(\sigma_1)} + f_3^{(\sigma_1)}\right)(a_1(x_1, \ldots, x_\ell)) \cdots \left(f_1^{(\sigma_m)} + f_3^{(\sigma_m)}\right)(a_m(x_1, \ldots, x_\ell)) \cdot \\ \mathbf{1}(\mathcal{B}'(a_j(x_1, \ldots, x_\ell)) = c'_j \ \forall j \in [m]) \end{array} \right] \tag{7}$$

Here $\mathbf{1}$ is the indicator function that is 1 when its input is true and 0 otherwise. In other words, what we are doing now is counting only patterns that arise from the selected subcells $c'_1, \ldots, c'_m$.

We next expand the product inside the expectation into 2^m terms. The main contribution will come from:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[f_1^{(\sigma_1)}(a_1(x_1,\ldots,x_\ell))\cdots f_1^{(\sigma_m)}(a_m(x_1,\ldots,x_\ell))\cdot\mathbf{1}(\mathcal{B}'(a_j(x_1,\ldots,x_\ell))=c'_j\ \forall j\in[m])\right] \tag{8}$$

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 be an arbitrary function with $\|g\|_\infty\le 1$, such a term is of the form:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[f_3^{(\sigma_k)}(a_k(x_1,\ldots,x_\ell))g(x_1,\ldots,x_\ell)\cdot\mathbf{1}(\mathcal{B}'(a_j(x_1,\ldots,x_\ell))=c'_j\ \forall j\in[m])\right] \tag{9}$$

We can assume without loss of generality that $a_k(x_1,\ldots,x_\ell)$ is of the form $x_1+\sum_{i\in[\ell]}\alpha_i x_i$ for some $\alpha\in\mathbb{F}_p^\ell$. By a change of variables $x_1\to x_1-\sum_{i\in[\ell]}\alpha_i x_i$ and letting $a'_1,\ldots,a'_m$ denote the linear forms after the change of variables, we can bound the square of (9) using Cauchy-Schwarz as:

$$\left(\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[f_3^{(\sigma_k)}(a_k(x_1,\ldots,x_\ell))g(x_1,\ldots,x_\ell)\cdot\mathbf{1}(\mathcal{B}'(a_j(x_1,\ldots,x_\ell))=c'_j\ \forall j\in[m])\right]\right)^2 \tag{10}$$

$$\le\left(\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[\left|f_3^{(\sigma_k)}(x_1)\right|\cdot\mathbf{1}(\mathcal{B}'(a'_j(x_1,\ldots,x_\ell))=c'_j\ \forall j\in[m])\right]\right)^2 \tag{11}$$

$$\le\mathop{\mathbb{E}}_{x_1}\left(|f_3^{(\sigma_k)}(x_1)|^2\cdot\mathbf{1}(\mathcal{B}'(x_1)=c'_k)\right)\cdot\mathop{\mathbb{E}}_{x_1}\left(\mathop{\mathbb{E}}_{x_2,\ldots,x_\ell}\left[\mathbf{1}(\mathcal{B}'(a'_j(x_1,\ldots,x_\ell))=c'_j\ \forall j\in[m])\right]\right)^2 \tag{12}$$

$$\le\Delta^2(|\mathcal{B}|)\cdot\Pr_{x_1}[\mathcal{B}'(x_1)=c'_k]\cdot\mathop{\mathbb{E}}_{x_1}\left(\mathop{\mathbb{E}}_{x_2,\ldots,x_\ell}\left[\mathbf{1}(\mathcal{B}'(a'_j(x_1,\ldots,x_\ell))=c'_j\ \forall j\in[m])\right]\right)^2 \tag{13}$$

$$\le\Delta^2(|\mathcal{B}|)\cdot(p^{-|\mathcal{B}'|}+\alpha(|\mathcal{B}'|))\cdot\mathop{\mathbb{E}}_{x_1}\left|\mathop{\mathbb{E}}_{x_2,\ldots,x_\ell}\prod_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\frac{1}{p}\sum_{\lambda_{i,j}\in\mathbb{F}_p}\mathsf{e}\left(\lambda_{i,j}\cdot(P_i(a'_j(x_1,\ldots,x_\ell))-c'_{i,j})\right)\right|^2 \tag{14}$$

$$\le\frac{2\Delta^2(|\mathcal{B}|)}{p^{2|\mathcal{B}'|m+|\mathcal{B}'|}}\mathop{\mathbb{E}}_{x_1}\left|\sum_{\substack{\lambda_{i,j}\in\mathbb{F}_p:\\ i\in[|\mathcal{B}'|],j\in[m]}}\mathsf{e}\left(-\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\lambda_{i,j}c'_{i,j}\right)\mathop{\mathbb{E}}_{x_2,\ldots,x_\ell}\mathsf{e}\left(\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\lambda_{i,j}P_i(a'_j(x_1,x_2,\ldots,x_\ell))\right)\right|^2 \tag{15}$$

$$\le\frac{2\Delta^2(|\mathcal{B}|)}{p^{2|\mathcal{B}'|m+|\mathcal{B}'|}}\sum_{\substack{\lambda_{i,j},\tau_{i,j}\in\mathbb{F}_p:\\ i\in[|\mathcal{B}'|],j\in[m]}}\mathsf{e}\left(-\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\lambda_{i,j}c'_{i,j}\right)\mathsf{e}\left(\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\tau_{i,j}c'_{i,j}\right)$$
$$\mathop{\mathbb{E}}_{\substack{x_1,x_2,\ldots,x_\ell\\ y_2,\ldots,y_\ell}}\mathsf{e}\left(\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\lambda_{i,j}P_i(a'_j(x_1,x_2,\ldots,x_\ell))\right)\mathsf{e}\left(-\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\tau_{i,j}P_i(a'_j(x_1,y_2,\ldots,y_\ell))\right) \tag{16}$$

$$\le\frac{2\Delta^2(|\mathcal{B}|)}{p^{2|\mathcal{B}'|m+|\mathcal{B}'|}}\sum_{\substack{\lambda_{i,j},\tau_{i,j}\in\mathbb{F}_p:\\ i\in[|\mathcal{B}'|],j\in[m]}}\left|\mathop{\mathbb{E}}_{\substack{x_1,x_2,\ldots,x_\ell\\ y_2,\ldots,y_\ell}}\mathsf{e}\left(\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\lambda_{i,j}P_i(a'_j(x_1,x_2,\ldots,x_\ell))-\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\tau_{i,j}P_i(a'_j(x_1,y_2,\ldots,y_\ell))\right)\right| \tag{17}$$

First, observe that the $(d_1, \ldots, d_{|\mathcal{B}'|})$-dimension of $\{a_1, \ldots, a_m\}$ equals the $(d_1, \ldots, d_{|\mathcal{B}'|})$-dimension of $\{a'_1, \ldots, a'_m\}$. For, if there's a linear dependency among $\{a_j^{\otimes d_i} : i \in [|\mathcal{B}'|], j \in [m]\}$, then the same linear dependency exists among the $\{(a'_j)^{\otimes d_i} : i \in [|\mathcal{B}'|], j \in [m]\}$ by making a change of variables, and vice versa.

Let q denote the $(d_1, \ldots, d_{|\mathcal{B}'|})$-dimension of $\{a_1, \ldots, a_m\}$. We can now make the following claim.

Lemma 36 *The $(d_1, \ldots, d_{|\mathcal{B}'|})$-dimension of the $2m$ linear forms*

$$\{a'_1(X_1, X_2, \ldots, X_\ell), \ldots, a'_m(X_1, X_2, \ldots, X_\ell), a'_1(X_1, Y_2, \ldots, Y_\ell), \ldots, a'_m(X_1, Y_2, \ldots, Y_\ell)\}$$

is $2q - |\mathcal{B}'|$.

Proof: Let q_i denote the d_i-dimension of $\{a'_1, \ldots, a'_m\}$, so that $q = \sum_{i=1}^{|\mathcal{B}'|} q_i$. We show that for each $i \in [|\mathcal{B}'|]$, the d_i-dimension of $\{a'_1(X_1, X_2, \ldots, X_\ell), \ldots, a'_m(X_1, X_2, \ldots, X_\ell), a'_1(X_1, Y_2, \ldots, Y_\ell), \ldots, a'_m(X_1, Y_2, \ldots, Y_\ell)\}$ is $2q_i - 1$. We recall that $a'_1(X_1, X_2, \ldots, X_\ell) = a'_1(X_1, Y_2, \ldots, Y_\ell) = X_1$ and all other linear forms are distinct.

Take $S \subseteq \{2, \ldots, m\}$ to be a subset of size $q_i - 1$ such that the q_i elements $\left\{\left(a'_j(X_1, X_2, \ldots, X_m)\right)^{\otimes d_i} : j \in S \cup \{1\}\right\}$ are linearly independent and span $\left\{\left(a'_j(X_1, X_2, \ldots, X_m)\right)^{\otimes d_i} : j \in [m]\right\}$. Clearly, also $\left\{\left(a'_j(X_1, Y_2, \ldots, Y_m)\right)^{\otimes d_i} : j \in S \cup \{1\}\right\}$ are linearly independent and span $\left\{\left(a'_j(X_1, Y_2, \ldots, Y_m)\right)^{\otimes d_i} : j \in [m]\right\}$. Thus, the d_i-rank of the $2m$ linear forms is at most $2q_i - 1$. To conclude, we will show that the d_i-rank is at least $2q_i - 1$. To this end, we will show that

$$\left\{(X_1)^{\otimes d_i}\right\} \cup \left\{\left(a'_j(X_1, X_2, \ldots, X_m)\right)^{\otimes d_i} : j \in S\right\} \cup \left\{\left(a'_j(X_1, Y_2, \ldots, Y_m)\right)^{\otimes d_i} : j \in S\right\}$$

are linearly independent. To see this, note that any element in

$$\operatorname{span}\left\{\left(a'_j(X_1, X_2, \ldots, X_m)\right)^{\otimes d_i} : j \in S \cup \{1\}\right\} \cap \operatorname{span}\left\{\left(a'_j(X_1, Y_2, \ldots, Y_m)\right)^{\otimes d_i} : j \in S \cup \{1\}\right\}$$

must be a multiple of $X_1^{\otimes d_i}$ since no other variable is shared between the forms in the two sets. But on the other hand, by construction of S, $X_1^{\otimes d_i}$ is not in $\operatorname{span}\left\{(a_j(X_1, X_2, \ldots, X_m))^{\otimes d_i} : j \in S\right\}$ nor in $\operatorname{span}\left\{(a_j(X_1, Y_2, \ldots, Y_m))^{\otimes d_i} : j \in S\right\}$. This shows that the d_i-dimension of the $2m$ linear forms is exactly $2q_i - 1$ as claimed. ■

Now, just as in the proof of Theorem 30, the above information is enough to upperbound (17). Lemma 36 and Lemma 27 allows to count the number of $\lambda_{i,j}$ and $\tau_{i,j}$ such that the quantity inside the expectation in (17) is identically 1, and Lemma 25 along with the high-rank condition on the polynomials P_i bounds the expectation otherwise. It follows that (17), and therefore the square of (9), is at most:

$$\frac{2\Delta^2(|\mathcal{B}|)}{p^{2m|\mathcal{B}'|+|\mathcal{B}'|}}\left(p^{2m|\mathcal{B}'|-(2q-|\mathcal{B}'|)} + p^{2m|\mathcal{B}'|}\alpha(|\mathcal{B}'|)\right) \leq 2\Delta^2(|\mathcal{B}|) \cdot (p^{-2q} + \alpha(|\mathcal{B}'|)) \tag{18}$$

Finally, we lowerbound the contribution from the main term (8). To begin with, we need to convince ourselves that f induces at least one copy of (A^i, σ^i) among the subcells $c'_1, \ldots, c'_m$.

Lemma 37 *The subcells $c'_1, \ldots, c'_m$ are consistent with respect to A^i and $d_1, \ldots, d_{|\mathcal{B}'|}$.*

Proof: Recall that for each $j \in [m]$, $c'_j = (c'_{1,j}, c'_{2,j}, \ldots, c'_{|\mathcal{B}'|,j}) \in \mathbb{F}^{|\mathcal{B}'|}$ satisfies $c'_{i,j} = c_{i,j}$ for all $1 \leq i \leq |\mathcal{B}|$ and $c'_{i,j} = s_{i-|\mathcal{B}|+1}$ for all $|\mathcal{B}| < i \leq |\mathcal{B}'|$. Also, recall that we have chosen $c_1, \ldots, c_m$ in such a way they index cells of $\mathcal{B}$ where (A^i, σ^i) is induced by f at least once.

From this last fact, we know that $\Pr_{x_1,\ldots,x_\ell}[\mathcal{B}(a_j(x_1, \ldots, x_\ell)) = c_j \ \forall j \in [m]] > 0$. By the last remark of Theorem 30, it follows that $c_1, \ldots, c_m$ are consistent with respect to A^i and $d_1, \ldots, d_{|\mathcal{B}|}$. This immediately implies that $c'_1, \ldots, c'_m$ are also consistent with respect to A^i and $d_1, \ldots, d_{|\mathcal{B}|}$.

To complete the argument for $i > |\mathcal{B}|$, recall that each a_j is of the form $X_1 + \sum_{r=2}^{\ell} c_r X_r$ for $c_r \in \mathbb{F}_p$. So, whenever $\sum_{j\in[m]} \lambda_{i,j}(a_j)^{\otimes d_i} = 0$ for any $d_i > 0$, we have that $\sum_{j\in[m]} \lambda_{i,j} = 0$, simply by looking at the sum along the coordinate of $a_j^{\otimes d_i}$ corresponding to $X_1^{\otimes d_i}$. Since for any $i > |\mathcal{B}|$, $c'_{i,j} = s_{i-|\mathcal{B}|+1}$ is independent of j, it follows that for any $i > |\mathcal{B}|$, if $\sum_{j\in[m]} \lambda_{i,j}(a_j)^{\otimes d_i} = 0$, then $\sum_{j\in[m]} \lambda_{i,j} c'_{i,j} = s_{i-|\mathcal{B}|+1} \sum_{j\in[m]} \lambda_{i,j} = 0$. ■

We can now lowerbound (8) as follows:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[f_1^{(\sigma_1)}(a_1(x_1,\ldots,x_\ell)) \cdots f_1^{(\sigma_m)}(a_m(x_1,\ldots,x_\ell)) \cdot \mathbf{1}(\mathcal{B}'(a_j(x_1,\ldots,x_\ell)) = c'_j \ \forall j \in [m])\right]$$

$$= \Pr[\mathcal{B}'(a_j(x_1,\ldots,x_\ell)) = c'_j \ \forall j \in [m]] \cdot \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[\begin{array}{l} f_1^{(\sigma_1)}(a_1(x_1,\ldots,x_\ell)) \cdots f_1^{(\sigma_m)}(a_m(x_1,\ldots,x_\ell))| \\ \mathcal{B}'(a_j(x_1,\ldots,x_\ell)) = c'_j \ \forall j \in [m] \end{array}\right] \tag{19}$$

$$\geq (p^{-q} - \alpha(|\mathcal{B}'|)) \cdot \left(\frac{\epsilon}{8R}\right)^m \tag{20}$$

Let us justify the last line. The first term is due to Lemma 37 and the lowerbound on the probability from Theorem 30. The second term in (20) is because each $f_1^{(\sigma_j)}$ is constant on the cells of $\mathcal{B}'$, and because by construction, the function F, on which (A^i, σ^i) was also induced, supports a value inside a cell c of $\mathcal{B}$ only if the original function f acquires the value on at least $\epsilon/(8R)$ fraction of the subcell (c, s).

Combining the bounds from (6), (18) and (20), and using our parameter settings, we get that (4) is at least:

$$(p^{-q} - \alpha(|\mathcal{B}'|)) \cdot \left(\frac{\epsilon}{8R}\right)^m - \sqrt{2\Delta^2(|\mathcal{B}|) \cdot (p^{-2q} + \alpha(|\mathcal{B}'|))} - 3^m \cdot \eta(|\mathcal{B}'|) \tag{21}$$

$$> \frac{p^{-q}}{2} \cdot \left(\frac{\epsilon}{8R}\right)^{\Psi_{\mathcal{A}}(|\mathcal{B}|)} - 2\Delta(|\mathcal{B}|) \cdot p^{-q} - 3^{\Psi_{\mathcal{A}}(|\mathcal{B}|)} \cdot \eta(|\mathcal{B}'|) \tag{22}$$

$$> \frac{p^{-\Psi_{\mathcal{A}}(|\mathcal{B}|)|\mathcal{B}'|}}{4} \cdot \left(\frac{\epsilon}{8R}\right)^{\Psi_{\mathcal{A}}(|\mathcal{B}|)} \tag{23}$$

where both $|\mathcal{B}|$ and $|\mathcal{B}'|$ are upperbounded by $C_{24}(\Delta, d, \rho, \zeta, \eta)$. ■

References

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

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

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

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

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

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

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

[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.

[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.

[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.

[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.

[CF11] David Conlon and Jacob Fox. Bounds for graph regularity and removal lemmas. Technical report, July 2011. http://arxiv.org/abs/1107.4829.

[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.

[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.

[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.

[Gow97] William T. Gowers. Lower bounds of tower type for Szemerédi's uniformity lemma. *Geometric and Functional Analysis*, 7:322–337, 1997.

[Gow98] William T. Gowers. A new proof of Szeméredi's theorem for arithmetic progressions of length four. *Geometric and Functional Analysis*, 8(3):529–551, 1998.

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

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

[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.

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

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

[GW10a] W. T. Gowers and J. Wolf. Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$. *Geom. Funct. Anal.*, to appear, 2010.

[GW10b] W. T. Gowers and J. Wolf. The true complexity of a system of linear equations. *Proc. Lond. Math. Soc. (3)*, 100(1):155–176, 2010.

[HK05] Bernard Host and Bryna Kra. Nonconventional ergodic averages and nilmanifolds. *Annals of Mathematics*, 161(1):397–488, 2005.

[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.

[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.

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

[KS08] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 403–412, 2008.

[KS11] Subrahmanyam Kalyanasundaram and Asaf Shapira. A Wowzer type lower bound for the Strong Regularity Lemma. Technical report, July 2011. http://arxiv.org/abs/1107.4896.

[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.

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

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

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

[Sha09] Asaf Shapira. Green's conjecture and testing linear-invariant properties. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 159–166, 2009.

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

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

[Tao11] Terence Tao. Higher order Fourier Analysis. Draft available at http://terrytao.files.wordpress.com/2011/03/higher-book.pdf, 2011. In preparation.

[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.

[VX11] Santosh Vempala and Ying Xiao. Structure from local optima: Learning subspace juntas via higher order PCA. Technical report, August 2011. http://arxiv.org/abs/1108.3329.

A Decomposition Theorems

Proof of Theorem 14: We need a standard "index increment" argument.

Lemma 38 *Suppose $\mathcal{B}$ is a polynomial factor of degree d and complexity C, and suppose $f : \mathbb{F}_p^n \to \{0,1\}$ is such that $\|f - \mathbb{E}[f|\mathcal{B}]\|_{U^{d+1}} \geq \delta$. Then, there exists a refined polynomial factor $\mathcal{B}'$ of degree d and complexity at most $C+1$ such that:*

$$\|\mathbb{E}[f|\mathcal{B}']\|_2^2 \geq \|\mathbb{E}[f|\mathcal{B}]\|_2^2 + (\epsilon_8(\delta))^2$$

where $\epsilon_8(\cdot)$ is the function in Theorem 8.

Proof: $g = f - \mathbb{E}[f|\mathcal{B}]$ is bounded to $[-1,1]$. So, applying Theorem 8 yields a degree-d polynomial q satisfying $|\mathbb{E}[g(x) \cdot \mathsf{e}(P(x))]| \geq \epsilon_8(\delta)$. The polynomial q generates a factor $\hat{\mathcal{B}}$ of complexity 1. Define $\mathcal{B}'$ to be the common refinement of $\mathcal{B}$ and $\hat{\mathcal{B}}$; it's of complexity $C+1$.

Observe that:

$$\|\mathbb{E}[g|\mathcal{B}']\|_1 = \mathop{\mathbb{E}}_x\left[|\mathbb{E}[g|\mathcal{B}'](x)|\right] \geq \left|\mathop{\mathbb{E}}_x\left[\mathbb{E}[g|\mathcal{B}'](x) \cdot \mathsf{e}(P(x))\right]\right| = \left|\mathop{\mathbb{E}}_x\left[g(x) \cdot \mathsf{e}(P(x))\right]\right| \geq \epsilon_8(\delta)$$

The last equality uses the fact that P is constant on each atom of $\mathcal{B}'$. Now:

$$\|\mathbb{E}[f|\mathcal{B}']\|_2^2 - \|\mathbb{E}[f|\mathcal{B}]\|_2^2 = \|\mathbb{E}[f|\mathcal{B}'] - \mathbb{E}[f|\mathcal{B}]\|_2^2 = \|\mathbb{E}[g|\mathcal{B}']\|_2^2 \geq \|\mathbb{E}[g|\mathcal{B}']\|_1^2 \geq \epsilon_8^2(\delta)$$

■

First, consider a weaker version of Theorem 14. For any *constant* η, we find a factor $\mathcal{B}' \preceq \mathcal{B}_0$ and a decomposition $f = g_1 + g_2$ where $g_1 = \mathbb{E}[f|\mathcal{B}']$ and $g_2 = f - \mathbb{E}[f|\mathcal{B}']$ satisfies $\|g_2\|_{U^{d+1}} \leq \eta$. This is straightforward to obtain by starting with the factor initially being $\mathcal{B}_0$ and then repeatedly applying Lemma 38 while the U^{d+1}-norm of g_2 is larger than η. Each time, the ℓ_2^2-norm of g_1 increases by $(\epsilon_8(\eta))^2$, which on the other hand, cannot increase to more than 1. Hence, the complexity of the final factor $\mathcal{B}'$ is bounded by a function of only η and C_0.

For Theorem 14, we iterate this argument. Let $\mathcal{B}^{(1)}$ denote $\mathcal{B}'$, and let C_1 be its complexity. For $i \geq 2$, let $\mathcal{B}^{(i)}$ be the refinement of $\mathcal{B}^{(i-1)}$ resulting from applying the above argument with η chosen to be smaller than $\eta(C_{i-1})$, and let C_i denote the complexity of $\mathcal{B}^{(i)}$. Because each $\mathcal{B}^{(i)}$ is a refinement of $\mathcal{B}^{(i-1)}$, the quantities $\|\mathbb{E}[f|\mathcal{B}^{(i)}]\|_2^2$ form a non-decreasing sequence. By the pigeonhole principle, there exists $i \leq 1/\delta^2$ such that $\|\mathbb{E}[f|\mathcal{B}^{(i+1)}]\|_2^2 - \|\mathbb{E}[f|\mathcal{B}^{(i)}]\|_2^2 \leq \delta^2$. Set $f_1 = \mathbb{E}[f|\mathcal{B}^{(i)}]$, $f_2 = f - \mathbb{E}[f|\mathcal{B}^{(i+1)}]$ and $f_3 = \mathbb{E}[f|\mathcal{B}^{(i+1)}] - \mathbb{E}[f|\mathcal{B}^{(i)}]$. The conditions of the theorem are now met.
■

Proof of Lemma 18: Suppose $\mathcal{B}$ is defined by the polynomials $P_1, \ldots, P_C$, each of degree at most d. Define the *degree vector* of $\mathcal{B}$ to be $(M_1, \ldots, M_d)$ where M_i is the number of polynomials in $P_1, \ldots, P_C$ of degree exactly i. Given two degree vectors $M = (M_1, \ldots, M_d)$ and $N = (N_1, \ldots, N_d)$, say $M \prec N$ if there exists $i \in [d]$ such that $M_i < N_i$ and $M_j = N_j$ for all $j \in [i+1, d]$. It is a standard fact in set theory that this is a well-ordering of the degree vectors. We perform induction on the degree vectors using this ordering.

Suppose $\mathcal{B}$ is of rank $\leq r(C)$ with degree vector $M = (M_1, \ldots, M_d)$. Then, there exists $\alpha_1, \ldots, \alpha_C$ not all zero such that $P_\alpha = \sum_{i=1}^{C} \alpha_i P_i$ can be expressed as a function of $r(C)$ many polynomials $Q_1, \ldots, Q_{r(C)}$ of degree $e-1$ where $e = \max_{i:\alpha_i \neq 0} \deg(P_i)$. Without loss of generality, assume $\alpha_1 \neq 0$ and $\deg(P_1) = e$. Consider the polynomial factor $\mathcal{B}'$ generated by the polynomials $\{P_2, \ldots, P_C\} \cup \{Q_1, \ldots, Q_{r(C)}\}$. $\mathcal{B}'$ has complexity $\leq C + r(C)$, and it clearly semantically refines $\mathcal{B}$. Its degree vector $N = (N_1, \ldots, N_d)$ has $N_e = M_e - 1$ and $\sum_{i=1}^{e-1} N_i \leq \sum_{i=1}^{e-1} M_i + r(C)$. Since $N \prec M$, we are done by induction.

Now, let's turn to the last remark about an existing $\hat{\mathcal{B}}$. We follow the same argument as above, but at each step of the induction, instead of replacing P_1, we find a polynomial P_{i_0} to replace where $i_0 > \hat{C} = |\hat{\mathcal{B}}|$. That is, we claim there exists $i_0 > \hat{C}$ with $\alpha_{i_0} \neq 0$ and $\deg(P_{i_0}) = e$. The reason is that if for all $i > \hat{C}$, either $\alpha_i = 0$ or $\deg(P_i) < e$, we would contradict the rank assumption on $\hat{\mathcal{B}}$. In particular, the e-degree polynomial $\sum_{i=1}^{\hat{C}} \alpha_i P_i$ would be expressible as a function of the lower degree polynomials $Q_1, \ldots, Q_{r(C)}$ and $\sum_{i=\hat{C}+1}^{C} \alpha_i P_i$. So, the polynomials $P_1, \ldots, P_{\hat{C}}$ will remain in the polynomial factor throughout the induction, and hence, the final factor $\mathcal{B}'$ will be a syntactic refinement of $\hat{\mathcal{B}}$. ■

Proof of Theorem 19: First, apply Theorem 14 to get a factor $\mathcal{B}$ syntactically refining $\mathcal{B}_0$ and a decomposition $f = f_1 + f_2 + f_3$ such that $f_1 = \mathbb{E}[f|\mathcal{B}]$, $\|f_2\|_{U^{d+1}} \leq \eta(\tau_r(C))$ and $\|f_3\|_2 \leq \delta/2$ where τ_r is the function from Lemma 18 and $C \leq C_{14}(\delta/2, \eta \circ \tau_r, k, C_0)$ is the complexity of $\mathcal{B}$. Now, apply Lemma 18 to get a semantically refined factor $\mathcal{B}' \preceq_{sem} \mathcal{B}$ with complexity $\tau_r(C)$ and rank at least $r(\tau_r(C))$. This yields a new decomposition: $f = f_1' + f_2' + f_3'$ with:

$$f_1' = \mathbb{E}[f|\mathcal{B}'] \qquad f_2' = f_2 \qquad f_3' = f_3 + \mathbb{E}[f|\mathcal{B}] - \mathbb{E}[f|\mathcal{B}']$$

The only problem is that we might have $\|f_3'\|_2 > \delta$. But then, $\|\mathbb{E}[f|\mathcal{B}] - \mathbb{E}[f|\mathcal{B}']\|_2 > \delta/2$, since $\|f_3\|_2 \leq \delta/2$. This means that:

$$\|\mathbb{E}[f|\mathcal{B}']\|_2^2 > \|\mathbb{E}[f|\mathcal{B}]\|_2^2 + \delta^2/4$$

So, if we keep repeating the entire argument for at most $O(1/\delta^2)$ times, it must be the case that we eventually satisfy all the claims of Theorem 19.

The last part of the Theorem is true because of the last part of Lemma 18. ■

ECCC ISSN 1433-8092

http://eccc.hpi-web.de