Testing Low Complexity Affine-Invariant Properties

Arnab Bhattacharyya*
Princeton University
arnabb@princeton.edu

Eldar Fischer†
Israel Institute of Technology
eldar@cs.technion.ac.il

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

March 28, 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 that 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.

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

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.

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], and for dense graphs it was addressed in a long series of works culminating in [AFNS06] and [BCL$^+$06].

In this work, we make steps 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 a *linear-invariant* property. We say that a property $\mathcal{P} \subseteq \{\mathbb{F}_p^n \to [R]\}$ is linear-invariant if it is the case that for any $f \in \mathcal{P}$ and for any linear transformation $L : \mathbb{F}_p^n \to \mathbb{F}_p^n$, it holds that $f \circ A \in \mathcal{P}$. Similarly, an *affine-invariant* property is closed under composition with affine transformations $A : \mathbb{F}_p^n \to \mathbb{F}_p^n$ (an affine transformation A is of the form $L + c$ where L is linear and c is a constant). The property of a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ 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], homogeneous 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 property testable with a 1-sided error test.

Definition 1.1 (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 1.2 ([BGS10]) *Any affine subspace hereditary property is testable with 1-sided error.*

Moreover, [BGS10] show that *every* affine-invariant property testable by a "natural" tester is very "close" to an affine subspace hereditary property[1]. In fact, resolving Conjecture 1.2 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 being affine, by which we mean here that the function is a polynomial of degree at most 1. This is clearly an affine-invariant hereditary property. As we remarked earlier, the property is known to be testable. Note that here, we could also have defined being affine 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 whether 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 (finitely or infinitely many) such local characterizations. To state this formally, let us put forth a useful definition.

[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 the analog definitions in a graph-theoretic context.

Definition 1.3 (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 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* $f : \mathbb{F}_p^n \to [R]$ *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 simple observation.

Observation 1.4 *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.*

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 $f : \mathbb{F}_p^n$ 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 $\mathbb{F}_p^n$ and contains $\{a_z(X_1, \ldots, X_{n+1}) = X_1 + \sum_{i=1}^{n} z_i X_{i+1} : z = (z_1, \ldots, z_n) \in \mathbb{F}_p^n\}$, and σ_f is just set to f. From here it is easy to see that the property defined by $\mathcal{A}$ is contained in $\mathcal{P}$, while containment in the other direction follows from $\mathcal{P}$ being affine-invariant and hereditary.

The other direction of the observation is trivial. ■

Thus, resolving Conjecture 1.2 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 1.5 (Cauchy-Schwartz complexity, [GT10b]) *Let* $\mathcal{L} = \{L_1, \ldots, L_m\}$ *be a set of linear forms. The* (Cauchy-Schwartz) 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 any subset.*

Given this, one can formulate a conjecture that is a weakened version of Conjecture 1.2:

Conjecture 1.6 *A property that is given by a collection of induced affine constraints with a global bound on their complexity is testable with a* 1*-sided error.*

The following is our main result, which shows the above when the complexity bound is strictly smaller than the field size.

Theorem 1.7 (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 being $\mathcal{A}$-free, that makes 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 in fact inherent in the nature of the problem.

Let us lastly note that Theorem 1.7 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 greater than 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

To show Theorem 1.7, we will in fact show the following statement. Note that it uses a yet undefined notion of "conciseness"; for now it suffices to know that every $\mathcal{A}$ is equivalent to a concise one, as we will later prove.

Theorem 1.8 *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 $\mathcal{A}$ is concise, every A^i is of complexity less than p and consists of m_i linear forms over ℓ_i variables, and $\sigma^i \in [R]^{m_i}$ for every i. Then, there are functions $\ell_{\mathcal{A}}(\cdot)$ 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 is ϵ-far from being $\mathcal{A}$-free, then f induces at least $\delta_{\mathcal{A}}(\epsilon) \cdot p^{n\ell_i}$ many copies of (A^i, σ^i) for some i such that $\ell_i < \ell_{\mathcal{A}}(\epsilon)$.*

Theorem 1.7 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 1.8, if f is ϵ-far from $\mathcal{A}$-freeness, then 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.

Let us now give an overview of our proof of Theorem 1.8. 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:

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

where crucially, δ is a positive function that 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 immediately 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 some d, we have $\|f_1\|_{U^{d+1}} < \epsilon$, then any expectation of the form $\mathbb{E}_{x_1,\ldots,x_\ell \in \mathbb{F}_p^n}\left[\prod_{i=1}^m f_i(L_i(x_1,\ldots,x_\ell))\right]$ is bounded by ϵ for any linear forms $L_1,\ldots,L_m$.

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 only 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 inspired by previous work on property testing in the dense model, 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 a new combinatorially flavored decomposition theorem, which may 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 $f : \mathbb{F}_p^n \to \mathbb{R}$ to 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. To get around we essentially use a sequence of two decompositions, and make the norm of the second one decrease as a function of the complexity of the first, where 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 [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]. However, implementing these graph-theoretic techniques using higher-order Fourier analysis required several new ideas which, we hope, can be extended to eventually prove Conjecture 1.6.

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, thus proving Conjecture 1.6; this requires non-trivial generalizations of several technical lemmas to small fields, and may require new methods.

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

A third generalization, which might be too strong to hold, is to completely remove the bounded complexity assumption on the linear forms, thus proving Conjecture 1.2. 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. It would be thus also be interesting if the full Conjecture 1.2 can be *disproved*.

2 Map of the proof

The rest of this paper will be devoted to constructing the building blocks required to prove Theorem 1.7, and by extension Theorem 1.8. We believe that some of these building blocks, and especially the "Super Decomposition" Theorem 4.9 that we describe below, will be of independent interest.

In Section 3 and Section 4 we develop the main technical tools that we will need for our testability proof. Some of the following lemmas and arguments were proved before: Decomposition lemmas (without rank) were implicit in previous works by Green and Tao and explicit in [HL11a]; the existence of a refinement of a given rank was first proved in [GT09] (which is combined here with a decomposition lemma); other prior works are cited along with the proofs below.

Our new contributions there lie in the following:

- Our final "Super Decomposition" Theorem 4.9, and its related "Subcell Selection" Corollary 4.10, are new. Their relation to the original decomposition lemma could be thought of as somewhat akin to the relation of the strong graph regularity lemma in [AFKS00] to the original regularity lemma of Szemerédi.
- For the subcell selection corollary to work at all, we need to take careful count of when is a refinement of a partition by polynomials syntactic (i.e. there is a containment relationship between the polynomials defining the two partitions) or merely semantic (i.e. the polynomials may be different but the partitions they define satisfy a combinatorial refinement relationship). We add the accounting of syntactical vs semantic refinements to all the arguments leading up to our super decomposition theorem.
- We set the entire analysis in a "robustness" framework akin to the one developed for graphs in [FN07]. This streamlines the argument (essentially allowing us to encapsulate and move away iterative refinement arguments), which could get very unwieldy by the time the super decomposition theorem is reached.

In Section 5 we then develop algebraic and combinatorial constructions, that allow us to *use* Corollary 4.10 to provide counting type theorems, and in our case the main "algebro-combinatorial" Theorem 1.8. The algebraic part mostly involve procedures that calculate the numbers of affine configuration of a given type that satisfy given polynomial constraints; we also prove, using basic algebra, that we can assume the technical condition of $\mathcal{A}$ being "concise", that is not having more variables than conditions in any of its constraints. The combinatorial part is the "cleanup" procedure that we describe below.

We now describe the main components of our proofs in detail.

2.1 Partition by Polynomial Factors

We generally deal with a function $f : \mathbb{F}_p^n \to \{0,1\}$ (where a larger fixed size range $[R]$ is handled by considering a sequence of functions rather than one function – see Subsection 4.3), and would like to partition its domain $\mathbb{F}_p$ into a small number of regions, so that f has certain "randomness" properties in every region (or at least most of them). In the broadest terms, we seek algebraic analogs to Szemerédi's regularity lemma and its derivatives that have revolutionized graph theory. Recall that Szemerédi's lemma partitions the vertex set of the graph so that most vertex set pairs exhibit random-like properties in the bipartite subgraphs that they induce.

The groundwork providing this started with the works of Green and Tao. In general, a function $f : \mathbb{F}_p^n \to \{0,1\}$ can be decomposed to a sum of three real-valued functions. One that is constant on large regions of the input, one that generally takes small values (in terms of its l_2 norm), and one that is "very random" (in the sense of the Gowers norm). The relevance of the Gowers norm to our arguments is highlighted in Subsection 3.1.

In an ideal world, the large regions of the input over which we have a constant function should come from a partition of $\mathbb{F}_p^n$ into affine subspaces, but in fact this cannot be the case. The next best thing is to have a partition based on the values of a fixed length sequence of low degree polynomials over $\mathbb{F}_p^n$. These are called *polynomial factors* as per Definition 3.4, and the regions of $\mathbb{F}_p^n$ of their respective partitions are called *cells*.

However, now we need to re-address the question of independence. Standard linear independence would be insufficient to even guarantee that all regions are of similar sizes, let alone provide other "randomness" features. For this we use the notion of polynomial *rank*, first developed in [GT09]. Subsection 3.2 provides the details about polynomial factors and their rank.

2.2 Refinements and the Robustness Framework

For our purpose it is not enough to prove the existence of certain factors, and we will consider a relationships between pairs of factors, namely the *refinement* relationship. There are two kinds of refinements. The "combinatorial" *semantic* refinement notion means that the partition induced by the second factor consists of subsets of the sets of the first factor, while the "explicit" *syntactic* refinement notion means that the second factor is in fact defined by a sequence of polynomials extending the sequence that defines the first factor. Definition 3.9 provides the details.

An important measure of a factor with respect to a function $f : \mathbb{F}_p^n \to \{0,1\}$ is its *density index*, as per Definition 3.11. This was used in previous decomposition proofs, and is analogous to the index of a graph partition used in the proof of Szemerédi's regularity lemma and its variants. In Subsection 3.3 we introduce and analyze the framework of factor *robustness*, where a factor is considered robust if it cannot be refined (with respect to a size bound given as a function of the current size) in a way that significantly increases its index. Robust factors, including ones that refine existing factors, exist by a simple argument, Observation 3.13.

The robustness framework greatly simplifies the arguments used to prove the decomposition theorems in Section 4. Where previously such proofs used an iterative argument, basically repeating a construction of a refining factor as long as the factor does not provide the required properties, in the proofs here we start with a robust factor and then show that it provides the required object.

However, we need a factor to be both robust and of high rank. The high rank requirement (also as a function of the factor size) is in fact also provided through an iterative argument resembling the proof of regularity. In Subsection 3.4 we integrate arguments similar to those originally made in [GT09] to provide Lemma 3.19, the driving engine of our decomposition theorems. This lemma provides factor that is both robust and of high rank. Moreover, if we start from an existing factor that is a syntactic refinement of a base factor that also has high rank, then our new robust factor will additionally be a syntactic refinement of the same base factor. This is crucial to our super decomposition theorem, that requires such a refinement to be provided.

2.3 Decompositions and Super Decompositions

Chronologically, decomposition theorems for functions $f : \mathbb{F}_p^n \to \{0,1\}$ have progressed in stages. First a weak decomposition theorem was shown, where a factor is found and f is decomposed into a sum of two functions, $f = f_1 + f_2$, where $f_1 : \mathbb{F}_n^p \to [0,1]$ is constant over every cell of the factor, and $f_2 : \mathbb{F}_p^n \to [-1,1]$ has a bounded Gowers norm. In an ideal world we would like f_2 to have a bounded l_2 norm, as it denotes an "error" of some kind, but this is not possible.

However, for the Gowers norm bound to be of any use, it has to be bounded as a decreasing function of the factor size C. The next step was then to find a factor and a decomposition $f = f_1 + f_2 + f_3$, where f_3 is an "error" term that is of bounded l_2 norm (as we originally intended), and f_2 now has a Gowers norm that is smaller than the required function of C. The proof "internally" uses a sequence of two factors, one refining the other, and a corresponding "iterated argument of iterated arguments". However here we can encapsulate it through a robustness requirement. We provide the full details in Subsection 4.1, which culminates in Theorem 4.4, providing also a rank requirement. It is similar to theorems proved in previous works, but here we also maintain a syntactic refinement relationship to a base factor, a feature that will be used later.

This brings us to our new super decomposition Theorem 4.9. Its motivation is that for our purpose, we would also need the l_2 norm of the error function f_3 to decrease as a function of the factor size. This is required because for our analysis of non-monotone properties, we cannot make do with most of the cells of the factor exhibiting a random-like behavior of f – we would like *all* of them to exhibit it. However, such a demand on f_3 is clearly not possible.

The solution is then to provide a sequence of two factors, where the second factor is a syntactic refinement of the first. We then decompose f with respect to the second factor, as a sum of a constant-over-cells function f_1, a small Gowers norm function f_2, and a function f_3 whose l_2 norm is not small as a function of the second factor, but at least it is small as a function of the first factor. Additionally, we want f_1 to be "faithful" also with respect to the first factor: That is, if we had decomposed f according to the first factor rather than the second, then the corresponding "f_1 function" would still be close in most places to the function we got by decomposing according to the second factor.

In the next step of the proof of our main testability theorem, we will pick one "subcell", a cell of the second factor, out of every cell of the first factor. We will want most of these cells to be faithful (with respect to f_1) and all of them to exhibit the randomness properties. The syntactic refinement relationship in our super decomposition theorem is what allows us to pick these cells in a "uniform" manner, as per our subcell selection Corollary 4.10.

We believe that Theorem 4.9 and its proof methods are of independent interest, as they could open up possibilities for more analogies to the big body of knowledge concerning the applications of Szemerédi's lemma and its variants for graphs.

2.4 Function Cleanup

To find many induced structures in f, we restrict ourselves to the "good" subcells chosen by use of Corollary 4.10. However, to find the correct configuration of subcells exhibiting the induced structures, we refer to a modification of f called a *cleanup*. The modified f will be close to the original, and hence will still contain an induced structure. This particular structure might not exist in the original f, but the way the cleanup is performed, as per Definition 5.14, ensures the existence of the corresponding subcell configuration which "mimics" the location of the points of the structure (even that it may not actually contain those points). We then use the configuration of subcells with respect to the original f to find our affine structures.

This argument is in fact somewhat analogous to the argument considering forbidden induced subgraphs that appeared first in [AFKS00]. The function closeness lemma is Lemma 5.15, while the mimicking subcell argument is found in the proof of Theorem 1.8 in Subsection 5.4.

2.5 Randomness and consistency

After we find the subcell configuration corresponding to an affine induced structure, we still need to lower-bound the number of actual copies of the structure that it guarantees for f. This requires giving a lower bound for the number of actual small affine sets that reside in this configuration, and within them the number of sets for which f has the corresponding values. The second task is in fact accomplished by the function decomposition that we have. For the first task, we build upon works of Hatami and Lovett [HL11b] and of Gowers and Wolf [GW10b, GW10a] in Subsection 5.1.

We use there the notion of *consistent values*, Definition 5.5, as an algebraic characterization of when is a configuration of cells feasible for a given affine structure. This allows us to regulate "all-or-nothing" lemmas from previous works in Theorem 5.7, to provide a calculated bound for the number of structures. We also utilize it for Lemma 5.8, showing that the subcell selection process does not "spoil" a good configuration.

2.6 Wrapping Up

There are some final ingredients that we need before finalizing the proof of Theorem 1.8. One of which is a compactness argument, analogous to the one made in [AS08a], to be able to bound the size of the constraints we need to test for, even when the property is defined by an infinite number of constraints. In our case, we also need to perform a slight "preprocessing" to representation of the property, to make it *concise* as per Definition 5.19, which is done through Lemma 5.18. Apart from this, Subsection 5.3 contains a few other algebraic tools that help with the calculations used in the proof.

Finally, Subsection 5.4 contains the proof of Theorem 1.8, tying it all together, from finding a factor with a subcell selection, through consistency and randomness arguments, to finally using the function cleanup to bound from below the number of copies of the corresponding induced structure.

3 Tools of the Proof

In this section we lay the groundwork for the decomposition theorems that follow. This include the formal definition of partition by polynomial factors, the definition of factor robustness and rank with proofs of their impact, and finally we prove the main lemma about the existence of partitions that are both robust and of high rank.

3.1 Functions and Norms

In the most general setting we consider functions $f : G \to \mathbb{C}$, where G is a finite Abelian group[2].

Unless stated otherwise, expectations are taken over the uniform probability space with respect to the relevant range, e.g. $\mathbb{E}_x[f(x)]$ is set to $|G|^{-1} \sum_{x \in G} f(x)$. Apart from the traditional norms such as $\|f\|_2^2 = \mathbb{E}_x[|f(x)|^2]$, we will make extensive use of Gowers norms.

[2]Later we would mostly consider $G = \mathbb{F}_p^n$. Our main theorem is formulated for functions whose range is $\{0, 1\}$, but its proof uses interim function with larger ranges.

Definition 3.1 (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, i.e. $\mathcal{C}^l(a+bi) = a + (-1)^l bi$ for $a, b \in \mathbb{R}$ and integer l.

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 3.2 (Gowers Inverse Theorem) *Given positive integers $d < p$, for every $\delta > 0$, there exists $\epsilon = \epsilon_{3.2}(\delta, p)$ 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-Schwartz inequality. Refer to Definition 1.5 for the term "complexity".

Lemma 3.3 *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:*

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

3.2 Polynomial Factors and their Rank

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 3.4 (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 $y \in \mathbb{F}_p^C$. On the other hand, given a cell of $\mathcal{B}$, the common value $y = \mathcal{B}(x) \in \mathbb{F}_p^C$ is called the* image *of the cell. When there is no ambiguity, we will in fact abuse notation and identify a cell of $\mathcal{B}$ with its image y.*

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 3.5 (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 $\mathbb{E}_{x:\mathcal{B}(x)=y}[f(x)]$, 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 every cell of the polynomial factor.*

In essence we would want to choose a polynomial factor so that, among other things, the restriction of f in every cell would essentially consist of a constant element and other elements of small norms. However, since we are not dealing with affine linear subspaces, for our arguments to follow we also need the factor itself to be "well behaved". This is exemplified in the notion of polynomial rank [GT09], in essence a strengthening of linear independence.

Definition 3.6 (Rank of polynomial factors) *Suppose that $\mathcal{B}$ 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 rank of a single polynomial P is defined similarly (but without needing to relate to linear combinations).

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 3.7 *For any $\epsilon > 0$ and integer $d \geq 1$, there exists $r = r_{3.7}(d, \epsilon)$ 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 3.7 is, consider the following simple lemma which states that every cell of a polynomial factor with large enough rank has approximately the same size.

Lemma 3.8 *Given a polynomial factor $\mathcal{B}$ of degree d, complexity C, and rank at least $r_{3.7}(d, \epsilon)$ 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:

$$\begin{aligned}
\Pr_{x \in \mathbb{F}_p^n}[\mathcal{B}(x) = b] &= \mathop{\mathbb{E}}_x \left[\prod_{i \in [C]} \frac{1}{p} \sum_{\lambda_i \in \mathbb{F}_p} \mathsf{e}\left(\lambda_i \cdot (P_i(x) - b_i)\right) \right] \\
&= p^{-C} \sum_{(\lambda_1, \ldots, \lambda_C) \in \mathbb{F}_p^C} \mathop{\mathbb{E}}_x \left[\mathsf{e}\left(\sum_{i \in [C]} \lambda_i (P_i(x) - b_i) \right) \right] \\
&= p^{-C} \left(1 \pm p^C \epsilon\right)
\end{aligned}$$

where the last line uses Theorem 3.7 whenever $(\lambda_1, \ldots, \lambda_C) \neq 0^C$. ■

3.3 Refinement and Robustness

The decomposition theorems will iteratively partition the domain $\mathbb{F}_p^n$ into finer and finer partitions (though we will use a mechanism that hides the refinements that do not have to be "visible" for the other proofs). We will need to be careful about distinguishing between two different types of refinements.

Definition 3.9 (Refinement of a 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 3.10 *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 $\mathcal{B}'$. ■

On the other hand, doing the above conversion can "destroy" the rank of a polynomial factor, and there will be indeed situations in what follows where we will have to carefully distinguish the two refinement types.

Next, we define the density index of a polynomial factor with respect to a function, and use it to define the notion of robustness, which is central to what follows.

Definition 3.11 *The* density index *of a factor $\mathcal{B}$ with respect to a function f is the squared l_2 norm of the conditional expectation of f, that is* $\mathrm{indd}(\mathcal{B}) = \mathbb{E}\left[(\mathbb{E}[f|\mathcal{B}])^2\right]$.

Given a function $h : \mathbb{N} \to \mathbb{N}$ and a real parameter γ, A factor $\mathcal{B}$ is (h, γ)-robust *(semantically) if there exists no $\mathcal{B}'$ which is a semantic refinement of $\mathcal{B}$ for which $|\mathcal{B}'| \leq h(|\mathcal{B}|)$ and* $\mathrm{indd}(\mathcal{B}') \geq \mathrm{indd}(\mathcal{B}) + \gamma$.

Robustness is somewhat preserved when moving to a small refinement.

Observation 3.12 *If $\mathcal{B}$ is $(g \circ h, \gamma)$-robust, and $\mathcal{B}'$ is a (syntactic or semantic) refinement of $\mathcal{B}$ for which $|\mathcal{B}'| \leq h(|\mathcal{B}|)$, then $\mathcal{B}'$ is (g, γ)-robust.*

Proof: If $\mathcal{B}''$ is any refinement of $\mathcal{B}'$ for which $|\mathcal{B}''| \leq g(|\mathcal{B}'|)$, then $|\mathcal{B}''| \leq g(h(|\mathcal{B}|))$ and so $\mathrm{indd}(\mathcal{B}'') \leq \mathrm{indd}(\mathcal{B}) + \gamma$. On the other hand by the Cauchy-Schwartz inequality $\mathrm{indd}(\mathcal{B}') \geq \mathrm{indd}(\mathcal{B})$, and so $\mathrm{indd}(\mathcal{B}'') \leq \mathrm{indd}(\mathcal{B}') + \gamma$, proving the robustness condition of $\mathcal{B}'$. ■

Existence of robust factors, also as syntactic refinements of a given factor, is easy to prove. Note that the function in its statement takes another function as one of its parameters.

Observation 3.13 *For an appropriate function $T_{3.13}(k, h, \gamma)$, for any $\mathcal{B}$, $h : \mathbb{N} \to \mathbb{N}$ and $\gamma > 0$ there exists a syntactic refinement $\mathcal{B}'$ which is (h, γ)-robust, and for which $|\mathcal{B}'| \leq T_{3.13}(|\mathcal{B}|, h, \gamma)$.*

Proof: Without loss of generality we assume that h is monotone non-decreasing (otherwise replace $h(k)$ with $\max_{j \leq k} h(j)$). Set $\mathcal{B}_0 = \mathcal{B}$. Inductively, if $\mathcal{B}_i$ is not already (h, γ)-robust then set $\mathcal{B}_i'$ to be a semantic refinement of $\mathcal{B}_i$ for which $|\mathcal{B}_i'| \leq h(|\mathcal{B}_i|)$ and $\mathrm{indd}(\mathcal{B}') \geq \mathrm{indd}(\mathcal{B}) + \gamma$, and by Observation 3.10 then set $\mathcal{B}_{i+1}$ to be a syntactic refinement of $\mathcal{B}$ and $\mathcal{B}_i'$ for which $|\mathcal{B}_{i+1}| \leq h(|\mathcal{B}_i|) + |\mathcal{B}|$.

Noting that the index can only increase while moving to a refinement (by the Cauchy-Schwartz inequality), this process must stop for some $j \leq 1/\gamma$. $\mathcal{B}_j$ is the required factor, and its size is bounded by $k_{1/\gamma}$, where we define $k_0 = k$ and by induction $k_{i+1} = h(k_i) + k$. ■

Note: From now on we assume that all our relevant functions are monotone in their corresponding variables, also when this is not stated explicitly. For example, a function h fed to Observation 3.13 will assumed to be monotone non-decreasing, and if $k \leq k'$, $\gamma \geq \gamma'$, and $h(m) \leq h'(m)$ for every $m \in \mathbb{N}$ (while both h and h' are monotone non-decreasing), then $T_{3.13}(k, h, \gamma) \leq T_{3.13}(k', h', \gamma')$. All our lemmas can indeed be made to provide such functions.

3.4 Robustness with Rank

The next item on the agenda is to show that polynomial factors can be refined to ones of high rank. The following index definition is used for analyzing rank.

Definition 3.14 *The* degree index *of a factor $\mathcal{B}$ is the (infinite but almost everywhere zero) sequence of non-negative integers* $\mathrm{indm}(\mathcal{B}) = I = (i_1, i_2, \ldots)$*, where i_k is the number of polynomials of degree k in the sequence of polynomials defining $\mathcal{B}$.*

Denote the set of all possible degree sequences as above by $\mathcal{I}$. Over $\mathcal{I}$ we define the anti-lexicographic *order, where $I < I'$ if $i_k < i'_k$ for the largest k on which those coordinates differ.*

The set $\mathcal{I}$ defined above is well-ordered in the sense that there exist no infinite strictly decreasing sequences of members of $\mathcal{I}$, but this still does not provide for "standard" induction, as the order is not isomorphic to $\mathbb{N}$. To replace induction we define the notion of a decrement.

Definition 3.15 *Let $\mathcal{I}$ denote the well-ordered set of all possible degree indexes. A function $\kappa : \mathbb{N} \times \mathcal{I} \to \mathcal{I}$ is called a* decrement *if for all $A \in \mathcal{I}$ and $n \in \mathbb{N}$ it satisfies $\kappa(n, A) < A$, for all n and $A \leq B$ it satisfies $\kappa(n, A) \leq \kappa(n, B)$, and for all $n < m$ and A it satisfies $\kappa(n, A) \leq \kappa(m, A)$. The inequalities are with respect to the anti-lexicographic ordering of $\mathcal{I}$.*

The following shows how, when we are given a decrement that "bounds" some process, we can use it to bound an iterative process.

Lemma 3.16 *There exist $T_{3.16}(k, d, h, \kappa)$ and $m_{3.16}(k, d, h, \kappa)$ that take numbers k and d, a monotone $h : \mathbb{N} \to \mathbb{N}$ and a decrement κ, and satisfy the following. If $\mathcal{B}_0, \mathcal{B}_1, \ldots \mathcal{B}_m$ is a sequence of factors of bounded degree d for which $|\mathcal{B}_0| \leq k$, $|\mathcal{B}_i| \leq h(|\mathcal{B}_{i-1}|)$ and $\mathrm{indm}(\mathcal{B}_i) \leq \kappa\left(|\mathcal{B}_{i-1}|, \mathrm{indm}(\mathcal{B}_{i-1})\right)$, then $|\mathcal{B}_m|$ is bounded by $T_{3.16}(|\mathcal{B}|, d, h, \kappa)$ and m is bounded by $m_{3.16}(|\mathcal{B}|, d, h, \kappa)$.*

Proof: Let I_0 be the maximal (with respect to order) degree index of any degree d factor of complexity k, which is the sequence $(i_1, i_2, \ldots)$ for which $i_d = k$ and $i_j = 0$ for any $j \neq d$, and let $h_0 = k$. Inductively define $h_i \in \mathbb{N}$ as $h_i = h(h_{i-1})$, and $I_i = \kappa(h_{i-1}, I_{i-1})$. Because $I_0, I_1, \ldots$ is a decreasing sequence over a well-ordered set, it must be of bounded length, which we denote as $m_{3.16}(k, d, h, \kappa)$. We then set $T_{3.16}(k, d, h, \kappa) = h_{m_{3.16}(k,d,h,\kappa)}$. For a sequence of factors as above, the monotonicity conditions of κ ensure that $|\mathcal{B}_i| \leq h_i$ and $\mathrm{indm}(\mathcal{B}_i) \leq I_i$, and so we are done. ■

The following provides a decrement that will bound the process of obtaining a high rank refinement of a given factor, as well as a bound on the size increment. Note that also if the required rank depends on the factor size, we can still get a bounding decrement.

Lemma 3.17 *For every $r : \mathbb{N} \to \mathbb{N}$ there exist $h^{(r)}_{3.17} : \mathbb{N} \to \mathbb{N}$ and a decrement $\kappa^{(r)}_{3.17} : \mathbb{N} \times \mathcal{I} \to \mathcal{I}$, satisfying the following for every d. If $\mathcal{B}$ is a factor of degree at most d whose rank is less than $r(|\mathcal{B}|)$, then there exists a semantic refinement $\mathcal{B}'$ of $\mathcal{B}$ for which $|\mathcal{B}'| \leq h^{(r)}_{3.17}(|\mathcal{B}|)$ and $\mathrm{indm}(\mathcal{B}') \leq \kappa^{(r)}_{3.17}(|\mathcal{B}|, \mathrm{indm}(\mathcal{B}))$.*

Moreover, if $\mathcal{B}$ is in itself a syntactic refinement of some $\hat{\mathcal{B}}$ that is of rank at least $r(|\mathcal{B}|) + 1$, then additionally $\mathcal{B}'$ will be a syntactic refinement of $\hat{\mathcal{B}}$.

Proof: We will deal with the first case, and then show how to modify the proof for the case where being a syntactic refinement of some $\hat{\mathcal{B}}$ of the appropriate rank must be preserved.

Let $p_1, \ldots, p_C$ be the defining polynomials for $\mathcal{B}$, where $C = |\mathcal{B}|$. Suppose there is a linear combination over $\mathbb{F}$ that shows that $\mathcal{B}$ has a rank smaller than $r(C)$. This means that for some $(\alpha_1, \ldots, \alpha_C) \in \mathbb{F}^C \setminus \{0^C\}$, some arbitrary function $B : \mathbb{F}^l \to \mathbb{F}$ and polynomials $q_1, \ldots, q_l$ we have $\sum_{j=1}^{C} \alpha_j p_j(x) = B(q_1(x), \ldots, q_l(x))$ for every $x \in \mathbb{F}^n$, where $l < r(C)$ and every q_i is of degree smaller than $\max\{\deg(p_j) | \alpha_j \neq 0\}$ (a possible special case is where $l = 0$ and B is a constant).

We select j_0 so that $\alpha_{j_0} \neq 0$ and $\deg(p_{j_0}) = \max\{\deg(p_j) | \alpha_j \neq 0\}$, and construct $\mathcal{B}'$ by replacing p_{j_0} with $q_1, \ldots, q_l$. This is clearly a semantic refinement of $\mathcal{B}$ of complexity bounded by $h(C) = C + r(C) - 1$. Also, if $I = (i_1, \ldots)$ was the degree index of $\mathcal{B}$, then the degree index of $\mathcal{B}'$ is bounded above by the following $\kappa(C, I) = (j_1, \ldots)$: Letting k be the smallest number such that $i_k > 0$, we set $j_k = i_k - 1$, and if $k > 1$ then we set $j_{k-1} = i_{k-1} + r(C) - 1$; all other coordinates of $\kappa(C, I)$ are set equal to the respective coordinates of I.

The above argument provides us with $h^{(r)}_{3.17}$ and $\kappa^{(r)}_{3.17}$ as required.

Now we deal with an existing $\hat{\mathcal{B}}$ as above. We follow the same argument, but argue that we can find j_0 for which $\alpha_{j_0} \neq 0$ that corresponds to a maximal degree polynomial, satisfying additionally $j_0 > \hat{C} = |\hat{\mathcal{B}}|$. Assuming otherwise, we would find a counter example to the rank assumption on $\hat{B}$: We would get that $\sum_{j=1}^{\hat{C}} \alpha_j p_j(x)$ can be expressed as a function of $q_1, \ldots, q_l$ and $q_{l+1} = \sum_{j=\hat{C}+1}^{C} \alpha_j p_j$, which would all be of lower degree than $\max\{\deg(p_j) | 1 \leq j \leq \hat{C}, \alpha_j \neq 0\} = \max\{\deg(p_j) | \alpha_j \neq 0\}$, and would hence violate the rank of $\hat{\mathcal{B}}$. ■

Now we can combine the above two lemmas and prove the existence of high rank refinements.

Lemma 3.18 *There exists $D^{(d,r)}_{3.18}(k)$ which takes two numbers k and d and a monotone function $r : \mathbb{N} \to \mathbb{N}$, and satisfies the following. For every factor $\mathcal{B}$ of bounded degree d there is a semantic refinement $\mathcal{B}'$ for which $|\mathcal{B}'| \leq D^{(d,r)}_{3.18}(|\mathcal{B}|)$, is of bounded degree d and has rank at least $r(|\mathcal{B}'|)$.*

Moreover, if $\mathcal{B}$ is in itself a syntactic refinement of some $\hat{\mathcal{B}}$ that is of rank at least $r(D^{(d,r)}_{3.18}(|\mathcal{B}|)) + 1$, then additionally $\mathcal{B}'$ will be a syntactic refinement of $\hat{\mathcal{B}}$.

Proof: We set $D^{(d,r)}_{3.18}(k) = T_{3.16}(k, d, h^{(r)}_{3.17}, \kappa^{(r)}_{3.17})$. We set $\mathcal{B}_0 = \mathcal{B}$, and as long as $\mathcal{B}_i$ is of rank less than $r(|\mathcal{B}_i|)$ we move to a semantic refinement $\mathcal{B}_{i+1}$ as guaranteed by Lemma 3.17. By Lemma 3.16 the sequence $\mathcal{B}_0, \mathcal{B}_1, \ldots$ has length bounded by $m_{3.16}(k, d, h^{(r)}_{3.17}, \kappa^{(r)}_{3.17})$, and the

final factor $\mathcal{B}_l$ is of rank at least $r(|\mathcal{B}_l|)$ (otherwise we could have continued the sequence) and of complexity bounded by $T_{3.16}(k, d, h^{(r)}_{3.17}, \alpha^{(r)}_{3.17})$.

For the case of a prior factor $\hat{\mathcal{B}}$ we just use the corresponding case of Lemma 3.17. ■

Now we finally state the main technical lemma that we will use for our decompositions. It will find a refinement that is both robust and of high rank, while not breaking a given syntactic refinement relation to a high rank factor if one exists.

Lemma 3.19 (main robustness lemma) *For an appropriate function $T_{3.19}(k, h, d, r, \gamma)$, for any $\mathcal{B}$ of degree bound d, monotone $h : \mathbb{N} \to \mathbb{N}$ and $r : \mathbb{N} \to \mathbb{N}$, and $\gamma \in (0, 1)$ there exists a semantic refinement $\mathcal{B}'$ of $\mathcal{B}$ which is of rank at least $r(|\mathcal{B}'|)$ and (h, γ)-robust, for which $|\mathcal{B}'| \leq T_{3.19}(|\mathcal{B}|, h, d, r, \gamma)$.*

Moreover, if $\mathcal{B}$ is in itself a syntactic refinement of some $\hat{\mathcal{B}}$ that is of rank at least $r(T_{3.19}(|\mathcal{B}|, h, d, r, \gamma)) + 1$, then additionally $\mathcal{B}'$ will be a syntactic refinement of $\hat{\mathcal{B}}$ (this holds also for the case where $\mathcal{B} = \hat{\mathcal{B}}$).

Proof: We set $T_{3.19}(k, h, d, r, \gamma) = D^{(d,r)}_{3.18}\left(T_{3.13}(k, h \circ D^{(d,r)}_{3.18}, \gamma)\right)$. Given $\mathcal{B}$, we first use Lemma 3.13 to find $\mathcal{B}_1$ that is a syntactic refinement of $\mathcal{B}$ and is $(h \circ D^{(d,r)}_{3.18}, \gamma)$-robust. We then let $\mathcal{B}'$ be its semantic refinement according to Lemma 3.18 that is of rank $r(|\mathcal{B}'|)$. The complexity of $\mathcal{B}'$ is at most $D^{(d,r)}_{3.18}(|\mathcal{B}_1|)$, and hence (apart from being bounded by the above $T_{3.19}(|\mathcal{B}|, h, d, r, \gamma)$) by Observation 3.12 it is (h, γ)-robust as required.

For the case where there is a prior factor $\hat{\mathcal{B}}$ of the stated rank, we just use the corresponding case of Lemma 3.18. ■

4 Decomposition Theorems

We use here the tools of the previous section to prove two decomposition theorems. First we state and prove the strong decomposition theorem (it is called "strong" on account of also guaranteeing high rank); similar theorems were proved in previous works, and we only make a seemingly small (yet crucial to what follows) addition that preserves a given syntactic refinement relation. Then we state and prove the super decomposition theorem, which uses the strong decomposition theorem (or more accurately the main lemma implying it) as a lemma.

Super decomposition provides us with two successive factors, one being a syntactic refinement of the other. For the testing proofs, instead of using it directly, we will use a corollary that "chooses" out of the finer factor only one representative for each of the cells of coarser factor. This is done in the subcell selection corollary. The resulting representatives will satisfy properties that are stronger than what *any* one factor can satisfy by itself.

4.1 Strong Decomposition

First, a corollary of Theorem 3.2.

Lemma 4.1 *For $d < p$, suppose that $\mathcal{B}$ 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:*

$$\|\mathbb{E}[f|\mathcal{B}']\|_2^2 \geq \|\mathbb{E}[f|\mathcal{B}]\|_2^2 + (\epsilon_{3.2}(\delta,p))^2$$

where $\epsilon_{3.2}$ is the function in Theorem 3.2.

Proof: $g = f - \mathbb{E}[f|\mathcal{B}]$ is bounded to $[-1,1]$. So, applying Theorem 3.2 yields a degree-d polynomial P satisfying $|\mathbb{E}[g(x) \cdot \mathsf{e}(P(x))]| \geq \epsilon_{3.2}(\delta,p)$. The polynomial P generates a factor $\hat{\mathcal{B}}$ of complexity 1. Define $\mathcal{B}'$ to be the common refinement of $\mathcal{B}$ and $\hat{\mathcal{B}}$ (by adding P to the polynomials defining $\mathcal{B}$); its complexity is $C+1$.

Observe that:

$$\|\mathbb{E}[g|\mathcal{B}']\|_1 = \mathop{\mathbb{E}}_x\left[|\mathbb{E}[g|\mathcal{B}'](x)|\right] = \mathop{\mathbb{E}}_x\left[|\mathbb{E}[g|\mathcal{B}'](x) \cdot \mathsf{e}(P(x))|\right]$$
$$\geq \left|\mathop{\mathbb{E}}_x\left[\mathbb{E}[g|\mathcal{B}'](x) \cdot \mathsf{e}(P(x))\right]\right| = \left|\mathop{\mathbb{E}}_x[g(x) \cdot \mathsf{e}(P(x))]\right| \geq \epsilon_{3.2}(\delta,p)$$

where the second equality is simply due to $|\mathsf{e}(P(x))| = 1$, and the third equality uses the fact that P is constant on each atom of $\mathcal{B}'$. Now finally:

$$\|\mathbb{E}[f|\mathcal{B}']\|_2^2 - \|\mathbb{E}[f|\mathcal{B}]\|_2^2 = \|\mathbb{E}[f|\mathcal{B}'] - \mathbb{E}[f|\mathcal{B}]\|_2^2 = \|\mathbb{E}[g|\mathcal{B}']\|_2^2 \geq \|\mathbb{E}[g|\mathcal{B}']\|_1^2 \geq \epsilon_{3.2}^2(\delta,p)$$

where the first equality uses the fact that $\mathcal{B}'$ is a refinement of $\mathcal{B}$. ■

The contra-positive of the above provides us with a function decomposition given a sufficiently robust polynomial factor.

Lemma 4.2 *For any η and $d < p$ there exist $h_{4.2} : \mathbb{N} \to \mathbb{N}$ and $\gamma_{4.2}(\eta,p)$, so that if $\mathcal{B}$ is $(h_{4.2}, \gamma_{4.2}(\eta,p))$-robust (with respect to f) among factors of degree bound d over $\mathbb{F}_p^n$, then there is a decomposition $f = f_1 + f_2$ where f_1 is constant over every atom of $\mathcal{B}$ and ranges in $[0,1]$, and f_2 satisfies $\|f_2\|_{U^k} \leq \eta$ and ranges in $[-1,1]$.*

Proof: We set simply $h_{4.2}(k) = k+1$ and $\gamma_{4.2}(\eta,p) = \epsilon_{3.2}(\eta,p)^2$. Given $\mathcal{B}$ as above we set $f_1 = \mathbb{E}[f|\mathcal{B}]$ and $f_2 = f - \mathbb{E}[f|\mathcal{B}]$. These functions clearly have the required ranges. The robustness condition of $\mathcal{B}$ implies the contra-positive of the conclusion of Lemma 4.1, and so we must have $\|f_2\|_{U^k} \leq \eta$ as required. ■

However, we would like to make the Gowers norm bound also a function of $|\mathcal{B}|$. For this we will decompose f into three functions, where the third "error term" function has a bound on its l_2 norm. In fact an l_2 norm bound is what we need for an error term, but to reach even a constant l_2 norm bound we cannot avoid having also the function that has "only" a Gowers norm bound.

Lemma 4.3 *For any $d < p$, δ and $\eta : \mathbb{N} \to \mathbb{R}^+$ there exist $h_{4.3}^{(\eta,p)} : \mathbb{N} \to \mathbb{N}$ and $\gamma_{4.3}(\delta)$, so that if $\mathcal{B}$ is $(h_{4.3}^{(\eta,p)}, \gamma_{4.3}(\delta))$-robust (with respect to f) among factors of degree bound d, then there is a decomposition $f = f_1 + f_2 + f_3$ where f_1 is constant over every atom of $\mathcal{B}$ and ranges in $[0,1]$, f_2 satisfies $\|f_2\|_{U^k} \leq \eta(|\mathcal{B}|)$ and ranges in $[-1,1]$, and f_3 ranges in $[-1,1]$ and satisfies $\|f_3\|_2 \leq \delta$, where $f_1 + f_3$ also ranges in $[0,1]$.*

Proof: We set $h_{4.3}^{(\eta,p)}(m) = T_{3.13}(m, h_{4.2}, \gamma_{4.2}(\eta(m), p))$ for every $m \in \mathbb{N}$ and $\gamma_{4.3}(\delta) = \delta^2$. Given $\mathcal{B}$ satisfying the robustness condition above, we let $\mathcal{B}'$ be its syntactic refinement which is $(h_{4.2}, \gamma_{4.2}(\eta(|\mathcal{B}|), p))$-robust and for which $|\mathcal{B}'| \leq T_{3.13}(|\mathcal{B}|, h_{4.2}, \gamma_{4.2}(\eta(|\mathcal{B}|), p))$. We let $f_1 = \mathbb{E}[f|\mathcal{B}]$, and $f_2 = f - \mathbb{E}[f|\mathcal{B}']$. As per Lemma 4.2 f_2 satisfies the required Gowers norm condition. This leaves us with $f_3 = \mathbb{E}[f|\mathcal{B}'] - \mathbb{E}[f|\mathcal{B}]$. The required l_2 condition on this function follows directly from $\mathcal{B}'$ not violating the robustness condition of $\mathcal{B}$. ■

We now have all the tools to quickly wrap up the proof of the existence of a strong decomposition.

Theorem 4.4 (Strong Decomposition Theorem) *Suppose $\delta > 0$ and $C_0, d \geq 1$ 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 exists $C = C_{4.4}(\delta, \eta, p, r, C_0)$ such that the following holds.*

Given $f : \mathbb{F}_p^n \to \{0,1\}$ 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:

- $f = f_1 + f_2 + f_3$
- $f_1 = \mathbb{E}[f|\mathcal{B}]$
- $\|f_2\|_{U^{d+1}} \leq 1/\eta(|\mathcal{B}|)$
- $\|f_3\|_2 \leq \delta$
- *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 $r(C)+1$, then $\mathcal{B}$ will also be a syntactic refinement of $\hat{\mathcal{B}}$ (in particular this also holds if $\mathcal{B}_0 = \hat{\mathcal{B}}$).

Proof: Set $C_{4.4}(\delta, \eta, p, r, C_0) = T_{3.19}(C_0, h_{4.3}^{(\eta,p)}, p, r, \gamma_{4.3}(\delta)) \geq T_{3.19}(C_0, h_{4.3}^{(\eta,p)}, d, r, \gamma_{4.3}(\delta))$. Given $\mathcal{B}_0$ and f, we set $\mathcal{B}$ to be the $(h_{4.3}^{(\eta,p)}, \gamma_{4.3}(\delta))$-robust refinement of $\mathcal{B}_0$ guaranteed by Lemma 3.19. Lemma 4.3 guarantees the required decomposition $f = f_1 + f_2 + f_3$, and the case of a prior $\hat{\mathcal{B}}$ is also handled seamlessly by Lemma 3.19. ■

4.2 Super Decomposition and Subcell Selection

What we would really like is that in some sense the δ of Theorem 4.4 would also be able to depend on $|\mathcal{B}|$, but this is clearly impossible. So instead, taking some inspiration from [AFKS00], we will strive to have a sequence of two factors $\mathcal{B}$ and $\mathcal{B}'$, the latter a syntactic refinement of the former, so that the δ of $\mathcal{B}'$ would be a function of $|\mathcal{B}|$. However, for this to mean anything we also need $\mathcal{B}'$ to "faithfully" represent $\mathcal{B}$, in the sense that we define now.

Definition 4.5 (Polynomial factor represents another factor) *Given a function $f : \mathbb{F}_p^n \to \{0,1\}$, a polynomial factor $\mathcal{B}'$ that syntactically 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 a ζ fraction of cells c of $\mathcal{B}$, more than ζ fraction of the cells c' lying inside c satisfy $|\mathbb{E}[f|c] - \mathbb{E}[f|c']| > \zeta$.*

To be able to infer that a refinement is representing, we will use the following well-known defect version of the Cauchy-Schwartz inequality:

Observation 4.6 *If $\sum_{i\in I}^{m}\alpha_i = 1$ where α_i are all non-negative, $f : I \to \mathbb{R}$ ranges over $[0,1]$, and for some $J \subseteq I$ we have $\left(\sum_{j\in J}\alpha_i f(i)\right) / \left(\sum_{j\in J}\alpha_i\right) = \sum_{i\in I}\alpha_i f(i) + \eta$ where $\eta \in [-1,1]$, then $\sum_{i\in I}\alpha_i(f(i)^2) \geq (\sum_{i\in I}\alpha_i f(i))^2 + \left(\sum_{j\in J}\alpha_i\right)\eta^2$.*

Proof: For ease of notation denote the average $a = \sum_{i\in I}\alpha_i f(i)$ of f and set $\xi = \sum_{j\in J}\alpha_i$. By the standard Cauchy-Schwartz inequality $\sum_{i\in I}\alpha_i(f(i)^2) \geq \sum_{i\in I}\alpha_i(f'(i)^2)$, where $f'(i) = \left(\sum_{j\in J}\alpha_i f(i)\right) / \left(\sum_{j\in J}\alpha_i\right) = a+\eta$ if $i \in J$ and $f'(i) = \left(\sum_{j\in I\setminus J}\alpha_i f(i)\right) / \left(\sum_{j\in I\setminus J}\alpha_i\right) = a - \xi\eta/(1-\xi)$ if $i \notin J$. The sum over f' now equals $\xi(a+\eta)^2 + (1-\xi)(a-\xi\eta/(1-\xi))^2 \geq a^2 + \xi\eta^2$. ■

We can now show that, under some rank assumptions, a non-representing refinement is evidence to a factor being non-robust.

Lemma 4.7 *There are functions $r_{4.7}(p,m)$ and $\gamma_{4.7}(\zeta)$ for which the following holds. For $f : \mathbb{F}_p^n \to \{0,1\}$, if $\mathcal{B}'$ is a factor of rank $r_{4.7}(p,|\mathcal{B}'|)$, and is a syntactic refinement of a factor $\mathcal{B}$ of rank $r_{4.7}(p,|\mathcal{B}|)$, both of degree $d < p$, and $\mathcal{B}'$ does not ζ-represent $\mathcal{B}$ with respect to f, then $\mathrm{indd}(\mathcal{B}') \geq \mathrm{indd}(\mathcal{B}) + \gamma_{4.7}(\zeta)$.*

Proof: We first set $r_{4.7}(p,m) = r_{3.7}(p, 1/2p^m)$. If $\mathcal{B}'$ does not ζ-represent $\mathcal{B}$, then it must be the case that there are at least $\zeta p^{|\mathcal{B}|}/2$ cells of $\mathcal{B}$, so that for every cell c of them, there are at least $\zeta p^{|\mathcal{B}'|-|\mathcal{B}|}/2$ cells c' of $\mathcal{B}'$ lying inside of it, so that $|\,\mathbb{E}[f|c] - \mathbb{E}[f|c']| > \zeta$.

Let us concentrate for now on one such cell c of $\mathcal{B}$. Either there are at least $\zeta p^{|\mathcal{B}'|-|\mathcal{B}|}/4$ cells c' inside c so that $\mathbb{E}[f|c'] - \mathbb{E}[f|c] > \zeta$, or there are more than $\zeta p^{|\mathcal{B}'|-|\mathcal{B}|}/4$ such cells so that $\mathbb{E}[f|c'] - \mathbb{E}[f|c] < -\zeta$. We will assume the first case, as the treatment of the second case is virtually identical and provides the same lower bound for the cell.

Now we refer to Observation 4.6, where I is identified with $\mathbb{F}_p^{|\mathcal{B}'|-|\mathcal{B}|}$, the set of cells of $\mathcal{B}'$ lying in c, and J is identified with the set of those cells c' satisfying $\mathbb{E}[f|c'] - \mathbb{E}[f|c] > \zeta$. The value of each α_i can easily be shown to be at least $p^{|\mathcal{B}'|-|\mathcal{B}|}/3$, by comparing the minimum possible size of the cell c' with the maximum possible size of the cell c. Inserting the other corresponding values in Observation 4.6, we obtain $\mathbb{E}[\mathbb{E}[(f(x))^2]|c] > (\mathbb{E}[(f(x))^2|c])^2 + \zeta^3/12$.

Summing up the above contribution for all cells c of $\mathcal{B}$, and noting that the relative size of every cell of $\mathcal{B}$ is at least $p^{-|\mathcal{B}|}/2$ by Lemma 3.8, we obtain that $\mathrm{indd}(\mathcal{B}') = \mathbb{E}[(f(x))^2|\mathcal{B}'] \geq \mathbb{E}[(f(x))^2|\mathcal{B}] + \zeta^3/24 = \mathrm{indd}(\mathcal{B}) + \gamma_{4.7}(\zeta)$, where we set $\gamma_{4.7}(\zeta) = \zeta^3/24$. ■

The following technical lemma shows that if the partition is robust enough, then it has a specified robust and representing *syntactic* refinement, where we also take a rank requirement into account.

Lemma 4.8 *For every $h : \mathbb{N} \to \mathbb{N}$, $\gamma : \mathbb{N} \to (0,1)$, $r : \mathbb{N} \to \mathbb{N}$, $p \in \mathbb{N}$ and $\zeta \in (0,1)$ there are $H_{4.8}^{(h,\gamma,p,r)} : \mathbb{N} \to \mathbb{N}$, $R_{4.8}^{(h,\gamma,p,r)} : \mathbb{N} \to \mathbb{N}$ and $\Gamma_{4.8}(\zeta) \in (0,1)$ satisfying the following among factors of degree bound $d < p$ over $\mathbb{F}_p^n$. If $\mathcal{B}$ is an $(H_{4.8}^{(h,\gamma,p,r)}, \Gamma_{4.8}(\zeta))$-robust partition of rank at least $R_{4.8}^{(h,\gamma,p,r)}(|\mathcal{B}|)$, then it has a ζ-representing syntactic refinement $\mathcal{B}'$ which is $(h,\gamma(|\mathcal{B}|))$-robust and is of rank at least $r(|\mathcal{B}'|)$, which satisfies also $|\mathcal{B}'| \leq S_{4.8}(|\mathcal{B}|,h,p,r,\gamma)$ for the appropriate function $S_{4.8}(m,h,p,r,\gamma)$.*

Proof: Set the following in order:

$$\begin{aligned} S_{4.8}(m,h,p,r,\gamma) &= T_{3.19}(m,h,p,r,\gamma(m)) \\ H_{4.8}^{(h,\gamma,p,r)}(m) &= S_{4.8}(m,h,p,r,\gamma) \\ R_{4.8}^{(h,\gamma,p,r)}(m) &= \max\{r(S_{4.8}(m,h,p,r,\gamma))+1, r_{4.7}(p,m)\} \\ \Gamma_{4.8}(\zeta) &= \gamma_{4.7}(\zeta) \end{aligned}$$

Assuming that $\mathcal{B}$ satisfies the requisites, we use Lemma 3.19 to find a refinement $\mathcal{B}'$ that is $(h,\gamma(|\mathcal{B}|))$-robust, of rank at least $r(|\mathcal{B}'|)$, and satisfying $|\mathcal{B}'| \leq T_{3.19}(|\mathcal{B}|,h,d,r,\gamma(|\mathcal{B}|)) \leq T_{3.19}(|\mathcal{B}|,h,p,r,\gamma(|\mathcal{B}|))$ – the required complexity bound (note that Lemma 3.19 is fed the number $\gamma(|\mathcal{B}|)$, not the function γ).

The condition that $\mathcal{B}$ is $(H_{4.8}^{(h,\gamma,p,r)}, \Gamma_{4.8}(\zeta))$-semantically-robust means that $\mathrm{indd}(\mathcal{B}') \leq \mathrm{indd}(\mathcal{B}) + \gamma_{4.7}(\zeta)$, and so $\mathcal{B}'$ is ζ-representing for $\mathcal{B}$ by Lemma 4.7 (as the partitions also satisfy the corresponding rank requirement).

The condition that $\mathcal{B}$ is of rank at least $R_{4.8}^{(h,\gamma,p,r)}(|\mathcal{B}|) \geq r\left(T_{3.19}(|\mathcal{B}|,h,p,r,\gamma(|\mathcal{B}|))\right)+1$ means that (setting $\hat{\mathcal{B}} = \mathcal{B}$) Lemma 3.19 provides the additional requirement that $\mathcal{B}'$ is a syntactic refinement of $\mathcal{B}$. ■

We can now put forth our final decomposition theorem.

Theorem 4.9 (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 exists $C = C_{4.9}(\delta,\eta,p,r,\zeta,C_0)$ such that the following holds.*

Given $f : \mathbb{F}_p^n \to \{0,1\}$ 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: Let the function γ be defined by $\gamma(m) = \gamma_{4.3}(\delta(m))$ and let h be defined by $h(m) = h_{4.3}^{(\eta,p)}(m)$. Then set:

$$C_{4.9}(\delta,\eta,p,r,\zeta,C_0) = S_{4.8}\left(T_{3.19}\left(C_0, H_{4.8}^{(h,\gamma,p,r)}, p, R_{4.8}^{(h,\gamma,p,r)}, \Gamma_{4.8}(\zeta)\right), h, p, r, \gamma\right).$$

Given $\mathcal{B}_0$, we set $\mathcal{B}$ to be the semantic refinement that is guaranteed by Lemma 3.19 that is $\left(H_{4.8}^{(h,\gamma,p,r)}, \Gamma_{4.8}(\zeta)\right)$-robust and is of rank at least $R_{4.8}^{(h,\gamma,p,r)}(|\mathcal{B}|)$. $|\mathcal{B}|$ will be bounded by $T_{3.19}\left(C_0, H_{4.8}^{(h,\gamma,p,r)}, p, R_{4.8}^{(h,\gamma,p,r)}, \Gamma_{4.8}(\zeta)\right)$. Note also that $R_{4.8}^{(h,\gamma,p,r)}(|\mathcal{B}|) \geq r(|\mathcal{B}|)$.

Now we can use Lemma 4.8 to provide us a ζ-representing syntactic refinement $\mathcal{B}'$ of $\mathcal{B}$, that is of rank at least $r(|\mathcal{B}'|)$, and is $(h, \gamma(|\mathcal{B}|))$-robust and thus $\left(h_{4.3}^{(\eta,p)}, \gamma_{4.3}(\delta(|\mathcal{B}|))\right)$-robust. The factor $\mathcal{B}'$ satisfies the required complexity upper bound by substituting the bound on $|\mathcal{B}|$ into the guaranteed complexity bound of Lemma 4.8. Finally Lemma 4.3 provides the required decomposition $f = f_1 + f_2 + f_3$ over $\mathcal{B}'$. ■

One could envision future applications in which we would need the whole of $\mathcal{B}'$. Here we will need a careful choice of one cell of $\mathcal{B}'$ for every cell of $\mathcal{B}$. This selection will satisfy the following:

- The choice of cells will be made in a "uniform" manner. This part is helped by $\mathcal{B}'$ being a syntactic refinement. We will in fact set the "subcell ID" (the values of the polynomials appearing in $\mathcal{B}'$ and not in $\mathcal{B}$) to be the same for all cells of $\mathcal{B}$.
- All the subcells will feature a "good" decomposition, in terms of the norm of f_3.
- Most subcells will "well-represent" their corresponding cells from $\mathcal{B}$, in terms of the corresponding conditional expectation of f.

Now we state this formally.

Corollary 4.10 (Subcell 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 $C = C_{4.10}(\delta, \eta, p, r, \zeta)$ 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 $c' = (c,s) \in \mathbb{F}_p^{|\mathcal{B}'|}$ has the property that $\mathbb{E}_{\mathcal{B}(x)=c'}[(f_3(x))^2] < (\delta(|\mathcal{B}|))^2$.*
- $\Pr_{c \in \mathbb{F}_p^{|\mathcal{B}|}}[|\mathbb{E}[f|c] - \mathbb{E}[f|(c,s)]| > \zeta] < \zeta$, *where we denote $\mathbb{E}[f|c] = \mathbb{E}[f(x)|\mathcal{B}(x) = c]$ and $\mathbb{E}[f|(c,s)] = \mathbb{E}[f(x)|\mathcal{B}'(x) = (c,s)]$.*

Proof: Let $r'(m) = r_{3.7}(p, p^{-m}/10)$, so that by Theorem 3.7, a polynomial factor $\mathcal{B}$ of degree d and rank at least $r'(|\mathcal{B}|)$ satisfies 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 $C_{4.10}(\delta, \eta, p, r, \zeta) = C_{4.9}(\Delta, \eta, p, r'', \zeta/4, 1)$, where $\Delta(m) = 0.1 \cdot \delta(m)/p^m$ and $r''(m) = \max(r(m), r'(m))$. Apply Theorem 4.9 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 a $0.1p^{-|\mathcal{B}|}$ fraction of the subcells $c' \in \{c\} \times \mathbb{F}_p^{|\mathcal{B}'|-|\mathcal{B}|}$ of $\mathcal{B}'$ have $\mathbb{E}_x[(f_3(x))^2|c'] > \delta(|\mathcal{B}|)^2$. To show this, assume on the contrary that even for one cell $c \in \mathbb{F}_p^{|\mathcal{B}|}$ this event does not occur, and denote by S the set of cells $c' \in \mathbb{F}_p^{|\mathcal{B}'|}$ of $\mathcal{B}'$ that lie in c for which $\mathbb{E}_x[(f_3(x))^2|c'] > \delta(|\mathcal{B}|)^2$. By our assumption $|S| \geq (0.1p^{-|\mathcal{B}|})p^{|\mathcal{B}'|-|\mathcal{B}|}$, and then $\|f_3\|_2^2 = \mathbb{E}_{x \in \mathbb{F}_p^n}[(f_3(x))^2] > \delta(|\mathcal{B}|)^2 \Pr_{x \in \mathbb{F}_p^n}[\mathcal{B}(x) = c \wedge \mathcal{B}'(x) \in S] \geq 0.09\, \delta(|\mathcal{B}|)^2/p^{2|\mathcal{B}|} > \Delta(|\mathcal{B}|)^2$, a contradiction to the guarantee of Theorem 4.9.

Hence, for any fixed c, the probability that s is such that $\mathbb{E}_x[(f_3(x))^2|(c,s)] > \delta(|\mathcal{B}|)^2$ is at most $0.1p^{-|\mathcal{B}|}$. 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 $\mathbb{E}_x[(f_3(x))^2|c'] \leq \delta(|\mathcal{B}|)^2$.

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, by the Markov inequality, 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 that an s exists with both the desired properties. ■

4.3 Extending to Multiple Functions

The theorems so far referred to only a single function $f : \mathbb{F}_p^n \to \{0,1\}$. However, we actually require decomposition theorems which work for several functions $f^{(1)}, \ldots, f^{(R)} : \mathbb{F}_p^n \to \{0,1\}$ simultaneously with a single polynomial factor; alternatively, this could be thought of as decomposing a single "vector" function $f : \mathbb{F}_p^n \to \{0,1\}^R$.

It is quite straightforward to adapt all the previous proofs to this framework. The main adaptation to be done is the following version of the definition of a density index.

Definition 4.11 *The* density index *of a factor $\mathcal{B}$ with respect to a vector function $f = (f^{(1)}, \ldots, f^{(R)}) : \mathbb{F}_p^n \to \{0,1\}^R$ is the sum of the squared l_2 norms of the conditional expectation of the $f^{(i)}$ functions, that is* $\mathrm{indd}(\mathcal{B}) = \sum_{i=1}^R \mathbb{E}\left[(\mathbb{E}[f^{(i)}|\mathcal{B}])^2\right]$.

Given a function $h : \mathbb{N} \to \mathbb{N}$ and a real parameter γ, A factor $\mathcal{B}$ is (h,γ)-robust *(semantically) if there exists no $\mathcal{B}'$ which is a semantic refinement of $\mathcal{B}$ for which $|\mathcal{B}'| \leq h(|\mathcal{B}|)$ and* $\mathrm{indd}(\mathcal{B}') \geq \mathrm{indd}(\mathcal{B}) + \gamma$.

From here we can follow nearly the exact same arguments. The main difference is that now all resulting bounds will depend on R, starting with the multiple functions analog analog of $T_{3.19}$, as the index is now bounded by R rather than 1. Eventually we can reach the following version of the subcell selection theorem.

Theorem 4.12 (Subcell 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 $C = C_{4.12}(\delta, \eta, p, r, \zeta, R)$ 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 every* $i \in [R]$.
- $f_1^{(i)} = \mathbb{E}[f^{(i)}|\mathcal{B}']$ *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* $[-1,1]$.
- $\mathcal{B}$ *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* $\mathbb{E}_x[(f_3^{(i)}(x))^2 \mid \mathcal{B}'(x) = (c,s)] < (\delta(|\mathcal{B}|))^2$ *for every* $i \in [R]$.
- $\Pr_{c \in \mathbb{F}_p^{|\mathcal{B}|}}[\exists_{i \in [R]} |\mathbb{E}[f^{(i)}|c] - \mathbb{E}[f^{(i)}|(c,s)]| > \zeta] < \zeta$, *where we denote* $\mathbb{E}[f|c] = \mathbb{E}[f(x)|\mathcal{B}(x) = c]$ *and* $\mathbb{E}[f|(c,s)] = \mathbb{E}[f(x)|\mathcal{B}'(x) = (c,s)]$.

5 Counting and Testability

5.1 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:

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

where $(a_1, \ldots, a_m)$ is an affine constraint of size m on ℓ variables. In Lemma 3.8, 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 re-express (2) in the following way:

$$\begin{aligned}
&\Pr_{x_1,\ldots,x_\ell}[\mathcal{B}(a_1(x_1,\ldots,x_\ell)) = b_1 \wedge \cdots \wedge \mathcal{B}(a_m(x_1,\ldots,x_\ell)) = b_m] \\
&= \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in \mathbb{F}_p^n}\left[\prod_{i\in[C]}\prod_{j\in[m]} \frac{1}{p}\sum_{\lambda_{i,j}\in\mathbb{F}_p} \mathsf{e}\left(\lambda_{i,j}\cdot(P_i(a_j(x_1,\ldots,x_\ell)) - b_{i,j})\right)\right] \\
&= p^{-mC}\sum_{\substack{\lambda_{i,j}\in\mathbb{F}_p:\\ i\in[C],j\in[m]}} \mathsf{e}\left(-\sum_{i\in[C]}\sum_{j\in[m]}\lambda_{i,j}b_{i,j}\right) \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[\mathsf{e}\left(\sum_{i\in[C]}\sum_{j\in[m]}\lambda_{i,j}P_i(a_j(x_1,\ldots,x_\ell))\right)\right] \quad (3)
\end{aligned}$$

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

Lemma 5.1 (Lemma 5.1 in [HL11b]) *Suppose we are given $\epsilon \in (0,1)$, positive integer $d < p$ and an affine constraint (A, σ) where $A = (a_1, \ldots, a_m)$ is of size m and over 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 $r_{3.7}(d, \epsilon)$. 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:*

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

then 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 5.1. To this end, let us make the following definition, following the works of Gowers and Wolf [GW10b, GW10a].

Definition 5.2 (Dimension of linear forms) *For a positive integer d and linear form $L(X_1, \ldots, X_\ell) = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_\ell X_\ell$ where $\alpha_1, \ldots, \alpha_\ell \in \mathbb{F}_p$, let the dth tensor power of L denote:*

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

*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:*

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

To show the relevance of the above definition, we first need an algebraic "all or nothing" lemma from [HL11b] that concerns linear and polynomials without explicitly referring to the dimension of the forms.

Lemma 5.3 (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 (A, σ) where $A = (a_1, \ldots, a_m)$ be an affine constraint, where every linear form a_j is over 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:*
$$\sum_{i=1}^{C} \sum_{j=1}^{m} \lambda_{i,j} P_i\left(a_j(X_1, \ldots, X_\ell)\right) \not\equiv 0$$

Now we can make the connection between the definition of the dimension of the linear forms, and their effect on a sequence of polynomials with given degrees.

Lemma 5.4 *Let the notation here be same as in Lemma 5.1. 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. For this we use Lemma 5.3, so that instead of having the polynomials $P_1, \ldots, P_C$, we can analyze a collection of much simpler linearly independent polynomials of respective degrees $d_1, \ldots, d_C$.

In particular, let us define $P'_i(x) = x_i^{d_i}$ for every $i \in [C]$ (we assume that $n > C$). 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 $\sum_{j=1}^{m} \lambda_{i,j} a_j^{\otimes d_i} = 0$ for every $i \in [C]$.

Standard 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 move to the main theorem of this section. Let us first make the following definition, that in some ways captures the essence of "polynomial feasibility" for a sequence of values.

Definition 5.5 *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 $b_j = (b_{1,j}, \ldots, b_{C,j}) \in \mathbb{F}_p^C$ for every $j \in [m]$, are* consistent with respect to A and $d_1, \ldots, d_C$ *if the following is true:*

- *For every 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 following is easy to observe using basic linear algebra:

Observation 5.6 *Being consistent is equivalent to satisfying the following condition: For every 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 $i \in [C]$, we have $\sum_{i \in [C]} \sum_{j \in [m]} \lambda_{i,j} b_{i,j} = 0$.*

The following theorem shows that the expectation in (2) is nonzero, and is in fact close to a calculated number, if and only if $b_1, \ldots, b_m$ are consistent.

Theorem 5.7 *Let $\epsilon \in (0,1)$, let (A, σ) where $A = (a_1, \ldots, a_m)$ be an affine constraint over ℓ variables, and let $\mathcal{B}$ be a polynomial factor of degree d, complexity C and rank at least $r_{3.7}(d, \epsilon)$ generated by the polynomials $P_1, \ldots, P_C : \mathbb{F}_p^n \to \mathbb{F}_p$. For every $i \in [C]$, 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 every $j \in [m]$, fix the image of a cell in $\mathcal{B}$, indexed by $b_j = (b_{1,j}, \ldots, b_{C,j}) \in \mathbb{F}_p^C$.*

If $b_1, \ldots, b_m$ are consistent with respect to A and $d_1, \ldots, d_C$, then:

$$\Pr_{x_1, \ldots, x_\ell \in \mathbb{F}_p^n} \left[\mathcal{B}\left(a_1(x_1, \ldots, x_\ell)\right) = b_1 \wedge \cdots \wedge \mathcal{B}\left(a_m(x_1, \ldots, x_\ell)\right) = b_m\right] = p^{-s} \pm \epsilon$$

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 5.1, 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 5.4, 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 5.4. Thus:

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

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 $\mathcal{B}\left(a_j(x_1,\ldots,x_\ell)\right) = b_{i,j}$ for all $i\in[C]$ and $j\in[m]$. Then, for all possible values of $\lambda_{i,j}$ we have $\sum_{i\in[C],j\in[m]}\lambda_{i,j}b_{i,j} = \sum_{i\in[C],j\in[m]}\lambda_{i,j}P_i\left(a_j(x_1,\ldots,x_\ell)\right)$. But, by the argument of Lemma 5.4, $\sum_{j\in[m]}\lambda_{i,j}P_i\left(a_j(X_1,\ldots,X_\ell)\right)\equiv 0$ if $\sum_{j\in[m]}\lambda_{i,j}(a_j^{\otimes d_i}) = 0$ for any $i\in[C]$, and so the supposition is true. ■

By now we know the importance of the definition of consistency. One more building block that we need shows why, when selecting cells from a refining partition as in Theorem 4.9, consistency will pass over from $\mathcal{B}$ to $\mathcal{B}'$.

Lemma 5.8 *Suppose that $\mathcal{B}'$ is a syntactic refinement of $\mathcal{B}$, that $A=(a_1,\ldots,a_m)$ is a sequence of linear forms where (A,σ) is some affine constraint, and that $c_1,\ldots,c_m\in\mathbb{F}_p^{|\mathcal{B}|}$ are consistent with A and the degrees $d_1,\ldots,d_{|\mathcal{B}|}$ of the polynomials defining $\mathcal{B}$. Given any fixed $s\in\mathbb{F}_p^{|\mathcal{B}'|-|\mathcal{B}|}$, the cells $c'_1,\ldots,c'_m\in\mathbb{F}_p^{|\mathcal{B}'|}$ defined by the concatenations $c'_j=(c_j,s)$ for all $j\in[m]$ are consistent with respect to A and the degrees $d_1,\ldots,d_{|\mathcal{B}'|}$ of the polynomials defining $|\mathcal{B}'|$.*

Proof: Coordinate-wise, for $j\in[m]$, $c'_j=(c'_{1,j},c'_{2,j},\ldots,c'_{|\mathcal{B}'|,j})\in\mathbb{F}^{|\mathcal{B}'|}$ satisfies $c'_{i,j}=c_{i,j}$ for all $1\le i\le|\mathcal{B}|$ and $c'_{i,j}=s_{i-|\mathcal{B}|+1}$ for all $|\mathcal{B}|<i\le|\mathcal{B}'|$.

To show the consistency condition for $i>|\mathcal{B}|$, recall that each a_j is of the form $X_1+\sum_{r=2}^{\ell}c_rX_r$ for $c_r\in\mathbb{F}_p$ (as it came from an affine constraint). 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 $X_1^{\otimes d_i}$ (i.e. the one labeled by the sequence $(1,\ldots,1)$; the vector $a_j^{\otimes d_i}$ will always be 1 at that coordinate). 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 $\sum_{j\in[m]}\lambda_{i,j}c'_{i,j}=s_{i-|\mathcal{B}|+1}\sum_{j\in[m]}\lambda_{i,j}=0$ for all $i>|\mathcal{B}|$. Since we already know (by the consistency of the c_i relative to $\mathcal{B}$) that $\sum_{j\in[m]}\lambda_{i,j}c'_{i,j}=0$ for all $i\le|\mathcal{B}|$, we can conclude the proof. ■

5.2 Big Picture Arguments

We will prove the existence of many copies of a given linear constraint by analyzing the existence of a particular configuration of cells of a factor $\mathcal{B}$, where in every cell we look at the entire set of values that f can take at once. The following is a formal definition of the function giving the "big picture".

Definition 5.9 *Given a function $f : \mathbb{F}_p^n \to [R]$ and a polynomial factor $\mathcal{B}$, the* big picture function *of f is the function $f_{\mathcal{B}} : \mathbb{F}_p^{|\mathcal{B}|} \to 2^{[R]}$, 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 cell of $\mathcal{B}$.*

On the other hand, given any function $g : \mathbb{F}_n^C \to 2^{[R]}$, and a set of degrees $d_1, \ldots, d_C$ (of which we think as corresponding to the degrees of some future polynomial factor of size C), we will define what it means for such a function to "induce" a copy of a given constraint.

Definition 5.10 (Partially induce) *Suppose we are given positive integers $d_1, \ldots, d_C$, a function $g : \mathbb{F}_p^C \to 2^{[R]}$, and an induced affine constraint (A, σ) of size m over ℓ variables. We say that g* 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 $d_1, \ldots, d_C$.*
- *$\sigma_j \in g(b_j)$ for every $j \in [m]$.*

The big picture function defined above extracts a finitary description of a function $f : \mathbb{F}_p^n \to [R]$ 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 5.11 *If $f : \mathbb{F}_p^n \to [R]$ induces a constraint (A, σ), then for a factor $\mathcal{B}$ with degree sequence $(d_1, \ldots, d_{|\mathcal{B}|})$ (where all degrees are smaller than p), the function $f_{\mathcal{B}} : \mathbb{F}_p^C \to 2^{[R]}$ partially $(d_1, \ldots, d_{|\mathcal{B}|})$-induces (A, σ).*

Proof: Let m be the size of A and ℓ be its number of variables. Suppose that F induces (A, σ) at $x_1, \ldots, x_\ell$, and let $c_1, \ldots, c_m \in \mathbb{F}_p^{|\mathcal{B}|}$ be the images of the m cells in $\mathcal{B}$ defined by $c_1 = \mathcal{B}(a_1(x_1, \ldots, x_\ell)), c_2 = \mathcal{B}(a_2(x_1, \ldots, x_\ell)), \ldots, c_m = \mathcal{B}(a_m(x_1, \ldots, x_\ell))$ where $A = (a_1, \ldots, a_m)$. Then, because of the last condition in Theorem 5.7, it must be the case that $c_1, \ldots, c_m$ are consistent with respect to A and $d_1, \ldots, d_{|\mathcal{B}|}$. This fulfills the first condition of Definition 5.10, and the second condition is true by the definition of every $f_{\mathcal{B}}(c_i)$ including all values that f takes in that cell. ■

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

$$\Psi_{\mathcal{A}}(C) = \max_{d_1,\dots,d_C<p} \max_{g \in \mathcal{G}(d_1,\dots,d_C)} \min_{\substack{(A^i,\sigma^i) \text{ partially} \\ \text{induced by } g}} m_i$$

Whenever $\mathcal{G}(d_1, \dots, d_C)$ is empty we set the corresponding maximum to 0.

Note that the above is indeed finite, as both the number of possible degree sequences (bounded by p^C) and the size of $\mathcal{G}(d_1, \dots, d_C)$ (bounded by $2^{|R|p^C}$) 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 5.13 *Let $d_1, \dots, d_C < p$ be a degree sequence, for which a function $g : \mathbb{F}_p^C \to 2^{[R]}$ partially 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 $\mathcal{G}(d_1, \dots, d_C)$. ■

For our proofs, we will refer first not to f itself, but to some small modification of f that will make it a "perfect" representation of some cells from f according to some factor, which will be selected as per Corollary 4.10.

Definition 5.14 (Function cleanup) *Suppose we have a factor $\mathcal{B}'$ that is a syntactic refinement of $\mathcal{B}$, and some $s \in \mathbb{F}_n^{|\mathcal{B}'|-|\mathcal{B}|}$. The ζ-*cleanup *F of $f : \mathbb{F}_p^n \to [R]$ according to $\mathcal{B}$, $\mathcal{B}'$ and s is constructed by executing the following steps in order (where as usual (c, s) denotes the concatenation of c and s):*

1. *For every $z \in \mathbb{F}_p^n$ that is not covered by the cases below, let $F(z) = f(z)$.*
2. *For every cell c of $\mathcal{B}$ for which $|\Pr[f(x) = i \mid c] - \Pr[f(x) = i \mid (c, s)]| > \zeta$ for any $i \in [R]$, do the following. For every $z \in \mathcal{B}^{-1}(c)$, set $F(z) = \arg\max_{j \in [R]} \Pr[f(x) = j \mid (c, s)]$, the most popular value inside the subcell (c, s) (breaking ties arbitrarily, but consistently within each cell c).*
3. *For every cell c of $\mathcal{B}$, for every $i \in [R]$ such that $\Pr[f(x) = i \mid (c, s)] < \zeta$, set $F(z) = \arg\max_{j \in [R]} \Pr[f(x) = j \mid (c, s)]$ for every $z \in f^{-1}(i) \cap \mathcal{B}^{-1}(c)$ (breaking ties arbitrarily, but consistently within each cell c).*

Lemma 5.15 *If f, $\mathcal{B}$, $\mathcal{B}'$ and s are such that $\mathcal{B}$ is of rank at least $r_{3.7}(p, \beta/p^{|\mathcal{B}|})$, and $\Pr_{c \in \mathbb{F}_p^{|\mathcal{B}|}}[|\mathbb{E}[f|c] - \mathbb{E}[f|(c, s)]| > \zeta] < \zeta$, then the corresponding ζ-cleanup F is $(2R+1+\beta)\zeta$-close to f.*

Proof: Observe that the second step changes the value of F on at most a ζ fraction of the cells, by the condition involving s in the statement of the lemma. By Lemma 3.8, each cell occupies at most a $(1+\beta)p^{-C}$ fraction of the entire domain. So, the fraction of points whose values changed in the second step is at most $\zeta p^C \cdot (1+\beta)p^{-C} = (1+\beta)\zeta$.

The third step does not apply to any cell of $\mathcal{B}$ affected by the second step. Therefore, in the third case, for every $i \in [R]$, if $\Pr[f(x) = i \mid \mathcal{B}'(x) = (c,s)] < \zeta$ then $\Pr[f(x) = i \mid \mathcal{B}(x) = c] < 2\zeta$. Hence, the total fraction of the domain modified in the third case is at most $2R\zeta$. The total distance of F from f is therefor bounded by $(2R+1+\beta)\zeta$. ■

5.3 More about Algebra of Linear Forms

A linear form $a(X_1, \ldots, X_\ell) = \sum_{i=1}^{\ell} \alpha_i X_i$ can be identified with a linear function over $\mathbb{F}_p^\ell$, and thus a transformation in the spirit of "a change of basis" can be formulated.

Definition 5.16 (Change of view) *We identify the form $a(X_1, \ldots, X_\ell) = \sum_{i=1}^{\ell} \alpha_i X_i$ with the linear function $a : \mathbb{F}_p^\ell \to \mathbb{F}_p$ given by $a(v) = \sum_{i=1}^{\ell} \alpha_i v_i$, where $v = (v_1, \ldots, v_\ell) \in \mathbb{F}_i^\ell$ (in essence this is obtained by letting the X_i range over scalars from $\mathbb{F}_p$ rather than vectors from some space $\mathbb{F}_p^n$).*

Given an invertible $\ell \times \ell$ matrix M over $\mathbb{F}_p$, the corresponding change of view *of a is the linear form $a'(X_1, \ldots, X_\ell) = \sum_{i=1}^{\ell} \alpha'_i X_i$ obtained by the following process: Consider the linear function corresponding to a, perform on its domain $\mathbb{F}_p^\ell$ the change of variables corresponding to M, and then take the linear form corresponding its representation a' in the new basis.*

The reason that we use the term "change of view" is to not confuse it with a change of basis of $\mathbb{F}_p^n$. The following observation is easy:

Observation 5.17 *If (A, σ) is an affine constraint, and A' is obtained by performing the same change of view over all linear forms of A, then a function $f : \mathbb{F}_p^n \to \mathbb{F}$ satisfies (A, σ) if and only is it satisfies (A', σ).*

Additionally, a change of view does not affect the complexity of the affine constraint.

This yields the following lemma:

Lemma 5.18 *Any affine constraint (A, σ) is equivalent to one whose number of variables is not more than the number constraints.*

Proof: Assume that $A = (a_1, \ldots, a_m)$ take ℓ variables for $\ell > m$, and consider the linear functions from $\mathbb{F}_p^\ell$ to $\mathbb{F}$ corresponding to $a_1, \ldots, a_m$. By a linear dimension argument there are $\ell - m$ linearly independent vectors $u_1, \ldots, u_{\ell-m} \in \mathbb{F}_p^\ell$ for which $a_i(v_j) = 0$ for all $i \in [m]$ and $j \in [\ell - m]$. Complete these vectors to a basis $u_1, \ldots, u_\ell$ of $\mathbb{F}_p^\ell$, making sure that u_ℓ equals the vector that is 1 on its first coordinate and zero everywhere else (this vector is not in the span of $u_1, \ldots, u_{\ell-m} \in \mathbb{F}_p^\ell$, because by the definition of an affine constraint a_1 sends it to 1).

Now perform on the members of A the change of view corresponding to the change to this basis of $\mathbb{F}_p^\ell$. Denoting the resulting linear forms by $A' = (a'_1, \ldots, a'_m)$, we note now that no a'_i has any

mention of the variables $X_1, \ldots, X_{\ell-m}$, and so the constraint (A', σ) in fact takes at most m variables. A' will also have the standard form of an affine constraint with X_ℓ taking the place of X_1. ■

We need the above because the test would eventually query a number of places that is a function of p and the maximum number of variables in a subset of the constraints of $\mathcal{A}$, where this subset is only guaranteed a bound on the number of linear forms per constraint; we thus need $\mathcal{A}$ to satisfy the following definition:

Definition 5.19 (Concise collections) *The collection $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots\}$ is called* concise *if for every A_i, the total number of its variables does not exceed the number of its linear forms.*

Lemma 5.18 implies that every collection of linear constraints is equivalent to a concise one.

We would also need to know the (lack of) affect that a change of view has on the d-dimension, and hence the $(d_1, \ldots, d_C)$-dimension, of A.

Lemma 5.20 *If $A = (a_1, \ldots, a_m)$ is a sequence of linear forms, and $A' = (a'_1, \ldots, a'_m)$ is a sequence of the resulting forms after a fixed change of view, then A and A' have the same d-dimension for any d.*

Proof: We use the identification of linear forms with linear functions from $\mathbb{F}_p^\ell$ to $\mathbb{F}_p$, and by extension for a linear form a we consider the vector $a^{\otimes d}$ as the multilinear function $a^{\otimes d} : (\mathbb{F}_p^\ell)^d \to \mathbb{F}_p$ that sends $(v^{(1)}, \ldots, v^{(d)})$ to $\prod_{i=1}^d a(v^{(i)})$; the representation of this multilinear function in the standard basis indeed corresponds to the vector originally defined as $a^{\otimes d}$.

The operation that takes a to $a^{\otimes d}$ is not linear in itself; however, a change of basis over $\mathbb{F}_p^\ell$ (corresponding to the change of view) can be extended to an invertible linear operation over the linear space of all multilinear functions of d vectors (not all of which come from linear forms). Namely, if M is the basis change matrix, then the change of view for a sends it to the function defined by $a'(v) = a(Mv)$, and $(a')^{\otimes d}$ in fact corresponds to $\prod_{i=1}^d a(Mv^{(i)})$. Now by basic linear algebra, the operation that sends any multilinear form $b : (\mathbb{F}_p^\ell)^d \to \mathbb{F}_p$ to the form b' defined by $b'(v^{(1)}, \ldots, v^{(d)}) = a(Mv^{(1)}, \ldots, Mv^{(d)})$ is linear and invertible; thus the d-dimension, and in fact the exact corresponding linear dependencies, do not change when moving from $A = (a_1, \ldots, a_m)$ to $A' = (a'_1, \ldots, a'_m)$. ■

We end this section with a lemma about a "juxtaposition" of two sets of identical forms while sharing one variable.

Lemma 5.21 *Suppose that $(a'_1, \ldots, a'_m)$ are linear forms over $(X_1, \ldots, X_\ell)$ of d-dimension q, where for some k the form a'_k sends $(X_1, \ldots, X_\ell)$ to X_1. The d-dimension of the following $2m$ linear forms over $(Z, X_2, \ldots, X_\ell, Y_2, \ldots, Y_\ell)$:*

$$(a'_1(Z, X_2, \ldots, X_\ell), \ldots, a'_m(Z, X_2, \ldots, X_\ell), a'_1(Z, Y_2, \ldots, Y_\ell), \ldots, a'_m(Z, Y_2, \ldots, Y_\ell))$$

is exactly $2q - 1$.

Proof: We note that $a'_k(Z, X_2, \ldots, X_\ell) = a'_k(Z, Y_2, \ldots, Y_\ell) = Z$, and that all other linear forms are distinct. Abusing notation somewhat, we let Z denote also the linear form that returns the value of Z from the variables $(Z, X_2, \ldots, X_\ell, Y_2, \ldots, Y_\ell)$; note that in particular $Z^{\otimes d}$ corresponds to the vector from $\mathbb{F}_p^{(2\ell-1)^d}$ that is 1 on its coordinate corresponding to $(1, \ldots, 1)$, and zero everywhere else.

Let $S \subseteq \{1, \ldots, m\} \setminus \{k\}$ be a set of size $q-1$ such that $\left\{ \left(a'_j(Z, X_2, \ldots, X_m)\right)^{\otimes d} : j \in S \cup \{k\} \right\}$ is a basis of size q for the linear space $\text{span}\left\{ \left(a'_j(Z, X_2, \ldots, X_m)\right)^{\otimes d} : j \in [m] \right\}$. Clearly, $\left\{ \left(a'_j(Z, Y_2, \ldots, Y_m)\right)^{\otimes d} : j \in S \cup \{k\} \right\}$ is a basis for $\text{span}\left\{ \left(a'_j(Z, Y_2, \ldots, Y_m)\right)^{\otimes d} : j \in [m] \right\}$. Thus, the d-rank of the $2m$ linear forms is at most $2q-1$. To conclude, we will show that the d-rank is at least $2q-1$. To this end, we analyze the intersection

$$\text{span}\left\{ \left(a'_j(Z, X_2, \ldots, X_m)\right)^{\otimes d} : j \in S \cup \{k\} \right\} \cap \text{span}\left\{ \left(a'_j(Z, Y_2, \ldots, Y_m)\right)^{\otimes d} : j \in S \cup \{k\} \right\}.$$

It is clearly contained in $\text{span}\left\{Z^{\otimes d}\right\}$, since no other coordinate can be non-zero in both sets (the left set can have only non-zero coordinates corresponding to sequences of length d over $\{1, \ldots, \ell\}$, and the right set can have only non-zero coordinates corresponding to sequences of length d over $\{1, \ell+1, \ldots, 2\ell-1\}$). On the other hand, the intersection contains (and hence is equal to) $\text{span}\left\{Z^{\otimes d}\right\}$, because this vector appears on both sides (as a'_k). This shows by a linear dimension argument that the d-dimension of the $2m$ linear forms is exactly $2q-1$ as claimed. ■

5.4 The Proof of Testability

We finally have all the building blocks in place to prove Theorem 1.8, which implies Theorem 1.7.

Proof of Theorem 1.8: 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 the compactness function of $\mathcal{A}$.

$$\begin{aligned}
\alpha(C) &= p^{-2\Psi_A(C)C} \\
\rho(C) &= r_{3.7}(d, \alpha(C)) \\
\Delta(C) &= \frac{1}{16}\left(\frac{\epsilon}{8R}\right)^{\Psi_{\mathcal{A}}(C)} \\
\eta(C) &= \frac{1}{8(3p)^{C\Psi_{\mathcal{A}}(C)}}\left(\frac{\epsilon}{8R}\right)^{\Psi_{\mathcal{A}}(C)} \\
\zeta &= \frac{\epsilon}{8R}
\end{aligned}$$

$\ell_{\mathcal{A}}$ and $\delta_{\mathcal{A}}$ will be defined, based on the above functions, in (4) and (13) below.

Next, apply Theorem 4.12 to the functions $f^{(1)}, f^{(2)}, \ldots, f^{(R)}$ in order to get polynomial factors $\mathcal{B}' \preceq_{syn} \mathcal{B}$ of degree d and size at most $C_{4.12}(\Delta, \eta, p, \rho, \zeta, R)$, 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 every $i \in [R]$. 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 $P_1, \ldots, P_{|\mathcal{B}|}$.

Let F be the ζ-cleanup of f with respect to $\mathcal{B}$, $\mathcal{B}'$ and s. By Lemma 5.15, and what we know of these partitions and s, F is $\epsilon/2$-close to f, and hence by our assumption on the farness of f, the function F will still include an induced constraint from $\mathcal{A}$.

By Observation 5.11, the big picture function $F_{\mathcal{B}}$ of F will $(d_1, \ldots, d_{|\mathcal{B}|})$-partially induce some constraint from $\mathcal{A}$, and hence by Observation 5.13 it will partially induce some (A^i, σ^i) for which $m_i \leq \Psi_{\mathcal{A}}(|\mathcal{B}|)$. This will be the constraint of which we will find many copies in the original f. Let $m \stackrel{\text{def}}{=} m_i$, let $\ell \stackrel{\text{def}}{=} \ell_i$, and let $\sigma_1, \ldots, \sigma_m$ denote $\sigma_1^i, \ldots, \sigma_m^i$ respectively. Since a concise $\mathcal{A}$ means that $\ell_i \leq m_i$, we can now define

$$\ell_{\mathcal{A}}(\epsilon) = \Psi_{\mathcal{A}}(C_{4.12}(\Delta, \eta, p, \rho, \zeta, R)). \tag{4}$$

Denote the linear forms in A^i by $a_1, \ldots, a_m$ and denote $\sigma^i = (\sigma_1, \ldots, \sigma_m)$. 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 partially induced by $F_{\mathcal{B}}$, the big picture function of the cleanup function F, i.e., $c_1, \ldots, c_m$ are consistent, and $\sigma_i \in F_{\mathcal{B}}(c_i)$ for every $j \in [m]$. Also, let $c'_1, \ldots, c'_m \in \mathbb{F}_p^{|\mathcal{B}'|}$ index the associated subcells of $\mathcal{B}'$, obtained by letting $c'_j = (c_j, s)$ for every $j \in [m]$.

Our goal will now be to lower bound:

$$\begin{aligned}&\Pr_{x_1,\ldots,x_\ell \in \mathbb{F}_p^n}[f(a_1(x_1, \ldots, x_\ell)) = \sigma_1 \wedge \cdots \wedge f(a_m(x_1, \ldots, x_\ell)) = \sigma_m] \\ &= \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in \mathbb{F}_p^n}\left[f^{(\sigma_1)}(a_1(x_1, \ldots, x_\ell)) \cdots f^{(\sigma_m)}(a_m(x_1, \ldots, x_\ell))\right] \end{aligned} \tag{5}$$

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

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell \in \mathbb{F}_p^n}\left[(f_1^{(\sigma_1)} + f_2^{(\sigma_1)} + f_3^{(\sigma_1)})(a_1(x_1, \ldots, x_\ell)) \cdots (f_1^{(\sigma_m)} + f_2^{(\sigma_m)} + f_3^{(\sigma_m)})(a_m(x_1, \ldots, x_\ell))\right] \tag{6}$$

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 $\|f_2^{(\sigma_j)}\|_{U^{d+1}} \leq \eta(|\mathcal{B}'|)$, because of Lemma 3.3 and the fact that the complexity of A^i is bounded by d. Hence, the expression (6) is at least:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[(f_1^{(\sigma_1)} + f_3^{(\sigma_1)})(a_1(x_1, \ldots, x_\ell)) \cdots (f_1^{(\sigma_m)} + f_3^{(\sigma_m)})(a_m(x_1, \ldots, x_\ell))\right] - 3^m \eta(|\mathcal{B}'|) \tag{7}$$

Before we continue, to ease notation, for the rest of the proof we will now define an indicator function. $\mathcal{I}^{(c'_1,\ldots,c'_m)}_{(a_1,\ldots,a_m)}(x_1, \ldots, x_\ell)$ will be set to 1 if $\mathcal{B}'(a_j(x_1, \ldots, x_\ell)) = c'_j$ 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 (7) is at least:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[\left(f_1^{(\sigma_1)} + f_3^{(\sigma_1)}\right)(a_1(x_1, \ldots, x_\ell)) \cdots \left(f_1^{(\sigma_m)} + f_3^{(\sigma_m)}\right)(a_m(x_1, \ldots, x_\ell)) \cdot \mathcal{I}^{(c'_1,\ldots,c'_m)}_{(a_1,\ldots,a_m)}(x_1, \ldots, x_\ell)\right]$$

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:

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[f_1^{(\sigma_1)}(a_1(x_1,\ldots,x_\ell))\cdots f_1^{(\sigma_m)}(a_m(x_1,\ldots,x_\ell))\cdot \mathcal{I}^{(c'_1,\ldots,c'_m)}_{(a_1,\ldots,a_m)}(x_1,\ldots,x_\ell)\right] \tag{8}$$

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

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[f_3^{(\sigma_k)}(a_k(x_1,\ldots,x_\ell))g(x_1,\ldots,x_\ell)\cdot \mathcal{I}^{(c'_1,\ldots,c'_m)}_{(a_1,\ldots,a_m)}(x_1,\ldots,x_\ell)\right] \tag{9}$$

By our definition of affine constraints, $a_k(x_1,\ldots,x_\ell)$ is of the form $x_1 + \sum_{i\in[\ell]}\alpha_i x_i$ for some $\alpha_i \in \mathbb{F}_p$. We now change the summation variables of the expectation by replacing x_1 with $z = x_1 + \sum_{i\in[\ell]}\alpha_i x_i$, affecting a change of view for $a_1,\ldots,a_m$. Letting $a'_1,\ldots,a'_m$ 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 (9) using Cauchy-Schwartz as:

$$\begin{aligned}
&\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 \mathcal{I}^{(c'_1,\ldots,c'_m)}_{(a_1,\ldots,a_m)}(x_1,\ldots,x_\ell)\right]\right)^2\\
&\le \left(\mathop{\mathbb{E}}_{z,x_2,\ldots,x_\ell}\left[\left|f_3^{(\sigma_k)}(z)\right|\cdot \mathcal{I}^{(c'_1,\ldots,c'_m)}_{(a'_1,\ldots,a'_m)}(z,x_2,\ldots,x_\ell)\right]\right)^2\\
&\le \mathop{\mathbb{E}}_z\left[|f_3^{(\sigma_k)}(z)|^2\cdot \mathcal{I}^{(c'_k)}_{(\mathrm{id})}(z)\right]\cdot \mathop{\mathbb{E}}_z\left[\left(\mathop{\mathbb{E}}_{x_2,\ldots,x_\ell}\left[\mathcal{I}^{(c'_1,\ldots,c'_m)}_{(a'_1,\ldots,a'_m)}(z,x_2,\ldots,x_\ell)\right]\right)^2\right]\\
&\le \Delta^2(|\mathcal{B}|)\cdot \Pr_z[\mathcal{B}'(z)=c'_k]\cdot \mathop{\mathbb{E}}_z\left[\left(\mathop{\mathbb{E}}_{x_2,\ldots,x_\ell}\left[\mathcal{I}^{(c'_1,\ldots,c'_m)}_{(a'_1,\ldots,a'_m)}(z,x_2,\ldots,x_\ell)\right]\right)^2\right]\\
&\le \Delta^2(|\mathcal{B}|)\cdot (p^{-|\mathcal{B}'|}+\alpha(|\mathcal{B}'|))\cdot \mathop{\mathbb{E}}_z\left[\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(z,x_2,\ldots,x_\ell))-c'_{i,j})\right)\right)^2\right]\\
&\le \frac{2\Delta^2(|\mathcal{B}|)}{p^{2|\mathcal{B}'|m+|\mathcal{B}'|}}\mathop{\mathbb{E}}_z\left[\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(z,x_2,\ldots,x_\ell))\right)\right)^2\right]\\
&\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(\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)\cdot\right.\\
&\qquad\left.\mathop{\mathbb{E}}_{\substack{z,x_2,\ldots,x_\ell\\ y_2,\ldots,y_\ell}}\left[\mathsf{e}\left(\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\lambda_{i,j}P_i(a'_j(z,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(z,y_2,\ldots,y_\ell))\right)\right]\right)\\
&\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{z,x_2,\ldots,x_\ell\\ y_2,\ldots,y_\ell}}\left[\mathsf{e}\left(\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\lambda_{i,j}P_i(a'_j(z,x_2,\ldots,x_\ell))-\sum_{\substack{i\in[|\mathcal{B}'|]\\ j\in[m]}}\tau_{i,j}P_i(a'_j(z,y_2,\ldots,y_\ell))\right)\right]\right|
\end{aligned} \tag{10}$$

Now, by Lemma 5.20, the $(d_1, \ldots, d_{|\mathcal{B}'|})$-dimension of $\{a_1, \ldots, a_m\}$ equals the $(d_1, \ldots, d_{|\mathcal{B}'|})$-dimension of $\{a'_1, \ldots, a'_m\}$.

Let q denote the $(d_1, \ldots, d_{|\mathcal{B}'|})$-dimension of $\{a_1, \ldots, a_m\}$. By Lemma 5.21, summing over all of $(d_1, \ldots, d_{|\mathcal{B}'|})$, we know that the $(d_1, \ldots, d_{|\mathcal{B}'|})$-dimension of

$$\big(a'_1(Z, X_2, \ldots, X_\ell), \ldots, a'_m(Z, X_2, \ldots, X_\ell), a'_1(Z, Y_2, \ldots, Y_\ell), \ldots, a'_m(Z, Y_2, \ldots, Y_\ell)\big)$$

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

Now, just as in the proof of Theorem 5.7, the above information is enough to upper-bound (10). The above $(d_1, \ldots, d_{|\mathcal{B}'|})$-dimension bound and Lemma 5.4 allow us to count the number of $\lambda_{i,j}$ and $\tau_{i,j}$ such that the quantity inside the expectation in (10) is identically 1, and Lemma 5.1 along with the high-rank condition on the polynomials P_i bounds the expectation otherwise. It follows that (10), and therefore the square of (9), is at most:

$$\frac{2\Delta^2(|\mathcal{B}|)}{p^{2m|\mathcal{B}'|+|\mathcal{B}'|}} \left(p^{2m|\mathcal{B}'|-(2q-|\mathcal{B}'|)} + p^{2m|\mathcal{B}'|}\alpha(|\mathcal{B}'|)\right) \leq 2\Delta^2(|\mathcal{B}|) \cdot (p^{-2q} + \alpha(|\mathcal{B}'|)) \tag{11}$$

Finally, we lower-bound the contribution from the main term (8). To begin with, we need to convince ourselves that f induces many copies of (A^i, σ^i) among the subcells $c'_1, \ldots, c'_m$. Recall that $c_1, \ldots, c_m$ are consistent with $d_1, \ldots, d_{|\mathcal{B}|}$ and A^i, and that $\sigma_i \in F_{\mathcal{B}}(c_i)$ for every $i \in [m]$. By Lemma 5.8 $c'_1, \ldots, c'_m$ are consistent with $d_1, \ldots, d_{|\mathcal{B}'|}$ and A^i as well.

We can now lower-bound (8) as follows:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[f_1^{(\sigma_1)}(a_1(x_1, \ldots, x_\ell)) \cdots f_1^{(\sigma_m)}(a_m(x_1, \ldots, x_\ell)) \cdot \mathcal{I}^{(c'_1,\ldots,c'_m)}_{(a_1,\ldots,a_m)}(x_1, \ldots, x_\ell)\right] \\
&= \Pr[\mathcal{B}'(a_1(x_1, \ldots, x_\ell)) = c'_1 \wedge \cdots \wedge \mathcal{B}'(a_m(x_1, \ldots, x_\ell)) = c'_m] \cdot \\
&\qquad \mathop{\mathbb{E}}_{x_1,\ldots,x_\ell}\left[f_1^{(\sigma_1)}(a_1(x_1, \ldots, x_\ell)) \cdots f_1^{(\sigma_m)}(a_m(x_1, \ldots, x_\ell)) | \forall j \in [m]\ \mathcal{B}'(a_j(x_1, \ldots, x_\ell)) = c'_j\right] \\
&\geq (p^{-q} - \alpha(|\mathcal{B}'|)) \cdot \left(\frac{\epsilon}{8R}\right)^m
\end{aligned} \tag{12}$$

Let us justify the last line. The first term is due to Lemma 5.8 and the lower bound on the probability from Theorem 5.7. The second term in (12) is because each $f_1^{(\sigma_j)}$ is constant on the cells of $\mathcal{B}'$, and because by construction, the big picture function $F_{\mathcal{B}}$ of the cleanup function F, on which (A^i, σ^i) was partially induced, supports a value inside a cell c of $\mathcal{B}$ only if the original function f acquires the value on at least an $\epsilon/(8R)$ fraction of the subcell (c, s).

Combining the bounds from (7), (11) and (12), and using our parameter settings, we get that (5) is at least:

$$\begin{aligned}
&(p^{-q} - \alpha(|\mathcal{B}'|)) \cdot \left(\frac{\epsilon}{8R}\right)^m - \sqrt{2\Delta^2(|\mathcal{B}|) \cdot (p^{-2q} + \alpha(|\mathcal{B}'|))} - 3^m \cdot \eta(|\mathcal{B}'|) \\
&> \frac{p^{-q}}{2} \cdot \left(\frac{\epsilon}{8R}\right)^{\Psi_{\mathcal{A}}(|\mathcal{B}|)} - 2\Delta(|\mathcal{B}|) \cdot p^{-q} - 3^{\Psi_{\mathcal{A}}(|\mathcal{B}|)} \cdot \eta(|\mathcal{B}'|) \\
&> \frac{p^{-\Psi_A(|\mathcal{B}|)|\mathcal{B}'|}}{4} \cdot \left(\frac{\epsilon}{8R}\right)^{\Psi_{\mathcal{A}}(|\mathcal{B}|)}
\end{aligned}$$

where both $|\mathcal{B}|$ and $|\mathcal{B}'|$ are upper-bounded by $C_{4.12}(\Delta, \eta, p, \rho, \zeta, R)$. We can now define

$$\delta_{\mathcal{A}}(\epsilon) = \frac{1}{4} p^{-\Psi_A(C_{4.12}(\Delta,\eta,p,\rho,\zeta,R))C_{4.12}(\Delta,\eta,p,\rho,\zeta,R)} \cdot \left(\frac{\epsilon}{8R}\right)^{\Psi_{\mathcal{A}}(C_{4.12}(\Delta,\eta,p,\rho,\zeta,R))} \tag{13}$$

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.

[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.

[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.

[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.

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

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *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.