Electronic Colloquium on Computational Complexity, Report No. 29 (2011)

# Correlation testing for affine invariant properties on $\mathbb{F}_p^n$ in the high error regime

Hamed Hatami
School of Computer Science, McGill University, Montréal, Canada
hatami@cs.mcgill.ca

Shachar Lovett
Institute of Advanced Study, Princeton, USA
slovett@ias.edu

**Abstract**

Recently there has been much interest in Gowers uniformity norms from the perspective of theoretical computer science. This is mainly due to the fact that these norms provide a method for testing whether the maximum correlation of a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ with polynomials of degree at most $d \leq p$ is non-negligible, while making only a constant number of queries to the function. This is an instance of *correlation testing*. In this framework, a fixed test is applied to a function, and the acceptance probability of the test is dependent on the correlation of the function from the property. This is an analog of *proximity oblivious testing*, a notion coined by Goldreich and Ron, in the high error regime.

We study in this work general properties which are affine invariant and which are correlation testable using a constant number of queries. We show that any such property (as long as the field size is not too small) can in fact be tested by the Gowers uniformity test, and hence having correlation with the property is equivalent to having correlation with degree $d$ polynomials for some fixed $d$. We stress that our result holds also for non-linear properties which are affine invariant. This completely classifies affine invariant properties which are correlation testable.

The proof is based on higher-order Fourier analysis, where we establish a new approximate orthogonality for structures defined by linear forms. In particular, this resolves an open problem posed by Gowers and Wolf. Another ingredient is a nontrivial extension of a graph theoretical theorem of Erdös, Lovász and Spencer to the context of additive number theory.

AMS Subject Classification:
Keywords: Gowers uniformity, property testing, higher-order Fourier analysis, true complexity;

ISSN 1433-8092

# 1 Introduction

Blum, Luby, and Rubinfeld [2] made a beautiful observation that given a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$, it is possible to inquire the value of $f$ on a few random points, and accordingly probabilistically distinguish between the case that $f$ is a linear function and the case that $f$ has to be modified on at least $\varepsilon > 0$ fraction of points to become a linear function. Inspired by this observation, Rubinfeld and Sudan [15] defined the concept of property testing which is now a major area of research in theoretical computer science. Roughly speaking to test a function for a property means to examine the value of the function on a few random points, and accordingly (probabilistically) distinguish between the case that the function has the property and the case that it is not too close to any function with that property. Interestingly and to some extent surprisingly these tests exist for various basic properties. The first substantial investigation of property testing occurred in Goldreich, Goldwasser, and Ron [4] who showed that several natural combinatorial properties are testable. Since then there has been a significant amount of research on classifying the testable properties in combinatorial and algebraic settings.

Studying property testing in algebraic setting requires understanding the density of linear structures in subsets of Abelian groups. It is possible to express the density of those structures by certain analytical averages. Analyzing these averages and understanding the relations between them is the core of many problems and results in additive combinatorics and analytic number theory, and many tools and theories are developed for this purpose. The theory of uniformity, initiated by the proof of Szemerédi's theorem [16], plays an important role in this area, and it was a major breakthrough when Gowers [6] introduced a new notion of uniformity in a Fourier-analytic proof for Szemerédi's theorem.

Gowers' work initiated an extension of Fourier analysis, called higher-order Fourier analysis. In the classical Fourier-analysis of $\mathbb{F}_p^n$, a function is expressed as a linear combination of the characters of $\mathbb{F}_p^n$ which are exponentials of linear polynomials. In higher-order Fourier analysis, the linear polynomials are replaced by higher degree polynomials. Higher-order Fourier expansions are very useful in studying averages that are defined through linear structures. However for these expansions to be useful one needs some kind of orthogonality, or at least an approximation of it. The works of Green and Tao [10] and Kaufman and Lovett [12] provide an approximate orthogonality that can be used to analyze averages such as $\mathbb{E}_{X\in\mathbb{F}_p^n}[f_1(X)\dots f_m(X)]$ in a straightforward manner when proper higher-order Fourier expansions of $f_1,\dots,f_m$ are known. However it is not a priori clear that these results can be applied to analyze more general averages of the form

$$\mathbb{E}_{X_1,\dots,X_k\in\mathbb{F}_p^n}\left[f_1\left(\sum_{i=1}^k \lambda_{1,i}X_i\right)\dots f_m\left(\sum_{i=1}^k \lambda_{m,i}X_i\right)\right], \tag{1}$$

where $\lambda_{i,j} \in \mathbb{F}_p$ are constants. Such averages arise naturally when one studies linear structures in subsets of Abelian groups.

Our first result is an extension of the results of Green and Tao [10] which can be applied to such general averages. Then, we apply these techniques to determine which Gowers norms of $f_1,\dots,f_m$ one needs to bound in order to guarantee that the average in (1) is small. Our result in particular improves a result of Gowers and Wolf [7] and settles a conjecture posed by them.

Gowers' notion of uniformity has an interesting implication in the context of property testing. The Gowers norm of a function can be expressed as an average of values of this function on a few random sample points. There are theorems which show that a bounded function has non-negligible Gowers uniformity norm if and only if it has a non-negligible correlation with a low degree polynomial. These two facts show that having correlation with low degree polynomials is testable. Our main result is that, roughly speaking, the only correlation testable families of functions are the ones that can be tested by Gowers uniformity norms.

The tests that we study in this article are slightly different in nature from the typical statements in the area of property testing. Typically in property testing the goal is to distinguish the functions that are in a set $\mathcal{D}$ (structured) from the functions that are in a non-negligible distance from every element in $\mathcal{D}$. For example linearity testing [2], probably the most celebrated result in this area, says that it is possible to test whether a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is linear or that it has to be modified on at least a non-negligible fraction of all points in order to become a linear function. In this article, we are interested in a different kind of tests. Here we seek a weaker structure in the function, and having correlation with an element in $D$ replaces the usual condition of actually being in $\mathcal{D}$. This is in the spirit of proximity oblivious testing, a notion introduced by Goldreich and Ron [5] in the context of graph properties.

## 1.1 Notations

The complex unit disk is denoted by $\mathbb{D} = \{z \in \mathbb{C} : |z| \leq 1\}$. We denote elements of $\mathbb{F}_p^n$ by $x, y, z$, where $x = (x(1), \ldots, x(n))$. We frequently need to work with the elements of $(\mathbb{F}_p^n)^k$ which we regard as vectors with $k$ coordinates. We denote these by $\mathbf{x} = (x_1, \ldots, x_k) \in (\mathbb{F}_p^n)^k$. Capital letters $X$, $Y$, *etc* are used to denote random variables. For an element $m \in \mathbb{F}_p$, let $\mathrm{e}_p(m) := e^{\frac{2\pi i}{p} m}$. We denote by $f, g, \ldots$ functions from $\mathbb{F}_p^n$ to $\mathbb{C}$, and by $\mathrm{f}, \mathrm{g}, \ldots$ functions from $\mathbb{F}_p^n$ to $\mathbb{F}_p$. By $f := \mathrm{e}_p(\mathrm{f})$ we mean $f(x) = \mathrm{e}_p(\mathrm{f}(x))$. The *bias* of a function $f : \mathbb{F}_p^n \to \mathbb{C}$ is defined as

$$\mathrm{bias}(f) := \left|\mathbb{E}_{X \in \mathbb{F}_p^n}[f(X)]\right|. \tag{2}$$

The *bias* of a function $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$ is defined to be $\mathrm{bias}(\mathrm{f}) := \mathrm{bias}(\mathrm{e}_p(\mathrm{f}))$. The inner product of two functions $f, g : \mathbb{F}_p^n \to \mathbb{C}$ is defined as

$$\langle f, g \rangle := \mathbb{E}_{X \in \mathbb{F}_p^n}[f(X)\overline{g(X)}]. \tag{3}$$

The *correlation* of a function $f : \mathbb{F}_p^n \to \mathbb{C}$ with a set $D$ of functions from $\mathbb{F}_p^n$ to $\mathbb{C}$ is defined as

$$\|f\|_{u(D)} := \sup_{g \in D} |\langle f, g \rangle|. \tag{4}$$

By an abuse of notation, if $D$ is a set of functions from $\mathbb{F}_p^n$ to $\mathbb{F}_p$, we define

$$\|f\|_{u(D)} := \sup_{\mathrm{g} \in D} |\langle f, \mathrm{e}_p(\mathrm{g}) \rangle|.$$

Note that $\| \cdot \|_{u(D)}$ is always a semi-norm.

For two vector spaces $V$ and $W$ over $\mathbb{F}_p$, let $\mathrm{Lin}(V, W)$ denote the set of all *linear transformations* from $V$ to $W$. Let $\mathrm{Aff}(n, \mathbb{F}_p)$ denote the group of all *invertible affine transformations* from $\mathbb{F}_p^n$ to itself. For a function $f : \mathbb{F}_p^n \to \mathbb{C}$, and an $A \in \mathrm{Aff}(n, \mathbb{F}_p)$, we denote by $Af$ the function that maps $x$ to $f(Ax)$.

## 1.2 Gowers Uniformity Norms

Gowers uniformity norms are defined in the more general setting of arbitrary finite Abelian groups.

**Definition 1.1** (Gowers uniformity norms)**.** *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*

$$\|f\|_{U^k}^{2^k} := \mathbb{E}\left[\prod_{S\subseteq[k]} \mathcal{C}^{k-|S|} f\left(X + \sum_{i\in S} Y_i\right)\right], \tag{5}$$

*where $\mathcal{C}$ denotes the complex conjugation operator, and $X, Y_1, \ldots, Y_k$ are independent random variables taking values in $G$ uniformly at random.*

In this article we are only interested in the case where $G = \mathbb{F}_p^n$. These norms were first defined in [6] in the case where $G$ is the group $\mathbb{Z}_N$. Note that $\|f\|_{U^1} = |\mathbb{E}[f(X)]|$, and thus $\|\cdot\|_{U^1}$ is a semi-norm rather than a norm. The facts that the right-hand side of (5) is always non-negative, and that for $k > 1$, $\|\cdot\|_{U^k}$ is actually a norm are easy to prove, but certainly not trivial (see [6] for a proof).

Let us explain the relevance of the uniformity norms to the area of property testing. The goal in property testing is to obtain certain information about a function by "reading" its values only on a small number of points. The simplest case is that of testing correlation with linear functions. For $a \in \mathbb{F}_p^n$, let $\ell_a : \mathbb{F}_p^n \to \mathbb{F}_p$ be the corresponding linear function, defined as $\ell_a(x) = \sum_{i=1}^n a(i)x(i)$. Let $\text{Linear} = \{\ell_a : a \in \mathbb{F}_p^n\}$ be the set of linear functions. Let $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$ be a function, and let $f = \mathrm{e}_p(\mathrm{f})$. The correlation of f with linear functions is given by

$$\|f\|_{u(\text{Linear})} = \max_{\ell_a\in\text{Linear}} |\langle f, \mathrm{e}_p(\ell_a)\rangle| = \max_{a\in\mathbb{F}_p^n} \text{bias}(\mathrm{f} - \ell_a),$$

which is the same as the absolute value of the maximal Fourier coefficient of $f$. It is known [2, 6] that the correlation of f with linear functions is related to the $U^2$ norm of $f$. Specifically, for every $\varepsilon > 0$,

- *Direct Theorem:* If $\|f\|_{u(\text{Linear})} \geq \varepsilon$, then $\|f\|_{U^2} \geq \varepsilon$.
- *Inverse Theorem:* If $\|f\|_{U^2} \geq \varepsilon$, then $\|f\|_{u(\text{Linear})} \geq \varepsilon^2$.

These two facts together show that $\|f\|_{U^2}$ gives a rough estimate for the maximum correlation of f with linear functions. Recall that

$$\|f\|_{U^2}^4 = \|\mathrm{e}_p(\mathrm{f})\|_{U^2}^4 = \mathbb{E}[\mathrm{e}_p(\mathrm{f}(X+Y+Z) - \mathrm{f}(X+Y) - \mathrm{f}(X+Z) + \mathrm{f}(X))],$$

where $X, Y, Z \in \mathbb{F}_p^n$ are uniformly chosen. In other words, the joint distribution of $\mathrm{f}(X+Y+Z), \mathrm{f}(X+Y), \mathrm{f}(X+Z), \mathrm{f}(X)$ allows to distinguish between the case that f has correlation at least $\varepsilon$ with linear functions, and the case where f has correlation at most $\delta = \varepsilon^2/2$ (say) with linear functions. Hence, correlation with linear functions is "testable with just 4 queries to f" (for every $\varepsilon > 0$).

Similar to the case of linear functions, correlation with degree $d$ polynomials can be tested by the $U^{d+1}$ uniformity norm. The main ingredient is the inverse theorem for $\mathbb{F}_p^n$ which was proved in [1, 18]. Let $\text{Poly}_d(\mathbb{F}_p^n)$ be the set of polynomials of degree at most $d$ over $\mathbb{F}_p^n$. These results show that, as long as $p > d$ (i.e. the field is not too small), for every $\varepsilon > 0$,

- *Direct Theorem:* If $\|f\|_{u(\text{Poly}_d(\mathbb{F}_p^n))} \geq \varepsilon$, then $\|f\|_{U^{d+1}} \geq \varepsilon$.
- *Inverse Theorem:* If $\|f\|_{U^{d+1}} \geq \varepsilon$, then $\|f\|_{u(\text{Poly}_d(\mathbb{F}_p^n))} \geq \delta(\varepsilon) > 0$.

The exact dependency of $\delta(\varepsilon)$ on $\varepsilon$ is currently unknown, but crucially $\delta$ does not depend on $n$. Analogously to the case of linear functions and $U^2$, we get that the joint distribution of $(\mathrm{f}(X + \sum_{i \in I} Y_i) : I \subseteq [d+1])$ where $X, Y_1, \ldots, Y_{d+1} \in \mathbb{F}_p^n$ are uniformly chosen, distinguishes the case where f has noticeable correlation ($\geq \varepsilon$) with degree $d$ polynomials, from the case where f has negligible correlation ($\leq \delta(\varepsilon)/2$) with degree $d$ polynomials, for every $\varepsilon > 0$. Hence, correlation with polynomials of total degree $d$ is testable with just $2^{d+1}$ queries to f (for every $\varepsilon > 0$).

An equivalent qualitative formulation of the direct and inverse theorems stated above is as follows. Let $(\mathrm{f}_n : \mathbb{F}_p^n \to \mathbb{F}_p)_{n \in \mathbb{N}}$ be a sequence of functions. Then as long as $d < p$ we have that

$$\lim_{n \to \infty} \|\mathrm{e}_p(\mathrm{f}_n)\|_{u(\text{Poly}_d(\mathbb{F}_p^n))} = 0 \Longleftrightarrow \lim_{n \to \infty} \|\mathrm{e}_p(\mathrm{f}_n)\|_{U^{d+1}} = 0.$$

It will be convenient to express our results in such terms.

### 1.3 Correlation testable properties

The above discussion motivates the general definition of correlation testable properties. Let $\mathcal{D} = \{D_n\}_{n \in \mathbb{N}}$ be a family of sets, where each $D_n$ is a set of functions from $\mathbb{F}_p^n$ to $\mathbb{F}_p$. Informally, $\mathcal{D}$ is correlation testable using $q$ queries if there exists a distribution over $x_1, \ldots, x_q \in \mathbb{F}_p^n$, and for every $\varepsilon > 0$ there exists $\delta(\varepsilon) \in (0, \varepsilon)$, such that the following holds. The joint distribution of $(\mathrm{f}(x_1), \ldots, \mathrm{f}(x_q))$ allows to distinguish between the case that f has noticeable correlation ($\geq \varepsilon$) with $\mathcal{D}$ and the case that f has negligible correlation ($\leq \delta(\varepsilon)$) with $\mathcal{D}$.

**Definition 1.2** (Correlation testable properties)**.** *A family $\mathcal{D} = (D_n)$ is correlation testable with q queries, if there exists a distribution $\mu$ taking values in $(\mathbb{F}_p^n)^q$ and a mapping $\Gamma : \mathbb{F}_p^q \to \{0, 1\}$, such that the following holds. For every $\varepsilon > 0$, there exist $\delta \in (0, \varepsilon)$, $0 \leq \theta^- < \theta^+ \leq 1$ and $n_0 \in \mathbb{N}$, such that for every $n > n_0$ and $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$ we have:*

- *If $\|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \geq \varepsilon$ then $\Pr_{(X_1, \ldots, X_q) \sim \mu}[\Gamma(\mathrm{f}(X_1), \ldots, \mathrm{f}(X_q)) = 1] \geq \theta^+$.*
- *If $\|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \leq \delta$ then $\Pr_{(X_1, \ldots, X_q) \sim \mu}[\Gamma(\mathrm{f}(X_1), \ldots, \mathrm{f}(X_q)) = 1] \leq \theta^-$.*

Following the discussion on Gowers uniformity norms, $\text{Poly}_d = \{\text{Poly}_d(\mathbb{F}_p^n)\}_{n \in \mathbb{N}}$ is correlation testable using $q = 2^{d+1}$ queries, as long as $d < p$.

Our goal is to study families $\mathcal{D} = \{D_n\}_{n \in \mathbb{N}}$ which are correlation testable using a constant number of queries. We shall require the sets $D_n$ to be *consistent* with each other:

- **A1: (Consistency)** For positive integers $m > n$ and $\mathrm{g} \in D_n$, the function $\mathrm{h} : \mathbb{F}_p^m \to \mathbb{F}_p$ defined as $\mathrm{h}(x_1, \ldots, x_m) = \mathrm{g}(x_1, \ldots, x_n)$ belongs to $D_m$.

We need to make also a more crucial assumption. In this general setting, the algebraic structure of $\mathbb{F}_p^n$ is ignored, and we are treating $\mathbb{F}_p^n$ as a generic set of size $p^n$. In order to take the algebraic structure of $\mathbb{F}_p^n$ into account, we shall require $D_n$ to be *affine invariant*:

- **A2: (Affine invariance)** For every positive integer $n$, if $\mathrm{g} \in D_n$, then for every $A \in \text{Aff}(n, \mathbb{F}_p)$, we have $A\mathrm{g} \in D_n$.

Invariance plays a crucial role in the area of property testing. We refer to the work of Kaufman and Sudan [13] for the role of invariance in algebraic property testing. We stress that we **do not** require $\mathcal{D}$ to be linear; i.e. we do not require that if $\mathrm{f}, \mathrm{g} \in D_n$ then also $\mathrm{f} + \mathrm{g} \in D_n$.

The last condition relates to the size of $D_n$. We study families where one can distinguish functions with noticeable correlation from function with negligible correlation. In order not to make this meaningless, we would like to have functions with negligible correlation. For example, we would like a random function not to have correlation with $\mathcal{D}$ with high probability. Fix a function $\mathrm{g} \in D_n$. The number of functions $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$ which have correlation at least $\delta$ with g is $p^{c(\delta) \cdot p^n}$ where $\lim_{\delta \to 0} c(\delta) = 1$. Thus, a sufficient condition for random functions not be correlated with $D_n$ is that the size of $D_n$ is $p^{o(p^n)}$. We thus add the following requirement.

- **A3: (Sparsity)** For every $\varepsilon > 0$ and large enough $n$, we have $|D_n| \leq p^{\varepsilon \cdot p^n}$.

We call every $\mathcal{D} = \{D_n\}_{n \in \mathbb{N}}$ satisfying assumptions $\mathbf{A1}, \mathbf{A2}, \mathbf{A3}$ a *proper dual.*

Our main result is the following theorem which roughly speaking says that the only correlation testable families of functions are the ones that can be tested by Gowers uniformity norms provided that the field size $p$ is not too small. Say a sequence of functions $(\mathrm{f}_n)_{n \in \mathbb{N}}$ is *unbiased* if $\lim_{n \to \infty} \mathrm{bias}(\mathrm{f}_n) = 0$.

**Theorem 1.3** (Main theorem)**.** *Consider a proper dual $\mathcal{D} = \{D_n\}_{n \in \mathbb{N}}$. If $\mathcal{D}$ is correlation testable with $q$ queries, and if $p \geq q - 1$, then there exists $1 \leq k \leq q - 1$ such that the following holds. For every unbiased sequence of functions $(\mathrm{f}_n : \mathbb{F}_p^n \to \mathbb{F}_p)_{n \in \mathbb{N}}$, we have*

$$\lim_{n \to \infty} \|\mathrm{e}_p(\mathrm{f}_n)\|_{u(D_n)} = 0 \Longleftrightarrow \lim_{n \to \infty} \|\mathrm{e}_p(\mathrm{f}_n)\|_{U^{k+1}} = 0.$$

Equivalently, Theorem 1.3 can be phrased as follows: the sequence $(\mathrm{f}_n)_{n \in \mathbb{N}}$ has correlation with $\mathcal{D}$ iff it has correlation with polynomials of degree at most $k$.

We remark on the seemingly odd requirement that $\mathrm{f}_n$ is unbiased. Let $\mu$ be some distribution over $\mathbb{F}_p$, and let $\mathrm{f}_n : \mathbb{F}_p^n \to \mathbb{F}_p$ be a random function, where each $\mathrm{f}_n(x)$ is sampled independently according to $\mu$. By condition **A3**, since $|D_n| = p^{o(p^n)}$, we have that with probability $1 - o_n(1)$,

$$\lim_{n \to \infty} \|\mathrm{e}_p(\mathrm{f}_n)\|_{u(D_n)} = 0.$$

However, if $\mu$ has a nonzero bias, then almost surely $\mathrm{f}_n$ will have correlation with constant functions. It turns out that ruling out sequences of functions which have correlation with constant functions is sufficient for establishing Theorem 1.3.

**Paper organization** We give a short overview of the proof of our results in Section 2. We discuss systems of linear forms in Section A. We survey higher-order Fourier analysis in Section B. We prove new strong orthogonality results for systems of linear forms in Section C which in particular answers a conjecture of Gowers and Wolf. We prove a slightly weaker version of the main theorem (for strongly correlation testable families) in Section D. We prove the main result for correlation testable families in Section E. We prove the extension of Erdös-Lovász-Spencer theorem in Section F. We give some concluding remarks and pose some open problems in Section G.

## 2 Proof overview

In this section we give an overview of the proofs of our main results. We skip most of the technicalities, and try to emphasis the essence of the proofs. The proof of Theorem 1.3 is based on studying averages of functions evaluated on linear forms.

### 2.1 Linear forms

A *linear form* in $k$ variables is a vector $L = (\lambda_1, \dots, \lambda_k) \in \mathbb{F}_p^k$ regarded as a linear function from $V^k$ to $V$, for every vector space $V$ over $\mathbb{F}_p$: If $\mathbf{x} = (x_1, \dots, x_k) \in V^k$, then $L(\mathbf{x}) := \lambda_1 x_1 + \ldots + \lambda_k x_k$. A *system of linear forms* in $k$ variables is a finite set $\mathcal{L} = \{L_1, \dots, L_m\}$ of distinct linear forms $L_i$ in $k$ variables. For a function $f : \mathbb{F}_p^n \to \mathbb{C}$, and a system of linear forms $\mathcal{L} = \{L_1, \dots, L_m\}$ in $k$ variables, define the average

$$t_{\mathcal{L}}(f) := \mathbb{E}\left[\prod_{i=1}^{m} f(L_i(\mathbf{X}))\right], \tag{6}$$

where $\mathbf{X}$ is a random variable taking values uniformly in $(\mathbb{F}_p^n)^k$. We define two generalizations of such averages. First, when $f : \mathbb{F}_p^n \to \mathbb{C}$, one may take conjugations. For $\alpha \in \{0,1\}^m$, define

$$t_{\mathcal{L},\alpha}(f) := \mathbb{E}\left[\prod_{i=1}^{m} \mathcal{C}^{\alpha(i)} f(L_i(\mathbf{X}))\right], \tag{7}$$

where $\mathcal{C}$ is the conjugation operator. Second, if $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$, one may take coefficients in $\mathbb{F}_p$. For $\beta \in \mathbb{F}_p^n$, define

$$t^*_{\mathcal{L},\beta}(\mathrm{f}) := \mathbb{E}\left[\mathrm{e}_p\left(\sum_{i=1}^{m} \beta(i)\mathrm{f}(L_i(\mathbf{X}))\right)\right]. \tag{8}$$

We note that for functions $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$, (8) is more general than (7). Indeed, if $\alpha \in \{0,1\}^m$, then setting $\beta(i) = (-1)^{\alpha(i)}$ gives $t_{\mathcal{L},\alpha}(\mathrm{e}_p(\mathrm{f})) = t^*_{\mathcal{L},\beta}(\mathrm{f})$.

Let $\mathcal{D}$ be a proper dual family which is correlation testable. Then since $\mathcal{D}$ is affine invariant, it is easy to show that correlation with $\mathcal{D}$ can be expressed in terms of averages $t^*_{\mathcal{L},\beta}$. We show in Lemma E.2 that for every $\varepsilon > 0$ there exist $\delta \in (0, \varepsilon)$, systems of linear forms $\mathcal{L}_1, \dots, \mathcal{L}_\ell$ and corresponding coefficients $\beta_1, \dots, \beta_\ell$, and $n_0 \in \mathbb{N}$, such that the closures of the following two sets are disjoint:

$$T_\varepsilon := \{(t^*_{\mathcal{L}_1,\beta_1}(\mathrm{f}), \dots, t^*_{\mathcal{L}_\ell,\beta_\ell}(\mathrm{f})) | n > n_0, \mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p, \|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \geq \varepsilon\},$$

and

$$S_\varepsilon := \{(t^*_{\mathcal{L}_1,\beta_1}(\mathrm{f}), \dots, t^*_{\mathcal{L}_\ell,\beta_\ell}(\mathrm{f})) | n > n_0, \mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p, \|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \leq \delta\}.$$

Moreover, each of the systems $\mathcal{L}_1, \dots, \mathcal{L}_\ell$ has at most $q$ linear forms.

Recall that $\mathbb{D}$ is the unit disk in $\mathbb{C}$. It turns out to be easier to first analyze sets with a stronger requirement which holds for all functions $f : \mathbb{F}_p^n \to \mathbb{D}$, and not just for functions of the form $f = \mathrm{e}_p(\mathrm{f})$ for $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$. Say a family $\mathcal{D}$ is *strongly correlation testable* if the closures of the following two sets are disjoint:

$$T_\varepsilon := \{(t_{\mathcal{L}_1,\alpha_1}(f), \dots, t_{\mathcal{L}_\ell,\alpha_\ell}(f)) | n > n_0, f : \mathbb{F}_p^n \to \mathbb{D}, \|f\|_{u(D_n)} \geq \varepsilon\},$$

and

$$S_\varepsilon := \{(t_{\mathcal{L}_1,\alpha_1}(f), \dots, t_{\mathcal{L}_\ell,\alpha_\ell}(f)) | n > n_0, f : \mathbb{F}_p^n \to \mathbb{D}, \|f\|_{u(D_n)} \leq \delta\}.$$

## 2.2 Complexity of systems of linear forms

We turn to analyze averages of the form $t_{\mathcal{L},\alpha}(f)$, and more generally averages of the form $\mathbb{E}\left[\prod_{i=1}^m \mathcal{C}^{\alpha(i)} f_i(L_i(\mathbf{X}))\right]$ where $\mathbf{X} \in (\mathbb{F}_p^n)^k$ is uniform. A crucial ingredient is that we can "approximate" $f_i$ by nice functions, and that these approximation are essentially undetected by averages of the above form.

Green and Tao [11] defined the notion of the *Cauchy-Schwarz complexity* of a system of linear forms (for precise definition, see Definition A.1). This gives a parameter $s \in \mathbb{N}$ for which the following lemma follows by a sequence of clever applications of the Cauchy-Schwarz inequality.

**Lemma 2.1** ([11]). *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms of Cauchy-Schwarz complexity $s$. Let $f_i, g_i : \mathbb{F}_p^n \to \mathbb{D}$ be functions for $1 \le i \le m$. Assume that $\|f_i - g_i\|_{U^{s+1}} \le \varepsilon/m$ for all $1 \le i \le m$. Then*

$$\left| \mathbb{E}_{\mathbf{X}}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right] - \mathbb{E}_{\mathbf{X}}\left[\prod_{i=1}^m g_i(L_i(\mathbf{X}))\right] \right| \le \varepsilon,$$

*where $\mathbf{X} \in (\mathbb{F}_p^n)^k$ is uniform.*

Gowers and Wolf [8, 7] asked for the minimal $d \in \mathbb{N}$ such that a bound on $\max_{i\in[m]} \|f_i - \mathbb{E}[f_i]\|_{U^{d+1}}$ will allow to approximate $\mathbb{E}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right]$ by $\prod_{i=1}^m \mathbb{E}[f_i]$. They denoted this $d$ as the *true complexity* of a system of linear forms. For a linear form $L = (\lambda_1, \dots, \lambda_k) \in \mathbb{F}_p^k$ and for $t \in \mathbb{N}$, define $L^t \in \mathbb{F}_p^{k^t}$ to be the $t$-tensor power of $L$, defined as

$$L^t = \left( \prod_{j=1}^t \lambda_{i_j} : i_1, \dots, i_t \in [k] \right).$$

Gowers and Wolf [8, 7] showed that if the field size is not too small, then the true complexity is the minimal $d$ for which $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent.

**Theorem 2.2** ([8, 7]). *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms of Cauchy-Schwarz complexity $s$, and assume $p \ge s$. Let $d \ge 0$ be such that $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $f_1, \dots, f_m : \mathbb{F}_p^n \to \mathbb{D}$ be functions where $\|f_i - \mathbb{E}[f_i]\|_{U^{d+1}} \le \delta$. Then*

$$\left| \mathbb{E}_{\mathbf{X}}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right] - \prod_{i=1}^m \mathbb{E}[f_i] \right| \le \varepsilon,$$

*where $\mathbf{X} \in (\mathbb{F}_p^n)^k$ is uniform.*

We prove in this paper that Gowers and Wolf definition of true complexity extends also to the more general cases where $g_i$ may be any functions, and not just constant functions.

**Corollary A.8 (restated).** *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms of true complexity $d$ and Cauchy-Schwarz complexity at most $p$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $f_i, g_i : \mathbb{F}_p^n \to \mathbb{D}$ for $1 \le i \le m$ be functions such that $\|f_i - g_i\|_{U^{d+1}} \le \delta$. Then*

$$\left| \mathbb{E}_{\mathbf{X}}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right] - \mathbb{E}_{\mathbf{X}}\left[\prod_{i=1}^m g_i(L_i(\mathbf{X}))\right] \right| \le \varepsilon,$$

*where $\mathbf{X} \in (\mathbb{F}_p^n)^k$ is uniform.*

In fact, Corollary A.8 follows immediately from the following stronger theorem which we prove, which in particular resolves an open problem posed by Gowers and Wolf (Problem 7.6 in [7].

**Theorem C.6 (restated).** *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms of Cauchy-Schwarz complexity at most $p$. Let $d \geq 0$, and assume that $L_1^{d+1}$ is not in the linear span of $L_2^{d+1}, \ldots, L_m^{d+1}$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that for any functions $f_1, \ldots, f_m : \mathbb{F}_p \to \mathbb{D}$ where $\|f_1\|_{U^{d+1}} \leq \delta$ we have*

$$\left| \mathbb{E}_{\mathbf{X} \in (\mathbb{F}_p^n)^k} \left[ \prod_{i=1}^m f_i(L_i(\mathbf{X})) \right] \right| \leq \varepsilon.$$

## 2.3 Approximation by averages over polynomial factors

We now define the notion of a "nice approximating function", which will play the role of $g_1, \ldots, g_m$ in the previous subsection. These functions will be averages of $f_1, \ldots, f_m$ on polynomial factors.

A *polynomial factor* $\mathcal{B}$ is a partition (sigma-algebra) of $\mathbb{F}_p^n$ defined by a collection of polynomials $P_1, \ldots, P_C$. The atoms of the partitions are

$$\{x \in \mathbb{F}_p^n : P_1(x) = a(1), \ldots, P_C(x) = a(C)\},$$

where $a(1), \ldots, a(C) \in \mathbb{F}_p$. The *degree* of $\mathcal{B}$ is the maximal degree of $P_1, \ldots, P_C$, and the *complexity* of $\mathcal{B}$ is $C$.

The *rank* of a polynomial $P$ of degree $k$ is the minimal number of lower degree polynomials required to compute it. That is, $\text{rank}(P) \leq r$ if there exists $r$ polynomials $Q_1, \ldots, Q_r$ of degree at most $k-1$ such that $P(x) = \Gamma(Q_1(x), \ldots, Q_r(x))$ where $\Gamma : \mathbb{F}_p^r \to \mathbb{F}_p$ is some function. The rank of a set of polynomials $P_1, \ldots, P_C$ is the minimal rank of a nonzero linear combination of them. The rank of a polynomial factor $\mathcal{B}$ is the rank of the set of polynomials defining it.

The average of a function $f : \mathbb{F}_p^n \to \mathbb{D}$ over a polynomial factor $\mathcal{B}$, denoted $\mathbb{E}(f|\mathcal{B}) : \mathbb{F}_p^n \to \mathbb{D}$, is

$$\mathbb{E}(f|\mathcal{B})(x) = \mathbb{E}_{\{y \in \mathbb{F}_p^n : P_1(y) = P_1(x), \ldots, P_C(y) = P_C(x)\}}[f(y)].$$

That is, $\mathbb{E}(f|\mathcal{B})(x)$ is the average of $f$ in the atom to which $x$ belongs. The usefulness of these averages is that they may be used to approximate the function $f$. It is known (see Theorem B.7) that given any $d < p$ and $\delta > 0$, for every function $f : \mathbb{F}_p^n \to \mathbb{D}$, there exists a polynomial factor $\mathcal{B}$ of degree $d$ and bounded complexity such that

$$\|f - \mathbb{E}(f|\mathcal{B})\|_{U^{d+1}} \leq \delta.$$

Combining this with the discussion in the previous subsection, we may replace any arbitrary function $f$ with a structured function $\mathbb{E}(f|\mathcal{B})$ without changing averages $t_{\mathcal{L},\alpha}$ by much. Note that $\mathbb{E}(f|\mathcal{B})$ is $\mathcal{B}$-measurable, that is, it is constant on the atoms of $\mathcal{B}$. Thus, it will suffice to study averages $t_{\mathcal{L},\alpha}(g)$ where $g$ is a $\mathcal{B}$-measurable function.

## 2.4 Strong orthogonality and an invariance theorem

Let $g : \mathbb{F}_p^n \to \mathbb{D}$ be a $\mathcal{B}$-measurable function. Equivalently, it can be expressed as $g(x) = \Gamma(P_1(x), \dots, P_C(x))$, where $\Gamma : \mathbb{F}_p^C \to \mathbb{D}$ is some function. By the Fourier decomposition of $\Gamma$,

$$g(x) = \sum_{\gamma \in \mathbb{F}_p^C} \widehat{\Gamma}(\gamma) \mathrm{e}_p \left( \sum_{j=1}^{C} \gamma(j) \cdot P_j(x) \right).$$

Thus, to analyze the average $t_{\mathcal{L},\alpha}(g)$, we need to understand averages of the form

$$\mathbb{E}_{\mathbf{X}} \left[ \mathrm{e}_p \left( \sum_{i=1}^{m} \sum_{j=1}^{C} \lambda_{i,j} \cdot P_j(L_i(\mathbf{X})) \right) \right], \tag{9}$$

where $\lambda_{i,j} \in \mathbb{F}_p$ are coefficients.

It is known that polynomials which have large rank must be almost unbiased (Theorem B.11). That is, if $P_1, \dots, P_C$ have large enough rank, then

$$\left| \mathbb{E}_{\mathbf{X}} \left[ \mathrm{e}_p \left( \sum_{j=1}^{C} \gamma(j) \cdot P_j(\mathbf{X}) \right) \right] \right| \le \varepsilon$$

for all nonzero $\gamma \in \mathbb{F}_p^C$. However, when studying averages of the form $\sum \lambda_{i,j} P_j(L_i(\mathbf{x}))$ there could be cancelations; for example, if $P$ is a linear function then $P(x+y+z) - P(x+y) - P(x+z) + P(x) \equiv 0$.

We prove in Lemma C.1 that this is the only bad case that can occur. If $P_1, \dots, P_C$ are polynomials of degree less than $p$, then for every $\varepsilon > 0$, if their rank is large enough, then for every setting of $\lambda_{i,j} \in \mathbb{F}_p$ one of the following two cases must hold:

$$\left| \mathbb{E}_{\mathbf{X}} \left[ \mathrm{e}_p \left( \sum_{i=1}^{m} \sum_{j=1}^{C} \lambda_{i,j} \cdot P_j(L_i(\mathbf{X})) \right) \right] \right| \le \varepsilon,$$

or

$$\sum_{i=1}^{m} \sum_{j=1}^{C} \lambda_{i,j} \cdot P_j(L_i(\mathbf{x})) \equiv 0.$$

We study when the latter case may hold, and prove relevant conditions on the polynomials in Lemmas C.2 and C.3. Essentially, we show that the only important parameter is the degree of the polynomials. We use these to prove the following invariance theorem.

**Proposition C.5 (restated).** *Let $\mathcal{P} = \{P_1, \dots, P_C\}, \mathcal{Q} = \{Q_1, \dots, Q_C\}$ be two collections of polynomials over $\mathbb{F}_p^n$ of degree at most $d < p$ such that $\deg(P_i) = \deg(Q_i)$ for every $1 \le i \le C$. Let $L_1, \dots, L_m$ be linear forms, and $\Gamma : \mathbb{F}_p^C \to \mathbb{D}$ be an arbitrary function. Define $f, g : \mathbb{F}_p^n \to \mathbb{D}$ by*

$$f(x) = \Gamma(P_1(x), \dots, P_C(x))$$

*and*

$$g(x) = \Gamma(Q_1(x), \dots, Q_C(x)).$$

*Then for every $\varepsilon > 0$, if* $\mathrm{rank}(\mathcal{P}), \mathrm{rank}(\mathcal{Q})$ *are large enough then for all* $\alpha \in \{0,1\}^m$,

$$|t_{\mathcal{L},\alpha}(f) - t_{\mathcal{L},\alpha}(g)| \leq \varepsilon,$$

*if at least one of the following two conditions hold:*

*(i) The polynomials $P_1, \ldots, P_C$ and $Q_1, \ldots, Q_C$ are homogenous.*

*(ii) The system of linear forms $\{L_1, \ldots, L_m\}$ is homogenous.*

As we shall see, we may assume that all the systems of linear forms arising in our proofs are *homogenous*. Informally, a system of linear forms is homogeneous if (maybe after some change of basis) we have $L_i(x_1, \ldots, x_k) = L'_i(x_1, \ldots, x_{k-1}) + x_k$.

## 2.5 Interior of a set of averages of linear forms

Another ingredient in the proof of Theorem 1.3 is an extension of a graph theoretical result of Erdös, Lovász, and Spencer [3] to the setting of additive combinatorics. We show (Theorem F.1) that if $\mathcal{L}_1, \ldots, \mathcal{L}_\ell$ are systems of linear forms, then (excluding some trivial obstructions) then the set

$$\left\{(t_{\mathcal{L}_1}(f), \ldots, t_{\mathcal{L}_\ell}(f)) : f : \mathbb{F}_p^N \to [0,1]\right\} \subset \mathbb{R}^\ell$$

has a nonempty interior for some finite $N \in \mathbb{N}$. Formally, for the theorem to hold, we require the systems of linear forms to be *non-isomorphic* and *connected*:

Two systems of linear forms $\mathcal{L}' = \{L'_1, \ldots, L'_m\}$ and $\mathcal{L}'' = \{L''_1, \ldots, L''_m\}$ are *isomorphic*, if the distributions $(L'_1(\mathbf{X}), \ldots, L'_m(\mathbf{X}))$ and $(L''_1(\mathbf{X}), \ldots, L''_m(\mathbf{X}))$ are identical. Note that if $\mathcal{L}', \mathcal{L}''$ are isomorphic then $t_{\mathcal{L}'}(f) = t_{\mathcal{L}''}(f)$ for all functions $f : \mathbb{F}_p^n \to \mathbb{D}$.

A system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\}$ is *connected* if it cannot be partitioned as $\mathcal{L} = \mathcal{L}_1 \uplus \mathcal{L}_2$ where $\mathrm{span}(\mathcal{L}_1) \cap \mathrm{span}(\mathcal{L}_2) = \{\vec{0}\}$. Note that if $\mathcal{L}$ is not connected, then $t_{\mathcal{L}}(f) = t_{\mathcal{L}_1}(f) t_{\mathcal{L}_2}(f)$ for all functions $f : \mathbb{F}_p^n \to \mathbb{D}$.

## 2.6 Proof for strongly correlation testable families

Let $\mathcal{D} = \{D_n\}_{n \in \mathbb{N}}$ be a family which is strongly correlation testable. Assume for simplicity of exposition in the proof overview that the following holds (in the actual proof we have generalized averages $t_{\mathcal{L}_i,\alpha_i}(f)$): there exists a system of linear forms $\mathcal{L}_1, \ldots, \mathcal{L}_\ell$ such that for every $\varepsilon > 0$ there exists $\delta \in (0, \varepsilon)$ and $n_0 \in \mathbb{N}$, such that the following two sets are disjoint:

$$T_\varepsilon := \{(t_{\mathcal{L}_1}(f), \ldots, t_{\mathcal{L}_\ell}(f)) | n > n_0, f : \mathbb{F}_p^n \to \mathbb{D}, \|f\|_{u(D_n)} \geq \varepsilon\},$$

and

$$S_\varepsilon := \{(t_{\mathcal{L}_1}(f), \ldots, t_{\mathcal{L}_\ell}(f)) | n > n_0, f : \mathbb{F}_p^n \to \mathbb{D}, \|f\|_{u(D_n)} \leq \delta\}.$$

We can furthermore assume that these systems are homogeneous, and that their Cauchy-Schwarz complexity is $s < p$ (this follows from the assumption the the number of queries is $q \leq p$).

Let $t \in \mathbb{N}$ be maximal such the following holds. There exists polynomials $P_n$ of degree exactly $t$ such that

- $P_n$ has noticeable correlation with $D_n$: $\liminf_{n\to\infty} \|\mathrm{e}_p(P_n)\|_{u(D_n)} > 0$.

- $P_n$ has "large enough rank" (exact definition is deferred to the actual proof).

We can show that $t \leq s$: if $P_n$ are polynomials of degree $> s$ and large enough rank, then $\|\mathrm{e}_p(P_n)\|_{U^{s+1}} \approx 0$ and hence $t_{\mathcal{L}_i}(\mathrm{e}_p(P_n)) \approx 0$, for all $1 \leq i \leq \ell$. Thus we must have $\|\mathrm{e}_p(P_n)\|_{u(D_n)} \approx 0$.

Let $f_n : \mathbb{F}_p^n \to \mathbb{D}$ be a sequence of functions. We establish the main result (for strongly correlation testable families) by showing that:

$$\lim_{n\to\infty} \|f_n\|_{u(D_n)} = 0 \Longleftrightarrow \lim_{n\to\infty} \|f_n\|_{U^{t+1}} = 0.$$

**Proof of $\Leftarrow$:** Assume by contradiction that $\lim_{n\to\infty} \|f_n\|_{U^{t+1}} = 0$ but that $\varepsilon := \liminf_{n\to\infty} \|f_n\|_{u(D_n)} > 0$. Let $\varepsilon' > 0$ be the $L_\infty$ distance between $T_\varepsilon$ and $S_\varepsilon$ (here we use the fact that their closures are disjoint, hence there is positive distance between the sets). Let $\mathcal{B}_n$ be a polynomial factor of degree $s$ such that $\|f_n - \mathbb{E}(f_n|\mathcal{B}_n)\|_{U^{s+1}} \leq \eta$. Choosing $\eta > 0$ small enough we can guarantee that $|t_{\mathcal{L}_i}(f_n) - t_{\mathcal{L}_i}(\mathbb{E}(f_n|\mathcal{B}_n))| \leq \varepsilon'/2$, hence we must have $\|\mathbb{E}(f_n|\mathcal{B}_n)\|_{u(D_n)} \geq \delta$.

Assume $\mathcal{B}_n$ is defined by polynomials $P_{n,1}, \ldots, P_{n,C}$, which we can assume to be of high enough rank. For $\gamma \in \mathbb{F}_p^C$, define $P_{n,\gamma}(x) := \sum \gamma(i) P_{n,i}(x)$. We have $\mathbb{E}(f_n|\mathcal{B}_n)(x) = \sum_{\gamma \in \mathbb{F}_p^C} c_\gamma \mathrm{e}_p(P_{n,\gamma}(x))$ where $|c_\gamma| \leq 1$. Using the assumption that $\lim_{n\to\infty} \|f_n\|_{U^{t+1}} = 0$, we show that if $\deg(P_\gamma) \leq t$, then its contribution to the sum is negligible, i.e. $c_\gamma$ can be assumed to be arbitrarily small. Thus, there must exist a polynomial $P_{n,\gamma}$ of degree at least $t+1$ such that $\|\mathrm{e}_p(P_{n,\gamma})\|_{u(D_n)} \geq \delta p^{-C}$. As this polynomial has "large enough rank", this contradicts the maximality of $t$.

**Proof of $\Rightarrow$:** Assume that $\varepsilon := \lim_{n\to\infty} \|f_n\|_{U^{t+1}} > 0$ (actually we have $\liminf$, which we can replace by an actual limit by Condition **A1**). Thus, there exist polynomials $P_n$ of degree at most $t$ such that $|\langle f_n, P_n \rangle| \geq \varepsilon$. We can assume these polynomials to have "large enough rank". Assume first that these polynomials are of degree exactly $t$. By the definition of $t$, there exist polynomials $Q_n$ of degree $t$ and "large enough rank" such that $\|Q_n\|_{u(D_n)} \geq \varepsilon'$. We use the invariance theorem to construct a new sequence of functions $f'_n : \mathbb{F}_p^n \to \mathbb{D}$ (essentially replacing $P_n$ with $Q_n$) such that

- $f'_n$ has correlation with $Q_n$. In fact, this correlation is strong enough to guarantee that $\|f'_n\|_{u(D_n)} \geq \varepsilon''$ for some $\varepsilon'' > 0$ independent of $n$.
- Averages $t_{\mathcal{L}_i}$ cannot distinguish $f_n$ from $f'_n$. So, we must also have $\|f_n\|_{u(D_n)} \geq \delta'' > 0$.

The case where $P_n$ have degrees less than $t$ is reduced to the case of degree $t$. We use the theorem on the interior of a set of averages of linear forms to argue that we can (essentially) tweak $f_n$ slightly, without changing averages $t_{\mathcal{L}_i}$ by much, but such that $f_n$ will have correlation with polynomials of degree exactly $t$ and large enough rank.

## 2.7 Proof for correlation testable families

The proof for the correlation testable families (i.e. with guarantees only for functions $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$) follows similar steps, albeit slightly more involved. Let $P(\mathbb{F}_p) \subset \mathbb{R}^p$ denote the convex set of probability distributions defined over $\mathbb{F}_p$. Then correlation testable families can in fact test randomized functions as well, i.e. functions $\mathrm{f} : \mathbb{F}_p^n \to P(\mathbb{F}_p)$. Once this is established the remainder of the proof follows identical lines to the case of strongly correlation testable families, where we employ the fact that $P(\mathbb{F}_p)$ is a convex set to appropriately define averages such as $\mathbb{E}(\mathrm{f}|\mathcal{B})$.

# A Systems of linear forms

Suppose that $p > 3$ is a prime, and let $A$ be a subset of $\mathbb{F}_p^n$. The density of the 3-terms arithmetic progressions in $A$ is given by

$$\mathbb{E}\left[1_A(X)1_A(X+Y)1_A(X+2Y) \mid Y \neq 0\right], \tag{10}$$

where $X$ and $Y$ are independent uniform random variables. If we also allow the degenerate 3-term arithmetic progressions $(x, x, x)$, then the formula will simply be

$$\mathbb{E}\left[1_A(X)1_A(X+Y)1_A(X+2Y)\right].$$

Here we used the linear forms $x$, $x + y$, and $x + 2y$ to express the number of 3-terms arithmetic progressions in $A$.

Generalizing these notions, a *linear form* in $k$ variables is a vector $L = (\lambda_1, \ldots, \lambda_k) \in \mathbb{F}_p^k$ regarded as a linear function from $V^k$ to $V$, for every vector space $V$ over $\mathbb{F}_p$: If $\mathbf{x} = (x_1, \ldots, x_k) \in V^k$, then $L(\mathbf{x}) := \lambda_1 x_1 + \ldots + \lambda_k x_k$. A *system of $m$ linear forms* in $k$ variables is a finite set $\mathcal{L} = \{L_1, \ldots, L_m\}$ of *distinct* linear forms, each in $k$ variables. For a function $f : \mathbb{F}_p^n \to \mathbb{C}$, and a system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\}$ in $k$ variables, define

$$t_{\mathcal{L}}(f) := \mathbb{E}\left[\prod_{i=1}^{m} f(L_i(\mathbf{X}))\right], \tag{11}$$

where $\mathbf{X}$ is a random variable taking values uniformly in $(\mathbb{F}_p^n)^k$. More generally, we consider two types of generalized averages. First, when $f : \mathbb{F}_p^n \to \mathbb{C}$ takes complex values, we allow also conjugation. Let $\mathcal{C}$ denote the conjugation operator $\mathcal{C}(z) = \overline{z}$. Then for $\alpha \in \{0,1\}^m$ define

$$t_{\mathcal{L},\alpha}(f) := \mathbb{E}\left[\prod_{i=1}^{m} \mathcal{C}^{\alpha(i)} f(L_i(\mathbf{X}))\right]. \tag{12}$$

Note that $t_{\mathcal{L},0^m}(\cdot) = t_{\mathcal{L}}(\cdot)$. Second, when $f := \mathrm{e}_p(\mathrm{f})$ where $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$, we also allow coefficients. For $\beta \in \mathbb{F}_p^m$, define

$$t^*_{\mathcal{L},\beta}(\mathrm{f}) := \mathbb{E}\left[\mathrm{e}_p\left(\sum_{i=1}^{m} \beta(i)\mathrm{f}(L_i(\mathbf{X}))\right)\right]. \tag{13}$$

We note that for functions $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$, the latter average is more general than the first one: for every $\alpha \in \{0,1\}^m$, let $\beta \in \{-1,1\}^m$ be set as $\beta(i) = (-1)^{\alpha(i)}$; then $t^*_{\mathcal{L},\beta}(\mathrm{f}) = t_{\mathcal{L},\alpha}(\mathrm{e}_p(\mathrm{f}))$.

Note that if $A \subseteq \mathbb{F}_p^n$ and $1_A : \mathbb{F}_p^n \to \{0,1\}$ is the indicator function for $A$, then $t_{\mathcal{L}}(1_A)$ is the probability that $L_1(\mathbf{X}), \ldots, L_m(\mathbf{X})$ all fall in $A$. Roughly speaking, we say $A \subseteq \mathbb{F}_p^n$ is pseudorandom with regards to $\mathcal{L}$, if

$$t_{\mathcal{L}}(1_A) \approx \left(\frac{|A|}{p^n}\right)^m,$$

that is if the probability that all $L_1(\mathbf{X}), \ldots, L_m(\mathbf{X})$ fall in $A$ is close to what we would expect if $A$ was a random subset of $\mathbb{F}_p^n$ of size $|A|$. Let $\alpha = |A|/p^n$ be the relative measure of $A$, and define $f(x) := 1_A(x) - \alpha$. We have

$$t_{\mathcal{L}}(1_A) = t_{\mathcal{L}}(\alpha + f) = \alpha^m + \sum_{S \subseteq [m], S \neq \emptyset} \alpha^{m-|S|} \cdot t_{\{L_i : i \in S\}}(f).$$

So, a sufficient condition for $A$ to be pseudorandom with regards to $\mathcal{L}$ is that $t_{\{L_i:i\in S\}}(f) \ll \alpha^m$ for all nonempty subsets $S \subseteq [m]$. Green and Tao [11] showed that a sufficient condition for this to occur is that $\|f\|_{U^{s+1}}$ is small enough, where $s$ is the *Cauchy-Schwarz complexity* of the system of linear forms.

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

The reason for the term *Cauchy-Schwarz complexity* is the following lemma due to Green and Tao [11], whose proof is based on a clever iterative application of the Cauchy-Schwarz inequality.

**Lemma A.2** ([11]). *Let $f_1, \dots, f_m : \mathbb{F}_p \to \mathbb{D}$. Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of $m$ linear forms in $k$ variables of Cauchy-Schwarz complexity $s$. Then*

$$\left| \mathbb{E}_{\mathbf{X}\in(\mathbb{F}_p^n)^k} \left[ \prod_{i=1}^m f_i(L_i(\mathbf{X})) \right] \right| \le \min_{1\le i\le m} \|f_i\|_{U^{s+1}}.$$

Note that the Cauchy-Schwarz complexity of any system of $m$ linear forms in which any two linear forms are linearly independent (i.e. one is not a multiple of the other) is at most $m-2$, since we can always partition $\{L_j\}_{j\in[m]\setminus\{i\}}$ into the $m-1$ singleton subsets.

The following is an immediate corollary of Lemma A.2.

**Corollary A.3.** *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms in $k$ variables of Cauchy-Schwarz complexity $s$. Let $f_i, g_i : \mathbb{F}_p^n \to \mathbb{D}$ be functions for $1 \le i \le m$. Assume that $\|f_i - g_i\|_{U^{s+1}} \le \frac{\varepsilon}{2^m m}$ for all $1 \le i \le m$. Then*

$$\left| \mathbb{E}_{\mathbf{X}} \left[ \prod_{i=1}^m f_i(L_i(\mathbf{X})) \right] - \mathbb{E}_{\mathbf{X}} \left[ \prod_{i=1}^m g_i(L_i(\mathbf{X})) \right] \right| \le \varepsilon,$$

*where $\mathbf{X} \in (\mathbb{F}_p^n)^k$ is uniform.*

In particular, if $A \subseteq \mathbb{F}_p^n$ of size $|A| = \alpha p^n$ satisfies $\|1_A - \alpha\|_{U^{s+1}} \ll \alpha^m$, then $t_{\mathcal{L}}(1_A) \approx \alpha^m$.

## A.1 The true complexity of linear forms

The Cauchy-Schwarz complexity of $\mathcal{L}$ gives an upper bound on $s$, such that if $\|1_A - \alpha\|_{U^{s+1}}$ is small enough then $A$ is pseudorandom with regards to $\mathcal{L}$. Gowers and Wolf [8] defined the *true complexity* of a system of linear forms as the minimal $s$ such that the above condition holds for all sets $A$.

**Definition A.4** (True complexity [8]). *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms over $\mathbb{F}_p$. The true complexity of $\mathcal{L}$ is the smallest $d \in \mathbb{N}$ with the following property. For every $\varepsilon > 0$, there exists $\delta > 0$ such that if $f : \mathbb{F}_p^n \to \mathbb{D}$ is any function with $\|f\|_{U^{d+1}} \le \delta$, then*

$$|t_{\mathcal{L}}(f)| \le \varepsilon.$$

An obvious bound on the true complexity is the Cauchy-Schwarz complexity of the system. However, there are cases where this is not tight. Gowers and Wolf [7] characterized the true complexity of systems of linear forms, assuming the field is not too small. For a linear form $L \in \mathbb{F}_p^m$, let $L^k \in \mathbb{F}_p^{m^k}$ be the $k$-tensor power of $L$. That is, if $L = (\lambda_1, \dots, \lambda_m)$, then

$$L^k = \left( \prod_{j=1}^{k} \lambda_{i_j} : i_1, \dots, i_k \in [m] \right).$$

**Theorem A.5** (Characterization of the true complexity of linear systems, Theorem 6.1 in [7])**.** *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms over $\mathbb{F}_p^n$ of Cauchy-Schwarz complexity $s$, and assume that $s \le p$. The true complexity of $\mathcal{L}$ is the minimal $d$ such that $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent over $\mathbb{F}_p$.*

A natural generalization is to allow for multiple sets. Let $A_1, \dots, A_m \subseteq \mathbb{F}_p^n$ be sets of relative measures $\alpha_1, \dots, \alpha_m$. Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms over $\mathbb{F}_p$. We say $A_1, \dots, A_m$ are pseudorandom with respect to $L_1, \dots, L_m$ if

$$\Pr_{\mathbf{X} \in (\mathbb{F}_p^n)^k}[L_1(\mathbf{X}) \in A_1, \dots, L_m(\mathbf{X}) \in A_m] \approx \alpha_1 \cdot \dots \cdot \alpha_m.$$

Analogously to the case of a single set, let $f_i(x) = 1_{A_i}(x) - \alpha_i$. Then a sufficient condition is that for all nonempty subsets $S \subseteq [m]$, we have

$$\mathbb{E}\left[ \prod_{i \in S} f_i(L_i(\mathbf{X})) \right] \approx 0.$$

In [7], Gowers and Wolf showed that if $\mathcal{L}$ has true complexity $d$ and $\|f_1\|_{U^{d+1}}, \dots, \|f_m\|_{U^{d+1}}$ are small enough then this stronger condition also holds.

**Theorem A.6** (Theorem 7.2 in [7])**.** *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms over $\mathbb{F}_p^n$ of Cauchy-Schwarz complexity $s$, and assume that $s \le p$. Assume that $\mathcal{L}$ has true complexity $d$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $f_1, \dots, f_m : \mathbb{F}_p^n \to \mathbb{D}$ be functions such that $\|f_i\|_{U^{d+1}} \le \delta$, for all $1 \le i \le m$. Then for all nonempty subsets $S \subseteq [m]$ we have*

$$\left| \mathbb{E}_{\mathbf{X} \in (\mathbb{F}_p^n)^k} \left[ \prod_{i \in S} f_i(L_i(\mathbf{X})) \right] \right| \le \varepsilon.$$

In particular, Gowers and Wolf used this to derive the following corollary.

**Corollary A.7** ([7])**.** *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms of true complexity $d$ and Cauchy-Schwarz complexity at most $p$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $f_i : \mathbb{F}_p^n \to \mathbb{D}$ for $1 \le i \le m$ be functions such that $\|f_i - \mathbb{E}[f_i]\|_{U^{d+1}} \le \delta$. Then*

$$\left| \mathbb{E}_{\mathbf{X}} \left[ \prod_{i=1}^{m} f_i(L_i(\mathbf{X})) \right] - \prod_{i=1}^{m} \mathbb{E}[f_i] \right| \le \varepsilon.$$

In this paper we are able to extend the results of Gowers and Wolf to arbitrary approximating functions, and not just constant functions. We prove the following result, which qualitatively improves both Lemma A.2 and Corollary A.7.

**Corollary A.8.** *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms of true complexity $d$ and Cauchy-Schwarz complexity at most $p$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $f_i, g_i : \mathbb{F}_p^n \to \mathbb{D}$ for $1 \le i \le m$ be functions such that $\|f_i - g_i\|_{U^{d+1}} \le \delta$. Then*

$$\left| \mathbb{E}_{\mathbf{X}} \left[ \prod_{i=1}^{m} f_i(L_i(\mathbf{X})) \right] - \mathbb{E}_{\mathbf{X}} \left[ \prod_{i=1}^{m} g_i(L_i(\mathbf{X})) \right] \right| \le \varepsilon,$$

*where $\mathbf{X} \in (\mathbb{F}_p^n)^k$ is uniform.*

In fact, we prove a stronger results, from which Corollary A.8 follows immediately. It suffices to have $L_1^{d+1}$ linearly independent of $L_2^{d+1}, \ldots, L_m^{d+1}$ and $\|f_1\|_{U^{d+1}}$ small enough in order to bound averages of the form $\mathbb{E}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right]$. A weaker version of this was posed as an open problem by Gowers and Wolf [7][1].

**Theorem C.6 (restated).** *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms of Cauchy-Schwarz complexity at most $p$. Let $d \ge 0$, and assume that $L_1^{d+1}$ is not in the linear span of $L_2^{d+1}, \ldots, L_m^{d+1}$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that for any functions $f_1, \ldots, f_m : \mathbb{F}_p \to \mathbb{D}$ with $\|f_1\|_{U^{d+1}} \le \delta$, we have*

$$\left| \mathbb{E}_{\mathbf{X} \in (\mathbb{F}_p^n)^k} \left[ \prod_{i=1}^{m} f_i(L_i(\mathbf{X})) \right] \right| \le \varepsilon,$$

*where $\mathbf{X} \in (\mathbb{F}_p^n)^k$ is uniform.*

## A.2 Equivalence of systems of linear forms

Let $S$ be a subset of $\mathbb{F}_p^k$. Here we think of $S$ as a linear structure, and we are interested in the number of affine copies of $S$ in a set $A \subseteq \mathbb{F}_p^n$. Pick a linear transformation $T \in \mathrm{Lin}(\mathbb{F}_p^k, \mathbb{F}_p^n)$ uniformly at random, and also independently and uniformly a random element $X \in \mathbb{F}_p^n$. Then the density of the structure $S$ in $A$ is the probability that $X + T(x) \in A$ for every $x \in S$. Note that a uniform random $T \in \mathrm{Lin}(\mathbb{F}_p^k, \mathbb{F}_p^n)$ can be defined by mapping each standard vector $e_i \in \mathbb{F}_p^k$ uniformly and independently to a point in $\mathbb{F}_p^n$. Hence if the elements of a linear structure $S$ are $(\lambda_{i,1}, \ldots, \lambda_{i,k}) \in \mathbb{F}_p^k$ where $1 \le i \le m$, then the density of the structure $S$ in $A$ is the probability that for every $1 \le i \le m$, $X + \sum_{j=1}^k \lambda_{i,j} Y_j \in A$, where $X, Y_1, \ldots, Y_m$ are i.i.d. random variables taking values uniformly in $\mathbb{F}_p^n$. So if we define the system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\}$ by letting

$$L_i = (1, \lambda_{i,1}, \ldots, \lambda_{i,k}), \tag{14}$$

for $1 \le i \le m$, then $t_{\mathcal{L}}(1_A)$ gives the density of the structure $S$ in a set $A \subseteq \mathbb{F}_p^n$. Note that since for a fixed $c \in \mathbb{F}_p^n$, a uniform random variable $X$ has the same distribution as $X + c$, for this system of linear forms $\mathcal{L}$ the distribution of $(L_1(\mathbf{X}), \ldots, L_m(\mathbf{X}))$ is the same as the distribution of $(L_1(\mathbf{X}) + c, \ldots, L_m(\mathbf{X}) + c)$.

**Definition A.9** (Homogeneous linear forms). *A system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\}$ in $k$ variables is called* homogeneous*, if for a uniform random variable $\mathbf{X} \in (\mathbb{F}_p^n)^k$, and every fixed $c \in \mathbb{F}_p^n$, $(L_1(\mathbf{X}), \ldots, L_m(\mathbf{X}))$ has the same distribution as $(L_1(\mathbf{X}) + c, \ldots, L_m(\mathbf{X}) + c)$.*

[1] Gowers and Wolf required that $L_1^{d+1}$ is linearly independent of $L_2^{d+1}, \ldots, L_m^{d+1}$, and that all $\|f_1\|_{U^{d+1}}, \ldots, \|f_m\|_{U^{d+1}}$ will be bounded by $\delta$.

We wish to identify two systems of linear forms $\mathcal{L}_0 = \{L_1, \ldots, L_m\}$ in $k_0$ variables, and $\mathcal{L}_1 = \{L'_1, \ldots, L'_m\}$ in $k_1$ variables, if $(L_1(\mathbf{X}), \ldots, L_m(\mathbf{X}))$ has the same distribution as $(L'_1(\mathbf{Y}), \ldots, L'_m(\mathbf{Y}))$ where $\mathbf{X}$ and $\mathbf{Y}$ are uniform random variables taking values in $(\mathbb{F}_p^n)^{k_0}$ and $(\mathbb{F}_p^n)^{k_1}$, respectively. The distribution of $(L_1(\mathbf{X}), \ldots, L_m(\mathbf{X}))$ depends exactly on the linear dependencies between $L_1, \ldots, L_m$, and two systems of linear forms lead to the same distributions if and only if they have the same linear dependencies.

**Definition A.10** (Isomorphic linear forms)**.** *Two systems of linear forms $\mathcal{L}_0$ and $\mathcal{L}_1$ are* isomorphic *if and only if there exists a* bijection *from $\mathcal{L}_0$ to $\mathcal{L}_1$ that can be extended to an* invertible *linear transformation $T : \mathrm{span}(\mathcal{L}_0) \to \mathrm{span}(\mathcal{L}_1)$.*

Note that if $\mathcal{L} = \{L_1, \ldots, L_m\}$ is a homogenous system of linear forms, then $(L_1(\mathbf{X}), \ldots, L_m(\mathbf{X}))$ has the same distribution as $(L_1(\mathbf{X}) + Y, \ldots, L_m(\mathbf{X}) + Y)$, where $Y$ is a uniform random variable taking values in $\mathbb{F}_p^n$ and is independent of $\mathbf{X}$. We conclude with the following trivial observation.

**Observation A.11.** *Every homogenous system of linear forms is isomorphic to a system of linear forms in which there is a variable that appears with coefficient exactly one in every linear form.*

Consider a system of linear forms $\mathcal{L}$ in $\mathbb{F}_p^k$, and note that for every $f : \mathbb{F}_p^n \to \mathbb{C}$,

$$t_{\mathcal{L}}(f) = \mathbb{E}\left[\prod_{L \in \mathcal{L}} f(T(L))\right], \tag{15}$$

where $T$ is a random variable taking values uniformly in $\mathrm{Lin}(\mathrm{span}(\mathcal{L}), \mathbb{F}_p^n)$. Suppose that there exists a non-trivial subset $S \subseteq \mathcal{L}$ such that

$$\mathrm{span}(S) \cap \mathrm{span}(\mathcal{L} \setminus S) = \{\vec{0}\}.$$

Then for every $f : \mathbb{F}_p^n \to \mathbb{C}$ we have

$$t_{\mathcal{L}}(f) = t_S(f) t_{\mathcal{L} \setminus S}(f).$$

This leads to the following definition.

**Definition A.12** (Connected linear forms)**.** *A system of linear forms $\mathcal{L}$ is called* connected*, if for every non-trivial subset $S \subsetneq \mathcal{L}$, we have*

$$\mathrm{span}(S) \cap \mathrm{span}(\mathcal{L} \setminus S) \neq \{\vec{0}\}.$$

# B Higher-sorder Fourier analysis

Although Fourier analysis is a powerful tool in arithmetic combinatorics, there are key questions that cannot be addressed by this method in its classical form. For example in 1953 Roth [14] used Fourier analysis to show that every dense subset of integers contains 3-term arithmetic progressions. For more than four decades generalizing Roth's Fourier-analytic proof remained an important unsolved problem until finally Gowers in [6] introduced an extension of the classical Fourier analysis which enabled him to obtain such a generalization. The work of Gowers initiated a theory which has now come to be known as higher-order Fourier analysis. Ever since several mathematicians contributed to major developments in this rapidly growing theory.

This section has two purposes. One is to review the main results that form the foundations of the higher-order Fourier analysis. A second is to establish some new facts that enable us to deal with the averages $t_{\mathcal{L}}$ conveniently by appealing to higher-order Fourier analysis. The work of Gowers and Wolf [8] plays a central role for us, and many ideas in the proofs and the new facts established in this section are hinted by their work.

The characters of $\mathbb{F}_p^n$ are exponentials of linear polynomials; that is for $\alpha \in \mathbb{F}_p^n$, the corresponding character is defined as $\chi_\alpha(x) = \mathrm{e}_p(\sum_{i=1}^n \alpha_i x_i)$. In higher-order Fourier analysis, the linear polynomials $\sum \alpha_i x_i$ are replaced by higher degree polynomials, and one would like to express a function $f : \mathbb{F}_p^n \to \mathbb{C}$ as a linear combination of the functions $\mathrm{e}_p(P)$, where $P$ is a polynomial of a certain degree.

Consider a function $f : \mathbb{F}_p^n \to \mathbb{C}$, and a system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\}$. The basic properties of characters enable us to express $t_{\mathcal{L}}(f)$ as a simple formula in terms of the Fourier coefficients of $f$. Indeed if $f := \sum_{\alpha \in \mathbb{F}_p^n} \widehat{f}(\alpha)\chi_\alpha$ is the Fourier expansion of $f$, then it is easy to see that

$$t_{\mathcal{L}}(f) = \sum \widehat{f}(\alpha_1) \ldots \widehat{f}(\alpha_m), \tag{16}$$

where the sum is over all $\alpha_1, \ldots, \alpha_m \in \mathbb{F}_p^n$ satisfying $\sum_{i=1}^m \alpha_i \otimes L_i \equiv 0$. The tools that we develop in this section enables us to obtain simple formulas similar to (16) when Fourier expansion is replaced by a proper higher-order Fourier expansion.

## B.1 Inverse theorems for Gowers uniformity norms

We start with some basic definitions.

*Polynomials:* Consider a function $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$. For an element $y \in \mathbb{F}_p^n$, define the *derivative* of f in the direction $y$ as $\Delta_y \mathrm{f}(x) = \mathrm{f}(x+y) - \mathrm{f}(x)$. Inductively we define $\Delta_{y_1,\ldots,y_k}\mathrm{f} = \Delta_{y_k}(\Delta_{y_1,\ldots,y_{k-1}}\mathrm{f})$, for directions $y_1, \ldots, y_k \in \mathbb{F}_p^n$. We say that f is a *polynomial of degree at most* $d$, if for every $y_1, \ldots, y_{d+1} \in \mathbb{F}_p$, we have $\Delta_{y_1,\ldots,y_{d+1}}\mathrm{f} \equiv 0$. The set of polynomials of degree at most $d$ is a vector space over $\mathbb{F}_p$ which we denote by $\mathrm{Poly}_d(\mathbb{F}_p^n)$. It is easy to see that the set of *monomials* $x_1^{i_1} \ldots x_n^{i_n}$ where $0 \leq i_1, \ldots, i_n < p$ and $\sum_{j=1}^n i_j \leq d$ form a basis for $\mathrm{Poly}_d(\mathbb{F}_p^n)$. So every polynomial $P \in \mathrm{Poly}_d(\mathbb{F}_p^n)$ is of the from $P(x) := \sum c_{i_1,\ldots,i_n} x_1^{i_1} \ldots x_n^{i_n}$, where the sum is over all $1 \leq i_1, \ldots, i_n < p$ with $\sum_{j=1}^n i_j \leq d$, and $c_{i_1,\ldots,i_n}$ are elements of $\mathbb{F}_p$. The *degree* of a polynomial $P : \mathbb{F}_p^n \to \mathbb{F}_p$, denoted by $\deg(P)$, is the smallest $d$ such that $P \in \mathrm{Poly}_d(\mathbb{F}_p^n)$. A polynomial $P$ is called *homogenous*, if all monomials with non-zero coefficients in the expansion of $P$ are of degree exactly $\deg(P)$.

*Phase Polynomials:* For a function $f : \mathbb{F}_p^n \to \mathbb{C}$, and a direction $y \in \mathbb{F}_p^n$ define the *multiplicative derivative* of $f$ in the direction of $y$ as $\tilde{\Delta}_y f(x) = f(x+y)\overline{f(x)}$. Inductively we define $\tilde{\Delta}_{y_1,\ldots,y_k} f = \tilde{\Delta}_{y_k}(\tilde{\Delta}_{y_1,\ldots,y_{k-1}} f)$, for directions $y_1, \ldots, y_k \in \mathbb{F}_p^n$. A function $f : \mathbb{F}_p^n \to \mathbb{C}$ is called a *phase polynomial of degree at most* $d$, if for every $y_1, \ldots, y_{d+1} \in \mathbb{F}_p$, we have $\tilde{\Delta}_{y_1,\ldots,y_{d+1}} f \equiv 1$. We denote the space of all phase polynomials of degree at most $d$ over $\mathbb{F}_p^n$ by $\mathcal{P}_d(\mathbb{F}_p^n)$. Note that for every $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$, and every $y \in \mathbb{F}_p^n$, we have that

$$\tilde{\Delta}_y \mathrm{e}_p(\mathrm{f}) = \mathrm{e}_p(\Delta_y \mathrm{f}).$$

This shows that if $\mathrm{f} \in \mathrm{Poly}_d(\mathbb{F}_p^n)$, then $\mathrm{e}_p(\mathrm{f})$ is a phase polynomial of degree at most $d$. The following simple lemma shows that the inverse is essentially true in high characteristic case:

**Lemma B.1** (Lemma 1.2 in [18]). *Suppose that $0 \leq d < p$. Every $f \in \mathcal{P}_d(\mathbb{F}_p^n)$ is of the form $f(x) = \mathrm{e}_p(\theta + \mathrm{f}(x))$, for some $\theta \in \mathbb{R}/\mathbb{Z}$, and $\mathrm{f} \in \mathrm{Poly}_d(\mathbb{F}_p^n)$.*

When $d \geq p$, more complicated phase polynomials arise. Nevertheless obtaining a complete characterization is possible [17].

Now let us describe the relation between the phase polynomials and the Gowers norms. First note that one can express Gowers uniformity norms using multiplicative derivatives:

$$\|f\|_{U^k}^{2^k} = \mathbb{E}\left[\tilde{\Delta}_{Y_1,\dots,Y_k} f(X)\right],$$

where $X, Y_1, \dots, Y_k$ are independent random variables taking values in $\mathbb{F}_p^n$ uniformly. This for example shows that every phase polynomial $g$ of degree at most $d$ satisfies $\|g\|_{U^{d+1}} = 1$.

Many basic properties of Gowers uniformity norms are implied by the Gowers-Cauchy-Schwarz inequality which is first proved in [6] by iterated applications of the classical Cauchy-Schwarz inequality.

**Lemma B.2** (Gowers-Cauchy-Schwarz). *Let $G$ be a finite Abelian group, and consider a family of functions $f_S : G \to \mathbb{C}$, where $S \subseteq [k]$. Then*

$$\left| \mathbb{E}\left[\prod_{S\subseteq[k]} \mathcal{C}^{k-|S|} f_S(X + \sum_{i\in S} Y_i)\right] \right| \leq \prod_{S\subseteq[k]} \|f_S\|_{U^k}, \tag{17}$$

*where $X, Y_1, \dots, Y_k$ are independent random variables taking values in $G$ uniformly at random.*

A simple application of Lemma B.2 is the following. Consider an arbitrary function $f : G \to \mathbb{C}$. Setting $f_\emptyset := f$ and $f_S := 1$ for every $S \neq \emptyset$ in Lemma B.2, we obtain

$$|\mathbb{E}[f(X)]| \leq \|f\|_{U^k}. \tag{18}$$

Equation (18) in particular shows that if $f, g : \mathbb{F}_p^n \to \mathbb{C}$, then one can bound their inner product with Gowers uniformity norms of $f\overline{g}$:

$$|\langle f, g\rangle| \leq \|f\overline{g}\|_{U^k}. \tag{19}$$

Consider an arbitrary $f : \mathbb{F}_p^n \to \mathbb{C}$ and a phase polynomial $g$ of degree at most $d$. Then for every $y_1, \dots, y_{d+1} \in \mathbb{F}_p^n$, we have

$$\tilde{\Delta}_{y_1,\dots,y_{d+1}}(fg) = (\tilde{\Delta}_{y_1,\dots,y_{d+1}} f)(\tilde{\Delta}_{y_1,\dots,y_{d+1}} g) = \tilde{\Delta}_{y_1,\dots,y_{d+1}} f,$$

which in turn implies that $\|fg\|_{U^{d+1}} = \|f\|_{U^{d+1}}$. Combining this with (19), we conclude that the correlation of $f$ with any phase polynomial of degree at most $d$ is a lower-bound for $\|f\|_{U^{d+1}}$:

$$\|f\|_{u(\mathcal{P}_d)} \leq \|f\|_{U^{d+1}}. \tag{20}$$

This provides us with a "direct theorem" for the $U^{d+1}$ norm: If $\|f\|_{u(\mathcal{P}_d)} \geq \delta$, then $\|f\|_{U^{d+1}} \geq \delta$. Recently Bergelson, Tao, and Ziegler [1, 18] established the corresponding inverse theorem in the high characteristic case.

**Theorem B.3** ([1, 18]). *If $1 \leq d < p$, then there exists a function $\delta : (0,1] \to (0,1]$ such that for every $f : \mathbb{F}_p^n \to \mathbb{D}$, and $\varepsilon > 0$,*

- Direct theorem: *If $\|f\|_{u(\mathcal{P}_d)} \geq \varepsilon$, then $\|f\|_{U^{d+1}} \geq \varepsilon$.*

- Inverse theorem: *If* $\|f\|_{U^{d+1}} \geq \varepsilon$*, then* $\|f\|_{u(\mathcal{P}_d)} \geq \delta(\varepsilon)$.

Theorem B.3 requires that $1 \leq d < p$. In this range, Lemma B.1 shows that the phase polynomials of degree at most $d$ can be described using polynomials of degree at most $d$. So Theorem B.3 shows that in this case if $\|f\|_{U^{d+1}} \geq \varepsilon$, then there exists a polynomial $\mathrm{g} : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree at most $d$ such that $|\langle f, \mathrm{e}_p(\mathrm{g})\rangle| \geq \delta(\varepsilon) > 0$. Tao and Ziegler [18] conjectured that Theorem B.3 remains valid if one removes the condition $1 \leq d < p$. They proved the following partial result in this direction.

**Theorem B.4** ([1, 18])**.** *For every integer* $d \geq 1$*, there exists a function* $\delta : (0,1] \to (0,1]$ *and a positive integer* $c(d)$ *such that for every* $f : \mathbb{F}_p^n \to \mathbb{D}$*, and* $\varepsilon > 0$*,*

- Direct theorem: *If* $\|f\|_{u(\mathcal{P}_d)} \geq \varepsilon$*, then* $\|f\|_{U^{d+1}} \geq \varepsilon$*.*
- Inverse theorem: *If* $\|f\|_{U^{d+1}} \geq \varepsilon$*, then* $\|f\|_{u(\mathcal{P}_{c(d)})} \geq \delta(\varepsilon)$*.*

## B.2 Decomposition theorems

An important application of the inverse theorems is that they imply "decomposition theorems". Roughly speaking these results say that under appropriate conditions, a function $f$ can be decomposed as $f_1 + f_2$, where $f_1$ is "structured" in some sense that enables one to handle it easily, while $f_2$ is "quasi-random" meaning that it shares certain properties with a random function, and can be discarded as random noise. In the following we will discuss decomposition theorems that follow from Theorem B.3, but first we need to define the polynomial factors on $\mathbb{F}_p^n$.

**Definition B.5** (Polynomial factors)**.** *Let* $p$ *be a fixed prime. Let* $P_1, \ldots, P_C \in \mathrm{Poly}_d(\mathbb{F}_p^n)$*. The sigma-algebra on* $\mathbb{F}_p^n$ *whose atoms are* $\{x \in \mathbb{F}_p^n : P_1(x) = a(1), \ldots, P_C(x) = a(C)\}$ *for all* $a \in \mathbb{F}_p^C$ *is called a* polynomial factor of degree at most $d$ and complexity at most $C$*.*

Let $\mathcal{B}$ be a polynomial factor defined by $P_1, \ldots, P_C$. For $f : \mathbb{F}_p^n \to \mathbb{C}$, the conditional expectation of $f$ with respect to $\mathcal{B}$, denoted $\mathbb{E}(f|\mathcal{B}) : \mathbb{F}_p^n \to \mathbb{C}$, is

$$\mathbb{E}(f|\mathcal{B})(x) = \mathbb{E}_{\{y \in \mathbb{F}_p^n : P_1(y) = P_1(x), \ldots, P_C(y) = P_C(x)\}}[f(y)].$$

That is, $\mathbb{E}(f|\mathcal{B})$ is constant on every atom of $\mathcal{B}$, and this constant is the average value that $f$ attains on this atom. A function $g : \mathbb{F}_p^n \to \mathbb{C}$ is $\mathcal{B}$-measurable, if it is constant on every atom of $\mathcal{B}$. Equivalently, we can write $g$ as $g(x) = \Gamma(P_1(x), \ldots, P_C(x))$ for some function $\Gamma : \mathbb{F}_p^C \to \mathbb{C}$. The following claim is quite useful, although its proof is immediate and holds for every sigma-algebra.

**Observation B.6.** *Let* $f : \mathbb{F}_p^n \to \mathbb{C}$*. Let* $\mathcal{B}$ *be a polynomial factor defined by polynomials* $P_1, \ldots, P_C$*. Let* $g : \mathbb{F}_p^n \to \mathbb{C}$ *be any* $\mathcal{B}$*-measurable function. Then*

$$\langle f, g\rangle = \langle \mathbb{E}(f|\mathcal{B}), g\rangle.$$

The following theorem that follows in a standard manner from Theorem B.3 gives a simple decomposition theorem.

**Theorem B.7** (Decomposition Theorem [9])**.** *Let* $p$ *be a fixed prime,* $0 \leq d < p$ *be an integer, let* $\delta > 0$*, and suppose that* $n > n_0(\delta)$ *is sufficiently large. Given any function* $f : \mathbb{F}_p^n \to \mathbb{D}$*, there* *exists a polynomial factor $\mathcal{B}$ of degree at most $d$ and complexity at most $C_{\max}(p,d,\delta)$ together with a decomposition*

$$f = f_1 + f_2,$$

*where*

$$f_1 := \mathbb{E}(f|\mathcal{B}) \text{ and } \|f_2\|_{U^{d+1}} \le \delta.$$

We sketch the standard proof of Theorem B.7 below, as we will need some extensions of it in this paper. For a full proof we refer the reader to [9].

*Proof sketch.* We create a sequence of polynomial factors $\mathcal{B}_1, \mathcal{B}_2, \dots$ as follows. Let $\mathcal{B}_1$ be the trivial factor (i.e. $\mathbb{E}(f|\mathcal{B}_1)$ is the constant function $\mathbb{E}[f]$). Let $g_i = f - \mathbb{E}(f|\mathcal{B}_i)$. If $\|g_i\|_{U^{d+1}} \le \delta$ we are done. Otherwise by Theorem B.3, since $\|g_i\|_\infty \le 2$, there exists a polynomial $P_i \in \mathrm{Poly}_d(\mathbb{F}_p^n)$ such that $\langle g_i, \mathrm{e}_p(P_i)\rangle \ge c(\delta)$. Let $\mathcal{B}_{i+1} = \mathcal{B}_i \cup \{P_i\}$. The key point is that one can show

$$\|g_{i+1}\|_2^2 \le \|g_i - \langle g_i, \mathrm{e}_p(P_i)\rangle\|_2^2 \le \|g_i\|_2^2 - c(\delta).$$

Thus, the process must stop after at most $1/c(\delta)$ steps. □

Suppose that the factor $\mathcal{B}$ is defined by $P_1, \dots, P_C \in \mathrm{Poly}_d(\mathbb{F}_p^n)$. Assume that $f_1(x) = \Gamma(P_1(x), \dots, P_C(x))$. Using the Fourier decomposition of $\Gamma$, we can express $f_1$ as

$$f_1(x) = \sum_{\gamma \in \mathbb{F}_p^C} \widehat{\Gamma}(\gamma)\mathrm{e}_p\left(\sum_{i=1}^C \gamma(i)P_i(x)\right). \tag{21}$$

Note that for every $\gamma \in \mathbb{F}_p^C$, $\sum_{i=1}^C \gamma(i)P_i(x) \in \mathrm{Poly}_d(\mathbb{F}_p^n)$. So (21) gives an expansion for $f_1$ which is similar to the Fourier expansion, but instead of characters $\mathrm{e}_p(\sum \alpha(i)x_i)$, we have exponential functions $\mathrm{e}_p\left(\sum_{i=1}^C \gamma(i)P_i(x)\right)$ which have polynomials of degree $d$ in the powers instead of linear functions. For this decomposition to be useful similar to the Fourier expansion, one needs some kind of orthogonality for the functions appearing in the expansion.

**Definition B.8** (Bias)**.** *The bias of a polynomial $P \in \mathrm{Poly}_d(\mathbb{F}_p^n)$ is defined as*

$$\mathrm{bias}(P) := \mathrm{bias}(\mathrm{e}_p(P)) = |\mathbb{E}_{X \in \mathbb{F}_p^n}[\mathrm{e}_p(P(X))]|.$$

We shall refine the set of polynomials $\{P_1, \dots, P_t\}$ to obtain a new set of polynomials with the desired “approximate orthogonality” properties. This will be achieved through the notion of the *rank* of a set of polynomials.

**Definition B.9** (Rank)**.** *We say a set of polynomials $\mathcal{P} = \{P_1, \dots, P_t\}$ is of rank greater than $r$, and denote this by* $\mathrm{rank}(\mathcal{P}) > r$*, if the following holds. For any non-zero $\alpha = (\alpha_1, \dots, \alpha_t) \in \mathbb{F}_p^t$, define $P_\alpha(x) := \sum_{j=1}^t \alpha_j P_j(x)$. For $d := \max\{\deg(P_j) : \alpha_j \neq 0\}$, the polynomial $P_\alpha$ cannot be expressed as a function of $r$ polynomials of degree at most $d-1$. More precisely, it is not possible to find $r$ polynomials $Q_1, \dots, Q_r$ of degree at most $d-1$, and a function $\Gamma : \mathbb{F}_p^r \to \mathbb{F}_p$ such that*

$$P(x) = \Gamma(Q_1(x), \dots, Q_r(x)).$$

*The rank of a single polynomial $P$ is defined to be* $\mathrm{rank}(\{P\})$.

The *rank* of a polynomial factor is the rank of the set of polynomials defining it. The following lemma follows from the definition of the rank. For a proof see [10].

**Lemma B.10** (Making factors high-rank). *Let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary growth function. Then there is another function $\tau : \mathbb{N} \to \mathbb{N}$ with the following property. Let $\mathcal{B}$ be a polynomial factor with complexity at most $C$. Then there is a refinement $\mathcal{B}'$ of $\mathcal{B}$ with complexity at most $C' \leq \tau(C)$ and rank at least $r(C')$.*

The following theorem due to Kaufman and Lovett [12] connects the notion of the rank to the bias of a polynomial. It was proved first by Green and Tao [10] for the case $d < p$, and then extended by Kaufman and Lovett [12] for the general case.

**Theorem B.11** (Regularity [12]). *Fix $p$ prime and $d \geq 1$. There exists a function $r_{p,d} : (0, 1] \to \mathbb{N}$ such that the following holds. If $P : \mathbb{F}_p^n \to \mathbb{F}_p$ is a polynomial of degree at most $d$ with $\mathrm{bias}(P) \geq \varepsilon$, then $\mathrm{rank}(P) \leq r_{p,d}(\varepsilon)$.*

Combining Lemma B.10 with Theorem B.7, it is possible to obtain a strong decomposition theorem.

**Theorem B.12** (Strong Decomposition Theorem [9]). *Let $p$ be a fixed prime, $0 \leq d < p$ be an integer, $\delta > 0$, and let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary growth functions, and suppose that $n > n_0(d, \delta, r(\cdot))$ is sufficiently large. Given any function $f : \mathbb{F}_p^n \to \mathbb{D}$, there exists a decomposition*

$$f = f_1 + f_2,$$

*such that*

$$f_1 := \mathbb{E}(f|\mathcal{B}), \qquad \|f_2\|_{U^{d+1}} \leq \delta,$$

*where $\mathcal{B}$ is a polynomial factor of degree at most $d$, complexity $C \leq C_{\max}$ (where $C_{\max}$ depends on $p, d, \delta, r(\cdot)$), and rank at least $r(C)$.*

We sketch the proof below. For a full proof we refer the reader to [9].

*Proof sketch.* The proof follows the same steps as the proof of Theorem B.7, except that at each step, we regularize each polynomial factor $\mathcal{B}_i$ to obtain $\mathcal{B}'_i$, and set $\mathcal{B}_{i+1} = \mathcal{B}'_i \cup \{P_i\}$. The only new insight is that as $\mathcal{B}'_i$ is a refinement of $\mathcal{B}_i$ we have

$$\|f - \mathbb{E}(f|\mathcal{B}'_i)\|_2 \leq \|f - \mathbb{E}(f|\mathcal{B}_i)\|_2.$$

□

Note that Theorem B.12 guarantees a strong approximate orthogonality. For every fixed function $\omega : \mathbb{N} \to \mathbb{N}$, by taking $r(\cdot)$ to be a sufficiently fast growing function, one can guarantee that the polynomials $P_1, \ldots, P_C$ that define the factor $\mathcal{B}$ have the property

$$\left| \mathbb{E}\left[ \mathrm{e}_p\left( \sum_{i=1}^{C} \gamma(i) P_i(X) \right) \right] \right| \leq 1/\omega(C), \tag{22}$$

for all nonzero $\gamma \in \mathbb{F}_p^C$. That is, the polynomials can be made "nearly orthogonal" to any required precision. The following lemma demonstrates the power of this technique. It shows that by choosing $r(\cdot)$ large enough, we can make sure that the distribution of the polynomials is almost independent.

**Lemma B.13.** *There exists a function $r : \mathbb{N} \to \mathbb{N}$ such that the following holds. Let $P_1, \ldots, P_C \in$ $Poly_d(\mathbb{F}_p^n)$ be polynomials of rank at least $r(C)$. Then for every $a \in \mathbb{F}_p^C$, we have*

$$\Pr_{X \in \mathbb{F}_p^n}[P_1(X) = a(1), \ldots, P_C(X) = a(C)] \leq 2p^{-C}.$$

*Proof.* Using the notations of (22), set $\omega(C) = p^{2C}$ . We can write the probability that $P_1(X) = a(1), \ldots, P_C(X) = a(C)$ as

$$\Pr_{X \in \mathbb{F}_p^n}[P_1(X) = a(1), \ldots, P_C(X) = a(C)] = \frac{1}{p^C} \sum_{\gamma \in \mathbb{F}_p^C} \mathbb{E}_{X \in \mathbb{F}_p^n} \left[ \mathrm{e}_p \left( \sum_{i=1}^{C} \gamma(i)(P_i(X) - a(i)) \right) \right].$$

The contribution of $\gamma = 0$ to the sum is exactly $p^{-C}$, and by the choice of $\omega(\cdot)$, the contribution of any other terms $\gamma \neq 0$ is at most $p^{-2C}$, and the lemma follows. □

The decomposition theorems stated to far referred to a single function. In this paper we require decomposition theorems which relate to several functions with a single polynomial factor. The proofs can be adapted in a straight-forward manner to prove the next result.

**Lemma B.14** (Strong Decomposition Theorem - multiple functions)**.** *Let $p$ be a fixed prime, $0 \leq d < p$ and $m$ be integers, let $\delta > 0$, and let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary growth functions, and suppose that $n > n_0(p, d, \delta, m, r(\cdot))$ is sufficiently large. Given every set of functions $f_1, \ldots, f_m : \mathbb{F}_p^n \to \mathbb{D}$, there exists a decomposition of each $f_i$ as*

$$f_i = f_{i,1} + f_{i,2},$$

*such that*

$$f_{i,1} := \mathbb{E}(f_i|\mathcal{B}), \qquad \|f_{i,2}\|_{U^{d+1}} \leq \delta,$$

*where $\mathcal{B}$ is a polynomial factor of degree at most $d$, complexity $C \leq C_{\max}$ (where $C_{\max}$ depends on $p, d, \delta, m, r(\cdot)$) and rank at least $r(C)$. Furthermore we can assume that $\mathcal{B}$ is defined by homogeneous polynomials.*

# C Strong orthogonality

Let $\mathcal{B}$ be a polynomial factor defined by polynomials $P_1, \ldots, P_C$, and let $f : \mathbb{F}_p^n \to \mathbb{D}$ be a $\mathcal{B}$-measurable function. We saw in (21) that $f$ can be expressed as a linear combination of $\mathrm{e}_p(\sum_{i=1}^{C} \gamma(i) P_i(x))$, for $\gamma \in \mathbb{F}_p^C$. Furthermore if we require the polynomials to be of high rank, then we obtain approximate orthogonality as in (22). This approximate orthogonality is sufficient for analyzing averages such as $\mathbb{E}[f_1(X) f_2(X) \ldots f_m(X)]$, where $f_1, \ldots, f_m$ (not necessarily distinct) are all measurable with respect to $\mathcal{B}$. However it is not a priori clear how this orthogonality can be used to deal with averages of the form $\mathbb{E}[f_1(L_1(\mathbf{X})) \ldots f_m(L_m(\mathbf{X}))]$, for linear forms $L_1, \ldots, L_m$. The difficulty arises when one has to understand exponential averages such as $\mathbb{E}[\mathrm{e}_p(P(X+Y) - P(X) - P(Y))]$. This average is 1 when $P$ is a homogeneous polynomial of degree one, but it does not immediately follow from what we have said so far that it is small when $P$ is of higher degree. In this section we develop the results needed to deal with such exponential averages.

Consider a set of homogeneous polynomials $\{P_1, \dots, P_k\}$, and a set of linear forms $\{L_1, \dots, L_m\}$. We need to be able to analyze exponential averages of the form:

$$\mathbb{E}_{\mathbf{X}}\left[\mathrm{e}_p\left(\sum_{i=1}^{k}\sum_{j=1}^{m}\lambda_{i,j}P_i(L_j(\mathbf{X}))\right)\right],$$

where $\lambda_{i,j} \in \mathbb{F}_p$. Lemma C.1 below shows that if $\{P_1, \dots, P_k\}$ are of sufficiently high rank, then it is either the case that $\sum_{i=1}^{k}\sum_{j=1}^{m}\lambda_{i,j}P_i(L_j(\mathbf{x})) \equiv 0$ which implies that the corresponding exponential average is exactly 1, or otherwise the exponential average is very small. Note that this is an "approximate" version of the case of characters. Namely if $\{\chi_{y_1}, \dots, \chi_{y_k}\}$ are characters of $\mathbb{F}_p^n$, then $\mathbb{E}\left[\prod_{i=1}^{k}\prod_{j=1}^{m}\chi_{y_i}(L_j(\mathbf{X}))\right]$ is either 1 or 0. In the case of polynomials of high rank, the "zero" case is approximated by a small number.

**Lemma C.1.** *Let $p$ be a fixed prime. There exists a function $r : \mathbb{N} \times (0,1] \to \mathbb{N}$ such that the following holds. Let $\{L_1, \dots, L_m\}$ be a system of linear forms. Let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a collection of polynomials of degree at most $d < p$, such that $\operatorname{rank}(\mathcal{P}) > r(d, \varepsilon)$. For every set of coefficients $\{\lambda_{i,j} \in \mathbb{F}_p : i \in [k], j \in [m]\}$, and*

$$P_\Lambda(\mathbf{x}) := \sum_{i=1}^{k}\sum_{j=1}^{m}\lambda_{i,j}P_i(L_j(\mathbf{x})),$$

*one of the following two cases holds:*

$$P_\Lambda \equiv 0 \qquad \textit{or} \qquad \operatorname{bias}(P_\Lambda) < \varepsilon.$$

Lemma C.1 shows that in order to estimate $\operatorname{bias}(P_\Lambda)$ for polynomials of high rank $\mathcal{P} = \{P_1, \dots, P_k\}$, it suffices to determine whether $P_\Lambda$ is identically 0 or not. Our next observation which is one of the key tools in this article says that when the polynomials are homogeneous and linearly independent, then $P_\Lambda \equiv 0$ depends only on the set of the coefficients $\lambda_{i,j}$, the linear forms $L_j$, and the *degrees* of the polynomials involved in $P_1, \dots, P_k$, and not the particular choice of the polynomials.

**Lemma C.2.** *Let $\{L_1, \dots, L_m\}$ be a system of linear forms over $\mathbb{F}_p^n$, $\lambda_{i,j} \in \mathbb{F}_p$ for $i \in [k], j \in [m]$, and $d_1, \dots, d_k \in [d]$. One of the following two cases hold*

- *For every collection $\mathcal{P} = \{P_1, \dots, P_k\}$ of linearly independent homogeneous polynomials of degrees $d_1, \dots, d_k$:*

$$\sum_{i=1}^{k}\sum_{j=1}^{m}\lambda_{i,j}P_i(L_j(\mathbf{x})) \equiv 0$$

- *For every collection $\mathcal{P} = \{P_1, \dots, P_k\}$ of linearly independent homogeneous polynomials of degrees $d_1, \dots, d_k$:*

$$\sum_{i=1}^{k}\sum_{j=1}^{m}\lambda_{i,j}P_i(L_j(\mathbf{x})) \not\equiv 0$$

In this article we shall need to deal with non-homogenous polynomials. The following lemma shows that if the set of the linear forms is homogeneous, then the case of the non-homogenous polynomials reduces to the homogenous polynomials. Let $P : \mathbb{F}_p^n \to \mathbb{F}_p$ be a polynomial of degree at most $d$. For $1 \leq l \leq d$, let $P^{(l)}$ denote the homogenous polynomial that is obtained from $P$ by removing all the monomials whose degrees are *not* equal to $l$.

**Lemma C.3.** *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a set of* homogeneous *linear forms. Furthermore let $\mathcal{P} = \{P_1, \ldots, P_k\}$ be a collection of polynomials of degrees $d_1, \ldots, d_k$ such that $\{P_i^{(d_i)} : i \in [k]\}$ are linearly independent over $\mathbb{F}_p$. Then for $\lambda_{i,j} \in \mathbb{F}_p$,*

$$\textstyle\sum_{i=1}^{k} \sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(\mathbf{x})) \equiv 0 \Longleftrightarrow \sum_{i=1}^{k} \sum_{j=1}^{m} \lambda_{i,j} P_i^{(d_i)}(L_j(\mathbf{x})) \equiv 0$$

## C.1 Proofs of Lemmas C.1, C.2 and C.3

Let $P(x)$ be a homogeneous polynomial of degree $d < p$. Let $B(x_1, \ldots, x_d)$ be the symmetric multi-linear form associated with $P$, that is $P(x) = B(x, \ldots, x)$ and $B(x_1, \ldots, x_d) \equiv B(x_{\sigma_1}, \ldots, x_{\sigma_d})$, for every permutation $\sigma$ of $[d]$. If $L(\mathbf{x}) = \sum_{i=1}^{k} c_i x_i$ is a linear form in $k$ variables, then we have

$$P(L(\mathbf{x})) = \sum_{i_1, \ldots, i_d \in [k]} c_{i_1} \ldots c_{i_d} B(x_{i_1}, \ldots, x_{i_d}).$$

For $\mathbf{u} = (u_1, \ldots, u_d) \in [k]^d$, denote $\mathbf{x_u} = (x_{u_1}, \ldots, x_{u_d})$. Let $U^d \subseteq [k]^d$ be defined as $U^d = \{(u_1, \ldots, u_d) \in [k]^d : u_1 \leq u_2 \leq \ldots \leq u_d\}$. For $\mathbf{u} \in U^d$, denote by $r_d(\mathbf{u})$ the number of distinct permutations[2] of $(u_1, \ldots, u_d)$, and let $c_d(\mathbf{u}, L) := c_{u_1} \ldots c_{u_d}$. Since $B$ is symmetric, we have

$$P(L(\mathbf{x})) = \sum_{\mathbf{u} \in U^d} r_d(\mathbf{u}) c_d(\mathbf{u}, L) B(\mathbf{x_u}). \tag{23}$$

Note that $r_d(\mathbf{u})$ depends only on $\mathbf{u}$ and $c_d(\mathbf{u}, L)$ depends only on the linear form $L$ and $\mathbf{u} \in [k]^d$. We need the following claim.

**Claim C.4.** *Let $B$ be homogeneous multi-linear form over $\mathbb{F}_p$ of degree $d < p$. Consider a linear combination*

$$Q(\mathbf{x}) = \sum_{\mathbf{u} \in U^d} c_{\mathbf{u}} B(\mathbf{x_u})$$

*where not all the coefficients $c_{\mathbf{u}}$ are zero. Then there exist $a_1, \ldots, a_k \in \mathbb{F}_p$ and $\alpha \neq 0$ such that for every $w \in \mathbb{F}_p^n$*

$$Q(a_1 w, \ldots, a_k w) = \alpha B(w, \ldots, w).$$

*Proof.* Consider $\mathbf{x} = (a_1 w, \ldots, a_k w)$. As $B$ is multi-linear, we have

$$B(\mathbf{x_u}) = \left( \prod_{i=1}^{d} a_{u_i} \right) \cdot B(w, \ldots, w),$$

[2] If the multiplicities of the elements of a multi-set are $i_1, \ldots, i_\ell$, then the number of distinct permutations of those elements is $\frac{(i_1 + \ldots + i_\ell)!}{i_1! \ldots i_\ell!}$.

for every $\mathbf{u} \in U^d$. Let $a = (a_1, \ldots, a_k)$ and let $a^{\mathbf{u}}$ denote the monomial $a^{\mathbf{u}} = \prod_{i=1}^d a_{u_i}$. We thus have

$$Q(a_1 w, \ldots, a_k w) = \left( \sum_{\mathbf{u} \in U^d} c_{\mathbf{u}} a^{\mathbf{u}} \right) \cdot B(w, \ldots, w).$$

Consider $g(a_1, \ldots, a_k) = \sum_{\mathbf{u} \in U^d} c_{\mathbf{u}} a^{\mathbf{u}}$. This is a polynomial in $a_1, \ldots, a_k$ which is not identically zero, as distinct $\mathbf{u} \in U^d$ correspond to distinct monomials $a^{\mathbf{u}}$. Hence there exist some assignment for $a$ for which $\alpha := g(a) \neq 0$. □

Consider a set of linearly independent homogeneous polynomials $\{P_1, \ldots, P_k\}$, a system of linear forms $\{L_1, \ldots, L_m\}$ and some coefficients $\lambda_{i,j} \in \mathbb{F}_p$ where $i \in [m], j \in [k]$. Let $B_i$ be the symmetric multi-linear form associated with $P_i$. Denoting $d_i := \deg(P_i)$ and using the notation of (23), we thus have

$$\begin{aligned} P_\Lambda(\mathbf{x}) &= \sum_{i=1}^k \sum_{j=1}^m \lambda_{i,j} P_i(L_j(\mathbf{x})) = \sum_{i=1}^k \sum_{j=1}^m \lambda_{i,j} \sum_{\mathbf{u} \in U^{d_i}} r_{d_i}(\mathbf{u}) c_{d_i}(\mathbf{u}, L_j) B_i(\mathbf{x}_\mathbf{u}) \\ &= \sum_{i=1}^k \sum_{\mathbf{u} \in U^{d_i}} b_i^{d_i}(\mathbf{u}) B_i(\mathbf{x}_\mathbf{u}), \end{aligned} \tag{24}$$

where

$$b_i^{d_i}(\mathbf{u}) := r_{d_i}(\mathbf{u}) \sum_{j=1}^m \lambda_{i,j} c_{d_i}(\mathbf{u}, L_j).$$

Note that the coefficients $b_i^{d_i}(\mathbf{u})$ do not depend on the specific set of polynomials $P_1, \ldots, P_k$. Among Lemmas C.1, C.2 and C.3 first we prove Lemma C.2 which has the simplest proof.

*Proof of Lemma C.2.* We will show that $P_\Lambda(\mathbf{x}) \equiv 0$ if and only if $b_i^{d_i}(\mathbf{u}) = 0$, for all $i \in [k]$ and $\mathbf{u} \in U^{d_i}$. Let $B_{i_0}$ be a polynomial appearing in $P_\Lambda$ with a nonzero coefficient, that is $b_{i_0}^{d_{i_0}}(\mathbf{u}) \neq 0$ for some $\mathbf{u} \in U^{d_{i_0}}$ in (24). Consider any assignment of the form $\mathbf{x} = (a_1 w, \ldots, a_k w)$. We have that

$$B_i(\mathbf{x}_\mathbf{u}) = \mathbf{a}^\mathbf{u} B_i(w, \ldots, w) = \mathbf{a}^\mathbf{u} P_i(w).$$

Hence we get that

$$P_\Lambda(a_1 w, \ldots, a_k w) = \sum_{i=1}^k \alpha_i P_i(w), \tag{25}$$

where $\alpha_i = \sum_{\mathbf{u} \in U^{d_i}} \mathbf{a}^\mathbf{u} b_i^{d_i}(\mathbf{u})$. Applying Claim C.4, there exists a choice of $a_1, \ldots, a_k$, such that $\alpha_{i_0} \neq 0$, and then the linear independence of $\{P_1, \ldots, P_k\}$ shows that $P_\Lambda \not\equiv 0$. □

The proof of Lemma C.2 shows that $P_\Lambda(\mathbf{x}) \not\equiv 0$ if and only if $b_i^{d_i}(\mathbf{u}) \neq 0$ for some $i$ and $\mathbf{u}$. Next we prove Lemma C.1 where we show that in this case under the stronger condition of high rank bias$(P_\Lambda)$ is small.

*Proof of Lemma C.1.* Suppose that $P_\Lambda(\mathbf{x}) \not\equiv 0$ so that $b_i^{d_i}(\mathbf{u}) \neq 0$, for some $i \in [k]$ and $\mathbf{u} \in U^{d_i}$. Let $d_{i_0} \leq d$ be the largest degree such that $b_{i_0}^{d_{i_0}}(\mathbf{u}) \neq 0$, for some $\mathbf{u} \in U^{d_{i_0}}$.

Assume for contradiction that $\text{bias}(P_\Lambda) \geq \varepsilon$. By the regularity theorem for polynomials (Theorem B.11) we get that $P_\Lambda(\mathbf{x})$ can be expressed as a function of $r \leq r(d_{i_0}, \varepsilon) \leq r(d, \varepsilon)$ polynomials of degree at most $d_{i_0} - 1$. We will show that this implies $\text{rank}(\mathcal{P}) \leq r$. We know that $P_\Lambda(\mathbf{x})$ can be expressed as a function of at most $r$ polynomials of degree at most $d_{i_0} - 1$. This continues to hold under any assignment $\mathbf{x} = (a_1 w, \ldots, a_k w)$. That is

$$P_\Lambda(a_1 w, \ldots, a_k w) = \sum_{i=1}^{k} \alpha_i P_i(w),$$

with $\alpha_i = \sum_{\mathbf{u} \in U^{d_i}} \mathbf{a}^{\mathbf{u}} b_i^{d_i}(\mathbf{u})$ is a function of at most $r$ polynomials of degree at most $d_{i_0} - 1$. By Claim C.4 there exists a choice of $a_1, \ldots, a_k$ such that $\alpha_{i_0} \neq 0$, and this shows that $\text{rank}(\mathcal{P}) \leq r$. □

Next we prove Lemma C.3 where we deal with the case that the polynomials are not necessarily homogeneous, but instead the system of linear forms $\{L_1, \ldots, L_m\}$ is homogeneous.

*Proof of Lemma C.3.* It follows from Observation A.11 that by a change of variables we can assume that $L_1, \ldots, L_m$ are linear forms over $\mathbb{F}_p^n$ in $s$ variables $x_1, \ldots, x_s$, and $x_1$ appears with coefficient 1 in all linear forms. For $1 \leq i \leq k$, let $B_i^1, \ldots, B_i^{d_i}$ be the symmetric multilinear forms associated with $P_i^{(1)}, \ldots, P_i^{(d_i)}$, respectively. Then (24) must be replaced by

$$\begin{aligned} P_\Lambda(\mathbf{x}) &= \sum_{i=1}^{k} \sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(\mathbf{x})) = \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{l=1}^{d_i} \lambda_{i,j} \sum_{\mathbf{u} \in U^l} r_l(\mathbf{u}) c_l(\mathbf{u}, L_j) B_i^l(\mathbf{x}_{\mathbf{u}}) \\ &= \sum_{i=1}^{k} \sum_{l=1}^{d_i} \sum_{\mathbf{u} \in U^l} b_i^l(\mathbf{u}) B_i^l(\mathbf{x}_{\mathbf{u}}), \end{aligned} \tag{26}$$

where

$$b_i^l(\mathbf{u}) := r_l(\mathbf{u}) \sum_{j=1}^{m} \lambda_{i,j} c_l(\mathbf{u}, L_j).$$

Suppose that $P_\Lambda(\mathbf{x}) \not\equiv 0$. We claim that in this case there exists an $i \in [k]$ and $\mathbf{u} \in U^{d_i}$ such that $b_i^{d_i}(\mathbf{u}) \neq 0$, and this establishes the lemma. In order to prove this claim it suffices to show that if $b_i^{d_i}(\mathbf{u}) = 0$, for every $\mathbf{u} \in U^{d_i}$, then $b_i^t(\mathbf{u}) = 0$ for every $0 \leq t \leq d_i$ and $\mathbf{u} \in U^t$. Let otherwise $t$ be the largest integer such that $b_i^t(\mathbf{u}) = r_t(\mathbf{u}) \sum_{j=1}^{m} \lambda_{i,j} c_t(\mathbf{u}, L_j) \neq 0$, for some $\mathbf{u} = (u_1, \ldots, u_t) \in U^t$. Consider $\mathbf{u}' = (1, u_1, \ldots, u_t) \in U^{t+1}$. Since $x_1$ appears with coefficient 1 in every $L_j$, we have that $c_{t+1}(\mathbf{u}', L_j) = c_t(\mathbf{u}, L_j)$. Also note that since $d_i < p$, and $p$ is a prime, $r_l(\mathbf{u}) \neq 0$ for every $1 \leq l \leq d_i$ and $\mathbf{u} \in U^l$. Hence we conclude that

$$b_i^{t+1}(\mathbf{u}') = r_{t+1}(\mathbf{u}') \sum_{j=1}^{m} \lambda_{i,j} c_t(\mathbf{u}, L_j) = \frac{r_{t+1}(\mathbf{u}')}{r_t(\mathbf{u})} b_i^t(\mathbf{u}) \neq 0$$

which contradicts the maximality of $t$. □

## C.2 An invariance result

The following proposition which is a consequence of the lemmas that we established in Section C is one of the key ingredients in the proof of our main result, Theorem 1.3.

**Proposition C.5.** *Let $r : \mathbb{N} \times (0,1]$ be the function in Lemma C.1. Let $\mathcal{P} = \{P_1, \ldots, P_k\}, \mathcal{Q} = \{Q_1, \ldots, Q_k\}$ be two collections of polynomials over $\mathbb{F}_p^n$ of degree at most $d < p$ such that $\deg(P_i) = \deg(Q_i)$ for every $1 \leq i \leq k$. Let $L_1, \ldots, L_m$ be linear forms, and $\Gamma : \mathbb{F}_p^k \to \mathbb{D}$ be an arbitrary function. Define $f, g : \mathbb{F}_p^n \to \mathbb{D}$ by*

$$f(x) = \Gamma(P_1(x), \ldots, P_k(x))$$

*and*

$$g(x) = \Gamma(Q_1(x), \ldots, Q_k(x)).$$

*Then for every $\varepsilon > 0$, if $\mathrm{rank}(\mathcal{P}), \mathrm{rank}(\mathcal{Q}) > r(d, \varepsilon)$, we have*

$$|t_{\mathcal{L}}(f) - t_{\mathcal{L}}(g)| \leq 2p^{mk} \cdot \varepsilon,$$

*provided that at least one of the following two conditions hold:*

- *(i) The polynomials $P_1, \ldots, P_k$ and $Q_1, \ldots, Q_k$ are homogenous.*
- *(ii) The system of linear forms $\{L_1, \ldots, L_m\}$ is homogenous.*

*Proof.* The Fourier expansion of $\Gamma$ shows

$$\Gamma(z(1), \ldots, z(k)) = \sum_{\gamma \in \mathbb{F}_p^k} \widehat{\Gamma}(\gamma) \mathrm{e}_p\left(\sum_{i=1}^{k} \gamma(i) \cdot z(i)\right).$$

We thus have

$$\prod_{i=1}^{m} f(L_i(\mathbf{x})) = \sum_{\gamma_1, \ldots, \gamma_m \in \mathbb{F}_p^k} \widehat{\Gamma}(\gamma_1) \cdot \ldots \cdot \widehat{\Gamma}(\gamma_m) \mathrm{e}_p\left(\sum_{i \in [k], j \in [m]} \gamma_j(i) \cdot P_i(L_j(\mathbf{x}))\right),$$

and

$$\prod_{i=1}^{m} g(L_i(\mathbf{x})) = \sum_{\gamma^1, \ldots, \gamma^m \in \mathbb{F}_p^k} \widehat{\Gamma}(\gamma_1) \cdot \ldots \cdot \widehat{\Gamma}(\gamma_m) \mathrm{e}_p\left(\sum_{i \in [k], j \in [m]} \gamma_j(i) \cdot Q_i(L_j(\mathbf{x}))\right).$$

By Lemma C.1 we know that for every $\gamma_1, \ldots, \gamma_m \in \mathbb{F}_p^k$ each one of the polynomials $\sum_{i \in [k], j \in [m]} \gamma_j(i) \cdot P_i(L_j(\mathbf{x}))$ and $\sum_{i \in [k], j \in [m]} \gamma_j(i) \cdot Q_i(L_j(\mathbf{x}))$ are either zero, or have bias of at most $\varepsilon$. But Lemmas C.2 and C.3 show that under each one of the Conditions (i) or (ii), since $\deg(P_i) = \deg(Q_i)$, we have that

$$\sum_{i \in [k], j \in [m]} \gamma_j(i) \cdot P_i(L_j(\mathbf{x})) \equiv 0 \Longleftrightarrow \sum_{i \in [k], j \in [m]} \gamma_j(i) \cdot Q_i(L_j(\mathbf{x})) \equiv 0.$$

Hence we have that

$$\left|\mathbb{E}_{\mathbf{X}}\left[\mathrm{e}_p\left(\sum_{i\in[k],j\in[m]}\gamma_j(i)\cdot P_i(L_j(\mathbf{X}))\right)\right]-\mathbb{E}_{\mathbf{X}}\left[\mathrm{e}_p\left(\sum_{i\in[k],j\in[m]}\gamma_j(i)\cdot Q_i(L_j(\mathbf{X}))\right)\right]\right|\le 2\varepsilon,$$

and

$$\left|\mathbb{E}_{\mathbf{X}}\left[\prod_{i=1}^{m}f(L_i(\mathbf{X}))\right]-\mathbb{E}_{\mathbf{X}}\left[\prod_{i=1}^{m}g(L_i(\mathbf{X}))\right]\right|\le 2\varepsilon\cdot\sum_{\gamma_1,\ldots,\gamma_m\in\mathbb{F}_p^k}|\widehat{\Gamma}(\gamma_1)|\ldots|\widehat{\Gamma}(\gamma_m)|.$$

Since $\|\widehat{\Gamma}\|_\infty\le 1$, we conclude

$$\left|\mathbb{E}_{\mathbf{X}}\left[\prod_{i=1}^{m}f(L_i(\mathbf{X}))\right]-\mathbb{E}_{\mathbf{X}}\left[\prod_{i=1}^{m}g(L_i(\mathbf{X}))\right]\right|\le 2\varepsilon p^{mk}.$$

□

### C.3 A solution to the Gowers-Wolf conjecture

In this subsection we prove the following theorem, which qualitatively strengthens both Lemma A.2 and Theorem A.6.

**Theorem C.6.** *Let $\mathcal{L}=\{L_1,\ldots,L_m\}$ be a system of $m$ linear forms in $k$ variables of Cauchy-Schwarz complexity at most $p$. Let $d\ge 0$, and assume that $L_1^{d+1}$ is not in the linear span of $L_2^{d+1},\ldots,L_m^{d+1}$. Then for every $\varepsilon>0$, there exists $\delta>0$ such that for every functions $f_1,\ldots,f_m:\mathbb{F}_p\to\mathbb{D}$ where $\|f_1\|_{U^{d+1}}\le\delta$ we have*

$$\left|\mathbb{E}_{\mathbf{X}\in(\mathbb{F}_p^n)^k}\left[\prod_{i=1}^{m}f_i(L_i(\mathbf{X}))\right]\right|\le\varepsilon.$$

Let $s\le p$ denote the Cauchy-Schwarz complexity of $\mathcal{L}$. Note that Lemma A.2 requires the stronger condition $\|f_1\|_{U^{s+1}}\le\delta$, while Theorem A.6 requires two stronger conditions: that $L_1^{d+1},\ldots,L_m^{d+1}$ are linearly independent, and that all $\|f_1\|_{U^{d+1}},\ldots,\|f_m\|_{U^{d+1}}$ are bounded by $\delta$. Before proving Theorem C.6 we give an immediate corollary of it which we find quite useful.

**Corollary C.7.** *Let $\mathcal{L}=\{L_1,\ldots,L_m\}$ be a system of $m$ linear forms in $k$ variables of true complexity $d$ and Cauchy-Schwarz complexity at most $p$. Then for every $\varepsilon>0$, there exists $\delta>0$ such that the following holds. Let $f_1,\ldots,f_m:\mathbb{F}_p\to\mathbb{D}$ be functions. Let $\mathcal{B}$ be a polynomial factor such that*

$$\|f_i-\mathbb{E}(f_i|\mathcal{B})\|_{U^{d+1}}\le\delta,$$

*for all $1\le i\le m$. Then*

$$\left|\mathbb{E}\left[\prod_{i=1}^{m}f_i(L_i(\mathbf{X}))\right]-\mathbb{E}\left[\prod_{i=1}^{m}\mathbb{E}(f_i|\mathcal{B})(L_i(\mathbf{X}))\right]\right|\le\varepsilon,$$

*where the average is over $\mathbf{X}\in(\mathbb{F}_p^n)^k$.*

We now turn to prove Theorem C.6. If $d = s$, then Theorem C.6 follows from Lemma A.2. Thus we assume $d < s$, hence also $d < p$. We will assume throughout the proof that $p, m, s, d, \varepsilon$ are constants, and we will not explicitly state dependencies on them.

Let $r : \mathbb{N} \to \mathbb{N}$ be a growth function to be specified later. Let $\eta > 0$ be a sufficiently small constant which will be specified later. By Lemma B.14 there exists a polynomial factor $\mathcal{B}$ of degree $s$, complexity $C \le C_{\max}(\eta, r(\cdot))$ and rank at least $r(C)$ such that we can decompose each function $f_i$ as

$$f_i = h_i + h_i'$$

where $h_i = \mathbb{E}(f_i|\mathcal{B})$ and $\|h_i'\|_{U^{s+1}} \le \eta$. We also have $\|h_i\|_\infty \le 1$ and $\|h_i'\|_\infty \le 2$. We first show that in order to bound $\mathbb{E}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right]$ it suffices to bound $\mathbb{E}\left[\prod_{i=1}^m h_i(L_i(\mathbf{X}))\right]$, if $\eta$ is chosen to be small enough.

**Claim C.8.** *If $\eta \le \varepsilon \cdot 4^{-m}/10$ then*

$$\left| \mathbb{E}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right] - \mathbb{E}\left[\prod_{i=1}^m h_i(L_i(\mathbf{X}))\right] \right| \le \varepsilon/10,$$

*where the average is over uniform $\mathbf{X} \in (\mathbb{F}_p^n)^k$.*

*Proof.* We have

$$\mathbb{E}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right] - \mathbb{E}\left[\prod_{i=1}^m h_i(L_i(\mathbf{X}))\right] = \sum_{I \subsetneq [m]} \mathbb{E}\left[\prod_{i \in I} h_i(L_i(\mathbf{X})) \prod_{i \in [m] \setminus I} h_i'(L_i(\mathbf{X}))\right].$$

Fix $I \subsetneq [m]$. Let $j \in [m] \setminus I$. Since the Cauchy-Schwarz complexity of $\{L_1, \ldots, L_m\}$ is $s$, we have by Lemma A.2 that

$$\left| \mathbb{E}\left[\prod_{i \in I} h_i(L_i(\mathbf{X})) \prod_{i \in [m] \setminus I} h_i'(L_i(\mathbf{X}))\right] \right| \le \|h_j'\|_{U^{s+1}} \prod_{i \in I} \|h_i\|_\infty \prod_{i \in [m] \setminus (I \cup \{j\})} \|h_i\|_\infty \le \eta \cdot 2^m.$$

As there are $2^m - 1$ different choices for $I$ the claim follows. □

We thus set $\eta = \varepsilon \cdot 4^{-m}/10$, and regard $\eta$ from now on as a constant, and we do not specify explicitly dependencies on $\eta$ as well.

Let $\{P_i\}_{1 \le i \le C}$ be the polynomials which define the polynomial factor $\mathcal{B}$, where we assume each $P_i$ is homogeneous of degree $\deg(P_i) \le s$. Since each $h_i$ is measurable with regards to $\mathcal{B}$, we have $h_i(x) = \Gamma_i(P_1(x), \ldots, P_C(x))$ where $\Gamma_i : \mathbb{F}_p^C \to \mathbb{D}$ is some function. Decompose $\Gamma_i$ to its Fourier decomposition as

$$\Gamma_i(z(1), \ldots, z(C)) = \sum_{\gamma \in \mathbb{F}_p^C} c_{i,\gamma} \cdot \mathrm{e}_p\left(\sum_{j=1}^C \gamma(j) z(j)\right),$$

where $|c_{i,\gamma}| \le 1$. Define for $\gamma \in \mathbb{F}_p^C$, the linear combination $P_\gamma(x) = \sum_{j=1}^C \gamma(j) P_j(x)$. We can express each $h_i$ as

$$h_i(x) = \sum_{\gamma \in \mathbb{F}_p^C} c_{i,\gamma} \cdot \mathrm{e}_p\left(P_\gamma(x)\right),$$

and we can express

$$\mathbb{E}\left[\prod_{i=1}^{m} h_i(L_i(\mathbf{X}))\right] = \sum_{\gamma_1,\ldots,\gamma_m \in \mathbb{F}_p^C} \Delta(\gamma_1,\ldots,\gamma_m), \tag{27}$$

where

$$\Delta(\gamma_1,\ldots,\gamma_m) = \left(\prod_{i=1}^{m} c_{i,\gamma_i}\right) \mathbb{E}\left[\mathrm{e}_p\left(P_{\gamma_1}(L_1(\mathbf{X})) + \ldots + P_{\gamma_m}(L_m(\mathbf{X}))\right)\right]. \tag{28}$$

We will bound each term $\Delta(\gamma_1,\ldots,\gamma_m)$ by $\tau := \tau(C) = p^{-mC}\varepsilon/10$, which will establish the result. Let $S = \{\gamma \in \mathbb{F}_p^C : \deg(P_\gamma) \le d\}$. We first bound the terms $\Delta(\gamma_1,\ldots,\gamma_m)$ with $\gamma_1 \in S$.

**Claim C.9.** *If the growth function $r(\cdot)$ is chosen large enough, and if $\delta > 0$ is chosen small enough, then for all $\gamma_1 \in S$ we have*

$$|c_{1,\gamma_1}| \le \tau.$$

*Consequently, for all $\gamma_1 \in S$ and $\gamma_2,\ldots,\gamma_m \in \mathbb{F}_p^C$ we have*

$$\Delta(\gamma_1,\ldots,\gamma_m) \le \tau.$$

*Proof.* The bound on $\Delta(\gamma_1,\ldots,\gamma_m)$ follows trivially from the bound on $c_{1,\gamma_1}$, since $|c_{2,\gamma_2}|,\ldots,|c_{m,\gamma_m}| \le 1$. To bound $c_{1,\gamma}$, note that

$$c_{1,\gamma} = \mathbb{E}\left[h_1(X)\mathrm{e}_p(-P_\gamma(X))\right] - \sum_{\gamma' \in \mathbb{F}_p^C, \gamma' \ne \gamma} c_{1,\gamma'}\mathbb{E}\left[\mathrm{e}_p(P_{\gamma'}(X) - P_\gamma(X))\right],$$

where the averages are over uniform $X \in \mathbb{F}_p^n$. We first bound $\mathbb{E}\left[h_1(X)\mathrm{e}_p(-P_\gamma(X))\right]$. Using the fact that $h_1 = \mathbb{E}(f_1|\mathcal{B})$ and that the function $\mathrm{e}_p(-P_\gamma(x))$ is $\mathcal{B}$-measurable, we have by Observation B.6 that

$$\left|\mathbb{E}\left[h_1(X)\mathrm{e}_p(-P_\gamma(X))\right]\right| = \left|\mathbb{E}\left[f_1(X)\mathrm{e}_p(-P_\gamma(X))\right]\right| \le \|f_1\|_{U^{d+1}} \le \delta.$$

Hence, by choosing $\delta < p^{-m\cdot C_{\max}(r(\cdot))}\varepsilon/20$ we guarantee that $\left|\mathbb{E}\left[h_1(X)\mathrm{e}_p(-P_\gamma(X))\right]\right| \le \delta < \tau/2$. Next for $\gamma \ne \gamma'$, we bound each term $\mathbb{E}\left[\mathrm{e}_p(P_{\gamma'}(X) - P_\gamma(X))\right]$ by $\tau p^{-C}/2$. Assume that for some $\gamma' \ne \gamma$ we have

$$\left|\mathbb{E}\left[\mathrm{e}_p(P_{\gamma'}(X) - P_\gamma(X))\right]\right| > \tau p^{-C}/2.$$

Then by Theorem B.11 we have that

$$\mathrm{rank}(P_{\gamma'} - P_\gamma) \le r'(\tau p^{-C}/2) = r_1(C).$$

Thus, as long as we choose $r(C) > r_1(C)$, we have that

$$\sum_{\gamma' \ne \gamma} \left|c_{1,\gamma'}\mathbb{E}\left[\mathrm{e}_p(P_{\gamma'}(X) - P_\gamma(X))\right]\right| \le \tau/2$$

and we achieve the bound $|c_{1,\gamma}| \le \tau$. □

Consider now any $\gamma_1 \notin S$. We will show that if we choose $r(\cdot)$ large enough we can guarantee that

$$\left|\mathbb{E}\left[\mathrm{e}_p\left(P_{\gamma_1}(L_1(\mathbf{X})) + \ldots + P_{\gamma_m}(L_m(\mathbf{X}))\right)\right]\right| \le \tau, \tag{29}$$

which will establish the result. Assume that this is not the case for some $\gamma_1 \notin S$ and $\gamma_2, \dots, \gamma_m \in \mathbb{F}_p^C$. By Lemma C.1 there is a rank $r''(\tau) = r_2(C)$ such that if we guarantee that $r(C) > r_2(C)$ and if (29) does not hold, then we must have

$$P_{\gamma_1}(L_1(\mathbf{x})) + \ldots + P_{\gamma_m}(L_m(\mathbf{x})) \equiv 0. \tag{30}$$

Let $t = \deg(P_{\gamma_1}) > d$. Let $P_\gamma^{(t)}$ be the degree $t$ homogeneous part of $P_\gamma$. Since the degrees of the polynomials are at most $p-1$, we must have that

$$P_{\gamma_1}^{(t)}(L_1(\mathbf{x})) + \ldots + P_{\gamma_m}^{(t)}(L_m(\mathbf{x})) \equiv 0. \tag{31}$$

The following claim concludes the proof. It shows that if (31) holds, then $L_1^t$ is linearly dependent on $L_2^t, \dots, L_m^t$. This immediately implies that also $L_1^{d+1}$ is linearly dependent on $L_2^{d+1}, \dots, L_m^{d+1}$ (since $t \geq d+1$) which contradicts our initial assumption.

**Claim C.10.** *Let $P_1, \dots, P_m$ be homogenous polynomials of degree $t < p$, where $P_1$ is not identically zero, such that*

$$P_1(L_1(\mathbf{x})) + \ldots + P_m(L_m(\mathbf{x})) \equiv 0.$$

*Then $L_1^t$ is linearly dependent on $L_2^t, \dots, L_m^t$.*

*Proof.* Let $M(x) = x_{i_1} \cdot \ldots \cdot x_{i_t}$ be a monomial appearing in $P_1$ with a nonzero coefficient $\alpha_1 \neq 0$. Let $\alpha_i$ be the coefficient of $M(x)$ in $P_i$ for $2 \leq i \leq m$. We have that

$$\alpha_1 M(L_1(\mathbf{x})) + \ldots + \alpha_m M(L_m(\mathbf{x})) \equiv 0.$$

Let $\mathbf{x} = (x_1, \dots, x_k)$ and $L_i(\mathbf{x}) = \lambda_{i,1}x_1 + \ldots + \lambda_{i,k}x_k$. We have

$$M(L_i(\mathbf{x})) = \prod_{j=1}^{t} (\lambda_{i,1}x_1(i_j) + \ldots + \lambda_{i,k}x_k(i_j)).$$

Consider the assignment $x_i = (z(i), \dots, z(i))$ where $z(1), \dots, z(k) \in \mathbb{F}_p$ are new variables. We thus have the polynomial identity

$$\sum_{i=1}^{m} \alpha_i(\lambda_{i,1}z(1) + \ldots + \lambda_{i,k}z(k))^t \equiv 0,$$

which as $t < p$ is equivalent to

$$\sum_{i=1}^{m} \alpha_i L_i^t \equiv 0.$$

□

# D Characterization of strongly correlation testable properties

Consider a family $\mathcal{D} := \{D_n\}_{n \in \mathbb{N}}$ where $D_n$ is a set of functions from $\mathbb{F}_p^n$ to $\mathbb{D}$. We recall some basic definitions from the introduction. The correlation of a function $f : \mathbb{F}_p^n \to \mathbb{D}$ with $D_n$ is

$$\|f\|_{u(D_n)} = \sup_{g \in D_n} |\langle f, g \rangle|.$$

Given a function $f : \mathbb{F}_p^n \to \mathbb{D}$ and a system of linear forms $\mathcal{L} = \{L_1, \dots, L_m\}$ in $k$ variables, recall that the average of $f$ over $\mathcal{L}$, with conjugations $\alpha \in \{0,1\}^m$, is

$$t_{\mathcal{L},\alpha}(f) = \mathbb{E}_{\mathbf{X}\in(\mathbb{F}_p^n)^k}\left[\prod_{i=1}^{m} \mathcal{C}^{\alpha_i} f(L_i(\mathbf{X}))\right]$$

where $\mathcal{C}$ is the conjugation operator. A family $\mathcal{D}$ is said to be correlation testable with linear forms if there exists a set of linear forms $\mathcal{L}_1, \dots, \mathcal{L}_\ell$ along with conjugations $\alpha_1, \dots, \alpha_\ell$, such that the collection of averages $(t_{\mathcal{L}_1,\alpha_1}(f), \dots, t_{\mathcal{L}_\ell,\alpha_\ell}(f))$ allows to distinguish whether $f$ has noticeable correlation with $D_n$ or negligible correlation with $D_n$. The true complexity (Cauchy-Schwarz complexity) of $\mathcal{D}$ is the maximal true complexity (Cauchy-Schwarz complexity) of $\{\mathcal{L}_i\}_{i=1,\dots,\ell}$.

**Definition D.1** (Strongly correlation testable properties)**.** *A family $\mathcal{D} = \{D_n\}_{n\in\mathbb{N}}$ is* strongly correlation testable *by linear systems with true complexity $d$ and Cauchy-Schwarz complexity $s$, if the following holds. For every $\varepsilon > 0$, there exist $\delta \in (0, \varepsilon)$, $n_0 \in \mathbb{N}$, and systems of homogeneous linear forms $\mathcal{L}_1, \dots, \mathcal{L}_\ell$ in $m_1, \dots, m_\ell$ variables, respectively, where each system has true complexity at most $d$ and Cauchy-Schwarz complexity at most $s$, along with conjugations $\alpha_1 \in \{0,1\}^{m_1}, \dots, \alpha_\ell \in \{0,1\}^{m_\ell}$ such that the closures of the following two sets are disjoint:*

$$T_\epsilon = \left\{((t_{\mathcal{L}_1,\alpha_1}(f), \dots, t_{\mathcal{L}_k,\alpha_k}(f)) : f : \mathbb{F}_p^n \to \mathbb{D}, n \geq n_0, \|f\|_{u(D_n)} \geq \varepsilon\right\}$$

*and*

$$S_\epsilon = \left\{((t_{\mathcal{L}_1,\alpha_1}(f), \dots, t_{\mathcal{L}_k,\alpha_k}(f)) : f : \mathbb{F}_p^n \to \mathbb{D}, n \geq n_0, \|f\|_{u(D_n)} \leq \delta\right\}.$$

A system $\mathcal{D} = \{D_n\}_{n\in\mathbb{N}}$ where $D_n$ is a set of functions from $\mathbb{F}_p^n$ to $\mathbb{D}$ is *consistent* if $D_n \subseteq D_{n+1}$, where we identify $\mathbb{F}_p^n$ with the subspace $\mathbb{F}_p^n \times \{0\}$ of $\mathbb{F}_p^{n+1}$. In this section, we prove the following theorem.

**Theorem D.2** (Main theorem: strongly correlation testable functions)**.** *Consider a consistent family $\mathcal{D} = \{D_n\}_{n\in\mathbb{N}}$ . If $\mathcal{D}$ is strongly correlation testable with true complexity $d$ and Cauchy-Schwarz complexity $s < p$, then there exists $0 \leq t \leq d$ such that the following holds. Let $(f_n : \mathbb{F}_p^n \to \mathbb{D})_{n\in\mathbb{N}}$ be a sequence of functions. Then*

$$\lim_{n\to\infty} \|f_n - \mathbb{E}[f_n]\|_{u(D_n)} = 0 \iff \lim_{n\to\infty} \|f_n - \mathbb{E}[f_n]\|_{U^{t+1}} = 0.$$

Define a set $\mathcal{S} \subseteq \mathbb{N}$ to be the set of all degrees $k \geq 1$ for which the following holds. For every growth function $r : \mathbb{N} \to \mathbb{N}$, there exists $n_0, a \in \mathbb{N}$, such that for every $n \geq n_0$ there exist a polynomial $P_n$ over $\mathbb{F}_p^n$ of degree exactly $k$ such that

(i) $\|\mathrm{e}_p(P_n)\|_{u(D_n)} \geq 1/a$;

(ii) $\mathrm{rank}(P_n) > r(a)$.

**Claim D.3.** *Unless all functions in $\mathcal{D}$ are constant functions, we have $1 \in \mathcal{S}$.*

*Proof.* Let $g_{n_0} \in D_{n_0}$ be a nonconstant function. There must exist a nonzero Fourier coefficient $\alpha \in \mathbb{F}_p^{n_0}$ such that

$$\widehat{g_{n_0}}(\alpha) = \eta \neq 0.$$

Since we assume that the family $\mathcal{D}$ is consistent, for every $n \geq n_0$, the function

$$g_n(x(1), \ldots, x(n)) = g_{n_0}(x(1), \ldots, x(n_0))$$

belongs to $D_n$. Let $P_n(x) = \sum_{i=1}^{n_0} \alpha(i)x(i)$ be a linear function. For every nonzero linear function we have $\mathrm{rank}(P_n) = \infty$. By construction, each linear function $P_n$ has correlation with $D_n$,

$$\|\mathrm{e}_p(P_n)\|_{u(D_n)} \geq |\langle \mathrm{e}_p(P_n), g_n \rangle| = |\eta| > 0.$$

□

The case where all functions in $\mathcal{D}$ are constants is easy to analyze, as in this case we have that

$$\|f\|_{u(D_n)} = \|f\|_{U^1},$$

for all function $f : \mathbb{F}_p^n \to \mathbb{D}$. Thus, from now on we assume that there are some nonconstant functions in $\mathcal{D}$, and hence $1 \in \mathcal{S}$.

**Claim D.4.** $\mathcal{S} \subseteq \{1, \ldots, s\}$.

*Proof.* Let $k > s$, and assume by contradiction that for every growth function $r : \mathbb{N} \to \mathbb{N}$, there exists $a, n_0 \in \mathbb{N}$, such that for every $n \geq n_0$, there exists a polynomial $P_n$ of degree $k$ with $\|\mathrm{e}_p(P_n)\|_{u(D_n)} \geq 1/a$ and $\mathrm{rank}(P_n) \geq r(a)$. We will show that in this case, the closures of $T_{1/a}$ and $S_{1/a}$ are not disjoint for all $a \in \mathbb{N}$, which will yield a contradiction. Assume to the contrary that they are disjoint for every $a \in \mathbb{N}$. Then for every $a \in \mathbb{N}$, there exists a minimal distance $\mu(a) > 0$, such that for every $z' \in T_{1/a}$ and $z'' \in S_{1/a}$ we have

$$\|z' - z''\|_\infty \geq \mu(a). \tag{32}$$

Applying Theorem B.3 for $s < p$ and Theorem B.11 for $k$, if we choose the rank bound $r(a)$ large enough, we can guarantee that

$$\|\mathrm{e}_p(P_n)\|_{U^{s+1}} \leq \mu(a)/2. \tag{33}$$

We first note that $0^\ell = (0, \ldots, 0)$ is in $S_{1/a}$ for every $a \in \mathbb{N}$, since for $f \equiv 0$ we have $t_{\mathcal{L},\alpha}(f) = 0$. Combining this with (32) we get that for every sequence of functions $f_n : \mathbb{F}_p^n \to \mathbb{D}$ with $\liminf_{n\to\infty} \|f_n\|_{u(D_n)} \geq 1/a$, we have

$$\liminf_{n\to\infty} \|(t_{\mathcal{L}_1,\alpha_1}(f_n), \ldots, t_{\mathcal{L}_\ell,\alpha_\ell}(f_n))\|_\infty \geq \mu(a). \tag{34}$$

Consider now the polynomials $P_n$. By Lemma A.2, since each system $\mathcal{L}_i$ has Cauchy-Schwarz complexity at most $s < p$, we have

$$|t_{\mathcal{L}_i,\alpha_i}(\mathrm{e}_p(P_n))| \leq \|\mathrm{e}_p(P_n)\|_{U^{s+1}} \leq \mu(a)/2,$$

for all $1 \leq i \leq \ell$. Thus we reached a contradiction. □

We now define $t := \max(\mathcal{S})$. Theorem D.2 follows from the following two lemmas.

**Lemma D.5.** *Let $(f_n : \mathbb{F}_p^n \to \mathbb{D})_{n\in\mathbb{N}}$ be a sequence of functions such that $\mathbb{E}[f_n] = 0$. If $\lim_{n\to\infty} \|f_n\|_{U^{t+1}} = 0$, then $\lim_{n\to\infty} \|f_n\|_{u(D_n)} = 0$.*

**Lemma D.6.** *Let $(f_n : \mathbb{F}_p^n \to \mathbb{D})_{n\in\mathbb{N}}$ be a sequence of functions such that $\mathbb{E}[f_n] = 0$. If $\lim_{n\to\infty} \|f_n\|_{u(D_n)} = 0$, then $\lim_{n\to\infty} \|f_n\|_{U^{t+1}} = 0$.*

We prove Lemma D.5 in Subsection D.1 and Lemma D.6 in Subsection D.2.

### D.1 Proof of Lemma D.5

Suppose that $\lim_{n\to\infty} \|f_n\|_{U^{t+1}} = 0$, but

$$c := \limsup_{n\to\infty} \|f_n\|_{u(D_n)} > 0.$$

Since $\mathcal{D}$ is consistent, we can replace the lim sup by an actual limit. Assume that $c = \lim_{n\to\infty} \|f_n\|_{u(D_n)}$ and set $\varepsilon := c/2$. Since $\mathcal{D}$ is strongly correlation testable with true complexity $d$ and Cauchy-Schwarz complexity $s < p$, there exist $\delta \in (0, \varepsilon)$, $\varepsilon' > 0$, and a family of homogenous systems of linear forms $\mathcal{L}_1, \ldots, \mathcal{L}_\ell$ of Cauchy-Schwarz complexity at most $s$ and true complexity at most $d$ along with conjugations $\alpha_1, \ldots, \alpha_\ell$ such that

$$\|(t_{\mathcal{L}_1,\alpha_1}(f), \ldots, t_{\mathcal{L}_\ell,\alpha_\ell}(f)) - (t_{\mathcal{L}_1,\alpha_1}(g), \ldots, t_{\mathcal{L}_\ell,\alpha_\ell}(g))\|_\infty \geq \varepsilon', \tag{35}$$

for every $f, g : \mathbb{F}_p^n \to \mathbb{D}$ (with $n > n_0$) satisfying $\|f\|_{u(D_n)} \geq \varepsilon$ and $\|g\|_{u(D_n)} \leq \delta$.

Let $r : \mathbb{N} \to \mathbb{N}$ be a growth function to be defined later. We apply Theorem B.7 and Corollary C.7 to deduce that there exists a polynomial factor $\mathcal{B}_n$ of degree $s$, complexity $C_n \leq C_{\max}(s, \varepsilon, r(\cdot))$ and rank at least $r(C_n)$, such that for $h_n := \mathbb{E}(f_n|\mathcal{B}_n)$ we have

$$|t_{\mathcal{L}_i,\alpha_i}(f_n) - t_{\mathcal{L}_i,\alpha_i}(h_n)| \leq \varepsilon'/2, \tag{36}$$

for all $1 \leq i \leq \ell$. Equations (35) and (36) imply that for large enough $n$ we have $\|h_n\|_{u(D_n)} \geq \delta$. So, for large enough $n$, there exists $g_n \in D_n$ such that

$$|\langle h_n, g_n\rangle| \geq \delta. \tag{37}$$

Let $\mathcal{B}_n$ be defined by polynomials $Q_{n,1}, \ldots, Q_{n,C_n}$ and define $Q_{n,\gamma} := \sum_{i=1}^{C_n} \gamma(i) Q_{n,i}(x)$ for every $\gamma \in \mathbb{F}_p^{C_n}$. By choosing the growth function $r(\cdot)$ large enough, we have by Theorem B.11 that for all $\gamma \neq \gamma'$,

$$|\text{bias}(Q_{n,\gamma} - Q_{n,\gamma'})| \leq p^{-2C_n}\delta/100. \tag{38}$$

As $h_n$ is $\mathcal{B}_n$-measurable, we can express it as $h_n(x) = F_n(Q_{n,1}(x), \ldots, Q_{n,C_n}(x))$ for some $F_n : \mathbb{F}_p^{C_n} \to \mathbb{D}$. Consider the Fourier decomposition of $F_n$,

$$F_n(z_1, \ldots, z_{C_n}) = \sum_{\gamma \in \mathbb{F}_p^{C_n}} \widehat{F_n}(\gamma) \mathrm{e}_p\left(\sum_{i=1}^{C_n} \gamma(i) z_i\right).$$

We thus have

$$h_n(x) = \sum_{\gamma \in \mathbb{F}_p^{C_n}} \widehat{F_n}(\gamma) \mathrm{e}_p(Q_{n,\gamma}(x)),$$

where $|\widehat{F_n}(\gamma)| \leq 1$. Define $W_n := \{\gamma \in \mathbb{F}_p^{C_n} : \deg(Q_{n,\gamma}) \leq t\}$. We now show that the assumption $\lim_{n\to\infty} \|f_n\|_{U^{t+1}} = 0$ implies that by taking $n$ large enough, we can make $|\widehat{F_n}(\gamma)|$ arbitrarily small for all $\gamma \in W_n$.

**Claim D.7.** *For large enough $n$, we have $|\widehat{F_n}(\gamma)| \leq p^{-C_n}\delta/10$ for all $\gamma \in W_n$.*

*Proof.* By expanding $\langle h_n, \mathrm{e}_p(Q_{n,\gamma})\rangle$ we have

$$\widehat{F_n}(\gamma) = \langle h_n, \mathrm{e}_p(Q_{n,\gamma})\rangle - \sum_{\gamma' \neq \gamma} \widehat{F_n}(\gamma') \cdot \mathrm{bias}(Q_{n,\gamma'} - Q_{n,\gamma}).$$

We bound each term individually. As $\mathrm{e}_p(Q_{n,\gamma})$ is $\mathcal{B}_n$-measurable and $h_n = \mathbb{E}(f_n|\mathcal{B}_n)$, we have that

$$\langle h_n, \mathrm{e}_p(Q_{n,\gamma})\rangle = \langle f_n, \mathrm{e}_p(Q_{n,\gamma})\rangle,$$

and since $\deg(Q_{n,\gamma}) \leq t$ we have

$$|\langle h_n, \mathrm{e}_p(Q_{n,\gamma})\rangle| \leq \|f_n\|_{U^{t+1}} \leq p^{-C_{\max}}\delta/100,$$

for large enough $n$. By (38) and the bound $\|\widehat{F_n}\|_\infty \leq 1$, we conclude that

$$|\widehat{F_n}(\gamma)| \leq p^{-C_n}\delta/10,$$

for all $\gamma \in W_n$. □

Now, Claim D.7 implies that

$$\left| \sum_{\gamma \in W_n} \widehat{F_n}(\gamma)\langle \mathrm{e}_p(Q_{n,\gamma}), g_n\rangle \right| \leq \delta/2.$$

However, since $|\langle h_n, g_n\rangle| \geq \delta$ we must have that there exists $\gamma^* \notin W_n$ such that

$$|\langle \mathrm{e}_p(Q_{n,\gamma^*}), g_n\rangle| \geq p^{-C_n}\delta/2. \tag{39}$$

We now show a contradiction to the assumption that $t = \max(\mathcal{S})$. Set $P_n := Q_{n,\gamma^*}$. As $Q_{n,\gamma} \notin W_n$, by 39 we have that

- $t + 1 \leq \deg(P_n) \leq s$;
- $\|\mathrm{e}_p(P_n)\|_{u(D_n)} \geq p^{-C_n}\delta/2$;
- $\mathrm{rank}(P_n) \geq \mathrm{rank}(\mathcal{B}_n) \geq r(C_n)$.

Let $n_1 < n_2 < \ldots$ be an infinite sequence such that $t' := \deg(P_{n_i}) \geq t + 1$ and $C_n = C$. Since the family $\mathcal{D}$ is consistent, we may assume (by refining the sequence) that $\deg(P_n) = t'$ and $\mathrm{rank}(P_n) \geq r(C)$ for all $n \in \mathbb{N}$. As $r(\cdot)$ is an arbitrary growth function, we must have $t' \in \mathcal{S}$ and the lemma follows.

## D.2 Proof of Lemma D.6

Let $(f_n : \mathbb{F}_p^n \to \mathbb{D})_{n\in\mathbb{N}}$ be a sequence of functions such that $\lim_{n\to\infty} \|f_n\|_{u(D_n)} = 0$, and assume by contradiction that

$$c := \limsup_{n\to\infty} \|f_n\|_{U^{t+1}} > 0.$$

Since $\mathcal{D}$ is consistent we may replace the lim sup by an actual limit by refining the sequence. So we assume that $\|f_n\|_{U^{t+1}} \geq c$ for all large enough $n > n_0$. By Theorem B.3 this implies (since $t \leq s < p$) that there exist polynomials $Q_n \in \mathrm{Poly}_t(\mathbb{F}_p^n)$ such that

$$|\langle f_n, \mathrm{e}_p(Q_n)\rangle| \geq \delta,$$

for some $\delta = \delta(c) > 0$. We first regularize the polynomials to have arbitrarily large rank.

**Claim D.8.** *There exists $1 \leq t_0 \leq t$ such that the following holds. For every growth function $r : \mathbb{N} \to \mathbb{N}$ there exists $b \in \mathbb{N}$ and an infinite subsequence $n_1 < n_2 < \ldots$, along with polynomials $Q_{n_i}$ of degree exactly $t_0$, such that*

- $|\langle f_{n_i}, \mathrm{e}_p(Q_{n_i})\rangle| \geq 1/b$;
- $\mathrm{rank}(Q_{n_i}) \geq r(b)$.

*Proof.* We first claim that it is enough to prove the claim where $t_0$ may depend on $r(\cdot)$. Otherwise, assume that for every $r(\cdot)$ there exists a $t_0(r)$ for which the claim holds, but there is no single $t_0$ which holds for all $r(\cdot)$. Then, for each possible $1 \leq i \leq t$, let $r_i(\cdot)$ be a growth function for which the claim does not hold for $t_0 = i$ and $r_i(\cdot)$. Set $r = \max(r_1, \ldots, r_t)$, and we reach a contradiction.

Thus, it is sufficient to show that for every growth function $r(\cdot)$, there exists $1 \leq t_0 \leq t$ for which the claim holds. Applying Theorem B.11, for every growth function $r : \mathbb{N} \to \mathbb{N}$, we can express each polynomial $Q_n$ as a function of $C_n \leq C_{\max}(t, r(\cdot))$ polynomials $Q'_{n,1}, \ldots, Q'_{n,C_n}$ of degree at most $t$ where $\mathrm{rank}(Q'_{n,1}, \ldots, Q'_{n,C_n}) \geq r(C_n)$. Assume that $Q_n(x) = F_n(Q'_{n,1}, \ldots, Q'_{n,C_n})$. Let $Q'_{n,\gamma} = \sum_{i=1}^{C_n} \gamma(i) Q'_{n,i}(x)$ for $\gamma \in \mathbb{F}_p^{C_n}$. By the Fourier decomposition of $F_n$, we have

$$\mathrm{e}_p(Q_n(x)) = \sum_{\gamma \in \mathbb{F}_p^{C_n}} \widehat{F_n}(\gamma) \mathrm{e}_p\left(Q'_{n,\gamma}(x)\right).$$

Since $\|\widehat{F_n}\|_\infty \leq 1$, we have that for every $n$ there exists $\gamma_n$ such that

$$|\langle f_n, \mathrm{e}_p(Q'_{n,\gamma_n})\rangle| \geq \delta p^{-C_n}.$$

Since we assumed $\mathbb{E}[f_n] = 0$, we cannot have that $Q'_{n,\gamma_n}$ is a constant, thus we have $\deg(Q'_{n,\gamma_n}) \geq 1$ and hence $\mathrm{rank}(Q'_{n,\gamma_n}) \geq r(C_n)$. Let $t_0$ be a number which is the degree of $Q'_{n,\gamma_n}$ for infinitely many $n$. Since $r(\cdot)$ is arbitrary, the claim follows. □

We fix $1 \leq t_0 \leq t$ given by Claim D.8. Let $r : \mathbb{N} \to \mathbb{N}$ be a growth function to be determined later. Since $\mathcal{D}$ is consistent, we can refine the sequence $(f_n)_{n \in \mathbb{N}}$ to the subsequence given by Claim D.8, and obtain there exists function $f_n : \mathbb{F}_p^n \to \mathbb{D}$ with $\lim_{n\to\infty} \|f_n\|_{u(D_n)} = 0$, and polynomials $Q_n$ of degree exactly $t_0$, such that for every $n > n_0$,

- $|\langle f_n, \mathrm{e}_p(Q_n)\rangle| \geq 1/b$;
- $\mathrm{rank}(Q_n) \geq r(b)$.

We first derive a contradiction when $t_0 \in \mathcal{S}$.

**Lemma D.9.** *If $t_0 \in \mathcal{S}$, then $\limsup_{n\to\infty} \|f_n\|_{u(D_n)} > 0$.*

*Proof.* In the proof, we think of $p, s$ as constants and do not explicitly mention dependencies on them. Given the value of $b$, let $\widetilde{r}_b : \mathbb{N} \to \mathbb{N}$ be a growth function to be determined later, where we will in particular have $\widetilde{r}_b(a) \geq r(a)$, for all $a, b \in \mathbb{N}$. Since $t_0 \in \mathcal{S}$, there exist polynomials $P_n$ of degree exactly $t_0$, functions $g_n \in D_n$, and $n_0, a \in \mathbb{N}$, such that for every $n > n_0$,

- $|\langle \mathrm{e}_p(P_n), g_n\rangle| \geq 1/a$;
- $\mathrm{rank}(P_n) \geq \widetilde{r}_b(a)$.

Set $\varepsilon := \frac{1}{1000a^2b}$. Since $\mathcal{D}$ is correlation testable with true complexity $d$ and Cauchy-Schwarz complexity $s < p$, there exist $\delta \in (0, \varepsilon)$, $\eta > 0$, and a family of homogenous systems of linear forms $\mathcal{L}_1, \dots, \mathcal{L}_\ell$ of true complexity at most $d$ and Cauchy-Schwarz complexity at most $s$, and conjugations $\alpha_1, \dots, \alpha_\ell$, such that

$$\|(t_{\mathcal{L}_1,\alpha_1}(f), \dots, t_{\mathcal{L}_\ell,\alpha_\ell}(f)) - (t_{\mathcal{L}_1,\alpha_1}(f'), \dots, t_{\mathcal{L}_\ell,\alpha_\ell}(f'))\|_\infty \geq 2\eta,$$

for every $f, f' : \mathbb{F}_p^n \to \mathbb{D}$ (with $n > n_0$) satisfying $\|f'\|_{u(D_n)} \geq \varepsilon$ and $\|f\|_{u(D_n)} \leq \delta$. We will prove the lemma by constructing a new sequence of functions $(f'_n : \mathbb{F}_p^n \to \mathbb{D})_{n\in\mathbb{N}}$ such that for large enough $n$, we will have

- $|\langle f'_n, g_n\rangle| \geq \varepsilon$ and hence $\|f'_n\|_{u(D_n)} \geq \varepsilon$;
- $|t_{\mathcal{L}_i,\alpha_i}(f_n) - t_{\mathcal{L}_i,\alpha_i}(f'_n)| \leq \eta$, for all $1 \leq i \leq k$.

This will conclude the lemma as it will show that $\|\mathrm{e}_p(f_n)\|_{u(D_n)} \geq \delta$ for large enough $n$.

Let $r_1 : \mathbb{N} \to \mathbb{N}$ be a growth function to be determined later (whose choice depends on the values of $a, b$). By Theorem C.6, there exists a polynomial factor $\mathcal{B}_n$ of degree $s$, complexity $C_n \leq C_{\max}(\eta, a, b, r_1(\cdot))$ and rank at least $r_1(C_n)$, such that for $h_n := \mathbb{E}(f_n|\mathcal{B}_n)$ we have

$$|t_{\mathcal{L}_i,\alpha_i}(f_n) - t_{\mathcal{L}_i,\alpha_i}(h_n)| \leq \eta/10, \tag{40}$$

for all $1 \leq i \leq k$, and also

$$\|f_n - h_n\|_{U^{s+1}} \leq 1/2b. \tag{41}$$

Recall that $Q_n$ is a polynomial of degree $t_0 \leq t \leq s$ such that $|\langle f_n, Q_n\rangle| \geq 1/b$, for $n \geq n_0$. By the choice of (41) we have that $h_n$ is also correlated to $Q_n$, as

$$|\langle h_n, \mathrm{e}_p(Q_n)\rangle| \geq |\langle f_n, \mathrm{e}_p(Q_n)\rangle| - |\langle f_n - h_n, \mathrm{e}_p(Q_n)\rangle| \geq \frac{1}{b} - \|f_n - h_n\|_{U^{s+1}} \geq 1/2b. \tag{42}$$

Assume $\mathcal{B}_n$ is defined by polynomials $Q'_{n,1}, \dots, Q'_{n,C_n}$. Define $Q'_{n,\gamma} = \sum_{i=1}^{C_n} \gamma(i) Q'_{n,i}$ for $\gamma \in \mathbb{F}_p^{C_n}$. The function $h_n = \mathbb{E}(f_n|\mathcal{B}_n)$ is $\mathcal{B}_n$-measurable, hence $h_n = F_n(Q'_{n,1}, \dots, Q'_{n,C_n})$ for some function $F_n : \mathbb{F}_p^{C_n} \to \mathbb{D}$, and we have

$$h_n(x) = \sum_{\gamma\in\mathbb{F}_p^{C_n}} \widehat{F_n}(\gamma)\mathrm{e}_p(Q'_{n,\gamma}(x))$$

where $\|\widehat{F_n}\|_\infty \leq 1$. Thus we have

$$1/2b \leq \left|\langle h_n, \mathrm{e}_p(Q_n)\rangle\right| \leq \sum_{\gamma\in\mathbb{F}_p^{C_n}} |\widehat{F_n}(\gamma)| \cdot |\mathrm{bias}(Q'_{n,\gamma} - Q_n)|. \tag{43}$$

We now show that when $r_1(\cdot)$ is chosen large enough (as a function of $a, b$), then almost all the contribution to the correlation in (43) comes from a single term $\gamma^*$. Let $r_1(C)$ be chosen large enough, such if $R$ is a polynomial on $\mathbb{F}_p^n$ of degree at most $s$ and rank at least $r_1(C)/2$, then for large enough $n$ we have

$$|\mathrm{bias}(R)| \leq \varepsilon \cdot p^{-C}. \tag{44}$$

Such a choice is guaranteed by Theorem B.11. Assuming such a choice for $r_1(\cdot)$, we have the following claim.

**Claim D.10.** *There exists $\gamma^* \in \mathbb{F}_p^{C_n}$ such that $|\widehat{F_n}(\gamma^*)| \geq 1/4b$ and $|\mathrm{bias}(Q'_{n,\gamma^*} - Q_n)| \geq 1/4b$. Moreover, for all $\gamma \neq \gamma^*$ we have $|\mathrm{bias}(Q'_{n,\gamma} - Q_n)| \leq \varepsilon \cdot p^{-C_n}$.*

*Proof.* Assume first that $\mathrm{rank}(Q'_{n,\gamma} - Q_n) \geq r_1(C_n)/2$ for all $\gamma \in \mathbb{F}_p^{C_n}$. This implies by (44) that

$$|\langle h_n, \mathrm{e}_p(Q_n)\rangle| \leq \sum_{\gamma \in \mathbb{F}_p^{C_n}} |\widehat{F_n}(\gamma)||\mathrm{bias}(Q'_{n,\gamma} - Q_n)| \leq \varepsilon \leq \frac{1}{10b},$$

which is a contradiction to (42). So there must exist at least one $\gamma$ such that $\mathrm{rank}(Q'_{n,\gamma} - Q_n) < r_1(C_n)/2$. Assume there were two distinct $\gamma' \neq \gamma''$ such that $\mathrm{rank}(Q'_{n,\gamma'} - Q_n) < r_1(C_n)/2$ and $\mathrm{rank}(Q'_{n,\gamma''} - Q_n) < r_1(C_n)/2$. Then by subadditivity of rank, this implies

$$\begin{aligned}\mathrm{rank}(Q'_{n,\gamma'-\gamma''}) &= \mathrm{rank}((Q'_{n,\gamma'} - Q_n) - (Q_{n,\gamma''} - Q_n)) \\ &\leq \mathrm{rank}(Q'_{n,\gamma'} - Q_n) + \mathrm{rank}(Q_{n,\gamma''} - Q_n) < r_1(C_n),\end{aligned}$$

which is a contradiction to the fact that $\{Q_{n,1}, \ldots, Q_{n,C_n}\}$ has rank at least $r_1(C_n)$. So, there exists a unique $\gamma^* \in \mathbb{F}_p^{C_n}$ such that $\mathrm{rank}(Q'_{n,\gamma^*} - Q_n) < r_1(C_n)/2$. Thus by (44) we have

$$\sum_{\gamma \neq \gamma^*} |\widehat{F_n}(\gamma)||\mathrm{bias}(Q'_{n,\gamma} - Q_n)| \leq \varepsilon \leq 1/10b.$$

However, by (42) we have $|\langle h_n, \mathrm{e}_p(Q_n)\rangle| \geq 1/2b$. Hence

$$|\widehat{F_n}(\gamma^*)||\mathrm{bias}(Q'_{n,\gamma^*} - Q_n)| \geq 1/4b.$$

The claim follows as $|\widehat{F_n}(\gamma^*)| \leq 1$ and $|\mathrm{bias}(Q'_{n,\gamma^*} - Q_n)| \leq 1$. □

The next claim shows that if we require $r_1(\cdot)$ and $r(\cdot)$ to be large enough, then we can guarantee that $\deg(Q'_{n,\gamma^*}) = \deg(Q_n)$.

**Claim D.11.** *If $r_1(\cdot), r(\cdot)$ are chosen large enough, then*

$$\deg(Q'_{n,\gamma^*}) = \deg(Q_n).$$

*Proof.* Recall that we have by assumption $\mathrm{rank}(Q_n) \geq r(b)$ and $\mathrm{rank}(Q'_{n,\gamma^*}) \geq r_1(C_n)$. We also have $|\mathrm{bias}(Q'_{n,\gamma^*} - Q_n)| \geq 1/4b$ by Claim D.10, which by Theorem B.11 implies that $\mathrm{rank}(Q'_{n,\gamma^*} - Q_n) \leq r'(b)$, for some function $r'(\cdot)$. We have two cases to consider:

- If $\deg(Q_n) < \deg(Q'_{n,\gamma^*})$, then we have $\mathrm{rank}(Q'_{n,\gamma}) \leq r'(b) + 1$. To avoid this case we choose $r_1(C) > r'(b) + 1$ for all $C \in \mathbb{N}$.
- If $\deg(Q_n) > \deg(Q'_{n,\gamma^*})$, then we have $\mathrm{rank}(Q_n) \leq r'(b) + 1$. To avoid this case we choose $r(b) > r'(b) + 1$ for all $b \in \mathbb{N}$.

If these two conditions are met, this guarantees that $\deg(Q'_{n,\gamma^*}) = \deg(Q_n)$ as claimed. □

We thus have $\deg(Q'_{n,\gamma^*}) = \deg(Q_n) = t_0$. We now claim that without loss of generality we can assume $\gamma^* = e_1 = (1, 0, \ldots, 0)$. This can be achieved by performing an invertible linear transformation on the basis $Q'_{n,1}, \ldots, Q'_{n,C_n}$ which preserves the polynomial factor defined by it as well as its rank. Summarizing Claim D.10 and Claim D.11 and the discussion above, we have established the following properties:

1. $|\widehat{F_n}(e_1)| \geq \frac{1}{4b}$;
2. $|\mathrm{bias}(Q'_{n,1} - Q_n)| \geq \frac{1}{4b}$ and $|\mathrm{bias}(Q'_{n,\gamma} - Q_n)| \leq \varepsilon \cdot p^{-C_n}$ for all $\gamma \neq e_1$;
3. $\deg(Q'_{n,1}) = \deg(Q_n) = t_0$;
4. $\mathrm{rank}(Q'_{n,1}) \geq r_1(C_n)$ and $\mathrm{rank}(Q_n) \geq r(b)$.

We now repeat the same process for $g_n$. Let $\lambda = \varepsilon p^{-C_{\max}}$. There exists a polynomial factor $\mathcal{B}'_n$ of degree $s$ and complexity $C'_n \leq C'_{max}$ defined by polynomials $\{P'_{n,i} : 1 \leq i \leq C'_n\}$, such that for $g'_n := \mathbb{E}(g_n|\mathcal{B}'_n)$ we have

$$\|g_n - g'_n\|_{U^{s+1}} \leq \lambda,$$

and also that $g'_n = G_n(P'_{n,1}, \ldots, P'_{n,C'_n})$ where

1. $|\widehat{G_n}(e_1)| \geq \frac{1}{4a}$;
2. $|\mathrm{bias}(P'_{n,1} - P_n)| \geq \frac{1}{4a}$ and $|\mathrm{bias}(P'_{n,\gamma} - P_n)| \leq \varepsilon \cdot p^{-C'_n}$ for all $\gamma \neq e_1$;
3. $\deg(P'_{n,1}) = \deg(P_n) = t_0$;
4. $\mathrm{rank}(P'_{n,1}) \geq r_1(C'_n)$ and $\mathrm{rank}(P_n) \geq \widetilde{r}_b(a)$.

We now define a new polynomial factor $\widetilde{\mathcal{B}}_n$ as follows. Let $R_{n,2}, \ldots, R_{n,C_n}$ be random polynomials in $\mathbb{F}_p^n$ chosen such that $\deg(R_{n,i}) = \deg(P'_{n,i})$. We note that for every rank bound $r^*$, for large enough $n$ the following holds with high probability. Let $P^1$ be some (possibly zero) linear combination of $P_n, P'_{n,1}, \ldots, P'_{n,C'_n}$ and let $R^1$ be a nonzero linear combination of $R_{n,2}, \ldots, R_{n,C_n}$. Then as long as $\deg(R^1) \geq \deg(P^1)$ we have that $\mathrm{rank}(P^1 + R^1) \geq r^*$.

We define $\widetilde{\mathcal{B}}_n = \{P_n, R_{n,2}, \ldots, R_{n,C_n}\}$. Define $R_{n,\gamma} = \gamma(1)P_n(x) + \sum_{i=2}^{C_n} \gamma(i)R_{n,i}(x)$ for $\gamma \in \mathbb{F}_p^{C_n}$. We define the new sequence of functions as

$$f'_n(x) = F_n(P_n, R_{n,2}, \ldots, R_{n,C_n}). \tag{45}$$

We conclude the proof by showing that $|t_{\mathcal{L}_i,\alpha_i}(f'_n) - t_{\mathcal{L}_i,\alpha_i}(f_n)| \leq \eta$ for all $1 \leq i \leq \ell$, but that $|\langle f'_n, g_n\rangle| \geq \varepsilon$.

**Claim D.12.** *$|t_{\mathcal{L}_i,\alpha_i}(f'_n) - t_{\mathcal{L}_i,\alpha_i}(f_n)| \leq \eta$ for all $1 \leq i \leq \ell$.*

*Proof.* Note that $\deg(R_{n,i}) = \deg(P_{n,i})$ for all $1 \leq i \leq \ell$. Since all the linear forms are homogenous, we can apply Proposition C.5. By the proposition, there exists a bound $r'(\eta)$ such that if $\mathrm{rank}(\mathcal{B}_n), \mathrm{rank}(\widetilde{\mathcal{B}}_n) \geq r'$ then the claim follows. To ensure this, we require for $\mathcal{B}_n$ that $r_1(C) \geq r'$ for all $C \in \mathbb{N}$; and for $\widetilde{\mathcal{B}}_n$ that $r^* \geq r'$ and that $\widetilde{r}_b(a) \geq r'$ (note that it is crucial to allow $\widetilde{r}$ to depend on both $a, b$, since $\eta$ depends on both $a, b$). □

**Claim D.13.** *$|\langle f'_n, g_n\rangle| \geq \varepsilon$.*

*Proof.* We first claim, it suffices to prove $|\langle f'_n, g'_n\rangle| \geq 2\epsilon$. Indeed

$$|\langle f'_n, g_n - g'_n\rangle| \leq \sum_{\gamma \in \mathbb{F}_p^{C_n}} |\widehat{F_n}(\gamma)||\langle \mathrm{e}_p(R_{n,\gamma}), g_n - g'_n\rangle| \leq p^{C_n}\|g_n - g'_n\|_{U^{s+1}} \leq \varepsilon,$$

where we used the fact that $\deg(R'_{n,\gamma}) \leq s$. Now note that

$$\langle f'_n, g'_n \rangle = \sum_{\gamma \in \mathbb{F}_p^{C_n}, \gamma' \in \mathbb{F}_p^{C'_n}} \widehat{F_n}(\gamma)\overline{\widehat{G_n}(\gamma')}\mathrm{bias}(R_{n,\gamma} - P'_{n,\gamma'}).$$

Consider first the term $\gamma = \gamma' = e_1$. We have $R_{n,e_1} = P_n$ and by our construction,

$$\left|\widehat{F_n}(e_1)\overline{\widehat{G_n}(e_1)}\mathrm{bias}(P_n - P'_{n,1})\right| \geq \frac{1}{4b} \cdot \frac{1}{4a} \cdot \frac{1}{4a} \geq 3\varepsilon.$$

We now bound all other terms $(\gamma, \gamma') \neq (e_1, e_1)$. By choosing $r_1(\cdot), r^*$ large enough, we can bound

$$|\mathrm{bias}(R_{n,\gamma} - P'_{n,\gamma'})| \leq \varepsilon \cdot p^{-(C_n + C'_n)}.$$

Putting all these together we conclude that $|\langle f'_n, g_n \rangle| \geq \varepsilon$ as claimed. □

This concludes the proof of Lemma D.9. □

Lemma D.9 shows that if $t_0 \in \mathcal{S}$, then we are done. We will next show that $\mathcal{S} = \{1, \ldots, t\}$ which will conclude the proof of Lemma D.6. Consider any $t' < t$ and systems of linear forms $\mathcal{L}_1, \ldots, \mathcal{L}_\ell$. For a function $f : \mathbb{F}_p^n \to \mathbb{R}$ we shorthand $t_{\mathcal{L}}(f) = t_{\mathcal{L},\alpha}(f)$ for all conjugates $\alpha$ since $f = \overline{f}$. In Lemma D.14 below we shall show that it is possible to construct two different families of functions $f_n, f'_n : \mathbb{F}_p^n \to [-1, 1]$ such that

(i) $f_n$ and $f'_n$ cannot be distinguished by averages $t_{\mathcal{L}_i}$;

(ii) $f_n$ has correlation with polynomials of degree exactly $t$ and of arbitrarily high rank;

(iii) $f'_n$ is a linear combination of a bounded number of exponentials of polynomials of degree exactly $t'$ and arbitrarily high rank.

We note that the combination of (i), (ii), (iii) implies that $t' \in \mathcal{S}$: Condition (ii) implies by Lemma D.9 that $\limsup_{n \to \infty} \|f_n\|_{u(D_n)} > 0$. By Condition (i) this implies that also $\limsup_{n \to \infty} \|f'_n\|_{u(D_n)} > 0$; so there exist $g_n \in D_n$ such that $\limsup_{n \to \infty} |\langle f'_n, g_n \rangle| = \delta > 0$. By Condition (iii) we can express $f'_n(x) = \sum_{i=1}^{C} a_i \mathrm{e}_p(Q_{n,i}(x))$ where $C$ is a uniform bound, $|a_i| \leq 1$ and $Q_{n,i}$ are polynomials of degree $t'$ and arbitrarily high rank. Hence, we must have for infinitely many $n$ that $|\langle \mathrm{e}_p(Q_{n,i}), g_n \rangle| \geq \delta p^{-C}$ for some $i$. Since $\mathcal{D}$ is consistent, we can extend this to all large enough $n$ and complete the proof.

It only remains to prove the following lemma.

**Lemma D.14.** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_\ell$ be systems of linear forms. Let $t' < t$. There exists functions $f_n, f'_n : \mathbb{F}_p^n \to [-1, 1]$ and a constant $C \in \mathbb{N}$ such that the following holds:*

*(i) For every $i \in [\ell]$,*

$$t_{\mathcal{L}_i}(f_n) = t_{\mathcal{L}_i}(f'_n).$$

*(ii) There exist polynomials $P_n : \mathbb{F}_p^n \to \mathbb{F}_p$ satisfying $\deg(P_n) = t$ and $\lim_{n \to \infty} \mathrm{rank}(P_n) = \infty$, and*

$$\liminf_{n \to \infty} |\langle \mathrm{e}_p(P_n), f_n \rangle| > 0.$$

*(iii)* $f'_n$ *is a linear combination of exponentials of* $C$ *high rank polynomials of degree exactly* $t'$*. That is, there exist polynomials* $Q_{n,1}, \dots, Q_{n,C} : \mathbb{F}_p^n \to \mathbb{F}_p$ *satisfying* $\deg(Q_{n,i}) = t'$ *and* $\lim_{n\to\infty} \operatorname{rank}(Q_{n,i}) = \infty$ *for all* $1 \le i \le C$*, and*

$$f'_n(x) = \sum_{i=1}^{C} a_i \mathrm{e}_p(Q_{n,i}(x)),$$

*where* $|a_i| \le 1$*.*

*Proof.* We first note that it is sufficient to prove the lemma for systems of linear forms which are non-isomorphic and connected, since we can decompose each system to its connected components and remove isomorphic copies.

First let us introduce some notations. For positive integers $m > n$, let $\pi_{m\to n}$ denote the natural projection from $\mathbb{F}_p^m$ to $\mathbb{F}_p^n$ defined as

$$\pi_{m\to n} : (x_1, \dots, x_m) \mapsto (x_1, \dots, x_n).$$

For functions $f : \mathbb{F}_p^m \to \mathbb{C}$ and $g : \mathbb{F}_p^n \to \mathbb{C}$, let $f \otimes g : \mathbb{F}^{m+n} \to \mathbb{C}$ denote the function

$$(f \otimes g)(x_1, \dots, x_{m+n}) = f(x_1, \dots, x_m) g(x_{m+1}, \dots, x_{m+n}),$$

and note that for every system of linear forms $\mathcal{L}$, we have

$$t_{\mathcal{L}}(f \otimes g) = t_{\mathcal{L}}(f) t_{\mathcal{L}}(g).$$

By Theorem F.1, there exist a constant $N \in \mathbb{N}$, an $\varepsilon > 0$, and a function $F : \mathbb{F}_p^N \to [0,1]$ such that

$$\left\{ z \in \mathbb{R}^\ell \mid \|z - (t_{\mathcal{L}_1}(F), \dots, t_{\mathcal{L}_\ell}(F))\|_\infty \le \varepsilon \right\} \subseteq \left\{ (t_{\mathcal{L}_1}(f), \dots, t_{\mathcal{L}_\ell}(f)) \mid f : \mathbb{F}_p^N \to [0,1] \right\}. \tag{46}$$

Consider two sequences of polynomials $P_n, Q_n : \mathbb{F}_p^n \to \mathbb{F}_p$ such that $\deg(P_n) = t$ and $\deg(Q_n) = t'$, and

$$\lim_{n\to\infty} \operatorname{rank}(P_n) = \lim_{n\to\infty} \operatorname{rank}(Q_n) = \infty.$$

Define $g_n : \mathbb{F}_p^n \to \{-1/p, 1 - 1/p\}$ as $g_n(x) = 1 - 1/p$ if and only if $Q_n(x) = 0$. Note that

$$g_n(x) = \sum_{\alpha \in \mathbb{F}_p \setminus \{0\}} \frac{1}{p} \mathrm{e}_p(\alpha Q_n(x)).$$

Define $h_n : \mathbb{F}_p^n \to \{0, 1\}$ as $h_n(x) = 1$ if and only if $P_n(x) = 0$.

Let $\delta > 0$ be sufficiently small so that for every $m > N$, and every $i \in [k]$,

$$|t_{\mathcal{L}_i}(F) - t_{\mathcal{L}_i}(\delta h_m + (1-\delta) F \circ \pi_{m\to N})| \le \varepsilon/2.$$

Then by (46), for every $m > N$, there exists a function $G_m : \mathbb{F}_p^N \to [0,1]$ such that for every $i \in [k]$,

$$t_{\mathcal{L}_i}(G_m) = t_{\mathcal{L}_i}(\delta h_m + (1-\delta) F \circ \pi_{m\to N}). \tag{47}$$

For $n > 2N$, let $m = \lfloor n/2 \rfloor$ and define $f_n, f_n' : \mathbb{F}_p^n \to [-1, 1]$ as

$$\begin{aligned} f_n &:= g_{n-m} \otimes (\delta h_m + (1-\delta) F \circ \pi_{m \to N}), \\ f_n' &:= g_{n-m} \otimes (G_m \circ \pi_{m \to N}). \end{aligned}$$

By (47) for every $i \in [k]$ and every $n > 2N$,

$$t_{\mathcal{L}_i}(f_n) = t_{\mathcal{L}_i}(g_{n-m}) t_{\mathcal{L}_i}(\delta h_m + (1-\delta) F \circ \pi_{m \to N}) = t_{\mathcal{L}_i}(g_{n-m}) t_{\mathcal{L}_i}(G_m \circ \pi_{m \to N}) = t_{\mathcal{L}_i}(f_n'),$$

which establishes (i). To establish (ii), let $R_n := Q_{n-m} \otimes P_m$. Note that $R_n$ is a polynomial of degree $t$ and $\lim_{n \to \infty} \operatorname{rank}(R_n) = \infty$. We have

$$\langle f_n, \mathrm{e}_p(R_n) \rangle = \langle g_{n-m}, \mathrm{e}_p(Q_{n-m}) \rangle \cdot \langle \delta h_m + (1-\delta) F \otimes \pi_{m \to N}, \mathrm{e}_p(P_m) \rangle.$$

We now lower bound the terms. By the definition of $g_{n-m}$, we have

$$|\langle g_{n-m}, \mathrm{e}_p(Q_{n-m}) \rangle| = \left| \frac{1}{p} + \frac{1}{p} \sum_{\alpha \in \mathbb{F}_p \setminus \{0,1\}} \operatorname{bias}((\alpha - 1) Q_{n-m}) \right| \geq \frac{1}{2p},$$

for large enough $n-m$ since $\operatorname{rank}(Q_{n-m}) \to \infty$. The function $F \otimes \pi_{m \to N}$ depends only on the first $N$ variables; hence we have $\lim_{m \to \infty} \langle F \otimes \pi_{m \to N}, \mathrm{e}_p(P_m) \rangle = 0$ since $\operatorname{rank}(P_m) \to \infty$ by Theorem B.11. Finally, by the definition of $h_m$ we have

$$\langle h_m, \mathrm{e}_p(P_m) \rangle = \Pr_X[P_m(X) = 0] \geq \frac{1}{2p}$$

for large enough $m$, since $\operatorname{rank}(P_m) \to \infty$. We thus conclude that

$$|\langle f_n, \mathrm{e}_p(R_n) \rangle| \geq \frac{1}{4p^2}$$

for large enough $n$, which establishes (ii). To conclude the proof we establish (iii). Let

$$G_m(x) = \sum_{\gamma \in \mathbb{F}_p^N} \widehat{G}(\gamma) \mathrm{e}_p \left( \sum_{i=1}^N \gamma(i) x(i) \right)$$

where $|\widehat{G}(\gamma)| \leq 1$. We thus have

$$\begin{aligned} f_n'(x) &= g_{n-m}(x(1), \ldots, x(n-m)) G_m(x(n-m+1), \ldots, x(n-m+N)) \\ &= \sum_{\gamma \in \mathbb{F}_p^N, \alpha \in \mathbb{F}_p \setminus \{0\}} \frac{1}{p} \widehat{G}(\gamma) \mathrm{e}_p \left( \alpha Q_{n-m}(x(1), \ldots, x(n-m)) + \sum_{i=1}^N \gamma(i) x(n-m+i) \right). \end{aligned}$$

Hence, we can express $f_n'$ as the linear combination of $C = (p-1)p^N$ exponentials of polynomials of degree exactly $t'$; and as $N$ is fixed and $n - m \to \infty$, their rank is unbounded. This established (iii) and concludes the proof. □

# E Characterization of correlation testable properties

Consider a family $\mathcal{D} := \{D_n\}_{n\in\mathbb{N}}$ where $D_n$ is a set of functions from $\mathbb{F}_p^n$ to $\mathbb{F}_p$. Given a function $\mathrm{f}_n : \mathbb{F}_p^n \to \mathbb{F}_p$, we want to probabilistically determine whether $\mathrm{f}_n$ has correlation with $D_n$. That is, we want to determine if $\|\mathrm{e}_p(\mathrm{f}_n)\|_{u(D_n)}$ is non-negligible or not, where we are only allowed to read the value of $\mathrm{f}_n$ on a few points. We recall the definition of correlation testable properties.

**Definition E.1** (Correlation testable properties)**.** *A family $\mathcal{D} = (D_n)$ is correlation testable with $q$ queries, if there exists a distribution $\mu$ taking values in $(\mathbb{F}_p^n)^q$ and a mapping $\Gamma : \mathbb{F}_p^q \to \{0,1\}$, such that the following holds. For every $\varepsilon > 0$, there exist $\delta \in (0,\varepsilon)$, $0 \le \theta^- < \theta^+ \le 1$ and $n_0 \in \mathbb{N}$, such that for every $n > n_0$ and $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$ we have:*

- *If $\|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \ge \varepsilon$ then $\Pr_{(X_1,\dots,X_q)\sim\mu}[\Gamma(\mathrm{f}(X_1),\dots,\mathrm{f}(X_q)) = 1] \ge \theta^+$.*
- *If $\|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \le \delta$ then $\Pr_{(X_1,\dots,X_q)\sim\mu}[\Gamma(\mathrm{f}(X_1),\dots,\mathrm{f}(X_q)) = 1] \le \theta^-$.*

We study proper dual families $\mathcal{D}$. We recall that a family $\mathcal{D}$ is *proper dual* if the following conditions hold:

- **A1: Consistency** For positive integers $m > n$ and $\mathrm{g} \in D_n$, the function $\mathrm{h} : \mathbb{F}_p^m \to \mathbb{F}_p$ defined as $\mathrm{h}(x_1,\dots,x_m) = \mathrm{g}(x_1,\dots,x_n)$ belongs to $D_m$.
- **A2: Affine invariance** For every positive integer $n$, if $\mathrm{g} \in D_n$, then for every $A \in \mathrm{Aff}(n,\mathbb{F}_p)$, we have $A\mathrm{g} \in D_n$.
- **A3: Sparsity** For every $\varepsilon > 0$ and large enough $n$, we have $|D_n| \le p^{\varepsilon p^n}$.

## E.1 Correlation testing by averages over linear forms

We first show that if $\mathcal{D}$ is a proper dual family which is correlation testable using $q$ queries, then it is in fact also testable using averages of linear forms. When arguing about functions to $\mathbb{F}_p$, one may allow more general types of averages. Let $\mathcal{L} = \{L_1,\dots,L_m\}$ be a system of linear forms in $k$ variables. Let $\beta \in \mathbb{F}_p^m$ be a vector of coefficients. Recall that for a function $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$, we define the average

$$t^*_{\mathcal{L},\beta}(\mathrm{f}) = \mathbb{E}_{\mathbf{X}\in(\mathbb{F}_p^n)^k}\left[\mathrm{e}_p\left(\sum_{i=1}^m \beta(i)\mathrm{f}(L_i(\mathbf{X}))\right)\right].$$

We note that for functions $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$, these averages generalize the previous averages $t_{\mathcal{L},\alpha}$ which were defined for bounded functions. Indeed, for $\alpha \in \{0,1\}^m$ let $\beta \in \{-1,1\}^m$ be defined as $\beta(i) = (-1)^{\alpha(i)}$, then

$$t^*_{\mathcal{L},\beta}(\mathrm{f}) = t_{\mathcal{L},\alpha}(\mathrm{e}_p(\mathrm{f})).$$

**Lemma E.2.** *Suppose that a proper dual family $\mathcal{D} = \{D_n\}$ is correlation testable with $q$ queries. Then for every $\varepsilon > 0$, there exists $\delta \in (0,\varepsilon)$, $n_0 \in \mathbb{N}$, and homogeneous systems of linear forms $\mathcal{L}_1,\dots,\mathcal{L}_\ell$ with $m_1,\dots,m_\ell$ linear forms, accordingly, and corresponding coefficients $\beta_1 \in \mathbb{F}_p^{m_1},\dots,\beta_\ell \in \mathbb{F}_p^{m_\ell}$, such that the closures of the following two sets are disjoint:*

$$T_\varepsilon := \{(t^*_{\mathcal{L}_1,\beta_1}(\mathrm{f}),\dots,t^*_{\mathcal{L}_\ell,\beta_\ell}(\mathrm{f})) | n > n_0, \mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p, \|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \ge \varepsilon\},$$

*and*

$$S_\varepsilon := \{(t^*_{\mathcal{L}_1,\beta_1}(\mathrm{f}),\dots,t^*_{\mathcal{L}_\ell,\beta_\ell}(\mathrm{f})) | n > n_0, \mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p, \|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \le \delta\}.$$

*Moreover, the systems $\mathcal{L}_1,\dots,\mathcal{L}_\ell$ have Cauchy-Schwarz complexity at most $q-2$.*

*Proof.* Since $D_n$ is a proper dual, by Condition **A2**, for every $\mathrm{f}_n : \mathbb{F}_p^n \to \mathbb{F}_p$ and every $A \in \mathrm{Aff}(n, \mathbb{F}_p)$, we have that $\|\mathrm{e}_p(A\mathrm{f}_n)\|_{u(D_n)} = \|\mathrm{e}_p(\mathrm{f}_n)\|_{u(D_n)}$. Let $A \in \mathrm{Aff}(n, \mathbb{F}_p)$ be a uniform random invertible affine transformation. Then by the assumption that $\mathcal{D}$ is correlation testable, we have that

- If $\|\mathrm{e}_p(\mathrm{f}_n)\|_{u(D_n)} \geq \varepsilon$, then $\Pr_{(X_1,\ldots,X_q)\sim\mu, A\in\mathrm{Aff}(n,\mathbb{F}_p)}[\Gamma(\mathrm{f}_n(AX_1), \ldots, \mathrm{f}_n(AX_q)) = 1] \geq \theta^+$.
- If $\|\mathrm{e}_p(\mathrm{f}_n)\|_{u(D_n)} \leq \delta$, then $\Pr_{(X_1,\ldots,X_q)\sim\mu,, A\in\mathrm{Aff}(n,\mathbb{F}_p)}[\Gamma(\mathrm{f}_n(AX_1), \ldots, \mathrm{f}_n(AX_q)) = 1] \leq \theta^-$.

We establish the lemma by showing that if we set $n_0$ large enough, then the probability

$$\Pr_{(X_1,\ldots,X_q)\sim\mu, A\in\mathrm{Aff}(n,\mathbb{F}_p)}[\Gamma(\mathrm{f}(AX_1), \ldots, \mathrm{f}(AX_q)) = 1]$$

can be approximated with an arbitrarily small error by linear combinations of $t^*_{\mathcal{L}_1,\beta_1}(\mathrm{f}), \ldots, t^*_{\mathcal{L}_\ell,\beta_\ell}(\mathrm{f})$, where $\{(\mathcal{L}_i, \beta_i)\}_{1\leq i\leq \ell}$ are all possible homogeneous systems of at most $q$ linear forms. Note that $\ell$ is a constant depending only on $p, q$, and that the Cauchy-Schwarz complexity of any homogeneous system of $t \leq q$ linear forms is at most $t - 2 \leq q - 2$. We start by decomposing $\Gamma$ to its Fourier decomposition

$$\Gamma(z(1), \ldots, z(q)) = \sum_{\gamma\in\mathbb{F}_p^q} \widehat{\Gamma}(\gamma)\mathrm{e}_p\left(\sum_{i=1}^{q} \gamma(i) \cdot z(i)\right).$$

We thus have that

$$\Pr_{(X_1,\ldots,X_q)\sim\mu, A\sim\mathrm{Aff}(n,\mathbb{F}_p)}[\Gamma(\mathrm{f}(AX_1), \ldots, \mathrm{f}(AX_q)) = 1] = \sum_{\gamma\in\mathbb{F}_p^q} \widehat{\Gamma}(\gamma)\mathbb{E}\left[\mathrm{e}_p\left(\sum_{i=1}^{q} \gamma(i) \cdot \mathrm{f}(AX_i)\right)\right],$$

where the expectation is taken over $(X_1, \ldots, X_q) \sim \mu$ and $A \sim \mathrm{Aff}(n, \mathbb{F}_p)$. Thus, it is enough to show that each term $\mathbb{E}\left[\mathrm{e}_p(\sum_{i=1}^q \beta(i) \cdot \mathrm{f}(AX_i))\right]$ can be approximated by linear combinations of $\{t^*_{\mathcal{L}_i,\beta_i}(\mathrm{f})\}_{1\leq i\leq k}$.

Fix $(x_1, \ldots, x_q) \in (\mathbb{F}_p^n)^q$. Suppose that the rank of $\mathrm{span}\{x_1, \ldots, x_q\}$ over $\mathbb{F}_p$ is $r$. Let $y_1, \ldots, y_r \in \mathbb{F}_p^n$ form a basis for $\mathrm{span}\{x_1, \ldots, x_q\}$, so that $x_i = \sum_{j=1}^r \lambda_{i,j} y_j$, for every $1 \leq i \leq q$. Then the distribution of $(Ax_1, \ldots, Ax_q)$ is the same as the distribution of $(Y_0 + \sum_{j=1}^r \lambda_{1,j} Y_j, \ldots, Y_0 + \sum_{j=1}^r \lambda_{q,j} Y_j)$, where $Y_0, Y_1, \ldots, Y_r$ are i.i.d. random variables taking values in $\mathbb{F}_p^n$ uniformly at random conditioned on $Y_1, \ldots, Y_\ell$ being linearly independent. However since if we pick $Y_1, \ldots, Y_r$ independently and uniformly at random, with probability $1 - o_{n\to\infty}(1)$ they will be linearly independent, by taking $n$ to be sufficiently large this distribution can be made arbitrarily close to the distribution of $(Y_0 + \sum_{j=1}^r \lambda_{1,j} Y_j, \ldots, Y_0 + \sum_{j=1}^r \lambda_{q,j} Y_j)$, where $Y_0, \ldots, Y_r$ are i.i.d. random variables taking values in $\mathbb{F}_p^n$ uniformly at random.

Thus, we can approximate each term $\mathbb{E}\left[\mathrm{e}_p(\sum_{i=1}^q \beta(i) \cdot \mathrm{f}(Ax_i))\right]$ by $\mathbb{E}\left[\mathrm{e}_p(\sum_{i=1}^q \beta(i) \cdot \mathrm{f}(Y_0 + \sum_{j=1}^r \lambda_{i,j} Y_j))\right]$, which is one of the averages $t^*_{\mathcal{L}_i,\beta_i}(\mathrm{f})$. We now conclude the proof, since $\mathbb{E}\left[\mathrm{e}_p(\sum_{i=1}^q \beta(i) \cdot \mathrm{f}(AX_i))\right]$ where $(X_1, \ldots, X_q) \sim \mu$ can be approximated by an appropriate weighted average of $t^*_{\mathcal{L}_i,\beta_1}(\mathrm{f}), \ldots, t^*_{\mathcal{L}_\ell,\beta_\ell}(\mathrm{f})$. □

## E.2 From field functions to distributional functions

The next step is to move from functions $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$ to functions whose output lies in some convex set. Once this is accomplished, we can use the same techniques used for studying functions $f : \mathbb{F}_p^n \to \mathbb{D}$, there were used in Section D.

Let $P(\mathbb{F}_p)$ denote the family of probability measures over $\mathbb{F}_p$. That is, $P(\mathbb{F}_p) \subset \mathbb{R}^p$ is given by

$$P(\mathbb{F}_p) = \{\mu : \mathbb{F}_p \to [0,1] : \sum_{c \in \mathbb{F}_p} \mu(c) = 1\}.$$

We identify every element $c \in \mathbb{F}_p$ with its corresponding dirac measure on $\mathbb{F}_p$. That is $c \in \mathbb{F}_p$ is corresponded with the probability measure $\mu_c$ where $\mu_c(c) = 1$ and $\mu_c(c') = 0$ for all $c' \neq c$. We refer to functions $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ as *distributional functions.* Note that they are a superfamily of functions from $\mathbb{F}_p^n$ to $\mathbb{F}_p$, which can be regarded as deterministic functions. Given a distributional function $\Gamma$, we identify it with a distribution over functions from $\mathbb{F}_p^n$ to $\mathbb{F}_p$: If $\mathrm{F} \sim \Gamma$, the value $\mathrm{F}(x)$ is independently chosen for every $x \in \mathbb{F}_p^n$ according to the distribution $\Gamma(x)$.

We extend the notion of averages $t^*_{\mathcal{L},\beta}$ to distributional functions $\Gamma$. For $c \in \mathbb{F}_p$ define the function $\mathfrak{a}_c : P(\mathbb{F}_p) \to \mathbb{D}$ to be

$$\mathfrak{a}_c(\mu) = \mathbb{E}_{z \sim \mu}[\mathrm{e}_p(c \cdot z)].$$

For a distributional function $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ we consider the functions $\mathfrak{a}_c \circ \Gamma : \mathbb{F}_p^n \to \mathbb{D}$, which can equivalently be defined as

$$(\mathfrak{a}_c \circ \Gamma)(x) = \mathbb{E}_{\mathrm{F} \sim \Gamma}[\mathrm{e}_p(c \cdot \mathrm{F}(x))].$$

Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of $m$ linear forms in $k$ variables, and let $\beta \in \mathbb{F}_p^m$. We define

$$t^*_{\mathcal{L},\beta}(\Gamma) = \mathbb{E}_{\mathbf{X} \in (\mathbb{F}_p^n)^k} \left[ \prod_{i=1}^{m} (\mathfrak{a}_{\beta(i)} \circ \Gamma)(L_i(\mathbf{X})) \right].$$

Note that for functions $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$ this definition identifies with our previous definition.

**Claim E.3.** *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms in $k$ variables. Let $\beta \in \mathbb{F}_p^m$ be a vector of corresponding coefficients. Then for every distributional function $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$, and every $\epsilon > 0$, we have*

$$\Pr_{\mathrm{F} \sim \Gamma}\left[ \left| t^*_{\mathcal{L},\beta}(\mathrm{F}) - t^*_{\mathcal{L},\beta}(\Gamma) \right| \le \epsilon \right] = 1 - o_n(1).$$

*Proof.* The proof follows by a first and second moment estimation, and then applying Chebyshev's inequality. Let $\mathrm{F} \sim \Gamma$. Fix $\mathbf{x} \in (\mathbb{F}_p^n)^k$, and consider the random variable

$$A(\mathbf{x}) = \mathrm{e}_p\left( \sum_{i=1}^{m} \beta(i) \mathrm{F}(L_i(\mathbf{x})) \right).$$

We have $t^*_{\mathcal{L},\beta}(\mathrm{F}) = \frac{1}{p^{nk}} \sum_{\mathbf{x} \in (\mathbb{F}_p^n)^k} A(\mathbf{x})$. Note that when $L_1(\mathbf{x}), \ldots, L_m(\mathbf{x})$ are all distinct, we have

$$\mathbb{E}_{\mathrm{F} \sim \Gamma}[A(\mathbf{x})] = \prod_{i=1}^{n} (\mathfrak{a}_{\beta(i)} \circ \Gamma)(L_i(\mathbf{x})).$$

Thus, we get that

$$\left| \mathbb{E}_{\mathrm{F} \sim \Gamma}[t^*_{\mathcal{L},\beta}(\mathrm{F})] - t^*_{\mathcal{L},\beta}(\Gamma) \right| \le \Pr_{\mathbf{X} \in (\mathbb{F}_p^n)^k}[L_1(\mathbf{X}), \ldots, L_m(\mathbf{X}) \text{ not all distinct}] \le m^2 p^{-n},$$

where the second inequality follows by the union bound. We now bound the variance of $t^*_{\mathcal{L},\beta}(\mathrm{F})$. Note that two random variables $A(\mathbf{x}'), A(\mathbf{x}'')$ are independent if $\{L_1(\mathbf{x}'), \ldots, L_m(\mathbf{x}')\}$ and $\{L_1(\mathbf{x}''), \ldots, L_m(\mathbf{x}'')\}$ are disjoint. We thus can bound

$$\mathrm{Var}_{\mathrm{F}\sim\Gamma}[t^*_{\mathcal{L},\beta}(\mathrm{F})] \leq \Pr_{\mathbf{x}',\mathbf{x}''\in(\mathbb{F}_p^n)^k}[\{L_1(\mathbf{x}'), \ldots, L_m(\mathbf{x}')\} \cap \{L_1(\mathbf{x}''), \ldots, L_m(\mathbf{x}'')\} \neq \emptyset] \leq m^2 p^{-n},$$

where the second inequality follows from the union bound. The claim follows for Chebychev's bound. □

We extend also the notion of correlation to distributional functions. We shorthand $\mathrm{e}_p(\Gamma) := \mathfrak{a}_1 \circ \Gamma$, and consider

$$\|\mathrm{e}_p(\Gamma)\|_{u(D_n)} = \sup_{\mathrm{g}\in D_n} |\langle \mathrm{e}_p(\Gamma), \mathrm{e}_p(\mathrm{g})\rangle|.$$

**Claim E.4.** *Let $\mathcal{D} = \{D_n\}_{n\in\mathbb{N}}$ be a proper dual family. Then for every distributional function $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ and any $\varepsilon > 0$ we have*

$$\Pr_{\mathrm{F}\sim\Gamma}\left[\left|\|\mathrm{e}_p(\mathrm{F})\|_{u(D_n)} - \|\mathrm{e}_p(\Gamma)\|_{u(D_n)}\right| \geq \varepsilon\right] = o_n(1).$$

*Proof.* Fix $\mathrm{g} \in D_n$, and consider the random variable $\langle \mathrm{e}_p(\mathrm{F}), \mathrm{e}_p(\mathrm{g})\rangle = \frac{1}{p^n}\sum_{x\in\mathbb{F}_p^n} \mathrm{e}_p(\mathrm{F}(x) - \mathrm{g}(x))$. Its expected value is $\langle \mathrm{e}_p(\Gamma), \mathrm{e}_p(\mathrm{g})\rangle$, and since the values $\{\mathrm{F}(x) : x \in \mathbb{F}_p^n\}$ are chosen independently, we can apply Chernoff's bound and get that

$$\Pr_{\mathrm{F}\sim\Gamma}\left[|\langle \mathrm{e}_p(\mathrm{F}), \mathrm{e}_p(\mathrm{g})\rangle - \langle \mathrm{e}_p(\Gamma), \mathrm{e}_p(\mathrm{g})\rangle| \geq \varepsilon\right] \leq 2e^{-c\cdot p^n}$$

for some constant $c = c(\varepsilon) > 0$. By the sparsity Condition **A3** we get that for every $c' > 0$ there exists $n_0$, such that for every $n > n_0$ we have $|D_n| \leq p^{c' p^n}$. We conclude the proof by choosing $c'$ small enough such that $p^{c'} < e^c$, and apply the union bound over all $\mathrm{g} \in D_n$. □

We thus obtain the following lemma, which allows us to consider distributional functions instead of field functions.

**Lemma E.5.** *Suppose that a proper dual family $\mathcal{D} = \{D_n\}$ is correlation testable with $q$ queries. Then for every $\varepsilon > 0$, there exists $\delta \in (0, \varepsilon)$, $n_0 \in \mathbb{N}$, and homogeneous systems of linear forms $\mathcal{L}_1, \ldots, \mathcal{L}_\ell$ with $m_1, \ldots, m_\ell$ linear forms, accordingly, and corresponding coefficients $\beta_1 \in \mathbb{F}_p^{m_1}, \ldots, \beta_\ell \in \mathbb{F}_p^{m_\ell}$, such that the closures of the following two sets are disjoint:*

$$T_\varepsilon := \{(t^*_{\mathcal{L}_1,\beta_1}(\Gamma), \ldots, t^*_{\mathcal{L}_\ell,\beta_\ell}(\Gamma)) | n > n_0, \Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p), \|\mathrm{e}_p(\Gamma)\|_{u(D_n)} \geq \varepsilon\},$$

*and*

$$S_\varepsilon := \{(t^*_{\mathcal{L}_1,\beta_1}(\Gamma), \ldots, t^*_{\mathcal{L}_\ell,\beta_\ell}(\Gamma)) | n > n_0, \Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p), \|\mathrm{e}_p(\Gamma)\|_{u(D_n)} \leq \delta\}.$$

*Moreover, the systems $\mathcal{L}_1, \ldots, \mathcal{L}_\ell$ have Cauchy-Schwarz complexity at most $q - 2$.*

*Proof.* Apply Lemma E.2 for $\mathcal{D}$ and $\varepsilon/2$. There exist $\delta \in (0, \varepsilon/4)$ and systems of linear forms $\mathcal{L}_1, \ldots, \mathcal{L}_\ell$ with $m_1, \ldots, m_\ell$ linear forms, accordingly, along with coefficients $\beta_1 \in \mathbb{F}_p^{m_1}, \ldots, \beta_\ell \in \mathbb{F}_p^{m_\ell}$, such that the closures of the sets

$$T_{\varepsilon/2} := \{(t^*_{\mathcal{L}_1,\beta_1}(\mathrm{f}), \ldots, t^*_{\mathcal{L}_\ell,\beta_\ell}(\mathrm{f})) | n > n_0, \mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p, \|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \geq \varepsilon/2\}$$

and

$$S_{\varepsilon/2} := \{(t^*_{\mathcal{L}_1,\beta_1}(\mathrm{f}),\ldots,t^*_{\mathcal{L}_\ell,\beta_\ell}(\mathrm{f}))|n > n_0, \mathrm{f}:\mathbb{F}_p^n \to \mathbb{F}_p, \|\mathrm{e}_p(\mathrm{f})\|_{u(D_n)} \le 2\delta\}$$

are disjoint. For a distributional function $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ define $z(\Gamma) := (t^*_{\mathcal{L}_1,\beta_1}(\Gamma),\ldots,t^*_{\mathcal{L}_\ell,\beta_\ell}(\Gamma))$. Let $c > 0$ be the $L_\infty$ distance between the closures of $T_{\varepsilon/2}$ and $S_{\varepsilon/2}$, and set $\varepsilon' = c/4$. We will show that for large enough $n > n_0$, if $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ has $\|\mathrm{e}_p(\Gamma)\|_{u(D_n)} \ge \varepsilon$ then $z(\Gamma)$ is within $L_\infty$ distance $\varepsilon'$ from $T_{\varepsilon/2}$; and if $\|\mathrm{e}_p(\Gamma)\|_{u(D_n)} \le \delta$ then $z(\Gamma)$ is within $L_\infty$ distance $\varepsilon'$ from $S_{\varepsilon/2}$; this will conclude the lemma.

Consider first the case where $\|\mathrm{e}_p(\Gamma)\|_{u(D_n)} \ge \varepsilon$. Let $\mathrm{F} \sim \Gamma$. By Claims E.3 and E.4 we have that there exists $n_0$, such that for every $n > n_0$ we have

$$\Pr_{\mathrm{F}\sim\Gamma}[\|\mathrm{e}_p(\mathrm{F})\|_{u(D_n)} \ge \varepsilon/2] \ge 0.99,$$

and that for every $1 \le i \le \ell$ we have

$$\Pr_{\mathrm{F}\sim\Gamma}[|t^*_{\mathcal{L}_i,\beta_i}(\mathrm{F}) - t^*_{\mathcal{L}_i,\beta_i}(\Gamma)| \le \varepsilon'] \ge 1 - \frac{1}{100\ell}.$$

By the union bound, there exists a specific $\mathrm{f} : \mathbb{F}_p^n \to \mathbb{F}_p$ such that both conditions hold. That is, $z(\mathrm{f}) \in S_{\varepsilon/2}$ and $\|z(\Gamma) - z(\mathrm{f})\|_\infty \le \varepsilon'$. The case where $\|\mathrm{e}_p(\Gamma)\|_{u(D_n)} \le \delta$ is completely analogous. □

We thus study from now on distributional functions $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$. The next step is to define averages of such functions with regards to polynomial factors. Let $\mathcal{B}$ be a polynomial factor. We define the average $\mathbb{E}(\Gamma|\mathcal{B}) : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ as follows. Assume $\mathcal{B}$ defines a partition $C_1 \uplus \ldots \uplus C_b$ of $\mathbb{F}_p^n$. For $x \in C_i$ define $\mathbb{E}(\Gamma|\mathcal{B})(x)$ to be the average of $\Gamma(y)$ over $y \in C_i$,

$$\mathbb{E}(\Gamma|\mathcal{B})(x) = \frac{1}{|C_i|}\sum_{y\in C_i}\Gamma(y).$$

Note that $\mathfrak{a}_c \circ \mathbb{E}(\Gamma|\mathcal{B}) \equiv \mathbb{E}(\mathfrak{a}_c \circ \Gamma|\mathcal{B})$.

We say a polynomial factor $\mathcal{B}$ is $(d,\delta)$-good for $\Gamma$ if, informally, $U^{d+1}$ norms cannot distinguish between $\Gamma$ and $\mathbb{E}(\Gamma|\mathcal{B})$. Formally, we say a polynomial factor $\mathcal{B}$ is $(d,\delta)$-good for $\Gamma$ if for all $c \in \mathbb{F}_p$ we have

$$\|\mathfrak{a}_c \circ \Gamma - \mathfrak{a}_c \circ \mathbb{E}(\Gamma|\mathcal{B})\|_{U^{d+1}} \le \delta. \tag{48}$$

We first argue that $(d,\delta)$-good polynomial factors exist for every distributional function.

**Claim E.6.** *Let $\delta > 0$, $d < p$ and $r(\cdot)$ be an arbitrary growth function. Then for every distributional function $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ there exists a $(d,\delta)$-good polynomial factor $\mathcal{B}$ with degree $d$, complexity $C \le C_{\max}(p,d,\delta,r(\cdot))$ and rank at least $r(C)$.*

*Proof.* Apply Lemma B.14 on the set of functions $\{\mathfrak{a}_c \circ \Gamma : c \in \mathbb{F}_p\}$. □

The next claim is an analog of Corollary C.7. It shows that if $\Gamma$ is a distributional function, and if $\mathcal{B}$ is a $(d,\delta)$-good polynomial factor for $\Gamma$ where $\delta > 0$ is small enough, then averages $t^*_{\mathcal{L},\beta}$ cannot distinguish between $\Gamma$ and $\mathbb{E}(\Gamma|\mathcal{B})$ if the true complexity of $\mathcal{L}$ is at most $d$.

**Claim E.7.** *Let $\mathcal{L} = \{L_1,\ldots,L_m\}$ be a linear system of true complexity $d$ and Cauchy-Schwarz complexity at most $p$. For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ be a distributional function and let $\mathcal{B}$ be a $(d,\delta)$-good polynomial factor for $\Gamma$. Then for every $\beta \in \mathbb{F}_p^m$,*

$$|t^*_{\mathcal{L},\beta}(\Gamma) - t^*_{\mathcal{L},\beta}(\mathbb{E}(\Gamma|\mathcal{B}))| \le \varepsilon.$$

*Proof.* Let $f_i := \mathfrak{a}_{\beta(i)} \circ \Gamma$ for $1 \le i \le m$. We have

$$t^*_{\mathcal{L},\beta}(\Gamma) = \mathbb{E}_{\mathbf{X}\in(\mathbb{F}_p^n)^k}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right],$$

and since $\mathfrak{a}_{\beta(i)} \circ \mathbb{E}(\Gamma|\mathcal{B}) \equiv \mathbb{E}(f_i|\mathcal{B})$, we also have

$$t^*_{\mathcal{L},\beta}(\mathbb{E}(\Gamma|\mathcal{B})) = \mathbb{E}_{\mathbf{X}\in(\mathbb{F}_p^n)^k}\left[\prod_{i=1}^m \mathbb{E}(f_i|\mathcal{B})(L_i(\mathbf{X}))\right].$$

The claim follows by Corollary C.7 and (48). □

We would also require an analog of Proposition C.5. Let $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ be a distributional function, and let $\mathcal{B}$ be a $(d, \delta)$-good polynomial factor for $\Gamma$ defined by polynomials $P_1, \dots, P_C$. Let $\mathcal{B}'$ be another polynomial factor defined by polynomials $Q_1, \dots, Q_C$. We define a new hybrid distribution, denoted $\mathbb{E}(\Gamma|\mathcal{B} \to \mathcal{B}')$ as follows: assume $\mathbb{E}(\Gamma|\mathcal{B})(x) = F(P_1(x), \dots, P_C(x))$ where $F : \mathbb{F}_p^C \to P(\mathbb{F}_p)$ is some function; we define

$$\mathbb{E}(\Gamma|\mathcal{B} \to \mathcal{B}')(x) := F(Q_1(x), \dots, Q_C(x)).$$

**Lemma E.8.** *Let $\mathcal{L}$ be a homogeneous system of $m$ linear forms. Let $d \ge 1$ be a degree bound, and $\varepsilon > 0$ a required error. There exists $r_{\min} = r_{\min}(m, d, \varepsilon)$ such that the following holds. Let $\mathcal{B}, \mathcal{B}'$ be polynomial factors of degree at most $d$ defined by $P_1, \dots, P_C$ and $Q_1, \dots, Q_C$, respectively. Assume that $\deg(P_i) = \deg(Q_i)$ for all $1 \le i \le C$ and $\operatorname{rank}(\mathcal{B}), \operatorname{rank}(\mathcal{B}') \ge r_{\min}$. Then for every distributional function $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ and any $\beta \in \mathbb{F}_p^m$ we have*

$$\left|t^*_{\mathcal{L},\beta}(\mathbb{E}(\Gamma|\mathcal{B})) - t^*_{\mathcal{L},\beta}(\mathbb{E}(\Gamma|\mathcal{B} \to \mathcal{B}'))\right| \le \varepsilon.$$

The proof is identical to the proof of Proposition C.5 and we do not repeat it.

## E.3 Proof of Theorem 1.3

The proof follows very similar lines to the proof of Theorem D.2. We will highlight the changes that need to be made in the proof, and avoid repetition wherever possible.

Let $\mathcal{D} = \{D_n\}_{n\in\mathbb{N}}$ be a proper dual family where $D_n$ is a family of functions from $\mathbb{F}_p^n \to \mathbb{F}_p$. By Lemma E.5 we get that for every $\varepsilon > 0$ there exists $\delta \in (0, \varepsilon)$, $n_0 \in \mathbb{N}$ and systems of homogeneous linear forms $\mathcal{L}_1, \dots, \mathcal{L}_\ell$ which have Cauchy-Schwarz complexity $\le q - 2$, along with coefficients $\beta_1, \dots, \beta_\ell$, such such that the closure of the following two sets are disjoint:

$$T_\varepsilon := \{(t^*_{\mathcal{L}_1,\beta_1}(\Gamma), \dots, t^*_{\mathcal{L}_\ell,\beta_\ell}(\Gamma)) | n > n_0, \Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p), \|\mathrm{e}_p(\Gamma)\|_{u(D_n)} \ge \varepsilon\},$$

and

$$S_\varepsilon := \{(t^*_{\mathcal{L}_1,\beta_1}(\Gamma), \dots, t^*_{\mathcal{L}_\ell,\beta_\ell}(\Gamma)) | n > n_0, \Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p), \|\mathrm{e}_p(\Gamma)\|_{u(D_n)} \le \delta\}.$$

Let $s \le q - 2 < p$ be a bound on the Cauchy-Schwarz complexity of $\mathcal{L}_1, \dots, \mathcal{L}_\ell$ and $d \le s$ be a bound on their true complexity. We define $\mathcal{S}$ in the same way as in the proof of Theorem D.2.

**Claim E.9.** *Unless all functions in $\mathcal{D}$ are constants, we have $1 \in \mathcal{S}$.*

The proof is identical to the proof of Claim D.3. The case where $\mathcal{D}$ consists of only constant functions is analyzed in the same way, hence we assume that $1 \in \mathcal{S}$ from now on.

**Claim E.10.** *$\mathcal{S} \subseteq \{1, \ldots, s\}$.*

*Proof.* The proof is identical to the proof of Claim D.4. The only difference is the argument why $0^\ell$ is in the closure of $S_\varepsilon$. Let $\Gamma$ map every element of $\mathbb{F}_p^n$ to the uniform probability distribution over $\mathbb{F}_p$. Then it is easy to verify that $|t_{\mathcal{L}_i,\beta_i}(\Gamma)| = O(p^{-n})$ for all $1 \le i \le \ell$. □

We denote $t := \max(\mathcal{S})$. Let $(\Gamma_n : \mathbb{F}_p^n \to P(\mathbb{F}_p))_{n\in\mathbb{N}}$ be a sequence distributional functions where $\lim_{n\to\infty} \mathbb{E}[\mathrm{e}_p(\Gamma_n)] = 0$. Similar to the proof of Theorem D.2, the proof of Theorem 1.3 follows from the following two lemmas.

**Lemma E.11.** *If $\lim_{n\to\infty} \|\mathrm{e}_p(\Gamma_n)\|_{U^{t+1}} = 0$ then $\lim_{n\to\infty} \|\mathrm{e}_p(\Gamma_n)\|_{u(D_n)} = 0$.*

**Lemma E.12.** *If $\lim_{n\to\infty} \|f_n\|_{u(D_n)} = 0$ then $\lim_{n\to\infty} \|f_n\|_{U^{t+1}} = 0$.*

The proof of both lemmas is identical to the proof of Lemmas D.5 and D.6, where the only difference is that one considers averages $t^*_{\mathcal{L}_i,\beta_i}(\Gamma_n)$ instead of $t_{\mathcal{L}_i,\alpha_i}(f_n)$ and apply the claims proved in the previous subsection for function $\Gamma : \mathbb{F}_p^n \to P(\mathbb{F}_p)$ instead of their analogs for functions $f : \mathbb{F}_p^n \to \mathbb{D}$. The only lemma whose proof needs to be slightly changed is Lemma D.14. We sketch below an analog version for distributional functions. First, note that for every function $F : \mathbb{F}_p^N \to [0,1]$ there exists a distributional function $\Gamma_F : \mathbb{F}_p^N \to P(\mathbb{F}_p)$ such that $\mathfrak{a}_c \circ \Gamma_F \equiv F$ for all $c \in \mathbb{F}_p \setminus \{0\}$: simply set $\Pr[\Gamma_F(x) = 0] = F(x) + (1 - F(x))\frac{1}{p}$ and $\Pr[\Gamma_F(x) = z] = (1 - F(x))\frac{1}{p}$ for $z \in \mathbb{F}_p \setminus \{0\}$. Moreover, this implies that for every system of $m$ linear forms $\mathcal{L}$ and coefficients $\beta \in (\mathbb{F}_p \setminus \{0\})^m$,

$$t^*_{\mathcal{L},\beta}(\Gamma_F) = t_{\mathcal{L}}(F).$$

The lemma now follows, when in the proof of Lemma D.14 one replaces $F$ with $\Gamma_F$, $g_n$ with $Q_n$ (a deterministic distributional function) and $h_n$ with $\Gamma_{h_n}$.

# F Non-isomorphic connected systems have nonempty interior

We prove in this section the following theorem.

**Theorem F.1.** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_k$ be non-isomorphic connected systems of linear forms. For sufficiently large $n \in \mathbb{N}$, the set of points*

$$\left\{(t_{\mathcal{L}_1}(f), \ldots, t_{\mathcal{L}_k}(f)) \mid f : \mathbb{F}_p^n \to [0,1]\right\} \subseteq \mathbb{R}^k \tag{49}$$

*has a non-empty interior.*

We divided this section into two parts. In Section F.1 we introduce new notations and develop some preliminary tools. The proof of Theorem F.1 is given in Section F.2.

### F.1 Flagged systems of linear forms

Consider a system of linear forms $\mathcal{L}$. Recall that $\mathcal{L}$ is connected if there does not exists nonempty $S \subsetneq \mathcal{L}$ such that $\mathrm{span}(S) \cap \mathrm{span}(\mathcal{L} \setminus S) = \{\vec{0}\}$. Suppose that there are subsets $S_1, S_2 \subsetneq \mathcal{L}$ such that

$$\mathrm{span}(S_i) \cap \mathrm{span}(\mathcal{L} \setminus S_i) = \{\vec{0}\},$$

for $i = 1, 2$. Then for $T = S_1 \cap S_2$, we have

$$\mathrm{span}(T) \cap \mathrm{span}(\mathcal{L} \setminus T) = \{\vec{0}\}.$$

This in particular shows that (up to the isomorphisms) there is a unique way to partition a system of linear forms $\mathcal{L}$ into disjoint *connected* systems of linear forms $\mathcal{L}_1, \ldots, \mathcal{L}_k$. We call each one of $\mathcal{L}_1, \ldots, \mathcal{L}_k$ a *connected component* of $\mathcal{L}$. Since connectivity is invariant under isomorphisms we have the following trivial observation.

**Observation F.2.** *Two systems of linear forms $\mathcal{L}_1$ and $\mathcal{L}_2$ are isomorphic if and only if there is a one to one isomorphic correspondence between their connected components.*

A 1-*flagged system of linear forms* is a system of linear forms $\mathcal{L}$ and a *non-zero* linear form $M \in \mathrm{span}(\mathcal{L})$. We use the notation $\mathcal{L}^M$ to denote such a 1-flagged system of linear forms. Here $\mathcal{L}$ is called the *underlying system of linear forms* of $\mathcal{L}^M$. We call $\mathcal{L}_0^{M_0}$ and $\mathcal{L}_1^{M_1}$ *isomorphic*, if there is an invertible linear transformation $T : \mathrm{span}(\mathcal{L}_0) \to \mathrm{span}(\mathcal{L}_1)$ that maps $M_0$ to $M_1$ and its restriction to $\mathcal{L}_0$ induces an isomorphism between $\mathcal{L}_0$ and $\mathcal{L}_1$.

Let $\mathcal{L}$ be a system of linear forms in $k$ variables. For a 1-flagged system of linear forms $\mathcal{L}^M$, and a function $f : \mathbb{F}_p^n \to \mathbb{C}$ define the function $f^{\mathcal{L}^M} : \mathbb{F}_p^n \to \mathbb{C}$ by

$$f^{\mathcal{L}^M} : x \mapsto \mathbb{E}_{\mathbf{X} \in (\mathbb{F}_p^n)^k} \left[ \prod_{L \in \mathcal{L}} f(L(\mathbf{X})) \middle| M(\mathbf{X}) = x \right]. \tag{50}$$

Note that we have

$$t_{\mathcal{L}}(f) = \mathbb{E}_{X \in \mathbb{F}_p^n} \left[ f^{\mathcal{L}^M}(X) \right]. \tag{51}$$

Let $\mathcal{L}_0^{M_0}$ and $\mathcal{L}_1^{M_1}$ be 1-flagged systems of linear forms in $\mathbb{F}_p^{k_0}$ and $\mathbb{F}_p^{k_1}$, respectively. We want to define an operation that "glues" these two systems to each other by identifying $M_0$ and $M_1$: To this end, first we consider the system of linear forms $\mathcal{L}'$ defined as

$$\mathcal{L}' = \{L \oplus \vec{0} \in \mathbb{F}_p^{k_0+k_1} : L \in \mathcal{L}_0\} \cup \{\vec{0} \oplus L \in \mathbb{F}_p^{k_0+k_1} : L \in \mathcal{L}_1\}.$$

Take any element $M \neq \vec{0}$ in $\mathbb{F}_p^{k_0+k_1}$, and any *surjective* linear transformation $T : \mathbb{F}_p^{k_0+k_1} \to \mathbb{F}_p^{k_0+k_1-1}$ that maps both $\vec{0} \oplus M_0$ and $M_1 \oplus \vec{0}$ to $M$. Then the *product* of $\mathcal{L}_0^{M_0}$ and $\mathcal{L}_1^{M_1}$ which is denoted by $\mathcal{L}_0^{M_0} \cdot \mathcal{L}_1^{M_1}$ is defined as the 1-flagged system of linear forms

$$\mathcal{L}_0^{M_0} \cdot \mathcal{L}_1^{M_1} := \left(T(\mathcal{L}')\right)^M.$$

Note that since $T$ is surjective, this definition does not depend (up to isomorphism) on the particular choices of the linear form $M$ and the map $T$. The restrictions of $T$ to each one of the sets

$$\mathrm{span}\{L \oplus \vec{0} \in \mathbb{F}_p^{k_0+k_1} : L \in \mathcal{L}_0\},$$

and

$$\mathrm{span}\{\vec{0} \oplus L \in \mathbb{F}_p^{k_0+k_1} : L \in \mathcal{L}_1\}$$

is invertible, and thus $T$ induces isomorphisms between these sets and their corresponding $\mathcal{L}_i$ ($i = 0$ or 1). Therefore we shall refer to $\{T(L \oplus \vec{0}) \in \mathbb{F}_p^{k_0+k_1} : L \in \mathcal{L}_0\}^{M_0}$ and $\{T(\vec{0} \oplus L) \in \mathbb{F}_p^{k_0+k_1} : L \in \mathcal{L}_1\}^{M_1}$ respectively as copies of $\mathcal{L}_0^{M_0}$ and $\mathcal{L}_1^{M_1}$ in $\mathcal{L}_0^{M_0} \cdot \mathcal{L}_1^{M_1}$. In the sequel, we will frequently identify 1-flagged system of linear forms with their copies in their product.

The definition of the product of 1-flagged systems of linear forms is motivated by the following fact: It follows from (50) that if $\mathcal{L}_0^{M_0}$ and $\mathcal{L}_1^{M_1}$ are 1-flagged systems of linear forms, then for every function $f : \mathbb{F}_p^n \to \mathbb{C}$, we have

$$f^{\mathcal{L}_0^{M_0} \cdot \mathcal{L}_1^{M_1}} = f^{\mathcal{L}_0^{M_0}} f^{\mathcal{L}_1^{M_1}}. \tag{52}$$

**Lemma F.3.** *Let $\mathcal{L}_0^{M_0}$ and $\mathcal{L}_1^{M_1}$ be 1-flagged systems of linear forms such that both $\mathcal{L}_0 \cup \{M_0\}$ and $\mathcal{L}_1 \cup \{M_1\}$ are connected. Then $\mathcal{L} \cup \{M\}$ is also connected, where $\mathcal{L}^M := \mathcal{L}_0^{M_0} \cdot \mathcal{L}_1^{M_1}$.*

*Proof.* Suppose that $\mathcal{L}_0^{M_0}$ and $\mathcal{L}_1^{M_1}$ are respectively in $k_0$ and $k_1$ variables. Let $T : \mathbb{F}_p^{k_0+k_1} \to \mathbb{F}_p^{k_0+k_1-1}$ be as in the definition of the product of two 1-flagged systems of linear forms given above. Consider a nonempty set $S \subsetneq \mathcal{L} \cup \{M\}$. Suppose to the contrary of the assertion that

$$\mathrm{span}(S) \cap \mathrm{span}((\mathcal{L} \cup \{M\}) \setminus S) = \{\vec{0}\}. \tag{53}$$

We identify $\mathcal{L}_0^{M_0}$ and $\mathcal{L}_1^{M_1}$ with their copies in $\mathcal{L}^M$. In particular, both $M_0$ and $M_1$ are identified with $M$. Since $\mathcal{L}_0 \cup \{M\}$ and $\mathcal{L}_1 \cup \{M\}$ are both connected with have $M \in \mathrm{span}(\mathcal{L}_0 \setminus \{M\})$ and $M \in \mathrm{span}(\mathcal{L}_1 \setminus \{M\})$. Then it follows from (53) that we have $S \neq \mathcal{L}_i \cup \{M\}$, for $i = 1, 2$. Also by (53) we have

$$\mathrm{span}(S \cap (\mathcal{L}_0 \cup \{M\})) \cap \mathrm{span}((\mathcal{L}_0 \cup \{M\}) \setminus S) = \{\vec{0}\},$$

and

$$\mathrm{span}(S \cap (\mathcal{L}_1 \cup \{M\})) \cap \mathrm{span}((\mathcal{L}_1 \cup \{M\}) \setminus S) = \{\vec{0}\}.$$

Thus at least one of $\mathcal{L}_0 \cup \{M\}$ or $\mathcal{L}_1 \cup \{M\}$ is not connected which contradicts our assumption. □

For a system of linear forms $\mathcal{L}$ and an $L \in \mathrm{span}(\mathcal{L})$ define $\deg_{\mathcal{L}}(L)$ to be the number of pairs $(x, y) \in \mathcal{L} \times \mathcal{L}$ satisfying $x + y = L$.

Let $\mathcal{L}^M := \mathcal{L}_0^M \cdot \mathcal{L}_1^M$ where $\mathcal{L}_0^M$ and $\mathcal{L}_1^M$ are 1-flagged systems of linear forms. It follows from the definition of the product that if $x + y \in \mathrm{span}(\{M\})$ with $x \in \mathcal{L}_0$ and $y \in \mathcal{L}_1$, then both $x, y$ belong to $\mathrm{span}(\{M\})$. Hence for every $L \in \mathrm{span}(\{M\})$, we have

$$\deg_{\mathcal{L}_0}(L) + \deg_{\mathcal{L}_1}(L) \leq \deg_{\mathcal{L}}(L) \leq \deg_{\mathcal{L}_0}(L) + \deg_{\mathcal{L}_1}(L) + |\mathrm{span}(\{M\}) \cap \mathcal{L}_0| + |\mathrm{span}(\{M\}) \cap \mathcal{L}_1|, \tag{54}$$

and similarly for $L \in \mathcal{L}_i \setminus \mathrm{span}(\{M\})$ where $i = 0, 1$, we have

$$\deg_{\mathcal{L}_i}(L) \leq \deg_{\mathcal{L}}(L) \leq \deg_{\mathcal{L}_i}(L) + 2|\mathrm{span}(\{M\}) \cap \mathcal{L}_{1-i}|. \tag{55}$$

**Lemma F.4.** *Let $\mathcal{L}_1^{M_1}, \ldots, \mathcal{L}_k^{M_k}$ be non-isomorphic 1-flagged systems of linear forms such that $\mathcal{L}_i \cup \{M_i\}$ are connected for all $i \in [k]$. For every $N > 0$, there exist a 1-flagged system of linear forms $\mathcal{L}^M$ such that $\mathrm{rank}(\mathrm{span}(\mathcal{L})) > N$, and the underlying systems of linear forms of $\mathcal{L}^M \cdot \mathcal{L}_i^{M_i}$ for $i \in [k]$ are connected and non-isomorphic.*

*Proof.* Let $d > p+N$ be larger than the size of $\mathcal{L}_i$, for every $i \in [k]$. Denote $e_1 = (1,0,\dots,0) \in \mathbb{F}_p^d$, and consider the system of linear forms

$$\mathcal{M} := \left(\{0\} \times \mathbb{F}_p^{d-1}\right) \cup \left(\{1\} \times \{0,1\}^{d-1}\right) \setminus \{\vec{0}, e_1\} \subseteq \mathbb{F}_p^d.$$

We claim that $\mathcal{M}$ is connected. Indeed $\mathbb{F}_p^{d-1} \setminus \{\vec{0}\} \equiv \left(\{0\} \times \mathbb{F}_p^{d-1}\right) \setminus \{\vec{0}\} \subsetneq \mathcal{M}$ is trivially connected, and hence if $S \subseteq \mathcal{M}$ is such that $\mathrm{span}(S) \cap \mathrm{span}(\mathcal{M} \setminus S) = \{\vec{0}\}$, then without loss of generality we can assume that

$$\left(\{0\} \times \mathbb{F}_p^{d-1}\right) \setminus \{\vec{0}\} \subseteq S.$$

One can easily verify that $S$ cannot be equal to $\left(\{0\} \times \mathbb{F}_p^{d-1}\right) \setminus \{\vec{0}\}$. Hence there exists at least one element $L \in S \cap \left(\{1\} \times \{0,1\}^{d-1}\right)$. Then $\mathcal{M} \subseteq \mathrm{span}(\{L\} \cup (\{0\} \times \mathbb{F}_p^{d-1})) \subseteq \mathrm{span}(S)$ which together with the assumption $\mathrm{span}(S) \cap \mathrm{span}(\mathcal{M} \setminus S) = \{\vec{0}\}$ shows that $S = \mathcal{M}$. Hence $\mathcal{M}$ is connected.

Note that $2^{d-1} \leq \deg_{\mathcal{M}}(L) \leq 4p^{d-1}$ for every $L \in \mathcal{M}$, and $\deg_{\mathcal{M}}(e_1) = 2(2^{d-1}-1)$. Furthermore for every $\lambda \in \mathbb{F}_p \setminus \{0,1\}$, we have $\deg_{\mathcal{M}}(\lambda e_1) = 0$. Also we have $\mathrm{span}(\{e_1\}) \cap \mathcal{M} = \emptyset$. Set

$$\widetilde{\mathcal{L}}^M := \underbrace{\mathcal{M}^{e_1} \cdot \ldots \cdot \mathcal{M}^{e_1}}_{10p^d \text{ times}}.$$

and

$$\mathcal{L}^M := (\widetilde{\mathcal{L}} \cup \{M\})^M.$$

By (54) and (55), and the above properties of $\mathcal{M}$, we have $2^{d-1} \leq \deg_{\mathcal{L}}(L) \leq 4p^{d-1}$, for every $L \in \mathcal{L} \setminus \{M\}$. Moreover $\deg_{\mathcal{L}}(M) \geq 10p^d$, and $\deg_{\mathcal{L}}(L) = 0$ for every $L \in \mathrm{span}(\{M\}) \setminus \{\vec{0}, M\}$. It also follows from $\mathrm{span}(\{e_1\}) \cap \mathcal{M} = \emptyset$ that $\mathrm{span}(\{M\}) \cap \mathcal{L} = \{M\}$.

For every $i \leq [k]$, set $\mathcal{N}_i^{W_i} := \mathcal{L}^M \cdot \mathcal{L}_i^{M_i}$. Then by (54) and (55), we have

(i) $\deg_{\mathcal{N}_i}(W_i) \geq 10p^d$;

(ii) $2^{d-1} \leq \deg_{\mathcal{N}_i}(L) \leq 4p^{d-1} + 2p \leq 5p^{d-1}$ for every $L \in \mathcal{N}_i \setminus (\mathcal{L}_i \cup \{W_i\})$;

(iii) $\deg_{\mathcal{N}_i}(L) \leq |\mathcal{L}_i| < 2^{d-1}$ for every $L \in \mathcal{L}_i \setminus \{W_i\}$.

Since $\mathcal{M}$ is connected and $e_1 \in \mathrm{span}(\mathcal{M})$, we have that $\mathcal{M} \cup \{e_1\}$ is also connected. Then Lemma F.3 shows that $\mathcal{L}$ is connected. Now since $\mathcal{L}_i$ are connected, Lemma F.3 implies that $\mathcal{N}_i = \mathcal{N}_i \cup \{W_i\}$ are connected. It remains to show that they are non-isomorphic. But if $\mathcal{N}_i$ is isomorphic to $\mathcal{N}_j$ for some $i,j \in [k]$, then there is a bijection between $\mathcal{N}_i$ and $\mathcal{N}_j$ that can be extended to an invertible $T : \mathrm{span}(\mathcal{N}_i) \to \mathrm{span}(\mathcal{N}_j)$. Since such a function, maps $\vec{0}$ to $\vec{0}$, and preserves the degrees, by (i), (ii), and (iii) above, we have $T(W_i) = W_j$ and $\{T(L) : L \in \mathcal{L}_i\} = \mathcal{L}_j$. Thus the restriction of $T$ to $\mathcal{L}_i^{M_i}$ is an isomorphism between $\mathcal{L}_i^{M_i}$ and $\mathcal{L}_j^{M_j}$ contradicting our assumption that $\mathcal{L}_i^{M_i}$ and $\mathcal{L}_j^{M_j}$ are non-isomorphic. □

## F.2 Finishing the proof.

We view averages $t_{\mathcal{L}}(f)$ as polynomials in the variables $\{f(x) : x \in \mathbb{F}_p^n\}$,

$$t_{\mathcal{L}}(f) = \frac{1}{p^{nk}} \sum_{\mathbf{x} \in (\mathbb{F}_p^n)^k} \prod_{i=1}^{m} f(L_i(x)).$$

We start by proving a technical lemma.

**Lemma F.5.** *Let $\mathcal{L}_1, \ldots, \mathcal{L}_k$ be non-isomorphic connected systems of linear forms. Let $P_1, \ldots, P_k$ be functions mapping every $f : \mathbb{F}_p^n \to \mathbb{C}$ to $\mathbb{C}$ in the following way. Every $P_i$ is a polynomial of degree at most $d$ in variables $\{f(x) : x \in \mathbb{F}_p^n\}$. If $n > d + \max_{i\in[k]} \text{rank}(\text{span}(\mathcal{L}_i))$, and for every $i \in [k]$, $\text{rank}(\text{span}(\mathcal{L}_i)) > d$, then*

$$P_1(f)t_{\mathcal{L}_1}(f) + \ldots + P_k(f)t_{\mathcal{L}_k}(f) \not\equiv 0,$$

*unless $P_i \equiv 0$ for all $i \in [k]$.*

*Proof.* We claim a stronger statement that if at least one of $P_i$ is not divisible by $f(\vec{0})$, then

$$P_1(f)t_{\mathcal{L}_1}(f) + \ldots + P_k(f)t_{\mathcal{L}_k}(f)|_{f(\vec{0})=0} \not\equiv 0.$$

Trivially it suffices to prove this statement for the case where for every $i \in [k]$, no monomial of $P_i$ is divisible by $f(\vec{0})$. Assume to the contrary that there exist polynomials $P_i$ of degree at most $d$ in variables $\{f(x) : x \in \mathbb{F}_p^n\}$ with monomials which are not divisible by $f(\vec{0})$ such that

$$P_1(f)t_{\mathcal{L}_1}(f) + \ldots + P_k(f)t_{\mathcal{L}_k}(f)|_{f(\vec{0})=0} \equiv 0.$$

Without loss of generality assume that every $\mathcal{L}_i$ is a system of linear forms in $\mathbb{F}_p^l$ for some positive integer $l$. Define the rank of a monomial $\prod_{x\in\mathbb{F}_p^n} f(x)^{\alpha_x}$ to be the rank of $\text{span}(\{x : \alpha_x \neq 0\})$. Let $r_i$ denote the largest rank of a monomial with a non-zero coefficient in $P_i$.

Set $i_0 := \text{argmax}_{i\in[k]} (r_i + \text{rank}(\text{span}(\mathcal{L}_i)))$. Let non-zero $x_1, \ldots, x_a \in \mathbb{F}_p^n$ be so that

$$\text{rank}(\text{span}(\{x_1, \ldots, x_a\})) = r_{i_0},$$

and $\prod_{i=1}^a f(x_i)^{\alpha_i}$ where $\alpha_i > 0$ appears with a non-zero coefficient in $P_{i_0}$. Since $n \geq d + \text{rank}(\text{span}(\mathcal{L}_{i_0}))$, there exists $\mathbf{x} \in (\mathbb{F}_p^n)^k$ such that $\{L(\mathbf{x}) : L \in \mathcal{L}_{i_0}\}$ are all distinct and $\text{span}(\{L(\mathbf{x}) : L \in \mathcal{L}_{i_0}\}) \cap \text{span}(\{x_1, \ldots, x_a\}) = \{\vec{0}\}$. Note that the connectivity of $\mathcal{L}_{i_0}$, and the assumption that $\deg(P_{i_0}) < \text{rank}(\text{span}(\mathcal{L}_{i_0}))$ implies that the monomial

$$\left(\prod_{i=1}^a x_i^{\alpha_i}\right) \prod_{L\in\mathcal{L}_{i_0}} f(L(\mathbf{x})),$$

appears with a non-zero coefficient in $P_{i_0}(f)t_{\mathcal{L}_{i_0}}(f)$ (i.e. there is no cancelation). This is because the maximal connected component

Suppose that this monomial appears with a non-zero coefficient also for some other $1 \leq j \leq k$ in $P_j(f)\prod_{L\in\mathcal{L}_j} f(L(\mathbf{x}'))$, where $\mathbf{x}' \in (\mathbb{F}_p^n)^k$. Then the maximality of $r_{i_0}$+$\text{rank}(\text{span}(\mathcal{L}_{i_0}))$, connectivity of $\mathcal{L}_j$, and the assumption that $\deg(P_j) < \text{rank}(\text{span}(\mathcal{L}_j))$ shows that $\{L(\mathbf{x}') : L \in \mathcal{L}_j\} = \{L(\mathbf{x}) : L \in \mathcal{L}_{i_0}\}$ as multisets. By the assumption that $\{L(\mathbf{x}) : L \in \mathcal{L}_{i_0}\}$ are all distinct we get that $\{L(\mathbf{x}') : L \in \mathcal{L}_j\}$ are also all distinct. It follows that $\mathcal{L}_j$ is isomorphic to $\mathcal{L}_{i_0}$, which is a contradiction. □

Consider a system of linear forms $\mathcal{L}$ and a function $f : \mathbb{F}_p^n \to \mathbb{C}$. Define the function $f^{\partial\mathcal{L}} : \mathbb{F}_p^n \to \mathbb{C}$, as

$$f^{\partial\mathcal{L}}(x) := \sum_{L\in\mathcal{L}} f^{(\mathcal{L}\setminus\{L\})^L}(x).$$

The following easy lemma which follows from linearity of expectation explains the motivation for this notation.

**Lemma F.6.** *For $f, g : \mathbb{F}_p^n \to \mathbb{C}$, and every system of linear forms $\mathcal{L}$, we have*

$$\frac{d}{dt} t_{\mathcal{L}}(f + tg)|_{t=0} = \mathbb{E}\left[g(X) f^{\partial \mathcal{L}}(X)\right],$$

*where $X$ is a random variable taking values in $\mathbb{F}_p^n$ uniformly at random.*

Consider connected non-isomorphic systems of linear forms $\mathcal{L}_1, \dots, \mathcal{L}_k$. We claim that in order to prove Theorem F.1 it suffices to shows that there exists $f : \mathbb{F}_p^n \to (0,1)$ such that $f^{\partial \mathcal{L}_1}, \dots, f^{\partial \mathcal{L}_k}$ are linearly independent over $\mathbb{R}$.

**Claim F.7.** *Let $f : \mathbb{F}_p^n \to (0,1)$ be such that $f^{\partial \mathcal{L}_1}, \dots, f^{\partial \mathcal{L}_k}$ are linearly independent over $\mathbb{R}$. Then there exists $\varepsilon > 0$ such that*

$$\left\{(t_{\mathcal{L}_1}(f), \dots, t_{\mathcal{L}_k}(f)) + z : z \in \mathbb{R}^k, \|z\|_\infty \le \varepsilon\right\} \subseteq \left\{(t_{\mathcal{L}_1}(g), \dots, t_{\mathcal{L}_k}(g)) : g : \mathbb{F}_p^n \to (0,1)\right\}.$$

*Proof.* Let $e_1, \dots, e_k \in \mathbb{R}^k$ denote the unit vectors. Since $f^{\partial \mathcal{L}_1}, \dots, f^{\partial \mathcal{L}_k}$ are linearly independent over $\mathbb{R}$, for every $e_i$ there exists $g_i : \mathbb{F}_p^n \to \mathbb{R}$ such that

$$\mathbb{E}_{X \in \mathbb{F}_p^n}\left[g_i(X) f^{\partial \mathcal{L}_j}(X)\right] = \delta_{i,j},$$

where $\delta_{i,j} = 1_{i=j}$ is the Kronecker delta function. For $z \in \mathbb{R}^k$ define $g_z(x) = \sum_{i=1}^k z_i g_i(x)$, and consider the map $T : \mathbb{R}^k \to \mathbb{R}^k$

$$T(z_1, \dots, z_k) = (t_{\mathcal{L}_1}(f + g_z), \dots, t_{\mathcal{L}_k}(f + g_z)).$$

The Jacobian of $T$ is the identity matrix, and hence invertible. By the inverse function theorem, for every $\eta > 0$, $\{T(z) : \|z\|_\infty < \eta\}$ contains a neighborhood of $T(0) = (t_{\mathcal{L}_1}(f), \dots, t_{\mathcal{L}_k}(f))$. We will choose $\eta > 0$ small enough such that $\|f + g_z\|_\infty < 1$ for all $\|z\|_\infty < \eta$. □

Suppose to the contrary that for every $f : \mathbb{F}_p^n \to (0,1)$, $f^{\partial \mathcal{L}_1}, \dots, f^{\partial \mathcal{L}_k}$ are linearly *dependent* over $\mathbb{R}$. Note that for every $1 \le i \le k$, and for every $x_0 \in \mathbb{F}_p^n$, $f^{\partial \mathcal{L}_i}(x_0)$ is a polynomial of degree $|\mathcal{L}_i| - 1$ in the variables $\{f(x) : x \in \mathbb{F}_p^n\}$. The linear dependency of $f^{\partial \mathcal{L}_1}, \dots, f^{\partial \mathcal{L}_k}$ shows that for every $f : \mathbb{F}_p^n \to (0,1)$, the $k \times k$ matrix whose $ij$-th entry is $\mathbb{E}\left[f^{\partial \mathcal{L}_i}(X) f^{\partial \mathcal{L}_j}(X)\right]$ is singular which in turn implies that the determinant of this matrix as a polynomial in $\{f(x) : x \in \mathbb{F}_p^n\}$ is the zero polynomial. So the functions $f^{\partial \mathcal{L}_i}$ considered as vectors with polynomial entries are dependent over the field of fractions of polynomials in variables $\{f(x) : x \in \mathbb{F}_p^n\}$. Furthermore since the degree of the $ij$-th entry of this matrix is at most $|\mathcal{L}_i| + |\mathcal{L}_j| - 2$ which does not depend on $n$, we conclude that there exists polynomials $P_1, \dots, P_k$ in the variables $\{f(x) : x \in \mathbb{F}_p^n\}$, and of degree at most some integer $C := C(|\mathcal{L}_1|, \dots, |\mathcal{L}_k|)$ (which does not depend on $n$) such that

$$P_1(f) f^{\partial \mathcal{L}_1} + \ldots + P_k(f) f^{\partial \mathcal{L}_k} \equiv 0. \tag{56}$$

Let $\mathcal{M}_1^{L_1}, \dots, \mathcal{M}_l^{L_l}$ be some representatives for all the isomorphism classes of 1-flagged systems of linear forms $\{(\mathcal{L}_i \setminus \{L\})^L : i \in [k], L \in \mathcal{L}_i\}$. Let $\{\alpha_{i,j} \in \mathbb{Z}_+ : i \in [k], j \in [l]\}$ be such that for every $i \in [k]$, we have

$$f^{\partial \mathcal{L}_i} = \sum_{j=1}^{l} \alpha_{i,j} f^{\mathcal{M}_j^{L_j}}.$$

By Lemma F.4 it is possible to find a 1-flagged system of linear form $\mathcal{L}^M$ of arbitrarily large rank such that for $\mathcal{N}_j^{W_j} := \mathcal{M}_j^{L_j} \cdot \mathcal{L}^M$ (where $j \in [l]$), the systems of linear forms $\mathcal{N}_j$ are non-isomorphic and connected. Note that by (52) we have $f^{\mathcal{N}_j^{W_j}} = f^{\mathcal{M}_j^{L_j}} f^{\mathcal{L}^M}$, and so we have

$$\sum_{i=1}^{k} \sum_{j \in [l]} \alpha_{i,j} P_i(f) f^{\mathcal{N}_j^{W_j}} \equiv 0,$$

which by (51)) implies that

$$\sum_{j \in [l]} \left( \sum_{i=1}^{k} \alpha_{i,j} P_i(f) \right) t_{\mathcal{N}_j}(f) \equiv 0. \tag{57}$$

By Lemma F.5 this shows that $\sum_{i=1}^{k} \alpha_{i,j} P_i(f) \equiv 0$ for every $j \in [l]$. Now since the $\mathcal{L}_i$ are non-isomorphic, $(\mathcal{L}_i \setminus \{L\})^L \not\equiv (\mathcal{L}_j \setminus \{M\})^M$ for every two distinct $i, j \in [l]$ and every $L \in \mathcal{L}_i$ and $M \in \mathcal{L}_j$. Hence for every $j \in [l]$, there is exactly one $i \in [k]$ such that $\alpha_{i,j} \neq 0$. It follows that $P_i \equiv 0$ for every $i \in [k]$ which is a contradiction.

# G Concluding remarks

In this paper we study affine invariant properties which are testable. We show that essentially every such property can be tested by an appropriate Gowers uniformity norm. One technical limitation of our techniques is that they hold only if the field size is not too small (i.e. if the Cauchy-Schwarz complexity is smaller than the field size). The main reason for this obstacle is that the inverse theorem for Gowers norm, which fully established for large fields, is only partially understood for small fields. Thus, we do not have corresponding decomposition theorems for small fields which are suitable for our needs.

Even if the conjectured form of the inverse theorem for the Gowers norm for small fields was proved, it would still not answer the following problem: is it possible to test, using a constant number of queries, whether a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ is correlated to a polynomial of degree $d$, where $d > p$? We know that the Gowers norm test fails, as it actually tests distance to a larger set of functions (non-classical polynomials). The simplest case which is unknown is the following:

**Problem G.1.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Does there exist a test which queries $f$ on a constant number of positions, and which can distinguish whether $f$ has noticeable or negligible correlation with cubic polynomials?*

In fact, even the following simpler problem is unknown:

**Problem G.2.** *Let $\varepsilon > 0$ and $f : \mathbb{F}_2^n \to \mathbb{F}_2$. Does there exist a test which queries $f$ on $q(\varepsilon)$ positions, and which can distinguish whether $f$ has correlation at least $\varepsilon$, or at most $\delta(\varepsilon)$, with cubic polynomials?*

# References

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

[2] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. In *STOC '90: Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pages 73–83, New York, NY, USA, 1990. ACM.

[3] Paul Erdős, László Lovász, and Joel Spencer. Strong independence of graphcopy functions. In *Graph theory and related topics (Proc. Conf., Univ. Waterloo, Waterloo, Ont., 1977)*, pages 165–172. Academic Press, New York, 1979.

[4] Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[5] Oded Goldreich and Dana Ron. On proximity oblivious testing. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 141–150, New York, NY, USA, 2009. ACM.

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

[7] 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.

[8] 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.

[9] Ben Green. Montréal notes on quadratic Fourier analysis. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 69–102. Amer. Math. Soc., Providence, RI, 2007.

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

[11] Ben Green and Terence Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010.

[12] Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. In *FOCS '08: Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 166–175, Washington, DC, USA, 2008. IEEE Computer Society.

[13] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 403–412, New York, NY, USA, 2008. ACM.

[14] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

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

[16] Endre Szemerédi. On sets of integers containing no $k$ elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. Collection of articles in memory of Juriui Vladimiroviv c Linnik.

[17] T. Tao. Some notes on non-classical polynomials in finite characteristic. Online blog. http://terrytao.wordpress.com/2008/11/13/some-notes-on-non-classical-polynomials-in-finite-characteristic.

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