Higher-order Fourier analysis of $\mathbb{F}_p^n$ and the complexity of systems of linear forms

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

In this article we are interested in the density of small linear structures (e.g. arithmetic progressions) in subsets A of the group $\mathbb{F}_p^n$. It is possible to express these densities as certain analytic averages involving 1_A, the indicator function of A. In the higher-order Fourier analytic approach, the function 1_A is decomposed as a sum $f_1 + f_2$ where f_1 is structured in the sense that it has a simple higher-order Fourier expansion, and f_2 is pseudorandom in the sense that the kth Gowers uniformity norm of f_2, denoted by $\|f_2\|_{U^k}$, is small for a proper value of k.

For a given linear structure, we find the smallest degree of uniformity k such that assuming that $\|f_2\|_{U^k}$ is sufficiently small, it is possible to discard f_2 and replace 1_A with f_1, affecting the corresponding analytic average only negligibly. Previously, Gowers and Wolf solved this problem for the case where f_1 is a constant function. Furthermore, our main result solves Problem 7.6 in [W. T. Gowers and J. Wolf. Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$. *Geom. Funct. Anal.*, to appear] regarding the analytic averages that involve more than one subset of $\mathbb{F}_p^n$.

AMS SUBJECT CLASSIFICATION: 11B30, 11T24

KEYWORDS: Gowers uniformity, higher-order Fourier analysis, true complexity;

1

Contents

1 Introduction

In additive combinatorics one is often interested in the density of small linear structures (e.g. arithmetic progressions) in subsets of Abelian groups. It is possible to express these densities as certain analytic averages. For example, consider a subset A of a finite Abelian group G. Then the density of the k-term arithmetic progressions in A is given by

$$\mathbb{E}\left[1_A(X)1_A(X+Y)\dots 1_A(X+(k-1)Y)\right], \tag{1}$$

where X and Y are random variables taking values in G uniformly at random, and 1_A is the indicator function of A. More generally, one is often interested in analyzing

$$\mathbb{E}_{X_1,\dots,X_k\in\mathbb{F}_p^n}\left[1_A\left(\sum_{i=1}^k \lambda_{1,i}X_i\right)\dots 1_A\left(\sum_{i=1}^k \lambda_{m,i}X_i\right)\right], \tag{2}$$

where $X_1,\dots,X_k$ are random variables taking values in G uniformly at random, $\lambda_{i,j}$ are constants, and $A\subseteq G$. Analyzing averages of this type and understanding the relations between them is the core of many problems and results in additive combinatorics and analytic number theory, and there are theories which are developed for this purpose. The theory of uniformity, initiated by the proof of Szemerédi's theorem [12], plays an important role in this area, and it was a major breakthrough when Gowers [2] introduced a new notion of uniformity in a Fourier-analytic proof for Szemerédi's theorem.

Gowers' work initiated an extension of the classical Fourier analysis, called higher-order Fourier analysis of Abelian groups. In this paper we are only interested in the case where the group is $\mathbb{F}_p^n$, where p is a fixed prime and n is large. 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; that is for $\alpha\in\mathbb{F}_p^n$, the corresponding character is defined as $\chi_\alpha(x)=\mathrm{e}_p(\sum_{i=1}^n \alpha_i x_i)$, where $\mathrm{e}_p(m):=e^{\frac{2\pi i}{p}m}$ for $m\in\mathbb{F}_p$. In higher-order Fourier analysis, the linear polynomials 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.

Higher-order Fourier expansions are extremely useful in studying averages that are defined through linear structures. To analyze the average in (3), one usually decomposes the function 1_A as f_1+f_2, where f_1 has a simple higher-order Fourier expansion, while f_2 is "quasi-random" meaning that it shares certain properties with a random function, and can be discarded as random noise. More precisely, $\|f_1\|_\infty\le 1$ and there is a small constant C such that $f_1=\sum_{i=1}^C c_i\mathrm{e}_p(P_i)$ where c_i are constants and P_i are low degree polynomials, and f_2 is quasi-random in the sense that for some proper constant k, its k-th Gowers uniformity norm $\|f_2\|_{U^k}$ is small.

The U^k norms increase as k increases, and thus the condition that $\|f_2\|_{U^k}$ is small becomes stronger. Therefore a question arises naturally: Given the average (2) and the decomposition $1_A=f_1+f_2$, what is the smallest k such that under the assumption that $\|f_2\|_{U^k}$ is sufficiently small, one can discard f_2 affecting the average only negligibly? This question was answered by Gowers and Wolf [4] in the case that f_1 is a constant function, provided that the field size p is not too small. In this work we extend their result to the case where f_1 is an arbitrary bounded function.

More concretely, a linear form $L=(\lambda_1,\dots,\lambda_k)\in\mathbb{F}_p^m$ maps every $\mathbf{x}=(x_1,\dots,x_k)\in(\mathbb{F}_p^n)^k$ to $L(\mathbf{x})=\sum_{i=1}^k \lambda_i x_i\in\mathbb{F}_p^n$. Let $\mathbb{D}$ denote the complex unit disk $\{z\in\mathbb{C}:|z|\le 1\}$. Gowers and Wolf [5]

defined the *true complexity* of a system of linear forms $\mathcal{L} = \{L_1, \dots, L_m\}$ as the minimal $d \geq 1$ such that the following holds: for every $\varepsilon > 0$, there exists $\delta > 0$ such that if $f : \mathbb{F}_p^n \to \mathbb{D}$ is a function with $\|f - \mathbb{E}[f]\|_{U^{d+1}} \leq \delta$, then

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

That is, as long as $\|f - \mathbb{E}[f]\|_{U^{d+1}}$ is small enough, we can approximate f by the constant function $\mathbb{E}[f]$, affecting the average only negligibly.

Gowers and Wolf [4] fully characterized the true complexity of a system of linear forms[1]. Let $L = (\lambda_1, \dots, \lambda_k) \in \mathbb{F}_p^k$ be a linear form in k variables. The dth tensor power of L is given by

$$L^d = \left(\prod_{j=1}^d \lambda_{i_j} : i_1, \dots, i_d \in [k] \right) \in \mathbb{F}_p^{k^d}.$$

Gower and Wolf [4] proved the following result, which characterizes the true complexity of a system of linear form as a simple linear algebraic condition.

Informal Statement (Theorem 3.5). *Provided that p is sufficiently large, the true complexity of a system $\mathcal{L} = \{L_1, \dots, L_m\}$ of linear forms is the minimal $d \geq 1$ such that $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent.*

In this work, we show that the true complexity in fact allows to approximate f by any other bounded function g, as long as $\|f - g\|_{U^{d+1}}$ is small. This removes the requirement that g is a constant function, which was present in the work of Gowers and Wolf. We already noted that any bounded function f can be decomposed as $f = f_1 + f_2$ where f_1 is "structured" and f_2 is "quasi-random" (see Theorem 4.7); our result thus shows that we can approximate f by f_1 without affecting averages significantly. In the context of sets, Gowers and Wolf's result allows one to handle only *uniform* sets A, for which $1_A - \mathbb{E}[1_A]$ is pseudorandom, while our result allows one to handle *all* sets.

Informal Statement (Theorem 3.8). *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms of true complexity d. If p is sufficiently large, then for every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $f, g : \mathbb{F}_p^n \to \mathbb{D}$ be functions such that $\|f - g\|_{U^{d+1}} \leq \delta$. Then*

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

More generally, one may consider averages over several functions. The average in (2) is the probability that for every $j \in \{1, \dots, m\}$, the linear combination $\sum_{i=1}^k \lambda_{j,i} X_i$ belongs to A. A more general case is the "off-diagonal" case, where instead of one subset A there are m subsets $A_1, \dots, A_m \subseteq \mathbb{F}_p^n$, and one is interested in estimating the probability that for every $j \in \{1, \dots, m\}$, we have $\sum_{i=1}^k \lambda_{j,i} X_i \in A_j$. Similar to the diagonal case, this can be expressed as an analytic average, and then one can decompose each function 1_{A_i} into structured and pseudorandom part $1_{A_i} = g_i + h_i$, and once again the question arises of what level of uniformity suffices for discarding

---

[1]Their result in fact requires the field size p not to be too small. Our results share the same requirement.

the pseudorandom parts without affecting the average significantly. Similar to the diagonal case, Gowers and Wolf [4] resolved this problem when all g_i are constant functions. In this work we extend also the off-diagonal case to the general case where g_i can be arbitrary bounded functions.

Informal Statement (Theorem 3.8). *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms of true complexity d. If p is sufficiently large, then for every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $f_i, g_i : \mathbb{F}_p^n \to \mathbb{D}$, $i \in [m]$ be functions such that $\|f_i - g_i\|_{U^{d+1}} \leq \delta$. Then*

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

It turns out that a more general phenomena holds, and all of the results above are immediate corollaries of the following theorem, which is the main technical contribution of this work. We show that in order to bound averages it suffices to have a single index i such that L_i^{d+1} is linearly independent of $\{L_j^{d+1} : j \neq i\}$ and that $\|f_i\|_{U^{d+1}}$ is small. This in particular resolves a conjecture of Gowers and Wolf [4].

Informal Statement (Theorem 3.9). *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms. Assume that L_1^{d+1} is not in the linear span of $L_2^{d+1}, \ldots, L_m^{d+1}$. If p is sufficiently large, 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 $\|f_1\|_{U^{d+1}} \leq \delta$, we have*

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

So far, we have only discussed which conditions allow discarding the pseudorandom terms. Note that after removing those terms, one arrives at an average of the following form

$$\mathbb{E}_{X_1, \ldots, X_k \in \mathbb{F}_p^n} \left[f_1 \left(\sum_{i=1}^k \lambda_{1,i} X_i \right) \ldots f_m \left(\sum_{i=1}^k \lambda_{m,i} X_i \right) \right], \tag{3}$$

where each f_j satisfies $\|f_j\|_\infty \leq 1$ and has a simple higher-order Fourier expansion. 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 [7] and Kaufman and Lovett [10] provide an approximate orthogonality that can be used to analyze averages such as $\mathbb{E}_{X \in \mathbb{F}_p^n}[f_1(X) \ldots f_m(X)]$ in a straightforward manner when proper higher-order Fourier expansions of $f_1, \ldots, f_m$ are known. However it is not a priori clear that these results can be applied to analyze more general averages of the form (3). In Lemmas 5.1, 5.2 and 5.3 we prove extensions of the results of Green and Tao [7] which are applicable to such general averages. These extensions allows us to approximate (6) with a simple formula in terms of the higher-order Fourier coefficients of $f_1, \ldots, f_m$.

Lemmas 5.1, 5.2 and 5.3 are quite useful, and in fact the proof of our main result, Theorem 3.9, heavily relies on them. We also apply these lemmas to prove an invariance result (Proposition 5.4 below), which is one of the key tools in our subsequent paper [9] which studies correlation testing for affine invariant properties on $\mathbb{F}_p^n$.

Paper organization We give some basic definitions and notations in Section 2. We discuss the complexity of systems of linear forms, and formally state our main theorems in Section 3. We give an overview of higher-order Fourier analysis in Section 4. We prove a strong orthogonality result in Section 5. We then use these to prove our main theorem, Theorem 3.9, in Section 6. We conclude with some open problems in Section 7.

2 Definitions and notations

For a natural number k, denote $[k] := \{1, \ldots, k\}$. The complex unit disk is denoted by $\mathbb{D} = \{z \in \mathbb{C} : |z| \le 1\}$. We will usually use the lower English letters x, y, z to denote elements of $\mathbb{F}_p^n$. For $x \in \mathbb{F}_p^n$, and $i \in [n]$, $x(i)$ denotes the i-th coordinate of x, i.e. $x = (x(1), \ldots, x(n))$. We frequently need to work with the elements of $(\mathbb{F}_p^n)^k$ which we regard as vectors with k coordinates. These elements are denoted with bold font e.g. $\mathbf{x} = (x_1, \ldots, x_k) \in (\mathbb{F}_p^n)^k$. For $1 \le i \le n$, e_i denotes the i-th standard vector in $\mathbb{F}_p^n$. Capital letters X, Y, *etc* are used to denote random variables. For an element $m \in \mathbb{F}_p$, we use the notation $\mathrm{e}_p(m) := e^{\frac{2\pi i}{p} m}$. We denote by f, g, h functions from $\mathbb{F}_p^n$ to $\mathbb{C}$. We denote n-variate polynomials over $\mathbb{F}_p^n$ by P, Q.

The *bias* of a function $f : \mathbb{F}_p^n \to \mathbb{C}$ is defined to be the quantity

$$\mathrm{bias}(f) := \left| \mathbb{E}_{X \in \mathbb{F}_p^n}[f(X)] \right|. \tag{4}$$

The inner product of two functions $f, g : \mathbb{F}_p^n \to \mathbb{C}$ is defined as

$$\langle f, g \rangle := \mathbb{E}_{X \in \mathbb{F}_p^n}[f(X)\overline{g(X)}]. \tag{5}$$

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

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

where $\mathbf{X}$ is a random variable taking values uniformly in $(\mathbb{F}_p^n)^k$.

Definition 2.1 (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 $(L_1(\mathbf{X}) + c, \ldots, L_m(\mathbf{X}) + c)$.*

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. Note that 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 2.2 (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 2.3. *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.*

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

Definition 2.4 (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*

$$\|f\|_{U^k}^{2^k} := \mathbb{E}\left[\prod_{S \subseteq [k]} \mathcal{C}^{k-|S|} f\left(X + \sum_{i \in S} Y_i\right)\right], \tag{7}$$

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 [2] 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 (7) 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 [2] for a proof).

3   Complexity of a system of linear forms

Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms in k variables. Note that if $A \subseteq \mathbb{F}_p^n$ and $1_A : \mathbb{F}_p^n \to \{0, 1\}$ is the indicator function of A, then $t_{\mathcal{L}}(1_A)$ is the probability that $L_1(\mathbf{X}), \ldots, L_m(\mathbf{X})$ all fall in A, where $\mathbf{X} \in (\mathbb{F}_p^n)^k$ is uniformly chosen. 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 $f := 1_A - \alpha$. We have

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

So, a sufficient condition for A to be pseudorandom with regards to $\mathcal{L}$ is that $t_{\{L_i : i \in S\}}(f) \approx 0$ for all nonempty subsets $S \subseteq [m]$. Green and Tao [8] 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 3.1 (Cauchy-Schwarz complexity)**.** *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms. The* Cauchy-Schwarz complexity *of $\mathcal{L}$ is the minimal s such that the following holds. For every $1 \leq i \leq m$, 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 [8], whose proof is based on a clever iterative application of the Cauchy-Schwarz inequality.

Lemma 3.2 ([8]). *Let $f_1, \ldots, f_m : \mathbb{F}_p \to \mathbb{D}$. Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of m linear forms in k variables of Cauchy-Schwarz complexity s. Then*

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

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 3.2.

Corollary 3.3. *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ 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 $1 \leq i \leq m$. Assume that $\|f_i - g_i\|_{U^{s+1}} \leq \frac{\varepsilon}{2m}$ for all $1 \leq i \leq m$. Then*

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

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 $\|1_A - \alpha\|_{U^{s+1}} \approx 0$, then $t_{\mathcal{L}}(1_A) \approx \alpha^m$.

3.1 The true complexity

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 [5] 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 3.4 (True complexity [5]). *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ 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 $\|f\|_{U^{d+1}} \leq \delta$, then*

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

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 [4] characterized the true complexity of systems of linear forms, assuming the field is not too small. For a linear form $L \in \mathbb{F}_p^k$, let $L^d \in \mathbb{F}_p^{k^d}$ be the dth tensor power of L. That is, if $L = (\lambda_1, \ldots, \lambda_k)$, then

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

Theorem 3.5 (Characterization of the true complexity of linear systems, Theorem 6.1 in [4]). *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms over $\mathbb{F}_p^n$ of Cauchy-Schwarz complexity $s \leq p$. The true complexity of $\mathcal{L}$ is the minimal d such that $L_1^{d+1}, \ldots, L_m^{d+1}$ are linearly independent over $\mathbb{F}_p$.*

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

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

Analogously to the case of a single set, let $f_i = 1_{A_i} - \alpha_i$. 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 [4], Gowers and Wolf showed that if $\mathcal{L}$ has true complexity d and $\|f_1\|_{U^{d+1}}, \ldots, \|f_m\|_{U^{d+1}}$ are small enough, then this stronger condition also holds.

Theorem 3.6 (Theorem 7.2 in [4]). *Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms over $\mathbb{F}_p^n$ of Cauchy-Schwarz complexity $s \leq p$ and true complexity d. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $f_1, \ldots, f_m : \mathbb{F}_p^n \to \mathbb{D}$ be functions such that $\|f_i\|_{U^{d+1}} \leq \delta$, for all $1 \leq i \leq m$. Then for all nonempty subsets $S \subseteq [m]$ we have*

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

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

Corollary 3.7 (Theorem 7.1 in [4]). *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 $f_i : \mathbb{F}_p^n \to \mathbb{D}$ for $1 \leq i \leq m$ be functions such that $\|f_i - \mathbb{E}[f_i]\|_{U^{d+1}} \leq \delta$. Then*

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

Note that Corollary 3.7 says if $\|f_i - \mathbb{E}[f_i]\|_{U^{d+1}}$ is small for all $i \in [m]$, then in the average $\mathbb{E}_{\mathbf{X}}[\prod_{i=1}^m f_i(L_i(\mathbf{X}))]$ one can replace the functions f_i with their expected values $\mathbb{E}[f_i]$ causing only an small error of ϵ. In other words if in the decomposition $f_i = \mathbb{E}[f_i] + (f_i - \mathbb{E}[f_i])$, the part $(f_i - \mathbb{E}[f_i])$ is sufficiently pseudorandom, then it is possible to discard it. As we shall see in Section 4.2, for every function $f_i : \mathbb{F}_p^n$, it is possible to find a "structured" function g_i, such that $f_i - g_i$ is pseudorandom in the sense that $\|f_i - g_i\|_{U^{d+1}}$ can be made arbitrarily small. However in the general case the function g_i will not necessarily be the constant function $\mathbb{E}[f_i]$. Hence it is important to obtain a version of Corollary 3.7 that can be applied in this general situation. We achieve this in the following theorem which qualitatively improves both Corollaries 3.3 and 3.7.

Theorem 3.8. *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 $1 \leq i \leq m$ be functions such that $\|f_i - g_i\|_{U^{d+1}} \leq \delta$. Then*

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

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

9

In fact, we prove a stronger result, from which Theorem 3.8 follows immediately. As Gowers and Wolf [4] correctly conjectured, in Theorem 3.6 the condition that the true complexity of $\mathcal{L}$ is at most d can be replaced by a much weaker condition. It suffices to assume that L_1^{d+1} linearly independent of $L_2^{d+1}, \ldots, L_m^{d+1}$. In fact it turns out that more is true[2], and the condition that all of $\|f_1\|_{U^{d+1}}, \ldots, \|f_m\|_{U^{d+1}}$ are small also can be replaced by the weaker condition that only $\|f_1\|_{U^{d+1}}$ is small.

Theorem 3.9 (Main theorem). *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 $f_1, \ldots, f_m : \mathbb{F}_p \to \mathbb{D}$ with $\|f_1\|_{U^{d+1}} \leq \delta$, we have*

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

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

Theorem 3.9 improves both Lemma 3.2 and Theorem 3.6: Let $s \leq p$ denote the Cauchy-Schwarz complexity of $\mathcal{L}$ in Theorem 3.9. Lemma 3.2 requires the stronger condition $\|f_1\|_{U^{s+1}} \leq \delta$, and Theorem 3.6 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 δ.

4 Higher-order 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 [11] 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 [2] 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 [5] 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 $\chi_\alpha(x) = e_p(\sum_{i=1}^n \alpha_i x_i)$. 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 $e_p(P)$, 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

---

[2]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 δ.

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

$$t_{\mathcal{L}}(f) = \sum \widehat{f}(\alpha_1)\dots\widehat{f}(\alpha_m), \tag{8}$$

where the sum is over all $\alpha_1, \dots, \alpha_m \in \mathbb{F}_p^n$ 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 (8) when Fourier expansion is replaced by a proper higher-order Fourier expansion.

4.1 Inverse theorems for Gowers uniformity norms

We start with some basic definitions.

Polynomials: Consider a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$. For an element $y \in \mathbb{F}_p^n$, define the *derivative* of f in the direction y as $\Delta_y f(x) = f(x + y) - f(x)$. Inductively we define $\Delta_{y_1,\dots,y_k}f = \Delta_{y_k}(\Delta_{y_1,\dots,y_{k-1}}f)$, for directions $y_1, \dots, y_k \in \mathbb{F}_p^n$. We say that f is a *polynomial of degree at most d*, if for every $y_1, \dots, y_{d+1} \in \mathbb{F}_p$, we have $\Delta_{y_1,\dots,y_{d+1}}f \equiv 0$. 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* $x_1^{i_1}\dots x_n^{i_n}$ where $0 \le i_1, \dots, i_n < p$ and $\sum_{j=1}^n i_j \le d$ 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 $P(x) := \sum c_{i_1,\dots,i_n} x_1^{i_1}\dots x_n^{i_n}$, where the sum is over all $1 \le i_1, \dots, i_n < p$ with $\sum_{j=1}^n i_j \le d$, and $c_{i_1,\dots,i_n}$ 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 $\widetilde{\Delta}_y f(x) = f(x + y)\overline{f(x)}$. Inductively we define $\widetilde{\Delta}_{y_1,\dots,y_k}f = \widetilde{\Delta}_{y_k}(\widetilde{\Delta}_{y_1,\dots,y_{k-1}}f)$, for directions $y_1, \dots, y_k \in \mathbb{F}_p^n$. A function $f : \mathbb{F}_p^n \to \mathbb{C}$ is called a *phase polynomial of degree at most d*, if for every $y_1, \dots, y_{d+1} \in \mathbb{F}_p$, we have $\widetilde{\Delta}_{y_1,\dots,y_{d+1}}f \equiv 1$. 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 $f : \mathbb{F}_p^n \to \mathbb{F}_p$, and every $y \in \mathbb{F}_p^n$, we have that

$$\widetilde{\Delta}_y e_p(f) = e_p(\Delta_y f).$$

This shows that if $f \in \mathrm{Poly}_d(\mathbb{F}_p^n)$, then $e_p(f)$ 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 4.1 (Lemma 1.2 in [14]). *Suppose that $0 \le d < p$. Every $f \in \mathcal{P}_d(\mathbb{F}_p^n)$ is of the form $f(x) = e_p(\theta + f(x))$, for some $\theta \in \mathbb{R}/\mathbb{Z}$, and $f \in \mathrm{Poly}_d(\mathbb{F}_p^n)$.*

When $d \ge p$, more complicated phase polynomials arise. Nevertheless obtaining a complete characterization is possible [13].

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:

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

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 [2] by iterated applications of the classical Cauchy-Schwarz inequality.

Lemma 4.2 (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*

$$\left| \mathbb{E}\left[\prod_{S \subseteq [k]} \mathcal{C}^{k-|S|} f_S\left(X + \sum_{i \in S} Y_i\right) \right] \right| \leq \prod_{S \subseteq [k]} \|f_S\|_{U^k}, \tag{9}$$

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

A simple application of Lemma 4.2 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 4.2, we obtain

$$|\mathbb{E}[f]| \leq \|f\|_{U^k}. \tag{10}$$

Equation (10) in particular shows that if $f, g : \mathbb{F}_p^n \to \mathbb{C}$, then one can bound their inner product by Gowers uniformity norms of $f\overline{g}$:

$$|\langle f, g \rangle| \leq \|f\overline{g}\|_{U^k}. \tag{11}$$

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

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

which in turn implies that $\|fg\|_{U^{d+1}} = \|f\|_{U^{d+1}}$. Let $\|f\|_{u(\mathcal{P}_d)} := \sup_{g \in \mathcal{P}_d} |\langle f, g \rangle|$ denote the maximal correlation of f with any phase polynomial of degree at most d. We conclude that

$$\|f\|_{u(\mathcal{P}_d)} \leq \|f\|_{U^{d+1}}. \tag{12}$$

This provides us with a "direct theorem" for the U^{d+1} norm: If $\|f\|_{u(\mathcal{P}_d)} \geq \varepsilon$, then $\|f\|_{U^{d+1}} \geq \varepsilon$. Recently Bergelson, Tao, and Ziegler [1, 14] established the corresponding inverse theorem in the high characteristic case. This gives the following result.

Theorem 4.3 ([1, 14]). *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 $\sup_{g \in \mathcal{P}_d} |\langle f, g \rangle| \geq \varepsilon$, then $\|f\|_{U^{d+1}} \geq \varepsilon$.*

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

Theorem 4.3 requires that $1 \leq d < p$. In this range, Lemma 4.1 shows that the phase polynomials of degree at most d can be described using polynomials of degree at most d. So Theorem 4.3 shows that in this case if $\|f\|_{U^{d+1}} \geq \varepsilon$, then there exists a polynomial $g : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree at most d such that $|\langle f, e_p(g) \rangle| \geq \delta(\varepsilon) > 0$. Tao and Ziegler [14] conjectured that Theorem 4.3 remains valid if one removes the condition $1 \leq d < p$. They proved the following partial result in this direction.

Theorem 4.4 ([1, 14]). *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 $\sup_{g \in \mathcal{P}_d} |\langle f, g \rangle| \geq \varepsilon$, then $\|f\|_{U^{d+1}} \geq \varepsilon$.*

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

4.2 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. They are discussed in this abstract form in [3]. In the following we will discuss decomposition theorems that follow from Theorem 4.3, but first we need to define polynomial factors on $\mathbb{F}_p^n$.

Definition 4.5 (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 4.6. *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 4.3 gives a simple decomposition theorem.

Theorem 4.7 (Decomposition Theorem [6]). *Let p be a fixed prime, $0 \le d < p$ be an integer, let $\varepsilon > 0$, and suppose that $n > n_0(p, d, \varepsilon)$ 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 $C_{\max}(p, d, \varepsilon)$ together with a decomposition*

$$f = f_1 + f_2,$$

where

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

We sketch the standard proof of Theorem 4.7 below, as we will need some extensions of it in this paper. For a full proof we refer the reader to [6].

Proof sketch. We create a sequence of polynomial factors $\mathcal{B}_1, \mathcal{B}_2, \ldots$ 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 $\|g_i\|_{U^{d+1}} \le \varepsilon$ we are done. Otherwise by Theorem 4.3, since $\|g_i\|_\infty \le 2$, there exists a polynomial $P_i \in \mathrm{Poly}_d(\mathbb{F}_p^n)$ such that $\langle g_i, e_p(P_i) \rangle \ge \delta(\varepsilon)$. Let $\mathcal{B}_{i+1} = \mathcal{B}_i \cup \{P_i\}$. The key point is that one can show

$$\|g_{i+1}\|_2^2 \le \|g_i - \langle g_i, e_p(P_i) \rangle\|_2^2 \le \|g_i\|_2^2 - \delta(\varepsilon).$$

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

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

$$f_1(x) = \sum_{\gamma \in \mathbb{F}_p^C} \widehat{\Gamma}(\gamma) \mathrm{e}_p \left(\sum_{i=1}^{C} \gamma(i) P_i(x) \right). \tag{13}$$

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 (13) 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 4.8 (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 $\{P_1, \ldots, P_t\}$ 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 4.9 (Rank). *We say a set of polynomials $\mathcal{P} = \{P_1, \ldots, P_t\}$ is of rank greater than r, and denote this by $\mathrm{rank}(\mathcal{P}) > r$, if the following holds. For any non-zero $\alpha = (\alpha_1, \ldots, \alpha_t) \in \mathbb{F}_p^t$, 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_α 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 $Q_1, \ldots, Q_r$ of degree at most $d - 1$, and a function $\Gamma : \mathbb{F}_p^r \to \mathbb{F}_p$ such that*

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

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 [7].

Lemma 4.10 (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 [10] connects the notion of the rank to the bias of a polynomial. It was proved first by Green and Tao [7] for the case $d < p$, and then extended by Kaufman and Lovett [10] for the general case.

Theorem 4.11 (Regularity [10]). *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 4.10 with Theorem 4.7, it is possible to obtain a strong decomposition theorem.

Theorem 4.12 (Strong Decomposition Theorem [6]). *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 function, and suppose that $n > n_0(p, d, \delta, r(\cdot))$ 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 [6].

Proof sketch. The proof follows the same steps as the proof of Theorem 4.7, 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

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

$\square$

Note that Theorem 4.12 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

$$\left| \mathbb{E} \left[e_p \left(\sum_{i=1}^{C} \gamma(i) P_i(X) \right) \right] \right| \leq 1/\omega(C), \tag{14}$$

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

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

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. $\square$

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.

15

Lemma 4.14 (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 function, 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*

$$f_i = h_i + h_i',$$

such that

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

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.

5 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 (13) that f can be expressed as a linear combination of $e_p(\sum_{i=1}^C \gamma(i) P_i(x))$, for $\gamma \in \mathbb{F}_p^C$. Furthermore if we require the polynomials to be of high rank, then we obtain an approximate orthogonality as in (14). 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 $\mathbb{E}[e_p(P(X+Y) - P(X) - P(Y))]$. 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 $\{P_1, \ldots, P_k\}$, and a set of linear forms $\{L_1, \ldots, L_m\}$. We need to be able to analyze exponential averages of the form:

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

where $\lambda_{i,j} \in \mathbb{F}_p$. Lemma 5.1 below shows that if $\{P_1, \ldots, P_k\}$ 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 $\{\chi_{y_1}, \ldots, \chi_{y_k}\}$ 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 5.1. *Fix p prime and $d < p$. Let $\{L_1, \ldots, L_m\}$ be a system of linear forms. Let $\mathcal{P} = \{P_1, \ldots, P_k\}$ be a collection of polynomials of degree at most d, such that $\mathrm{rank}(\mathcal{P}) > r_{p,d}(\varepsilon)$. For every set of coefficients $\Lambda = \{\lambda_{i,j} \in \mathbb{F}_p : i \in [k], j \in [m]\}$, and*

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

16

one of the following two cases holds:

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

Lemma 5.1 shows that in order to estimate $\mathrm{bias}(P_\Lambda)$ for polynomials of high rank $\mathcal{P} = \{P_1, \dots, P_k\}$, it suffices to determine whether P_Λ 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 5.2. *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 $d_1, \dots, d_k \in [d]$. Then 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 $d_1, \dots, d_k$:*
$$\sum_{i=1}^{k} \sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(\mathbf{x})) \not\equiv 0$$

Sometimes one needs 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 5.3. *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a set of* homogeneous *linear forms. Furthermore let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a collection of polynomials of degrees $d_1, \dots, d_k$ such that $\{P_i^{(d_i)} : i \in [k]\}$ are linearly independent over $\mathbb{F}_p$. Then for $\lambda_{i,j} \in \mathbb{F}_p$,*

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

Lemmas 5.1, 5.2 and 5.3 show that when studying the averages defined by systems of linear forms, under some homogeneity conditions (either for polynomials or for system of linear forms), high rank polynomials of the same degree behave in a similar manner. The following proposition captures this.

Proposition 5.4 (An invariance result). *Let $\mathcal{P} = \{P_1, \dots, P_k\}, \mathcal{Q} = \{Q_1, \dots, Q_k\}$ be two collections of polynomials over $\mathbb{F}_p^n$ of degree at most $d < p$ such that $\deg(P_i) = \deg(Q_i) = d_i$ for every $1 \leq i \leq k$. Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of 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*

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

17

and

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

Let $r_{p,d} : (0,1] \to \mathbb{N}$ be as in Lemma 5.1. Then for every $\varepsilon > 0$, if $\mathrm{rank}(\mathcal{P}), \mathrm{rank}(\mathcal{Q}) > r_{p,d}(\varepsilon)$, we have

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

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 $\mathcal{L}$ is homogenous.*

Proof. The Fourier expansion of Γ 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 5.1 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 ε. But Lemmas 5.2 and 5.3 show that under each one of the Conditions (i) or (ii), since $\deg(P_i) = \deg(Q_i)$, we have that

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

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

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

□

18

5.1 Proofs of Lemmas 5.1, 5.2 and 5.3

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 σ 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 $\ell_d(\mathbf{u})$ the number of distinct permutations[3] 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

$$P(L(\mathbf{x})) = \sum_{\mathbf{u} \in U^d} \ell_d(\mathbf{u}) c_d(\mathbf{u}, L) B(\mathbf{x_u}). \tag{15}$$

Note that $\ell_d(\mathbf{u})$ 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 5.5. *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

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

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 exists some assignment for a for which $\alpha := g(a) \neq 0$. $\square$

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

---

[3]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 $\frac{(i_1 + \ldots + i_\ell)!}{i_1! \ldots i_\ell!}$.

multi-linear form associated with P_i. Denoting $d_i := \deg(P_i)$ and using the notation of (15), we thus have

$$P_\Lambda(\mathbf{x}) = \sum_{i=1}^k \sum_{j=1}^m \lambda_{i,j} P_i(L_j(\mathbf{x})) = \sum_{i=1}^k \sum_{j=1}^m \lambda_{i,j} \sum_{\mathbf{u} \in U^{d_i}} \ell_{d_i}(\mathbf{u}) c_{d_i}(\mathbf{u}, L_j) B_i(\mathbf{x_u})$$

$$= \sum_{i=1}^k \sum_{\mathbf{u} \in U^{d_i}} b_i^{d_i}(\mathbf{u}) B_i(\mathbf{x_u}), \tag{16}$$

where

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

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 5.1, 5.2 and 5.3 first we prove Lemma 5.2 which has the simplest proof.

Proof of Lemma 5.2. 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_Λ 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 (16). Consider any assignment of the form $\mathbf{x} = (a_1 w, \ldots, a_k w)$. We have that

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

Hence we get that

$$P_\Lambda(a_1 w, \ldots, a_k w) = \sum_{i=1}^k \alpha_i P_i(w), \tag{17}$$

where $\alpha_i = \sum_{\mathbf{u} \in U^{d_i}} \mathbf{a^u} b_i^{d_i}(\mathbf{u})$. Applying Claim 5.5, 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 $P_\Lambda \not\equiv 0$. $\qquad \square$

The proof of Lemma 5.2 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 5.1 where we show that in this case under the stronger condition of high rank $\text{bias}(P_\Lambda)$ is small.

Proof of Lemma 5.1. 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 4.11) we get that $P_\Lambda(\mathbf{x})$ can be expressed as a function of $r \leq r_{p,d_{i_0}}(\varepsilon) \leq r_{p,d}(\varepsilon)$ 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

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

is a function of at most r polynomials of degree at most $d_{i_0} - 1$, where $\alpha_i = \sum_{\mathbf{u} \in U^{d_i}} \mathbf{a^u} b_i^{d_i}(\mathbf{u})$. By Claim 5.5 there exists a choice of $a_1, \ldots, a_k$ such that $\alpha_{i_0} \neq 0$, and this shows that $\text{rank}(\mathcal{P}) \leq r$. $\qquad \square$

Next we prove Lemma 5.3 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 5.3. It follows from Observation 2.3 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 (16) must be replaced by

$$
\begin{aligned}
P_\Lambda(\mathbf{x}) & = \sum_{i=1}^{k} \sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(\mathbf{x})) = \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{l=1}^{d_i} \lambda_{i,j} \sum_{\mathbf{u} \in U^l} \ell_l(\mathbf{u}) c_l(\mathbf{u}, L_j) B_i^l(\mathbf{x_u}) \\
& = \sum_{i=1}^{k} \sum_{l=1}^{d_i} \sum_{\mathbf{u} \in U^l} b_i^l(\mathbf{u}) B_i^l(\mathbf{x_u}),
\end{aligned}
\tag{18}
$$

where

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

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 $b_i^t(\mathbf{u}) = \ell_t(\mathbf{u}) \sum_{j=1}^{m} \lambda_{i,j} c_t(\mathbf{u}, L_j) \neq 0$, 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, $\ell_l(\mathbf{u}) \neq 0$ for every $1 \leq l \leq d_i$ and $\mathbf{u} \in U^l$. Hence we conclude that

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

which contradicts the maximality of t. $\qquad \square$

6 Proof of the main theorem

In this section we prove Theorem 3.9. For the convenience of the reader, we restate the theorem.

Theorem 3.9 (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 $f_1, \ldots, f_m : \mathbb{F}_p \to \mathbb{D}$ with $\|f_1\|_{U^{d+1}} \leq \delta$, we have*

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

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

If $d \geq s$, then Theorem 3.9 follows from Lemma 3.2. Thus we assume $d < s$, hence also $d < p$. We will assume throughout the proof that p, m, s, d, ε 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 4.14 there exists a polynomial factor $\mathcal{B}$ of degree s, complexity $C \leq C_{\max}(\eta, r(\cdot))$ 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 $\|h_i'\|_{U^{s+1}} \leq \eta$. Trivially $\|h_i\|_\infty \leq 1$ and $\|h_i'\|_\infty \leq 2$. 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 η is chosen to be small enough.

Claim 6.1. *If $\eta \leq \frac{\varepsilon 4^{-m}}{2}$ then*

$$\left| \mathbb{E}\left[\prod_{i=1}^m f_i(L_i(\mathbf{X}))\right] - \mathbb{E}\left[\prod_{i=1}^m h_i(L_i(\mathbf{X}))\right] \right| \leq \varepsilon/2,$$

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 3.2 that

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

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

We thus set $\eta = \frac{\varepsilon 4^{-m}}{2}$, and regard η from now on as a constant, and we do not specify explicitly dependencies on η as well.

Let $\{P_i\}_{1 \leq i \leq C}$ be the polynomials which define the polynomial factor $\mathcal{B}$, where we assume each P_i is homogeneous of degree $\deg(P_i) \leq s$. 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 Γ_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 $|c_{i,\gamma}| \leq 1$. 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

$$\mathbb{E}\left[\prod_{i=1}^m h_i(L_i(\mathbf{X}))\right] = \sum_{\gamma_1, \ldots, \gamma_m \in \mathbb{F}_p^C} \Delta(\gamma_1, \ldots, \gamma_m), \tag{19}$$

22

where
$$\Delta(\gamma_1, \ldots, \gamma_m) = \left(\prod_{i=1}^m c_{i,\gamma_i}\right) \mathbb{E}\left[e_p\left(P_{\gamma_1}(L_1(\mathbf{X})) + \ldots + P_{\gamma_m}(L_m(\mathbf{X}))\right)\right]. \tag{20}$$

We will bound each term $\Delta(\gamma_1, \ldots, \gamma_m)$ by $\tau := \tau(C) = p^{-mC}\varepsilon/2$, 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 6.2. *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
$$c_{1,\gamma} = \mathbb{E}\left[h_1(X)e_p(-P_\gamma(X))\right] - \sum_{\gamma' \in \mathbb{F}_p^C, \gamma' \ne \gamma} c_{1,\gamma'}\mathbb{E}\left[e_p(P_{\gamma'}(X) - P_\gamma(X))\right],$$

where the averages are over uniform $X \in \mathbb{F}_p^n$. We first bound $\mathbb{E}\left[h_1(X)e_p(-P_\gamma(X))\right]$. Using the fact that $h_1 = \mathbb{E}(f_1|\mathcal{B})$ and that the function $e_p(-P_\gamma(x))$ is $\mathcal{B}$-measurable, we have by Observation 4.6 that
$$\left|\mathbb{E}\left[h_1(X)e_p(-P_\gamma(X))\right]\right| = \left|\mathbb{E}\left[f_1(X)e_p(-P_\gamma(X))\right]\right| \le \|f_1\|_{U^{d+1}} \le \delta.$$

Hence, by choosing $\delta < p^{-mC_{\max}(r(\cdot))}\varepsilon/4$ we guarantee that $\left|\mathbb{E}\left[h_1(X)e_p(-P_\gamma(X))\right]\right| \le \delta < \tau/2$. Next for $\gamma \ne \gamma'$, we bound each term $\mathbb{E}\left[e_p(P_{\gamma'}(X) - P_\gamma(X))\right]$ by $\tau p^{-C}/2$. Assume that for some $\gamma' \ne \gamma$ we have
$$\left|\mathbb{E}\left[e_p(P_{\gamma'}(X) - P_\gamma(X))\right]\right| > \tau p^{-C}/2.$$
Then by Theorem 4.11 we have that
$$\operatorname{rank}(P_{\gamma'} - P_\gamma) \le r_{p,s}(\tau p^{-C}/2) = r_1(C).$$

Thus, as long as we choose $r(C) > r_1(C)$ for all $C \in \mathbb{N}$, we have that
$$\sum_{\gamma' \ne \gamma} \left|c_{1,\gamma'}\mathbb{E}\left[e_p(P_{\gamma'}(X) - P_\gamma(X))\right]\right| \le \tau/2$$

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

Consider now any $\gamma_1 \notin S$. We will show that if we choose $r(\cdot)$ large enough we can guarantee that
$$\left|\mathbb{E}\left[e_p\left(P_{\gamma_1}(L_1(\mathbf{X})) + \ldots + P_{\gamma_m}(L_m(\mathbf{X}))\right)\right]\right| \le \tau, \tag{21}$$

which will establish the result. Assume that this is not the case for some $\gamma_1 \notin S$ and $\gamma_2, \ldots, \gamma_m \in \mathbb{F}_p^C$. By Lemma 5.1 there is a rank $r_{p,s}(\tau) = r_2(C)$ such that if we guarantee that $r(C) > r_2(C)$ for all $C \in \mathbb{N}$ and if (21) does not hold, then we must have
$$P_{\gamma_1}(L_1(\mathbf{x})) + \ldots + P_{\gamma_m}(L_m(\mathbf{x})) \equiv 0. \tag{22}$$

23

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

$$P_{\gamma_1}^{(t)}(L_1(\mathbf{x})) + \ldots + P_{\gamma_m}^{(t)}(L_m(\mathbf{x})) \equiv 0. \tag{23}$$

The following claim concludes the proof. It shows that if (23) holds, then L_1^t is linearly dependent on $L_2^t, \ldots, L_m^t$. This immediately implies that also L_1^{d+1} is linearly dependent on $L_2^{d+1}, \ldots, L_m^{d+1}$ (since $t \geq d + 1$) which contradicts our initial assumption.

Claim 6.3. *Let $P_1, \ldots, P_m$ 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 $L_2^t, \ldots, L_m^t$.

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 α_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 $\mathbf{x} = (x_1, \ldots, x_k)$ 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 $x_i = (z(i), \ldots, z(i))$ where $z(1), \ldots, z(k) \in \mathbb{F}_p$ 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

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

$\square$

7   Summary and open problems

We study the complexity of structures defined by linear forms. Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms of Cauchy-Schwarz complexity s. Our main technical contribution is that as long as $s \leq p$, if L_1^{d+1} is not linearly dependent on $L_2^{d+1}, \ldots, L_m^{d+1}$, then averages $\mathbb{E}[\prod_{i=1}^{m} f_i(L_i(\mathbf{X}))]$ are controlled by $\|f_1\|_{U^{d+1}}$. We conjecture that such a result should hold for *all* systems of linear forms, removing the dependence on the Cauchy-Schwarz complexity of the system.

Conjecture 7.1. *Fix a prime p and $d \geq 1$. Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms over $\mathbb{F}_p$ such that L_1^{d+1} is not linearly dependent on $L_2^{d+1}, \ldots, L_m^{d+1}$. Then for every $\varepsilon > 0$ there exists $\delta = \delta(p, m, d, \varepsilon)$ such that if $f_i : \mathbb{F}_p^n \to \mathbb{D}$ are functions such that $\|f_1\|_{U^{d+1}} \leq \delta$ then*

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

The main technical ingredient that seems to be required to extend our techniques to prove Conjecture 7.1 is an inverse Gowers theorem for small fields which achieves the correct degree. Another ingredient is a generalization of Lemmas 5.1, 5.2 and 5.3 to small fields.

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] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[3] W. T. Gowers. Decompositions, approximate structure, transference, and the hahn-banach theorem. *Bull. London Math. Soc.*, 42(4):573–606, 2010.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] Ben Green and Terence Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010.

[9] Hamed Hatami and Shachar Lovett. Correlation testing for affine invariant properties on $\mathbb{F}_p^n$ in the high error regime. *manuscript*.

[10] 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.

[11] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[12] 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.

[13] 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.

[14] 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.