Advances in Mathematics 292 (2016) 446–477

Contents lists available at ScienceDirect

Advances in Mathematics

www.elsevier.com/locate/aim

General systems of linear forms: Equidistribution and true complexity

Hamed Hatami [a,1], Pooya Hatami [b,2], Shachar Lovett [c,*,3]

[a] *McGill University, Canada*
[b] *Institute for Advanced Study, United States*
[c] *UC San Diego, United States*

ARTICLE INFO

Article history:
Received 22 July 2014
Received in revised form 11 November 2015
Accepted 27 January 2016
Communicated by Benjamin Sudakov

Keywords:
Higher-order Fourier analysis
Additive combinatorics
Linear patterns
Nonclassical polynomials

ABSTRACT

Higher-order Fourier analysis is a powerful tool that can be used to analyze the densities of linear systems (such as arithmetic progressions) in subsets of Abelian groups. We are interested in the group F_p^n, for fixed p and large n, where it is known that analyzing these averages reduces to understanding the joint distribution of a family of sufficiently pseudorandom (formally, high-rank) nonclassical polynomials applied to the corresponding system of linear forms.

In this work, we give a complete characterization for these distributions for arbitrary systems of linear forms. This extends previous works which accomplished this in some special cases. As an application, we resolve a conjecture of Gowers and Wolf on the true complexity of linear systems. Our proof deviates from that of the previously known special cases and requires several new ingredients. One of which, which may be of independent interest, is a new theory of homogeneous nonclassical polynomials.

[*] Corresponding author.
E-mail addresses: hatami@cs.mcgill.ca (H. Hatami), pooyahat@math.ias.edu (P. Hatami), slovett@ucsd.edu (S. Lovett).

[1] Supported by an NSERC, and an FQRNT grant.

[2] Part of this work was done when the author was a student at University of Chicago. Research supported by the NSF grant No. CCF-1412958.

[3] Supported by NSF CAREER award 1350481.

http://dx.doi.org/10.1016/j.aim.2016.01.019

1. Introduction

Gowers' seminal work in combinatorial number theory [5] initiated an extension of the classical Fourier analysis, called *higher-order Fourier analysis* of Abelian groups. Higher-order Fourier analysis has been very successful in dealing with problems regarding the densities of small linear structures (e.g. arithmetic progressions) in subsets of Abelian groups. It is possible to express such densities as certain analytic averages. For example, the density of the three term arithmetic progressions in a subset A of an Abelian group G can be expressed as $\mathbf{E}_{x,y\in G}\left[\mathbf{1}_A(x)\mathbf{1}_A(x+y)\mathbf{1}_A(x+2y)\right]$. More generally, one is often interested in analyzing

$$\mathop{\mathbf{E}}_{x_1,\dots,x_k\in G}\big[\mathbf{1}_A(L_1(x_1,\dots,x_k))\cdots\mathbf{1}_A(L_m(x_1,\dots,x_k))\big], \tag{1}$$

where each L_i is a linear form on k variables. Averages of this type are of interest in computer science, additive combinatorics, and analytic number theory.

In this paper we are only interested in the group $\mathbb{F}^n$ where $\mathbb{F}=\mathbb{F}_p$ for a fixed prime p and n is large. In the classical Fourier analysis of $\mathbb{F}^n$, a function is expressed as a linear combination of the characters of $\mathbb{F}^n$. Note that the characters of $\mathbb{F}^n$ are exponentials of linear polynomials: for $\alpha\in\mathbb{F}^n$, the corresponding character is defined as $\chi_\alpha(x)=\mathsf{e}\left(1/p\cdot\sum_{i=1}^n\alpha_i x_i\right)$, where $\mathbb{T}=\mathbb{R}/\mathbb{Z}$ is the torus and $\mathsf{e}:\mathbb{T}\to\mathbb{C}$ is given by $\mathsf{e}(a)=\exp(2\pi i\cdot a)$. In higher-order Fourier analysis, the linear polynomials are replaced by higher degree "nonclassical polynomials" (a generalization of classical polynomials), and one would like to approximate a function $f:\mathbb{F}^n\to\mathbb{C}$ by a linear combination of such higher-order terms. A thorough treatment of nonclassical polynomials is presented in Section 2.1. For now, it suffices to say that a degree-d nonclassical polynomial of depth $k\geqslant 0$ is given by a function $P:\mathbb{F}^n\to\mathbb{U}_{k+1}$, where $\mathbb{U}_{k+1}=\frac{1}{p^{k+1}}\mathbb{Z}/\mathbb{Z}$ is a discrete subgroup of $\mathbb{T}$, such that P vanishes after taking $d+1$ additive derivatives. The case $k=0$ corresponds to classical polynomials. The existence of such approximations is a consequence of the so-called "inverse theorems" for Gowers norms which are established in a sequence of papers by Bergelson, Green, Samorodnitsky, Szegedy, Tao, and Ziegler [19,20,17,13,11,1,16].

Higher-order Fourier expansions are extremely useful in studying averages that are defined through linear structures. To analyze the average in (1), one approximates $1_A\approx\Gamma(P_1,\dots,P_C)$ where C is a constant which depends only on the approximation guarantees, $P_1,\dots,P_C$ are bounded degree polynomials of corresponding depths $k_1,\dots,k_C$, and $\Gamma:\prod_{i=1}^C\mathbb{U}_{k_i+1}\to\mathbb{R}$ is some composition function. Then applying the classical Fourier transform to Γ yields the higher-order Fourier expansion

$$\mathbf{1}_A\approx\Gamma(P_1,\dots,P_C)\approx\sum_\alpha\widehat{\Gamma}(\alpha)\,\mathsf{e}\left(\sum_{i=1}^C\alpha_i P_i\right),$$

where the coefficients $\widehat{\Gamma}(\alpha)$ are complex numbers, and α takes values in $\prod_{i=1}^C\mathbb{Z}_{p^{k_i+1}}\cong\prod_{i=1}^C\mathbb{U}_{k_i+1}$.

Counting linear patterns and equidistribution. One of the important and useful properties of the classical Fourier characters is that they form an orthonormal basis. For higher-order Fourier expansions to be useful, one needs a similar orthogonality for the higher-order characters $\mathsf{e}\left(\sum_{i=1}^{n} \alpha_i P_i\right)$ appearing in the expansion, or at least an approximation of it. This approximate orthogonality is established by Green and Tao [10] and Kaufman and Lovett [15], where it is shown that the joint distribution of a high-rank family of polynomials is nearly equidistributed: the polynomials in the approximation $1_A \approx \Gamma(P_1, \dots, P_C)$ can be chosen in such a way that the distribution of $(P_1(x), \dots, P_C(x))$ is close to the uniform distribution on their range $\prod_{i=1}^{C} \mathbb{U}_{k_i+1}$ when x is chosen uniformly at random from $\mathbb{F}^n$.

However, this is not completely satisfactory, as to study the averages of the form (1), one needs to understand the distribution of the more sophisticated random variable

$$A_{\mathcal{P},\mathcal{L}}(X) := \begin{pmatrix} P_1(L_1(X)) & P_2(L_1(X)) & \dots & P_C(L_1(X)) \\ P_1(L_2(X)) & P_2(L_2(X)) & \dots & P_C(L_2(X)) \\ \vdots & & & \vdots \\ P_1(L_m(X)) & P_2(L_m(X)) & \dots & P_C(L_m(X)) \end{pmatrix},$$

where $X = (x_1, \dots, x_k)$ is the uniform random variable taking values in $(\mathbb{F}^n)^k$, $\mathcal{P} = \{P_1, ..., P_C\}$ is a family of nonclassical polynomials and $\mathcal{L} = \{L_1, ..., L_m\} \subset \mathbb{F}^k$ is an arbitrary system of linear forms.

Since (nonclassical) polynomials of a given degree satisfy various linear identities (e.g. every degree-one polynomial P satisfies $P(x+y+z) = P(x+y) + P(x+z) - P(x)$), it is no longer possible to choose the polynomials in a way that this random matrix is almost uniformly distributed over all $m \times C$ matrices $A \in \prod_{i=1}^{m} \prod_{j=1}^{C} \mathbb{U}_{k_j+1}$. Therefore, in this case one would like to obtain an almost uniform distribution on the points of the configuration space that are consistent with these linear identities. Note that [10,15] address only the case where the system of linear forms consists only of a single linear form, in other words they show that the entries in each row of this matrix are nearly independent. The stronger near-equidistribution would in particular imply that the columns of this matrix are nearly independent, with each column uniform modulo the required linear dependencies. The previous works leading to inverse theorems for Gowers norms allow one to handle only the special case when the system of linear forms corresponds to a constant dimensional parallelepiped.

Hatami and Lovett [14] established this strong near-equidistribution in the case where the characteristic of the field $\mathbb{F}$ is greater than the degree of the involved polynomials, however their proof technique completely breaks down once $|\mathbb{F}|$ is allowed to be small. Bhattacharyya et al. [2] later extended the result of [14] to the general characteristic case, but under the extra assumption that the system of linear forms is affine, i.e. there is a variable that appears with coefficient 1 in all the linear forms. Finally, in the present paper we fully characterize the joint distribution of the matrix $A_{\mathcal{P},\mathcal{L}}$ without any extra assumptions on the linear forms.

Theorem. *(Theorem 3.12 – informal.) Let $\mathcal{P} = \{P_1, ..., P_C : \mathbb{F}^n \to \mathbb{T}\}$ be a sufficiently high-rank family of "homogeneous" nonclassical polynomials of degrees $d_1, ..., d_C$ and depths $k_1, ..., k_C$, respectively, and let $\mathcal{L} = \{L_1, ..., L_m\} \subset \mathbb{F}^k$ be a system of linear forms. For a uniformly chosen $X \in (\mathbb{F}^n)^k$, the matrix $A_{\mathcal{P},\mathcal{L}}(X)$ is distributed almost uniformly over all matrices in $\prod_{i=1}^m \prod_{j=1}^C \mathbb{U}_{k_j+1}$ whose columns satisfy the linear dependencies that are imposed by the degree and depth of the polynomial corresponding to the column.*

Homogeneous nonclassical polynomials are another contribution of the current work. They are introduced in Section 3.1 where we also prove that homogeneity can be assumed without any loss of generality.

Algebraic property testing. One of the examples where understanding the joint distribution of polynomials over linear structures is useful is the study of testability of algebraic properties. In [2], the authors use higher-order Fourier analysis and their near-equidistribution result for affine systems of linear forms in order to give a characterization of affine-invariant testable properties. An interesting future project would be to use our stronger near-equidistribution theorem to generalize such results to the linear-invariant properties.

A conjecture of Gowers and Wolf. In dealing with the averages of the form (1) a question arises naturally: Given such an average, what is the smallest k such that there is an approximation of $\mathbf{1}_A$ with a linear combination of a few higher-order characters of degree at most k that affects the average only negligibly? This question was asked and studied by Gowers and Wolf [7] who conjectured a simple characterization for this value, and verified it for the case of large $|\mathbb{F}|$ in [8]. Since Gowers and Wolf's work several of the tools in the framework of higher-order Fourier analysis have been extended to the low characteristic case, such as inverse theorems for Gowers norms by Bergelson, Tao and Ziegler [1]. However, it turns out that simply plugging in the extended tools in the arguments from [7] does not suffice to establish the Gowers–Wolf conjecture. As an application of our near-orthogonality result, we settle the Gowers–Wolf conjecture in full generality on $\mathbb{F}^n$. As such, our work extends previous work by Hatami and Lovett [14] who proved it under the assumption that the field is sufficiently large; and by Green and Tao [11] who established similar results and characterizations in the setting of functions on $\mathbb{Z}_N$.

Proof ingredients. In order to prove our near-orthogonality result, we use a derivative technique where we repeatedly take the derivative of a linear combination of the family of polynomials over linear forms in certain directions so that we can kill unwanted terms. The aim is to reduce to the case of a single polynomial applied to a specific system of linear forms where we know how to solve the problem using standard techniques. Such a derivative technique was introduced in previous works in this area, however previous approaches could only apply it when the field was large enough [14] or when the system of linear forms was affine [2]. In this work, we extend it to handle any system of linear forms. We achieve that by developing a new theory of homogeneous nonclassical polynomials. We show that homogeneous nonclassical polynomials exist, that they form a basis for all nonclassical polynomials, and that the derivative technique can be extended to the case of high-rank homogeneous nonclassical polynomials applied to arbitrary systems of linear forms.

Homogeneous nonclassical polynomials. As mentioned above the main difficulty in dealing with fields of low characteristic is that in the higher-order Fourier expansions, instead of the exponentials of classical polynomials, one has to work with exponentials of a generalization of them which are referred to as "nonclassical" polynomials. Recall that a classical polynomial is homogeneous if all of its monomials are of the same degree. A useful property of a homogeneous classical polynomial $P(x)$ of degree d is that $P(cx) = c^d P(x)$, for every $c \in \mathbb{F}$. We use this property to extend the definition of homogeneity to nonclassical polynomials. An ingredient of the proof of our near-equidistribution result is a statement about nonclassical polynomials which we believe is of independent interest. In Theorem 3.4 we show that homogeneous multivariate (nonclassical) polynomials span the space of multivariate (nonclassical) polynomials. We later use this to prove our near-equidistribution results for homogeneous polynomials.

2. Notation and preliminaries

Fix a prime field $\mathbb{F} = \mathbb{F}_p$ for a prime $p \geqslant 2$. Throughout the paper, we fix $\zeta \in \mathbb{F}^*$, a generator of $\mathbb{F}^*$. Define $|\cdot|$ to be the standard map from $\mathbb{F}$ to $\{0, 1, \dots, p-1\} \subset \mathbb{Z}$. Let $\mathbb{D}$ denote the complex unit disk $\{z \in \mathbb{C} : |z| \leqslant 1\}$.

For integers a, b, we let $[a]$ denote the set $\{1, 2, \dots, a\}$ and let $[a, b]$ denote the set $\{a, a+1, \dots, b\}$. For real numbers, $\alpha, \sigma, \varepsilon$, we use the shorthand $\sigma = \alpha \pm \varepsilon$ to denote $\alpha - \varepsilon \leqslant \sigma \leqslant \alpha + \varepsilon$. The zero element in $\mathbb{F}^n$ is denoted by $\underline{0}$. We will denote by lower-case letters, e.g. x, y, elements of $\mathbb{F}^n$. We use capital letters, e.g. $X = (x_1, \dots, x_\ell) \in (\mathbb{F}^n)^\ell$, to denote tuples of variables.

Definition 2.1. A linear form in ℓ variables is a vector $L = (\lambda_1, \dots, \lambda_\ell) \in \mathbb{F}^\ell$ and it maps $X = (x_1, \dots, x_\ell) \in (\mathbb{F}^n)^\ell$ to $L(X) = \sum_{i=1}^{\ell} \lambda_i x_i \in \mathbb{F}^n$.

For a linear form $L = (\lambda_1, \dots, \lambda_\ell) \in \mathbb{F}^\ell$ we define $|L| \overset{\text{def}}{=} \sum_{i=1}^{\ell} |\lambda_i|$.

2.1. Higher-order Fourier analysis

We need to recall some definitions and results about higher-order Fourier analysis. Most of the material in this section is directly quoted from the full version of [2].

Definition 2.2 *(Multiplicative derivative).* Given a function $f : \mathbb{F}^n \to \mathbb{C}$ and an element $h \in \mathbb{F}^n$, define the *multiplicative derivative in direction* h of f to be the function $\Delta_h f : \mathbb{F}^n \to \mathbb{C}$ satisfying $\Delta_h f(x) = f(x+h)\overline{f(x)}$ for all $x \in \mathbb{F}^n$.

The *Gowers norm* of order d for a function $f : \mathbb{F}^n \to \mathbb{C}$ is the expected multiplicative derivative of f in d random directions at a random point.

Definition 2.3 *(Gowers norm).* Given a function $f : \mathbb{F}^n \to \mathbb{C}$ and an integer $d \geqslant 1$, the *Gowers norm of order* d for f is given by

$$\|f\|_{U^d} = \left| \mathop{\mathbf{E}}_{y_1,\dots,y_d,x\in\mathbb{F}^n} [(\Delta_{y_1}\Delta_{y_2}\cdots\Delta_{y_d} f)(x)] \right|^{1/2^d}.$$

Note that as $\|f\|_{U^1} = |\mathbf{E}[f]|$ the Gowers norm of order 1 is only a semi-norm. However for $d > 1$, it is not difficult to show that $\|\cdot\|_{U^d}$ is indeed a norm.

If $f = e^{2\pi i P/p}$ where $P : \mathbb{F}^n \to \mathbb{F}$ is a polynomial of degree $< d$, then $\|f\|_{U^d} = 1$. If $d < p$ and $\|f\|_\infty \leqslant 1$, then in fact, the converse holds, meaning that any function $f : \mathbb{F}^n \to \mathbb{C}$ satisfying $\|f\|_\infty \leqslant 1$ and $\|f\|_{U^d} = 1$ is of this form. But when $d \geqslant p$, the converse is no longer true. In order to characterize functions $f : \mathbb{F}^n \to \mathbb{C}$ with $\|f\|_\infty \leqslant 1$ and $\|f\|_{U^d} = 1$, one needs to define the notion of *nonclassical polynomials.*

Nonclassical polynomials are not $\mathbb{F}$-valued. We need to introduce some notation. Let $\mathbb{T}$ denote the circle group $\mathbb{R}/\mathbb{Z}$. This is an Abelian group with group operation denoted $+$. For an integer $k \geqslant 0$, consider the subgroup $\frac{1}{p^k}\mathbb{Z}/\mathbb{Z} \subseteq \mathbb{T}$. Let $\mathsf{e} : \mathbb{T} \to \mathbb{C}$ denote the character $\mathsf{e}(x) = e^{2\pi i x}$.

Definition 2.4 *(Additive derivative).* Given a function[4] $P : \mathbb{F}^n \to \mathbb{T}$ and an element $h \in \mathbb{F}^n$, define the *additive derivative in direction* h of f to be the function $D_h P : \mathbb{F}^n \to \mathbb{T}$ satisfying $D_h P(x) = P(x+h) - P(x)$ for all $x \in \mathbb{F}^n$.

Definition 2.5 *(Nonclassical polynomials).* For an integer $d \geqslant 0$, a function $P : \mathbb{F}^n \to \mathbb{T}$ is said to be a *nonclassical polynomial of degree* $\leqslant d$ (or simply a *polynomial of degree* $\leqslant d$) if for all $y_1, \dots, y_{d+1}, x \in \mathbb{F}^n$, it holds that

$$(D_{y_1} \cdots D_{y_{d+1}} P)(x) = 0. \tag{2}$$

The *degree* of P is the smallest d for which the above holds. A function $P : \mathbb{F}^n \to \mathbb{T}$ is said to be a *classical polynomial of degree* $\leqslant d$ if it is a nonclassical polynomial of degree $\leqslant d$ whose image is contained in $\frac{1}{p}\mathbb{Z}/\mathbb{Z}$.

[4] We try to adhere to the following convention: upper-case letters (e.g. F and P) denote functions mapping from $\mathbb{F}^n$ to $\mathbb{T}$ or to $\mathbb{F}$, lower-case letters (e.g. f and g) denote functions mapping from $\mathbb{F}^n$ to $\mathbb{C}$, and upper-case Greek letters (e.g. Γ and Σ) denote functions mapping $\mathbb{T}^C$ to $\mathbb{T}$.

It is a direct consequence of the definition that a function $f : \mathbb{F}^n \to \mathbb{C}$ with $\|f\|_\infty \leqslant 1$ satisfies $\|f\|_{U^{d+1}} = 1$ if and only if $f = \mathsf{e}(P)$ for a (nonclassical) polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$. We denote by $\mathrm{Poly}(\mathbb{F}^n \to \mathbb{T})$ and $\mathrm{Poly}_{\leqslant d}(\mathbb{F}^n \to \mathbb{T})$, respectively, the set of all nonclassical polynomials, and the ones of degree at most d.

The following lemma of Tao and Ziegler [20] shows that a classical polynomial P of degree d must always be of the form $x \mapsto \frac{|Q(x)|}{p}$, where $Q : \mathbb{F}^n \to \mathbb{F}$ is a polynomial (in the usual sense) of degree d, and $|\cdot|$ is the standard map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\}$. This lemma also characterizes the structure of nonclassical polynomials.

Lemma 2.6. *(Lemma 1.7 in [20].) A function $P : \mathbb{F}^n \to \mathbb{T}$ is a nonclassical polynomial of degree $\leqslant d$ if and only if P can be represented as*

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

for a unique choice of $c_{d_1, \ldots, d_n, k} \in \{0, 1, \ldots, p-1\}$ and $\alpha \in \mathbb{T}$. The element α is called the shift *of P, and the largest integer k such that there exist $d_1, \ldots, d_n$ for which $c_{d_1, \ldots, d_n, k} \neq 0$ is called the* depth *of P. A depth-k polynomial P takes values in a coset of the subgroup $\mathbb{U}_{k+1} \overset{\text{def}}{=} \frac{1}{p^{k+1}}\mathbb{Z}/\mathbb{Z}$. Classical polynomials correspond to polynomials with 0 shift and 0 depth.*

Note that Lemma 2.6 immediately implies the following important observation[5]:

Remark 2.7. If $Q : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree d and depth k, then pQ is a polynomial of degree $\max(d - p + 1, 0)$ and depth $k - 1$. In other words, if Q is classical, then pQ vanishes, and otherwise, its degree decreases by $p - 1$ and its depth by 1. Also, if $\lambda \in [1, p-1]$ is an integer, then $\deg(\lambda Q) = d$ and $\mathrm{depth}(\lambda Q) = k$.

For convenience of exposition, we will assume throughout this paper that the shifts of all polynomials are zero. This can be done without affecting any of the results in this work. Hence, all polynomials of depth k take values in $\mathbb{U}_{k+1}$.

Given a degree-d nonclassical polynomial P, it is often useful to consider the properties of its d-th derivative. Motivated by this, we give the following definition.

Definition 2.8 *(Derivative polynomial).* Let $P : \mathbb{F}^n \to \mathbb{T}$ be a degree-d polynomial, possibly nonclassical. Define the derivative polynomial $\partial P : (\mathbb{F}^n)^d \to \mathbb{T}$ by the following formula:

$$\partial P(h_1, \ldots, h_d) \overset{\text{def}}{=} D_{h_1} \cdots D_{h_d} P(0),$$

[5] Recall that $\mathbb{T}$ is an additive group. If $n \in \mathbb{Z}$ and $x \in \mathbb{T}$, then nx is shorthand for $x + \cdots + x$ if $n \geqslant 0$ and $-x - \cdots - x$ otherwise, where there are $|n|$ terms in both expressions.

where $h_1, \dots, h_d \in \mathbb{F}^n$.[6] Moreover for $k < d$ define

$$\partial_k P(x, h_1, \dots, h_k) \overset{\text{def}}{=} D_{h_1} \cdots D_{h_k} P(x).$$

The following lemma shows some useful properties of the derivative polynomial.

Lemma 2.9. *Let $P : \mathbb{F}^n \to \mathbb{T}$ be a degree-d (nonclassical) polynomial. Then the polynomial $\partial P(h_1, \dots, h_d)$ is*

(i) *multilinear: ∂P is additive in each h_i.*
(ii) *invariant under permutations of $h_1, \dots, h_d$.*
(iii) *a classical nonzero polynomial of degree d.*
(iv) *homogeneous: All its monomials are of degree d.*

Notice that by multilinear we mean additive in each direction h_i, which is not the usual use of the term “multilinear”.

Proof. The proof follows by the properties of the additive derivative D_h. Multilinearity of ∂P follows from linearity of the additive derivative, namely for every function Q and directions h_1, h_2 we have the identity $D_{h_1+h_2}Q(x) = D_{h_1}Q(x) + D_{h_2}Q(x + h_1)$. The invariance under permutations of $h_1, \dots, h_d$ is a result of commutativity of the additive derivatives. Since P is a degree-d (nonclassical) polynomial, ∂P is nonzero by definition. Notice that since $D_{\underline{0}}Q \equiv 0$ for any function Q, we have $\partial P(h_1, \dots, h_d) = 0$ if any of h_i is equal to zero. Hence every monomial of ∂P must depend on all h_i's. The properties (iii) and (iv) now follow from this and the fact that $\deg(\partial P) \leqslant d$ and thus each monomial has exactly one variable from each h_i. □

2.2. Rank of a polynomial

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

Definition 2.10 *(Rank of a polynomial).* Given a polynomial $P : \mathbb{F}^n \to \mathbb{T}$ and an integer $d > 1$, the *d-rank* of P, denoted $\text{rank}_d(P)$, is defined to be the smallest integer r such that there exist polynomials $Q_1, \dots, Q_r : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d - 1$ and a function $\Gamma : \mathbb{T}^r \to \mathbb{T}$ satisfying $P(x) = \Gamma(Q_1(x), \dots, Q_r(x))$. If $d = 1$, then 1-rank is defined to be ∞ if P is non-constant and 0 otherwise.

[6] Notice since P is a degree-d polynomial, $D_{h_1} \dots D_{h_d} P(x)$ does not depend on x and thus we have the identity $\partial P(h_1, \dots, h_d) = D_{h_1} \dots D_{h_d} P(x)$ for any choice of $x \in \mathbb{F}^n$.

The *rank* of a polynomial $P : \mathbb{F}^n \to \mathbb{T}$ is its $\deg(P)$-rank. We say P is r-regular if $\mathrm{rank}(P) \geqslant r$.

Note that for integer $\lambda \in [1, p-1]$, $\mathrm{rank}(P) = \mathrm{rank}(\lambda P)$. We also define the following weaker analytical notion of uniformity for a polynomial.

Definition 2.11 *(Uniformity).* Let $\varepsilon > 0$ be a real. A degree-d polynomial $P : \mathbb{F}^n \to \mathbb{T}$ is said to be ε-uniform if

$$\|\mathsf{e}(P)\|_{U^d} < \varepsilon.$$

The following theorem of Tao and Ziegler shows that high-rank polynomials have small Gowers norm.

Theorem 2.12. *(Theorem 1.20 of [20].) For any $\varepsilon > 0$ and integer $d > 0$, there exists an integer $r(d, \varepsilon)$ such that the following is true. For any polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$, if $\|\mathsf{e}(P)\|_{U^d} \geqslant \varepsilon$, then $\mathrm{rank}_d(P) \leqslant r$.*

This immediately implies that a regular polynomial is also uniform.

Corollary 2.13. *Let ε, d, and $r(d, \varepsilon)$ be as in Theorem 2.12. Every r-regular polynomial P of degree d is also ε-uniform.*

2.3. Polynomial factors

A high-rank polynomial of degree d is, intuitively, a "generic" degree-d polynomial. There are no unexpected ways to decompose it into lower-degree polynomials. Next, we will formalize the notion of a generic collection of polynomials. Intuitively, it should mean that there are no unexpected algebraic dependencies among the polynomials. First, we need to set up some notation.

Definition 2.14 *(Factors).* If X is a finite set then by a *factor* $\mathcal{B}$ we mean simply a partition of X into finitely many pieces called *atoms*.

A function $f : X \to \mathbb{C}$ is called *$\mathcal{B}$-measurable* if it is constant on atoms of $\mathcal{B}$. For any function $f : X \to \mathbb{C}$, we may define the conditional expectation

$$\mathbf{E}[f|\mathcal{B}](x) = \mathop{\mathbf{E}}_{y \in \mathcal{B}(x)}[f(y)],$$

where $\mathcal{B}(x)$ is the unique atom in $\mathcal{B}$ that contains x. Note that $\mathbf{E}[f|\mathcal{B}]$ is $\mathcal{B}$-measurable.

A finite collection of functions $\phi_1, \dots, \phi_C$ from X to some other space Y naturally defines a factor $\mathcal{B} = \mathcal{B}_{\phi_1,\dots,\phi_C}$ whose atoms are sets of the form $\{x : (\phi_1(x), \dots, \phi_C(x)) = (y_1, \dots, y_C)\}$ for some $(y_1, \dots, y_C) \in Y^C$. By an abuse of notation we also use $\mathcal{B}$ to denote the map $x \mapsto (\phi_1(x), \dots, \phi_C(x))$, thus also identifying the atom containing x with $(\phi_1(x), \dots, \phi_C(x))$.

Definition 2.15 *(Polynomial factors).* If $P_1, \dots, P_C : \mathbb{F}^n \to \mathbb{T}$ is a sequence of polynomials, then the factor $\mathcal{B}_{P_1,\dots,P_C}$ is called a *polynomial factor*. The *complexity* of $\mathcal{B}$, denoted $|\mathcal{B}| := C$, is the number of the defining polynomials. The *degree* of $\mathcal{B}$ is the maximum degree among its defining polynomials $P_1, \dots, P_C$.

Note that if $P_1, \dots, P_C$ are of depths $k_1, \dots, k_C$, respectively, then the number of atoms of $\mathcal{B}$ is at most $\prod_{i=1}^{C} p^{k_i+1}$.

Definition 2.16 *(Rank and regularity).* A polynomial factor $\mathcal{B}$ defined by a sequence of polynomials $P_1, \dots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective depths $k_1, \dots, k_C$ is said to have rank r if r is the least integer for which there exists $(\lambda_1, \dots, \lambda_C) \in \mathbb{Z}^C$, with $(\lambda_1 \bmod p^{k_1+1}, \dots, \lambda_C \bmod p^{k_C+1}) \neq 0^C$, such that $\mathrm{rank}_d(\sum_{i=1}^{C} \lambda_i P_i) \leqslant r$, where $d = \max_i \deg(\lambda_i P_i)$.

Given a polynomial factor $\mathcal{B}$ and a function $r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$, we say that $\mathcal{B}$ is r-regular if $\mathcal{B}$ is of rank larger than $r(|\mathcal{B}|)$.

Notice that by the above definition of rank for a degree-d polynomial P of depth k we have

$$\mathrm{rank}(\mathcal{B}_P) = \min \left\{ \mathrm{rank}_d(P), \mathrm{rank}_{d-(p-1)}(pP), \dots, \mathrm{rank}_{d-k(p-1)}(p^k P) \right\}.$$

We also define the following weaker analytical notion of uniformity for a factor along the same lines as Definition 2.11.

Definition 2.17 *(Uniform factor).* Let $\varepsilon > 0$ be a real. A polynomial factor $\mathcal{B}$ defined by a sequence of polynomials $P_1, \dots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective depths $k_1, \dots, k_C$ is said to be ε-uniform if for every collection $(\lambda_1, \dots, \lambda_C) \in \mathbb{Z}^C$, with $(\lambda_1 \bmod p^{k_1+1}, \dots, \lambda_C \bmod p^{k^C+1}) \neq 0^C$

$$\left\| \mathsf{e}\left(\sum_i \lambda_i P_i \right) \right\|_{U^d} < \varepsilon,$$

where $d = \max_i \deg(\lambda_i P_i)$.

Remark 2.18. Similar to Corollary 2.13 it also follows from Theorem 2.12 that an r-regular degree-d factor $\mathcal{B}$ is also ε-uniform when $r = r(d, \varepsilon)$ is as in Theorem 2.12.

2.3.1. Regularization of factors

Due to the generic properties of regular factors, it is often useful to *refine* a given polynomial factor to a regular one [20,3,2]. We will first formally define what we mean by refining a polynomial factor.

Definition 2.19 *(Refinement).* A factor $\mathcal{B}'$ is called a *refinement* of $\mathcal{B}$, and denoted $\mathcal{B}' \succeq \mathcal{B}$, if the induced partition by $\mathcal{B}'$ is a combinatorial refinement of the partition induced by $\mathcal{B}$. In other words, if for every $x, y \in \mathbb{F}^n$, $\mathcal{B}'(x) = \mathcal{B}'(y)$ implies $\mathcal{B}(x) = \mathcal{B}(y)$. We will write $\mathcal{B} \succeq_{\mathrm{syn}} \mathcal{B}'$, if the polynomials defining $\mathcal{B}'$ extend that of $\mathcal{B}$.

The following lemma from [2] which uses a regularization theorem of [20] allows one to regularize a given factor to any desired regularity.

Lemma 2.20 *(Polynomial regularity lemma [2]). Let $r : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a non-decreasing function and let $d > 0$ be an integer. Then, there is a function $C_{2.20}^{(r,d)} : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ such that the following is true. Suppose $\mathcal{B}$ is a factor defined by polynomials $P_1, \dots, P_C : \mathbb{F}^n \to \mathbb{T}$ of degree at most d. Then, there is an r-regular factor $\mathcal{B}'$ consisting of polynomials $Q_1, \dots, Q_{C'} : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$ such that $\mathcal{B}' \succeq \mathcal{B}$ and $C' \leqslant C_{2.20}^{(r,d)}(C)$.*

2.4. Decomposition theorems

An important application of the inverse theorems are the "decomposition theorems" [6,18,10]. These theorems allow one to express a given function f with certain properties as a sum $\sum_{i=1}^{k} g_i$, where each g_i has certain desired structural properties. We refer the interested reader to [6] and [9] for a detailed discussion of this subject. Recall that the complexity of a polynomial factor $\mathcal{B}$, denoted $|\mathcal{B}|$, is the number of polynomials that define the factor. The following decomposition theorem is a consequence of an inverse theorem for Gowers norms [20, Theorem 1.11].

Theorem 2.21 *(Strong decomposition theorem for multiple functions). Let $m, d \geqslant 1$ be integers, $\delta > 0$ a parameter, and let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary growth function. Given any functions $f_1, \dots, f_m : \mathbb{F}^n \to \mathbb{D}$, there exists a decomposition*

$$f_i = g_i + h_i,$$

such that for every $1 \leqslant i \leqslant m$,

1. *$g_i = \mathbf{E}[f_i|\mathcal{B}]$, where $\mathcal{B}$ is an r-regular polynomial factor of degree at most d and complexity $|\mathcal{B}| \leqslant C_{\max}(p, m, d, \delta, r(\cdot))$,*
2. *$\|h_i\|_{U^{d+1}} \leqslant \delta$.*

3. Main results

3.1. Homogeneous polynomials

Recall that a classical polynomial is called homogeneous if all of its monomials are of the same degree. Trivially a homogeneous classical polynomial $P(x)$ satisfies $P(cx) =$

$|c|^d P(x)$ for every $c \in \mathbb{F}$. We will use this property to define the class of nonclassical homogeneous polynomials.

Definition 3.1 *(Homogeneity).* A (nonclassical) polynomial $P : \mathbb{F}^n \to \mathbb{T}$ is called homogeneous if for every $c \in \mathbb{F}$ there exists a $\sigma_c \in \mathbb{Z}$ such that $P(cx) = \sigma_c P(x) \bmod 1$ for all x.

Remark 3.2. It is not difficult to see that $P(cx) = \sigma_c P(x) \bmod 1$ implies that $\sigma_c = |c|^{\deg(P)} \bmod p$, a property that we will use later. Indeed for $d = \deg(P)$, we have $0 = \partial_d(P(cx) - \sigma_c P(x)) = (|c|^d - \sigma_c)\partial_d P(x) \bmod 1$. This, since $\partial_d P(x)$ is a nonzero degree-d classical polynomial, implies $\sigma_c = |c|^d \bmod p$.

Notice that for a polynomial P to be homogeneous it suffices that there exists $\sigma \in \mathbb{Z}$ for which $P(\zeta x) = \sigma P(x) \bmod 1$, where ζ is a generator of $\mathbb{F}^*$. If P has depth k, then we can assume that $\sigma \in \mathbb{Z}_{p^{k+1}}$, as $p^{k+1}P \equiv 0$. Obviously, the value of σ depends on the choice of the generator ζ. However, the following lemma shows that for a fixed ζ, the value of σ is uniquely determined for all homogeneous polynomials of degree d and depth k. Henceforth, we will denote this unique value by $\sigma(d, k)$.

In fact, if we denote by $\mathbf{Z}_p$ the p-adic integers, then there exists $\sigma(d) \in \mathbf{Z}_p$ such that $\sigma(d, k) = \sigma(d) \bmod p^{k+1}$. This is directly related to the so-called Teichmüller characters (see e.g. section 4.3 of [4]), however in order to keep the presentation elementary, we avoid using this connection directly.

Lemma 3.3. *For every d and k, there is a unique $\sigma = \sigma(d, k) \in \mathbb{Z}_{p^{k+1}}$, such that for every homogeneous polynomial P of degree d and depth k, $P(\zeta x) = |\sigma|P(x) \bmod 1$, where $|\cdot|$ is the natural map from $\mathbb{Z}_{p^{k+1}}$ to $\{0, 1, \dots, p^{k+1} - 1\} \subset \mathbb{Z}$. Moreover, there exists $\sigma(d) \in \mathbf{Z}_p$ such that $\sigma(d, k) = \sigma(d) \bmod p^{k+1}$.*

Proof. Let P be a homogeneous polynomial of degree d and depth k, and let $\sigma \in \mathbb{Z}_{p^{k+1}}$ be such that $P(\zeta x) = |\sigma|P(x) \bmod 1$. By Remark 3.2 we know that $|\sigma| = |\zeta|^d \bmod p$. We also observe that $P(x) = P(\zeta^{p-1}x) = |\sigma|^{p-1}P(x) \bmod 1$ from which it follows that $\sigma^{p-1} = 1$. We claim that $\sigma \in \mathbb{Z}_{p^{k+1}}$ is uniquely determined by the two properties:

i. $|\sigma| = |\zeta|^d \bmod p$, and
ii. $\sigma^{p-1} = 1 \bmod p^{k+1}$.

Suppose to the contrary that there are two nonzero values $\sigma_1, \sigma_2 \in \mathbb{Z}_{p^{k+1}}$ that satisfy the above two properties, and choose $t \in \mathbb{Z}_{p^{k+1}}$ such that $\sigma_1 = t\sigma_2$. It follows from (i) that $t = 1 \bmod p$ and from (ii) that $t^{p-1} = 1$. We will show that $t = 1$ is the only possible such value in $\mathbb{Z}_{p^{k+1}}$.

Let $a_1, \dots, a_{p^k} \in \mathbb{Z}_{p^{k+1}}$ be all the possible solutions to $x = 1 \bmod p$ in $\mathbb{Z}_{p^{k+1}}$. Note that $ta_1, \dots, ta_{p^k}$ is just a permutation of the first sequence and thus

$$t^{p^k}\prod a_i = \prod a_i.$$

Consequently $t^{p^k} = 1 \bmod p^{k+1}$, which combined with $t^p = t$ implies $t = 1 \bmod p^{k+1}$.

For the last assertion, note that $\sigma(d,k)^{p-1} = 1 \bmod p^{k+1}$ implies $\sigma(d,k)^{p-1} = 1 \bmod p^{t+1}$ for every $t \leqslant k$. By the uniqueness of $\sigma(d,t)$, this implies that $\sigma(d,k) = \sigma(d,t) \bmod p^{t+1}$. We can thus take $\sigma(d)$ in the p-adic integers given by $\sigma(d) \bmod p^{k+1} = \sigma(d,k)$. □

Lemma 2.6 allows us to express every (nonclassical) polynomial as a linear span of monomials of the form $\frac{|x_1|^{d_1}\cdots|x_n|^{d_n}}{p^{k+1}}$. Unfortunately, unlike in the classical case, these monomials are not necessarily homogeneous, and for some applications it is important to express a polynomial as a linear span of homogeneous polynomials. We show that this is possible as homogeneous multivariate (nonclassical) polynomials linearly span the space of multivariate (nonclassical) polynomials. We will present the proof of this theorem in Section 4.1.

Theorem 3.4. *There is a basis for* $\mathrm{Poly}(\mathbb{F}^n \to \mathbb{T})$ *consisting only of homogeneous multivariate polynomials.*

Theorem 3.4 allows us to make the extra assumption in the strong decomposition theorem (Theorem 2.21) that the resulting polynomial factor $\mathcal{B}$ consists only of homogeneous polynomials.

Corollary 3.5. *Let $d \geqslant 1$ be an integer, $\delta > 0$ a parameter, and let $r : \mathbb{N} \to \mathbb{N}$ be an arbitrary growth function. Given any functions $f_1, \dots, f_m : \mathbb{F}^n \to \mathbb{D}$, there exists a decomposition*

$$f_i = g_i + h_i,$$

such that or every $1 \leqslant i \leqslant m$,

1. $g_i = \mathbf{E}[f_i|\mathcal{B}]$, *where $\mathcal{B}$ is an r-regular polynomial factor of degree at most d and complexity $C \leqslant C_{\max}(p, d, \delta, r(\cdot))$, moreover $\mathcal{B}$ only consists of homogeneous polynomials.*
2. $\|h_i\|_{U^{d+1}} \leqslant \delta$.

3.2. Strong near-orthogonality

As mentioned in the introduction the main result of this paper is a new near-orthogonality result for polynomial factors of high rank. Such a statement was proved in [10,20] for systems of linear forms corresponding to repeated derivatives or equivalently Gowers norms, in [14] for the case when the field is of high characteristic but with arbitrary system of linear forms and in [2] for systems of affine linear forms. In Theorem 3.8

we establish the near-orthogonality over any arbitrary system of linear forms. Before stating this theorem we need to introduce the notion of consistency.

Definition 3.6 *(Consistency).* Let $\mathcal{L} = \{L_1, \ldots, L_m\}$ be a system of linear forms. A vector $(\beta_1, \ldots, \beta_m) \in \mathbb{T}^m$ is said to be (d, k)-*consistent with* $\mathcal{L}$ if there exist a homogeneous polynomial P of degree d and depth k and a point X such that $P(L_i(X)) = \beta_i$ for every $i \in [m]$. Let $\Phi_{d,k}(\mathcal{L})$ denote the set of all such vectors.

It is immediate from the definition that $\Phi_{d,k}(\mathcal{L}) \subseteq \mathbb{U}_{k+1}^m$ is a subgroup of $\mathbb{T}^m$, or more specifically, a subgroup of $\mathbb{U}_{k+1}^m$. Let

$$\Phi_{d,k}(\mathcal{L})^\perp := \left\{(\lambda_1, \ldots, \lambda_m) \in \mathbb{Z}^m \ : \ \forall (\beta_1, \ldots, \beta_m) \in \Phi_{d,k}(\mathcal{L}),\ \sum \lambda_i \beta_i = 0\right\}.$$

Equivalently $\Phi_{d,k}(\mathcal{L})^\perp$ is the set of all $(\lambda_1, \ldots, \lambda_m) \in \mathbb{Z}^m$ such that $\sum_{i=1}^m \lambda_i P(L_i(X)) \equiv 0$ for every homogeneous polynomial P of degree d and depth k. As we will later see, $\Phi_{d,k}(\mathcal{L})^\perp \subseteq \Phi_{d,t}(\mathcal{L})^\perp$ for all $t \leqslant k$.

Example 3.7. Consider the system of linear forms

$$\mathcal{L} = \{(1,0,0), (0,1,0), (0,0,1), (1,1,0), (1,0,1), (0,1,1), (1,1,1)\}$$

over $\mathbb{F}_2^n$ corresponding to $X_1, X_2, X_3, X_1+X_2, X_1+X_3, X_2+X_3, X_1+X_2+X_3$. Let P be a homogeneous polynomial of degree 1 and depth 0. Let $w_1, w_2, w_3, w_{12}, w_{13}, w_{23}, w_{123}$ denote the value of P on $X_1, X_2, X_3, X_1+X_2, X_1+X_3, X_2+X_3, X_1+X_2+X_3$, respectively. Since P is of degree 1 it satisfies $P(x) - P(x+y) - P(x+z) + P(x+y+z) = 0$ for every $x, y, z \in \mathbb{F}_2^n$, and by substituting proper values for x, y, z (and recalling that degree 1 depth 0 polynomials over $\mathbb{F}_2^n$ have their image in $\frac{1}{2}\mathbb{Z}/\mathbb{Z}$) we obtain $w_{13} = w_{12} + w_2 + w_3$, $w_{23} = w_{12} + w_1 + w_3$, $w_{123} = w_1 + w_2 + w_3$. It turns out that these identities give a complete description of $\Phi_{2,0}(\mathcal{L})^\perp$ and $\Phi_{2,0}(\mathcal{L})$. Indeed, it is not difficult to see that

$$\begin{aligned}\Phi_{2,0}(\mathcal{L}) = \Big\{ \frac{1}{2}(w_1, w_2, w_3, w_{12}, w_{13}, w_{23}, w_{123}) \mod 1 \ :\ & \\ w_1, w_2, w_3, w_{12}, w_{13}, w_{23}, w_{123} \in \mathbb{Z} & \\ w_{13} = w_{12} + w_2 + w_3,\ w_{23} = w_{12} + w_1 + w_3,\ w_{123} = w_1 + w_2 + w_3 & \Big\}\end{aligned}$$

and

$$\Phi_{2,0}(\mathcal{L})^\perp = (2\mathbb{Z})^7 + \{(0,1,1,1,1,0,0), (1,0,1,1,0,1,0), (1,1,1,0,0,0,1)\}.$$

Theorem 3.8 *(Near-orthogonality over linear forms). Let $L_1, \ldots, L_m$ be linear forms on ℓ variables and let $\mathcal{B} = (P_1, \ldots, P_C)$ be an ε-uniform polynomial factor defined only by* *homogeneous polynomials, where $\varepsilon \in (0,1)$. For every tuple Λ of integers $(\lambda_{i,j})_{i\in[C],j\in[m]}$, define $P_\Lambda : (\mathbb{F}^n)^\ell \to \mathbb{T}$ as*

$$P_\Lambda(X) = \sum_{i\in[C],j\in[m]} \lambda_{i,j} P_i(L_j(X)).$$

Then one of the following two statements holds:

- $P_\Lambda \equiv 0$.
- P_Λ *is non-constant and* $\left|\mathbf{E}_{X\in(\mathbb{F}^n)^\ell}[\mathsf{e}(P_\Lambda)]\right| < \varepsilon$.

Furthermore $P_\Lambda \equiv 0$ if and only if for every $i \in [C]$, we have $(\lambda_{i,j})_{j\in[m]} \in \Phi_{d_i,t_i}(\mathcal{L})^\perp$ for every $t_i \leqslant k_i$, where d_i, k_i are the degree and depth of P_i, respectively.

We will present the proof of Theorem 3.8 in Section 4.2.

Remark 3.9. Applying Theorem 3.8 to a single homogeneous polynomial P of degree d and depth k, we see that indeed $\Phi_{d,k}(\mathcal{L})^\perp \subseteq \Phi_{d,t}(\mathcal{L})^\perp$ for all $t \leqslant k$. Indeed, if $(\lambda_1, \dots, \lambda_m) \in \Phi_{d,k}(\mathcal{L})^\perp$ then $\sum \lambda_i P(L_i(X)) = 0$, and hence by Theorem 3.8 we have that $(\lambda_1, \dots, \lambda_m) \in \Phi_{d,t}(\mathcal{L})^\perp$ for all $t \leqslant k$.

Remark 3.10. By Corollary 2.13 the assumption of ε-uniformity in Theorem 3.8 is satisfied for every factor $\mathcal{B}$ of rank at least $r_{2.12}(d,\varepsilon)$. However, we would like to point out that in Theorem 3.8 by using the assumption of ε-uniformity instead of the assumption of high rank, we are able to achieve the quantitative bound of ε on the bias of P_Λ.

Remark 3.11. In Theorem 3.8 in the second case where P_Λ is non-constant, it is possible to deduce a more general statement that $\|\mathsf{e}(P_\Lambda)\|_{U^t}^{2^t} < \varepsilon$ for every $t \leqslant \deg(P_\Lambda)$. Indeed assume that P_Λ is non-constant, and consider the derivative

$$D_{Y_1} \dots D_{Y_t} P_\Lambda(X) = \sum_{i\in[C],j\in[m]} \lambda_{i,j} \sum_{S\subseteq[t]} (-1)^{|S|} P_i(L_j(X + \sum_{r\in S} Y_r)), \tag{3}$$

where $Y_i := (y_{i,1}, \dots, y_{i,\ell}) \in (\mathbb{F}^n)^\ell$. Notice that for every choice of $j \in [m]$ and $S \subseteq [t]$, $L_j(X + \sum_{r\in S} Y_r)$ is an application of a linear form on the vector $(x_1, \dots, x_\ell, y_{1,1}, \dots, y_{1,\ell}, \dots, y_{t,1}, \dots, y_{t,\ell}) \in (\mathbb{F}^n)^{(t+1)\ell}$, and since by Lemma 2.9 the polynomial ∂P_Λ is nonzero, Theorem 3.8 implies

$$\|\mathsf{e}(P_\Lambda)\|_{U^t}^{2^t} = \mathop{\mathbf{E}}_{h_1,\dots,h_t} [\mathsf{e}(\partial_t P_\Lambda(h_1, \dots, h_t))] \leqslant \varepsilon.$$

It is well known that statements similar to that of Theorem 3.8 imply “near-equidistributions” of the joint distribution of the polynomials applied to linear forms. Consider a highly uniform polynomial factor of degree $d > 0$, defined by a tuple of homogeneous polynomials $P_1, \dots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective degrees $d_1, \dots, d_C$ and depths $k_1, \dots, k_C$, and let $\mathcal{L} = (L_1, \dots, L_m)$ be a collection of linear forms on ℓ variables. As we mentioned earlier, we are interested in the distribution of the random matrix

$$\begin{pmatrix} P_1(L_1(X)) & P_2(L_1(X)) & \dots & P_C(L_1(X)) \\ P_1(L_2(X)) & P_2(L_2(X)) & \dots & P_C(L_2(X)) \\ \vdots & & & \vdots \\ P_1(L_m(X)) & P_2(L_m(X)) & \dots & P_C(L_m(X)) \end{pmatrix}, \tag{4}$$

where X is the uniform random variable taking values in $(\mathbb{F}^n)^\ell$. Note that by the definition of consistency, for every $1 \leqslant i \leqslant C$, the i-th column of this matrix must belong to $\Phi_{d_i,k_i}(\mathcal{L})$. Theorem 3.12 below says that (4) is "almost" uniformly distributed over the set of all matrices satisfying this condition. The proof of Theorem 3.12 is standard and is identical to the proof of [2, Theorem 3.10] with the only difference that it uses Theorem 3.8 instead of the weaker near-orthogonality theorem of [2].

Theorem 3.12 *(Near-equidistribution). Given $\varepsilon > 0$, let $\mathcal{B}$ be an ε-uniform polynomial factor of degree $d > 0$ and complexity C, that is defined by a tuple of homogeneous polynomials $P_1, \dots, P_C : \mathbb{F}^n \to \mathbb{T}$ having respective degrees $d_1, \dots, d_C$ and depths $k_1, \dots, k_C$. Let $\mathcal{L} = (L_1, \dots, L_m)$ be a collection of linear forms on ℓ variables.*

Suppose $(\beta_{i,j})_{i\in[C],j\in[m]} \in \mathbb{T}^{C\times m}$ is such that $(\beta_{i,1}, \dots, \beta_{i,m}) \in \Phi_{d_i,k_i}(\mathcal{L})$ for every $i \in [C]$. Then

$$\mathop{\mathbf{Pr}}_{X\in(\mathbb{F}^n)^\ell}[P_i(L_j(X)) = \beta_{i,j}\ \forall i \in [C], j \in [m]] = \frac{1}{K} \pm \varepsilon,$$

where $K = \prod_{i=1}^{C} |\Phi_{d_i,k_i}(\mathcal{L})|$.

Proof. We have

$$\begin{aligned} &\mathbf{Pr}[P_i(L_j(X)) = \beta_{i,j}\ \forall i \in [C], \forall j \in [m]] \\ &\quad = \mathbf{E}\left[\prod_{i,j} \frac{1}{p^{k_i+1}} \sum_{\lambda_{i,j}=0}^{p^{k_i+1}-1} \mathsf{e}\left(\lambda_{i,j}\left(P_i(L_j(X)) - \beta_{i,j}\right)\right)\right] \\ &\quad = \left(\prod_{i\in[C]} p^{-(k_i+1)}\right)^m \sum_{(\lambda_{i,j})} \mathsf{e}\left(-\sum_{i,j} \lambda_{i,j}\beta_{i,j}\right) \mathbf{E}\left[\mathsf{e}\left(\sum_{i\in[C],j\in[m]} \lambda_{i,j} P_i(L_j(X))\right)\right], \end{aligned}$$

where the outer sum is over $(\lambda_{i,j})_{i\in[C],j\in[m]}$ with $\lambda_{i,j} \in [0, p^{k_i+1} - 1]$. Let $\Lambda_i = \Phi_{d_i,k_i}(\mathcal{L})^\perp \cap [0, p^{k+1} - 1]^m$, and note that $|\Lambda_i||\Phi_{d_i,k_i}| = p^{m(k_i+1)}$. Since $(\beta_{i,1}, \dots, \beta_{i,m}) \in \Phi_{d_i,k_i}(\mathcal{L})$ for every $i \in [C]$, it follows that $\sum_{i,j} \lambda_{i,j}\beta_{i,j} = 0$ if $(\lambda_{i,1}, \dots, \lambda_{i,m}) \in \Lambda_i$ for all $i \in [C]$. If the latter holds, then the expected value in the above expression is 0, and otherwise by Theorem 3.8, it is bounded by ε. Hence the above expression can be approximated by

$$p^{-m\sum_{i=1}^{C}(k_i+1)} \cdot \left(\prod_{i=1}^{C}|\Lambda_i| \ \pm\ \varepsilon p^{m\sum_{i=1}^{C}(k_i+1)}\right) = \frac{1}{K} \pm \varepsilon. \qquad \square$$

3.3. On a theorem of Gowers and Wolf

Let A be a subset of $\mathbb{F}^n$ with the indicator function $\mathbf{1}_A : \mathbb{F}^n \to \{0,1\}$. As mentioned in the introduction, (1) equals the probability that $L_1(X), \dots, L_m(X)$ all fall in A, where $X \in (\mathbb{F}^n)^k$ is chosen uniformly at random. Roughly speaking, we say $A \subseteq \mathbb{F}^n$ is *pseudorandom* with regards to $\mathcal{L}$ if

$$\mathop{\mathbf{E}}_{X}\left[\prod_{i=1}^{m}\mathbf{1}_A(L_i(X))\right] \approx \left(\frac{|A|}{p^n}\right)^m;$$

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

$$\begin{aligned}\mathop{\mathbf{E}}_{X}\left[\prod_{i=1}^{m}\mathbf{1}_A(L_i(X))\right] &= \mathop{\mathbf{E}}_{X}\left[\prod_{i=1}^{m}(\alpha + f(L_i(X)))\right] \\ &= \alpha^m + \sum_{S\subseteq[m], S\neq\emptyset} \alpha^{m-|S|}\mathop{\mathbf{E}}_{X}\left[\prod_{i\in S} f(L_i(X))\right].\end{aligned}$$

Therefore, a sufficient condition for A to be pseudorandom with regards to $\mathcal{L}$ is that $\mathbf{E}_X\left[\prod_{i\in S} f(L_i(X))\right]$ is negligible for all nonempty subsets $S \subseteq [m]$. Green and Tao [12] 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 3.13 *(Cauchy–Schwarz complexity [12]).* 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 $1 \leqslant i \leqslant m$, we can partition $\{L_j\}_{j\in[m]\setminus\{i\}}$ into $s+1$ subsets, such that L_i does not belong to the linear span of any of the subsets.

The reason for the term *Cauchy–Schwarz complexity* is the following lemma due to Green and Tao [12] whose proof is based on a clever iterative application of the Cauchy–Schwarz inequality.

Lemma 3.14. *(See [12], also [7, Theorem 2.3].) Let $f_1, \dots, f_m : \mathbb{F}^n \to \mathbb{D}$. Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of m linear forms in ℓ variables of Cauchy–Schwarz complex-*

ity s. Then

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

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 Cauchy–Schwarz complexity of $\mathcal{L}$ gives an upper bound on s, such that if $\|f\|_{U^{s+1}}$ is small enough for some function $f:\mathbb{F}^n\to\mathbb{D}$, then f is pseudorandom with regards to $\mathcal{L}$. Gowers and Wolf [7] defined the *true complexity* of a system of linear forms as the minimal s such that the above condition holds for all $f:\mathbb{F}^n\to\mathbb{D}$.

Definition 3.15 *(True complexity [7]).* Let $\mathcal{L}=\{L_1,\dots,L_m\}$ be a system of linear forms over $\mathbb{F}$ in ℓ variables. The true complexity of $\mathcal{L}$ is the smallest $d\in\mathbb{N}$ with the following property. For every $\varepsilon>0$ and $\alpha\geqslant 0$, there exists $\delta>0$ such that if $f:\mathbb{F}^n\to\mathbb{D}$ is any function with $\|f-\alpha\|_{U^{d+1}}\leqslant\delta$, then

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

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 conjectured that the true complexity of a system of linear forms can be characterized by a simple linear algebraic condition. Namely, that it is equal to the smallest $d\geqslant 1$ such that $L_1^{d+1},\dots,L_m^{d+1}$ are linearly independent, where the d-th tensor power of a linear form $L=(\lambda_1,\dots,\lambda_\ell)$ is defined as

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

Example 3.16. Consider the collection of linear forms $L_1=(1,0,0)$, $L_2=(1,1,0)$, $L_3=(1,0,1)$, $L_4=(1,1,1)$, $L_5=(1,1,-1)$, $L_6=(1,-1,1)$. It is easy to check that here the Cauchy–Schwarz complexity is 2, while as observed by Gowers and Wolf the true complexity of this system of linear forms is 1.

Later in [8, Theorem 6.1] they verified their conjecture in the case where $|\mathbb{F}|$ is sufficiently large; more precisely when $|\mathbb{F}|$ is at least the Cauchy–Schwarz complexity of the system of linear form. In this paper we verify the Gowers–Wolf conjecture in full generality by proving the following theorem.

Theorem 3.17. *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms in ℓ variables, 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 for any collection of functions $f_1, \dots, f_m : \mathbb{F}^n \to \mathbb{D}$ with $\min_{i \in [m]} \|f_i\|_{U^{d+1}} \leqslant \delta$, we have*

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

We will present the proof of Theorem 3.17 in Section 4.3.

Remark 3.18. Note that if $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent, then for any $e \geqslant d$ we also have that $L_1^{e+1}, \dots, L_m^{e+1}$ are linearly independent. Thus, for a system of linear forms $L_1, \dots, L_m$ it makes sense to define the smallest d for which $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent. By Theorem 3.17, this equals the true complexity of the system.

The following corollary to Theorem 3.17 is very useful when combined with the decomposition theorems such as Theorem 2.21. In particular, choosing $f_1 = \dots = f_m = f$ and $g_1 = \dots = g_m = \alpha$ proves the Gowers–Wolf conjecture on the true complexity of a system of linear forms in its full generality.

Corollary 3.19. *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms. Assume that $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent. For every $\varepsilon > 0$, there exists $\delta > 0$ such that for any functions $f_1, \dots, f_m, g_1, \dots, g_m : \mathbb{F}^n \to \mathbb{D}$ with $\|f_i - g_i\|_{U^{d+1}} \leqslant \delta$, we have*

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

Proof. Choosing $\delta = \delta(\varepsilon')$ as in Theorem 3.17 for $\varepsilon' := \varepsilon/m$, we have

$$\begin{aligned} &\left| \mathop{\mathbf{E}}_{X} \left[\prod_{i=1}^{m} f_i(L_i(X)) \right] - \mathop{\mathbf{E}}_{X} \left[\prod_{i=1}^{m} g_i(L_i(X)) \right] \right| \\ &= \left| \sum_{i=1}^{m} \mathop{\mathbf{E}}_{X} \left[(f_i - g_i)(L_i(X)) \cdot \prod_{j=1}^{i-1} g_j(L_j(X)) \cdot \prod_{j=i+1}^{m} f_j(L_j(X)) \right] \right| \\ &\leqslant \sum_{i=1}^{m} \left| \mathop{\mathbf{E}}_{X} \left[(f_i - g_i)(L_i(X)) \cdot \prod_{j=1}^{i-1} g_j(L_j(X)) \cdot \prod_{j=i+1}^{m} f_j(L_j(X)) \right] \right| \\ &\leqslant m \cdot \varepsilon' \leqslant \varepsilon, \end{aligned}$$

where the second inequality follows from Theorem 3.17 since $\|f_i - g_i\|_{U^{d+1}} \leqslant \delta$. □

4. Main proofs

In this section we will present the proofs of Theorem 3.4, Theorem 3.8, and Theorem 3.17.

4.1. Homogeneity: Proof of Theorem 3.4

Theorem 3.4 (restated). *There is a basis for* $\mathrm{Poly}(\mathbb{F}^n \to \mathbb{T})$ *consisting only of homogeneous multivariate polynomials.*

To simplify the notation, in this section we will omit writing "mod 1" in the description of the defined nonclassical polynomials. We start by proving the following simple observation.

Claim 4.1. *Let* $P : \mathbb{F} \to \mathbb{T}$ *be a univariate polynomial of degree* d*. Then for every* $c \in \mathbb{F}\backslash\{0\}$,

$$\deg\left(P(cx) - |c|^d P(x)\right) < d.$$

Proof. By Lemma 2.6 it suffices to prove the claim for a monomial $q(x) := \frac{|x|^s}{p^{k+1}}$ with $k(p-1)+s = d$. Note that $q(cx) - |c|^d q(x)$ takes values in $\frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$ as by Fermat's little theorem

$$|cx|^s - |c|^d |x|^s \equiv |x|^s |c|^s (1 - |c|^{k(p-1)}) \equiv 0 \mod p.$$

It follows then from Lemma 2.6 that $q(cx) - |c|^d q(x)$ is of depth at most $k-1$, and hence

$$\deg\left(q(cx) - |c|^d q(x)\right) \leqslant (p-1)(k-1) + (p-1) < d. \qquad \square \tag{5}$$

Remark 4.2. It is not difficult to show that the above claim holds for any multivariate polynomial $P : \mathbb{F}^n \to \mathbb{T}$. However, since the univariate case suffices for our purpose, we do not prove the general case.

First we prove Theorem 3.4 for univariate polynomials.

Lemma 4.3. *There is a basis of homogeneous univariate polynomials for* $\mathrm{Poly}(\mathbb{F} \to \mathbb{T})$.

Proof. We will prove by induction on d that there is a basis $\{h_1, \dots, h_d\}$ of homogeneous univariate polynomials for $\mathrm{Poly}_{\leqslant d}(\mathbb{F} \to \mathbb{T})$ for every d. Let ζ be a fixed generator of $\mathbb{F}^*$. For any degree $d > 0$, we will build a degree-d homogeneous polynomial $h_d(x)$ such that $h_d(\zeta x) = \sigma_d h_d(x)$ for some integer σ_d. The base case of $d \leqslant p-1$ is trivial as $\mathrm{Poly}_{\leqslant p-1}(\mathbb{F} \to \mathbb{T})$ consists of only classical polynomials, and those are spanned by $h_0(x) := \frac{1}{p}$, $h_1(x) := \frac{|x|}{p}$, $\dots$, $h_{p-1}(x) := \frac{|x|^{p-1}}{p}$. Now suppose that $d = s + (p-1)(k-1)$ with $0 < s \leqslant p-1$, and $k > 1$. It suffices to show that the degree-d monomial $\frac{|x|^s}{p^k}$ can be expressed as a linear combination of homogeneous polynomials. Consider the function

$$f(x) := \frac{|\zeta x|^s}{p^k} - \frac{|\zeta|^d |x|^s}{p^k}.$$

Claim 4.1 implies that $\deg(f) < d$. Using the induction hypothesis, we can express $f(x)$ as a linear combination of $\frac{|x|^s}{p^\ell}$ for $\ell = 0, \dots, k-1$, and h_e for $e < d$ with $e \neq s \bmod (p-1)$:

$$f(x) = \sum_{\ell=1}^{k-1} a_\ell \frac{|x|^s}{p^\ell} + \sum_{\substack{e<d, \\ e\neq d \bmod (p-1)}} b_e h_e(x).$$

Set $A := |\zeta|^d + \sum_{\ell=1}^{k-1} a_\ell p^{k-\ell}$, so that

$$\frac{|\zeta x|^s}{p^k} - A\frac{|x|^s}{p^k} = \sum_{\substack{e<d, \\ e\neq d \bmod (p-1)}} b_e h_e(x). \tag{6}$$

By the induction hypothesis, for $e < d$, $h_e(\zeta x) = \sigma_e h(x)$ where $\sigma_e = |\zeta|^e \bmod p$, and thus as $A = |\zeta|^d \bmod p$, we have $\sigma_e \neq A \bmod p$ when $e \neq d \bmod (p-1)$. Consequently,

$$\begin{aligned}\sum_{\substack{e<d, \\ e\neq d \bmod (p-1)}} b_e h_e(x) &= \sum_{\substack{e<d, \\ e\neq d \bmod (p-1)}} \frac{b_e}{\sigma_e - A}(\sigma_e - A)h_e(x) \\ &= \sum_{\substack{e<d, \\ e\neq d \bmod (p-1)}} \frac{b_e}{\sigma_e - A}(h_e(\zeta x) - Ah_e(x)).\end{aligned}$$

Combing this with (6) we conclude that

$$h_d(x) := \frac{|x|^s}{p^k} - \sum_{\substack{e<d, \\ e\neq d \bmod (p-1)}} \frac{b_e}{\sigma_e - A} h_e(x)$$

satisfies

$$h_d(\zeta x) = Ah_d(x). \qquad \square$$

Proof of Theorem 3.4. We will show by induction on the degree d and k, that every degree d monomial of depth k can be written as a linear combination of homogeneous polynomials. The base case of $d < p$, $k = 0$ is trivial as such monomials are classical and thus homogeneous themselves. Consider a (nonclassical) monomial $M(x_1, \dots, x_n) = \frac{|x_1|^{s_1} \cdots |x_n|^{s_n}}{p^k}$ of degree $d = s_1 + \cdots + s_n + (p-1)(k-1)$. For every $i \in [n]$, let $g_i(x_i) := h_{s_i+(p-1)(k-1)}(x_i)$ where $h_{s_i+(p-1)(k-1)}(\cdot)$ is the homogeneous univariate polynomial from Lemma 4.3. Every g_i takes values in $\frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$, and thus corresponds to a polynomial $G_i : \mathbb{F} \to \mathbb{Z}_{p^k}$. Define $F : \mathbb{F}^n \to \mathbb{Z}_{p^k}$ as

$$F(x_1, \dots, x_n) := G_1(x_1) \cdots G_n(x_n),$$

and $f : \mathbb{F}^n \to \mathbb{T}$ as

$$f(x_1, \dots, x_n) := \frac{F(x_1, \dots, x_n)}{p^k}.$$

It is simple to verify that $\deg(f) = s_1 + \ldots + s_n + (p-1)(k-1) = d$, it has only one monomial of $\deg(f)$ and depth $k-1$, which is $\frac{|x_1|^{s_1} \cdots |x_n|^{s_n}}{p^k} = M(x_1, \dots, x_n)$, and it is homogeneous. Thus $M(x_1, \dots, x_n) - f(x_1, \dots, x_n)$ is of either of degree less than d or is of lower depth and by the induction hypothesis can be written as a linear combination of homogeneous polynomials. □

4.2. Near-orthogonality: Proof of Theorem 3.8

Theorem 3.8 (restated). *Let $L_1, \dots, L_m$ be linear forms on ℓ variables and let $\mathcal{B} = (P_1, \dots, P_C)$ be an ε-uniform polynomial factor defined only by homogeneous polynomials, where $\varepsilon \in (0, 1)$. For every tuple Λ of integers $(\lambda_{i,j})_{i \in [C], j \in [m]}$, define $P_\Lambda : (\mathbb{F}^n)^\ell \to \mathbb{T}$ as*

$$P_\Lambda(X) = \sum_{i \in [C], j \in [m]} \lambda_{i,j} P_i(L_j(X)).$$

Then one of the following two statements holds:

- $P_\Lambda \equiv 0$.
- P_Λ *is non-constant and* $\left|\mathbf{E}_{X \in (\mathbb{F}^n)^\ell}[\mathsf{e}(P_\Lambda)]\right| < \varepsilon$.

Furthermore $P_\Lambda \equiv 0$ if and only if for every $i \in [C]$, we have $(\lambda_{i,j})_{j \in [m]} \in \Phi_{d_i,t_i}(\mathcal{L})^\perp$ for every $t_i \le k_i$, where d_i, k_i are the degree and depth of P_i, respectively.

We prove Theorem 3.8 in this section. Our proof uses similar derivative techniques as used in [2], but in order to handle the general setting we will need a few technical claims which we present first. Recall that $|L| = \sum_{i=1}^{\ell} |\lambda_i|$ for a linear form $L = (\lambda_1, \dots, \lambda_\ell)$.

Claim 4.4. *Let $d > 0$ be an integer, and let $L = (\lambda_1, \dots, \lambda_\ell) \in \mathbb{F}^\ell$ be a linear form on ℓ variables. There exist linear forms $L_i = (\lambda_{i,1}, \dots, \lambda_{i,\ell}) \in \mathbb{F}^\ell$ for $i = 1, \dots, m$, and coefficients $a_1, \dots, a_m \in \mathbb{Z}$ with $m \leqslant |\mathbb{F}|^\ell$ such that*

- $P(L(X)) = \sum_{i=1}^{m} a_i P(L_i(X))$ *for every degree-d polynomial* $P : \mathbb{F}^n \to \mathbb{T}$;
- $|L_i| \leqslant d$ *for every* $i \in [m]$;
- $|\lambda_{i,j}| \leqslant |\lambda_j|$ *for every* $i \in [m]$ *and* $j \in [\ell]$.

Proof. The proof proceeds by simplifying $P(L(X))$ using identities that are valid for every polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree d.

We prove the statement by induction on $|L|$. For the base case $|L| \leqslant d$ there is nothing to prove. Consider $|L| > d$. Since P is of degree d and $|L| > d$, for every choice of $y_1, \dots, y_{|L|} \in \mathbb{F}^n$, we have

$$\sum_{S \subseteq [|L|]} (-1)^{|S|} P\left(\sum_{i \in S} y_i\right) = (D_{y_1} \dots D_{y_{|L|}} P)(0) \equiv 0. \tag{7}$$

Let $X = (x_1, \dots, x_\ell) \in (\mathbb{F}^n)^\ell$. Setting $|\lambda_i|$ of the vectors $y_1, \dots, y_{|L|}$ to x_i for every $i \in [\ell]$, (7) implies

$$P(L(X)) = \sum_i \alpha_i P(M_i(X)), \tag{8}$$

where $M_i = (\tau_{i,1}, \dots, \tau_{i,\ell})$, $|M_i| \leqslant |L| - 1$ and for every $j \in [\ell]$, $|\tau_{i,j}| \leqslant |\lambda_j|$. Since $|M_i| \leqslant |L| - 1$, we can apply the induction hypothesis to obtain the identities

$$P(M_i(X)) = \sum_{j=1}^{m_j} a_{ij} P(M_{ij}(X))$$

where the linear forms $M_{ij} = (\rho_{ij1}, \dots, \rho_{ij\ell})$ satisfy $|M_{ij}| \leqslant d$, and $|\rho_{ijk}| \leqslant |\tau_{ij}| \leqslant |\lambda_j|$ for every i, j, k. Substituting these identities in (8) yields the desired result. □

The next claim shows that we can further simplify the expression given in Claim 4.4. Let $\mathcal{L}_d \subseteq \mathbb{F}^\ell$ denote the set of nonzero linear forms L with $|L| \leqslant d$ and with the first (left-most) nonzero coefficient equal to 1, e.g. $(0, 1, 0, 2) \in \mathcal{L}_3$ but $(2, 1, 0, 0) \notin \mathcal{L}_3$.

Claim 4.5. *For any linear form $L \in \mathbb{F}^\ell$ and integer $d > 0$, there is a collection of coefficients $\{a_{M,c} \in \mathbb{Z}\}_{M \in \mathcal{L}_d, c \in \mathbb{F}^*}$ such that for every degree-d polynomial $P : \mathbb{F}^n \to \mathbb{T}$,*

$$P(L(X)) = \sum_{M \in \mathcal{L}_d, c \in \mathbb{F}^*} a_{M,c} P(cM(X)). \tag{9}$$

Proof. Similar to the proof of Claim 4.4 we simplify $P(L(X))$ using identities that are valid for every polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree d.

We use induction on the number of nonzero entries of L. The case when L has only one nonzero entry is trivial. For the induction step, choose $c \in \mathbb{F}$ so that the leading nonzero coefficient of $L' = c \cdot L$ is equal to 1. Assume that $L' = (\lambda_1, \dots, \lambda_\ell)$. If $|L'| \leqslant d$ we are done. Assume otherwise that $|L'| > d$. Applying Claim 4.4 for the degree-d polynomial $R(x) := P(c^{-1}x)$ and the linear form L' we can write

$$P(L(X)) = P(c^{-1}L'(X)) = \sum_i \beta_i P(c^{-1}M_i(X)), \tag{10}$$

where for every i, $M_i = (\lambda_{i,1}, \dots, \lambda_{i,\ell})$ satisfies $|M_i| \leqslant d$, and for every $j \in [\ell]$, $\lambda_{i,j} \leqslant \lambda_j$. Let $\mathcal{I}$ denote the set of indices i such that the leading nonzero entry of M_i is one. Then

$$P(L(x)) = \sum_{i \in \mathcal{I}} \alpha_i P(c^{-1} M_i(X)) + \sum_{j \notin \mathcal{I}} \alpha_j P(c^{-1} M_j(X)). \tag{11}$$

Notice that since the leading coefficient of L' is 1, for every $j \notin \mathcal{I}$, M_j has smaller support than L and thus applying the induction hypothesis to the linear forms $c^{-1}M_j$ with $j \notin \mathcal{I}$ concludes the claim. □

Claim 4.5 applies to all polynomials of degree d. If we also specify the depth then we can obtain a stronger statement. However, in this statement we shall need to assume that the polynomials are homogeneous. In fact, the only place in the proof of Theorem 3.8 that we use the homogeneity assumption is when we apply the following claim.

Claim 4.6. *For every d, k, every system of linear forms $\{L_1, \dots, L_m\}$, and constants $\{\lambda_i \in \mathbb{Z}\}_{i \in [m]}$, there exists $\{a_M \in \mathbb{Z}\}_{M \in \mathcal{L}_d}$ such that the following is true for every $t \leqslant k$ and every (d,t)-homogeneous polynomial $P : \mathbb{F}^n \to \mathbb{T}$:*

- $\sum_{i=1}^m \lambda_i P(L_i(X)) \equiv \sum_{M \in \mathcal{L}_d} a_M P(M(X))$*;*
- *For every M with $a_M \neq 0$, we have $|M| \leqslant \deg(a_M P)$.*

Proof. The proof is similar to that of Claim 4.5, except that now we repeatedly apply Claim 4.5 to every term of the form $\lambda P(L(X))$ to express it as a linear combination of $P(cM(X))$ for $M \in \mathcal{L}_{\deg(\lambda P)}$ and $c \in \mathbb{F}^*$. Then we use homogeneity to replace $P(cM(X))$ with $\sigma_c P(M(X))$, where if $c = \zeta^i$ for the fixed generator $\zeta \in \mathbb{F}^*$ then $\sigma_c = \sigma(d,k)^i$. By repeating this procedure we arrive at the desired expansion. □

We are now ready for the proof of our main theorem. For a linear form $L = (\lambda_1, \dots, \lambda_\ell)$, let $\mathrm{lc}(L)$ denote the index of its first nonzero entry, namely $\mathrm{lc}(L) \overset{\mathrm{def}}{=} \min_{i:\lambda_i \neq 0} i$.

Proof of Theorem 3.8. Let d' be the degree of the factor. For every $i \in [C]$, by Claim 4.6 we have

$$\sum_{j=1}^{m} \lambda_{i,j} P_i(L_j(X)) = \sum_{M \in \mathcal{L}_{d'}} \lambda'_{i,M} P_i(M(X)) \tag{12}$$

for some integers $\lambda'_{i,M}$ such that $|M| \leqslant \deg(\lambda'_{i,M} P_i)$ if $\lambda'_{i,M} \neq 0$. The simplifications of Claim 4.6 depend only on the degrees and depths of the polynomials. Hence, if $\lambda'_{i,M} P_i \equiv 0$ for all $M \in \mathcal{L}_{d'}$, then $(\lambda_{i,1}, \dots, \lambda_{i,m}) \in \Phi(d_i, t_i)^\perp$ for any $t_i \leqslant k_i$. So to prove the theorem, it suffices to show that P_Λ has small bias if $\lambda'_{i,M} P_i \not\equiv 0$ for some $M \in \mathcal{L}_{d'}$. Suppose this is true, and thus there exists a nonempty set $\mathcal{M} \subseteq \mathcal{L}_{d'}$ such that

$$P_\Lambda(X) = \sum_{i\in[C], M\in\mathcal{M}} \lambda'_{i,M} P_i(M(X)),$$

and for every $M \in \mathcal{M}$, there is at least one index $i \in [C]$ for which $\lambda'_{i,M}P_i \not\equiv 0$. Choose $i^* \in [C]$ and $M^* \in \mathcal{M}$ in the following manner:

- First, let $M^* \in \mathcal{M}$ be such that $\operatorname{lc}(M^*) = \min_{M\in\mathcal{M}} \operatorname{lc}(M)$, and among these, $|M^*|$ is maximal.
- Then, let $i^* \in [C]$ be such that $\deg(\lambda'_{i^*,M^*}P_{i^*})$ is maximized.

Without loss of generality assume that $i^* = 1$, $\operatorname{lc}(M^*) = 1$, and let $d := \deg(\lambda'_{1,M^*}P_1)$. We claim that if $\sum_{j\in[m]} \lambda_{1,j} P_1(L_j(X))$ is not the zero polynomial, then $\deg(P_\Lambda) \geqslant d$, and moreover, P_Λ has small bias. We prove this by deriving P_Λ in specific directions in a manner that all the terms but $\lambda'_{1,M^*}P_1(M^*(X))$ vanish.

Given a vector $\alpha \in \mathbb{F}^\ell$, an element $y \in \mathbb{F}^n$, and a function $P : (\mathbb{F}^n)^\ell \to \mathbb{T}$, define the derivative of P according to the pair (α, y) as

$$D_{\alpha,y}P(x_1,\ldots,x_\ell) \overset{\text{def}}{=} P(x_1+\alpha_1 y,\cdots,x_\ell+\alpha_\ell y) - P(x_1,\ldots,x_\ell). \tag{13}$$

Note that for every $M \in \mathcal{M}$,

$$\begin{aligned} D_{\alpha,y}(P_i \circ M)(x_1,\cdots,x_\ell) &= P_i(M(x_1,\ldots,x_\ell) + M(\alpha)y) - P_i(M(x_1,\ldots,x_\ell)) \\ &= (D_{\langle M,\alpha\rangle\cdot y}P_i)(M(x_1,\ldots,x_\ell)). \end{aligned}$$

Thus if α is chosen such that $\langle M,\alpha\rangle = 0$ then $D_{\alpha,y}(P_i \circ M) \equiv 0$.

Assume that $M^* = (w_1,\ldots,w_\ell)$, where $w_1 = 1$. Let $t := |M^*|$, $\alpha_1 := e_1 = (1,0,0,\ldots,0) \in \mathbb{F}^\ell$, and let $\alpha_2,\ldots,\alpha_t \in \mathbb{F}^\ell$ be the set of all vectors of the form $(-w,0,\ldots,0,1,0,\ldots,0)$ where 1 is in the i-th coordinate for $i \in [2,\ell]$ and $0 \leqslant w \leqslant w_i - 1$. In addition, pick $\alpha_{t+1} = \cdots = \alpha_d = e_1 \in \mathbb{F}^\ell$.

Claim 4.7.

$$\begin{aligned} &D_{\alpha_1,y_1}\cdots D_{\alpha_d,y_d}P_\Lambda(X) \\ &\quad = \left(D_{\langle M^*,\alpha_1\rangle y_1}\cdots D_{\langle M^*,\alpha_d\rangle y_d} \sum_{\substack{i\in[C]:\\ \deg(\lambda'_{i,M^*}P_i)=d}} \lambda'_{i,M^*}P_i\right)(M^*(X)). \end{aligned} \tag{14}$$

Proof. Deriving according to $(\alpha_1,y_1),\ldots,(\alpha_t,y_t)$ gives

$$\begin{aligned} &D_{\alpha_1,y_1}\cdots D_{\alpha_t,y_t}P_\Lambda(X) \\ &\quad = \left(D_{\langle M^*,\alpha_1\rangle y_1}\cdots D_{\langle M^*,\alpha_t\rangle y_t}\left(\sum_{i=1}^{C} \lambda'_{i,M^*}P_i\right)\right)(M^*(X)). \end{aligned} \tag{15}$$

This is because for every $M = (w'_1, \dots, w'_\ell) \in \mathcal{M} \setminus \{M^*\}$, either $w'_1 = 0$ in which case $\langle M, \alpha_1 \rangle = 0$, or otherwise $w'_1 = 1$ and $|M| \leqslant t$, which implies that $|M| \leqslant |M^*|$ as in this case $\mathrm{lc}(M) = 1$. Thus there exists an index $\xi \in [\ell]$ such that $w'_\xi < w_\xi$; By our choice of $\alpha_2, \dots, \alpha_t$ there is e, $2 \leqslant e \leqslant t$, such that $\alpha_e = (-w'_\xi, 0, \dots, 0, 1, 0, \dots, 0)$ where 1 is in the ξ-th coordinate and thus $\langle M, \alpha_e \rangle = 0$. Now, the claim follows after additionally deriving according to $(\alpha_{t+1}, y_{t+1}), \dots, (\alpha_d, y_d)$, which annihilates all polynomials of degree less than d. □

Claim 4.7 implies that

$$\begin{aligned}
&\mathop{\mathbf{E}}_{y_1,\dots,y_d,X} \left[\mathsf{e}\left(D_{\alpha_1,y_1} \cdots D_{\alpha_d,y_d} P_\Lambda)(X)\right)\right] \\
&= \mathop{\mathbf{E}}_{y_1,\dots,y_d,X} \mathsf{e}\left(\left(D_{\langle M^*,\alpha_1\rangle y_1} \cdots D_{\langle M^*,\alpha_d\rangle y_d} \sum_{\substack{i\in[C]: \\ \deg(\lambda'_{i,M^*} P_i)=d}} \lambda'_{i,M^*} P_i\right)(M^*(X))\right)
\end{aligned}$$

Since X is chosen uniformly at random, by a change of variable we can replace $M^*(X)$ by X. Similarly, since y_j's are chosen uniformly at random and independently, and since $\langle M^*, \alpha_j \rangle \neq 0$ for $j = 1, \dots, d$, we can replace $\langle M^*, \alpha_j \rangle y_j$ by y_j via a change of variables. In other words,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{y_1,\dots,y_d,X} \left[\mathsf{e}\left(D_{\alpha_1,y_1} \cdots D_{\alpha_d,y_d} P_\Lambda)(X)\right)\right] \\
&= \mathop{\mathbf{E}}_{y_1,\dots,y_d,X} \mathsf{e}\left(\left(D_{y_1} \cdots D_{y_d} \sum_{\substack{i\in[C]: \\ \deg(\lambda'_{i,M^*} P_i)=d}} \lambda'_{i,M^*} P_i\right)(X)\right).
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathop{\mathbf{E}}_{y_1,\dots,y_d,X} \left[\mathsf{e}\left(D_{\alpha_1,y_1} \cdots D_{\alpha_d,y_d} P_\Lambda)(x_1, \dots, x_\ell)\right)\right] \\
&= \left\| \mathsf{e}\left(\sum_{\substack{i\in[C]: \\ \deg(\lambda'_{i,M^*} P_i)=d}} \lambda'_{i,M^*} P_i\right) \right\|_{U^d}^{2^d} \leqslant \varepsilon^{2^d},
\end{aligned}$$

where the last inequality holds by the ε-uniformity of the polynomial factor. Now the theorem follows from the next claim from [2] which is a repeated application of the Cauchy–Schwarz inequality. We include a proof for self-containment.

Claim 4.8. *(See [2, Claim 3.4].) For any $\alpha_1, \ldots, \alpha_d \in \mathbb{F}^\ell \backslash \{\underline{0}\}$,*

$$\mathop{\mathbf{E}}_{\substack{y_1,\ldots,y_d,\\ x_1,\ldots,x_\ell}} [\mathsf{e}\left((D_{\alpha_1,y_1} \cdots D_{\alpha_d,y_d} P_\Lambda)(x_1, \ldots, x_\ell)\right)] \geqslant \left(\left| \mathop{\mathbf{E}}_{x_1,\ldots,x_\ell} \mathsf{e}\left(P_\Lambda(x_1, \ldots, x_\ell)\right) \right| \right)^{2^d}.$$

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

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

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

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

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

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

The above proof also implies the following proposition just by omitting the application of Claim 4.6.

Proposition 4.9. *Let $L_1, \ldots, L_m$ be linear forms on ℓ variables and let $\mathcal{B} = (P_1, \ldots, P_C)$ be an ε-uniform polynomial factor of degree $d > 0$ for some $\varepsilon \in (0, 1)$ which is defined by only homogeneous polynomials. For every tuple Λ of integers $(\lambda_{i,j})_{i \in [C], j \in [m]}$,* *define*

$$P_\Lambda(X) = \sum_{i\in[C],j\in[m]} \lambda_{i,j} P_i(L_j(X)),$$

where $P_\Lambda : (\mathbb{F}^n)^\ell \to \mathbb{T}$. Moreover, assume that for every $i \in [C]$, $j \in [m]$, $L_j \in \mathcal{L}_{\deg(\lambda_{i,j}P_i)}$. Then, P_Λ is of degree $d = \max_{i,j} \deg(\lambda_{i,j}P_i)$ and $\|\mathsf{e}(P_\Lambda)\|_{U^d} < \varepsilon$.

4.3. The Gowers–Wolf conjecture: Proof of Theorem 3.17

Theorem 3.17 (restated). *Let $\mathcal{L} = \{L_1, \dots, L_m\}$ be a system of linear forms in ℓ variables, such that $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent. For every $\varepsilon > 0$, there exists $\delta > 0$ such that for any collection of functions $f_1, \dots, f_m : \mathbb{F}^n \to \mathbb{D}$ with $\min_{i\in[m]} \|f_i\|_{U^{d+1}} \leq \delta$, we have*

$$\left| \mathop{\mathbf{E}}_{X\in(\mathbb{F}^n)^k} \left[\prod_{i=1}^m f_i(L_i(X)) \right] \right| \leq \varepsilon. \tag{16}$$

We prove Theorem 3.17 in this section. First, note that since $L_1^{d+1}, ..., L_m^{d+1}$ are linearly independent, it must be the case that $L_1, ..., L_m$ are pairwise linearly independent. Consequently, $\mathcal{L} = \{L_1, \dots, L_m\}$ is of finite Cauchy–Schwarz complexity s for some $s \leqslant m - 2 < \infty$. Indeed, for every $i \in [m]$, the partition of $\{L_j\}_{j\in[m]\setminus\{i\}}$ into $m - 1$ singletons satisfies the requirement of Definition 3.13. The case when $s \leqslant d$ follows from Lemma 3.14, thus we are left with the case when $s > d$. We will prove that $\|f_1\|_{U^{d+1}} \leqslant \delta$ where δ is sufficiently small, implies (16). The theorem then follows by symmetry in the choice of f_1.

We use Corollary 3.5 to decompose $f_i = g_i + h_i$ with

1. $g_i = \mathbf{E}[f_i|\mathcal{B}]$, where $\mathcal{B}$ is an r-regular polynomial factor of degree at most s and complexity $C \leqslant C_{\max}(p, s, \eta, \delta, r(\cdot))$ defined only by homogeneous polynomials, where r is a sufficiently fast growing growth function so that $\mathcal{B}$ is εp^{-5m^2C}-uniform (see Remark 2.18);
2. $\|h_i\|_{U^{s+1}} \leqslant \eta \leqslant \frac{\varepsilon}{2m}$.

We first show that by choosing a sufficiently small η we may replace f_i's in (16) with g_i's.

Claim 4.10. *Choosing $\eta \leqslant \frac{\varepsilon}{2m}$ we have*

$$\left| \mathop{\mathbf{E}}_{X\in(\mathbb{F}^n)^k} \left[\prod_{i=1}^m f_i(L_i(X)) \right] - \mathop{\mathbf{E}}_{X\in(\mathbb{F}^n)^k} \left[\prod_{i=1}^m g_i(L_i(X)) \right] \right| \leqslant \frac{\varepsilon}{2}.$$

Proof. We have

$$\begin{aligned}
&\left|\mathop{\mathbf{E}}_{X}\left[\prod_{i=1}^{m} f_i(L_i(X))\right]-\mathop{\mathbf{E}}_{X}\left[\prod_{i=1}^{m} g_i(L_i(X))\right]\right| \\
&=\left|\sum_{i=1}^{m}\mathop{\mathbf{E}}_{X}\left[h_i(L_i(X))\cdot\prod_{j=1}^{i-1} g_j(L_j(X))\cdot\prod_{j=i+1}^{m} f_j(L_j(X))\right]\right| \\
&\leqslant\sum_{i=1}^{m}\left|\mathop{\mathbf{E}}_{X}\left[h_i(L_i(X))\cdot\prod_{j=1}^{i-1} g_j(L_j(X))\cdot\prod_{j=i+1}^{m} f_j(L_j(X))\right]\right| \\
&\leqslant\sum_{i=1}^{m}\|h_i\|_{U^{s+1}}\leqslant m\cdot\eta\leqslant\frac{\varepsilon}{2},
\end{aligned}$$

where the second inequality follows from Lemma 3.14 since the Cauchy–Schwarz complexity of $\mathcal{L}$ is s. □

Thus it is sufficient to bound $\left|\mathbf{E}_{X\in(\mathbb{F}^n)^k}\left[\prod_{i=1}^{m} g_i(L_i(X))\right]\right|$ by $\varepsilon/2$. For each i, $g_i = \mathbf{E}[f_i|\mathcal{B}]$ and thus

$$g_i(x)=\Gamma_i(P_1(x),\dots,P_C(x)),$$

where $P_1,\dots,P_C$ are the (nonclassical) homogeneous polynomials of degree $\leqslant s$ defining $\mathcal{B}$ and $\Gamma_i:\mathbb{T}^C\to\mathbb{D}$ is a function. Let k_i denote the depth of the polynomial P_i so that by Lemma 2.6, each P_i takes values in $\mathbb{U}_{k_i+1}=\frac{1}{p^{k_i+1}}\mathbb{Z}/\mathbb{Z}$. Let $\Sigma:=\mathbb{Z}_{p^{k_1+1}}\times\cdots\times\mathbb{Z}_{p^{k_C+1}}$. Using the Fourier transform on $\mathbb{U}_{k_1+1}\times\ldots\times\mathbb{U}_{k_C+1}\cong\Sigma$, for every $\tau=(\tau_1,\dots,\tau_C)\in\Sigma$ we have

$$\Gamma_i(\tau)=\sum_{\Lambda=(\lambda_1,\dots,\lambda_C)\in\Sigma}\widehat{\Gamma}_i(\Lambda)\cdot\mathsf{e}\left(\sum_{j=1}^{C}\lambda_j\tau_j\right),\tag{17}$$

where

$$\widehat{\Gamma}_i(\Lambda):=\mathop{\mathbf{E}}_{\tau}\left[\Gamma_i(\tau)\mathsf{e}\left(\sum_{j=1}^{C}\lambda_j\tau_j\right)\right]$$

is the Fourier coefficient of Γ_i corresponding to Λ. Consequently,

$$g_i(x)=\sum_{\Lambda=(\lambda_1,\dots,\lambda_C)\in\Sigma}\widehat{\Gamma}_i(\Lambda)\cdot\mathsf{e}\left(\sum_{j=1}^{C}\lambda_jP_j(x)\right).\tag{18}$$

Let $P_\Lambda:=\sum_{j=1}^{C}\lambda_jP_j(x)$ for the sake of brevity so that we may write

$$\mathop{\mathbf{E}}_{X}\left[\prod_{i=1}^{m} g_i(L_i(X))\right] = \sum_{\Lambda_1,\ldots,\Lambda_m \in \Sigma} \left(\prod_{i=1}^{m} \widehat{\Gamma}_i(\Lambda_i)\right) \cdot \mathop{\mathbf{E}}_{X}\left[\mathsf{e}\left(\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X))\right)\right]. \tag{19}$$

We will show that each term in (19) can be bounded by $\sigma := \frac{\varepsilon}{2|\Sigma|^m}$, thus concluding the proof by the triangle inequality. We will first show that the terms for which $\deg(P_{\Lambda_1}) \leqslant d$ are small.

Claim 4.11. *Let $\Lambda \in \Sigma$ be such that $\deg(P_\Lambda) \leqslant d$. If $\delta \leqslant \frac{\sigma}{2}$, then*

$$\left|\widehat{\Gamma}_1(\Lambda)\right| < \sigma.$$

Proof. It follows from (18) that

$$\widehat{\Gamma}_1(\Lambda) = \mathop{\mathbf{E}}_{x}\left[g_1(x)\mathsf{e}\left(-P_\Lambda(x)\right)\right] - \sum_{\Lambda' \in \Sigma \setminus \{\Lambda\}} \widehat{\Gamma}_1(\Lambda') \cdot \mathop{\mathbf{E}}_{x}\left[\mathsf{e}\left(P_{\Lambda'}(x) - P_\Lambda(x)\right)\right].$$

Note that $\|f_1\|_{U^{d+1}} \leqslant \delta$ and thus

$$\begin{aligned}\left|\mathop{\mathbf{E}}_{x}\left[g_1(x)\mathsf{e}\left(-P_\Lambda(x)\right)\right]\right| &= \left|\mathop{\mathbf{E}}_{x}\left[f_1(x)\mathsf{e}\left(-P_\Lambda(x)\right)\right]\right| \leqslant \left\|f_1\mathsf{e}\left(-P_\Lambda\right)\right\|_{U^{d+1}} \\ &= \|f_1\|_{U^{d+1}} \leqslant \delta \leqslant \sigma/2,\end{aligned}$$

where we used of the fact that $g_1 = \mathbf{E}[f_1|\mathcal{B}]$ and the fact that Gowers norms are increasing in d. Finally, each term of the form $\widehat{\Gamma}_1(\Lambda')\cdot\mathbf{E}_x[\mathsf{e}\left(P_{\Lambda'}(x) - P_\Lambda(x)\right)]$ with $\Gamma' \neq \Gamma$ is bounded in magnitude by $\varepsilon p^{-5m^2C} \leqslant \frac{\sigma}{2|\Sigma|^m}$ as $P_{\Lambda'} - P_\Lambda = P_{\Lambda'-\Lambda}$ is a nonzero linear combination of the polynomials defining the εp^{-5m^2C}-uniform factor $\mathcal{B}$, and $|\widehat{\Gamma}_1(\Lambda')| \leqslant 1$ since f_1 take values in $\mathbb{D}$. □

The above claim allows us to bound the terms from (19) corresponding to tuples $(\Lambda_1, \ldots, \Lambda_m) \in \Sigma^m$ with $\deg(P_{\Lambda_1}) \leqslant d$. This is because for such terms $|\widehat{\Gamma}_1(\Lambda_1)| < \sigma$ by the above claim, and $|\widehat{\Gamma}_i(\Lambda_i)| \leqslant 1$ since f_i's take values in $\mathbb{D}$. It remains to bound the terms for which $\deg(P_{\Lambda_1}) > d$. We will need the following claim.

Lemma 4.12. *Assume that $L_1^{d+1}, \ldots, L_m^{d+1}$ are linearly independent, and let $(\Lambda_1, \ldots, \Lambda_m) \in \Sigma^m$ be such that $\deg(P_{\Lambda_1}) \geqslant d + 1$. Then*

$$\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X)) \not\equiv 0.$$

Since $\mathcal{B}$ is εp^{-5m^2C}-uniform, Lemma 4.12 combined with Theorem 3.8 implies that $\mathbf{E}_X\left[\mathsf{e}\left(\sum_{i=1}^{m} P_{\Lambda_i}(L_i(X))\right)\right] \leqslant \varepsilon p^{-5m^2C} < \sigma$. Hence, every term (whether $\deg(P_{\Lambda_1}) \leqslant d$

or $\deg(P_{\Lambda_1}) \geqslant d+1$) in (19) is bounded in magnitude by σ, and thus by the triangle inequality, the sum in (19) is bounded by $\sigma|\Sigma|^m \leqslant \frac{\varepsilon}{2}$. Combining this with Claim 4.10 concludes the proof of Theorem 3.17. Thus, we are left with proving Lemma 4.12.

Proof of Lemma 4.12. Assume to the contrary that $\sum_{i=1}^m P_{\Lambda_i}(L_i(X)) \equiv 0$. Denoting the coordinates of Λ_i by $(\lambda_{i,1},\dots,\lambda_{i,C}) \in \Sigma^C$ we have

$$\sum_{i=1}^m P_{\Lambda_i}(L_i(X)) = \sum_{i\in[m],j\in[C]} \lambda_{i,j} P_j(L_i(X)) \equiv 0.$$

The last assertion in the statement of Theorem 3.8 implies that for every $j \in [C]$ we must have

$$\sum_{i=1}^m \lambda_{i,j} P_j(L_i(X)) \equiv 0. \tag{20}$$

Since $\deg(P_{\Lambda_1}) \geqslant d+1$, there must exist $j \in [C]$ such that $\lambda_{1,j} \neq 0$ and $\deg(\lambda_{1,j}P_j) \geqslant d+1$. Let $j^* \in [C]$ be such that $\deg(\lambda_{1,j^*}P_{j^*})$ is maximized. Let $t \geqslant 0$ be the largest integer such that p^t divides λ_{i,j^*} for all $i \in [m]$. Note that $d^* := \deg(p^t P_{j^*}) \geqslant d+1$. Moreover, since P_{j^*} belongs to an εp^{-5m^2C}-uniform factor, $\{p^t P_{j^*}\}$ is εp^{-5m^2C}-uniform. Thus applying the last assertion of Theorem 3.8 to the degree-d^* polynomial $p^t P_{j^*}$, we have that for every classical homogeneous polynomial Q of degree d^*,

$$\sum_{i=1}^m \frac{\lambda_{i,j^*}}{p^t} Q(L_i(X)) \equiv 0.$$

Now choosing the polynomial $Q(x_1,\dots,x_n) = x_1 \cdots x_{d^*}$ and looking at the coefficients of the monomials of degree d^*, we have

$$\sum_{i=1}^m \frac{\lambda_{i,j^*}}{p^t} L_i^{d^*} \equiv 0 \mod p.$$

By our choice of t, there exists $i \in [m]$, such that $\frac{\lambda_{i,j^*}}{p^t} \neq 0 \mod p$, and the above is a nonzero linear combination. Since $d^* \geqslant d+1$, this contradicts the assumption that $L_1^{d+1},\dots,L_m^{d+1}$ are linearly independent. □

Acknowledgments

We thank the anonymous referees for carefully reading the paper and correcting minor errors. We also thank an anonymous referee for pointing out the connection of Lemma 3.3 to the notion of Teichmüller characters.

References

[1] V. Bergelson, T. Tao, T. Ziegler, An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}_p^\infty$, Geom. Funct. Anal. 19 (6) (2010) 1539–1596. MR2594614 (2011b:37009).

[2] A. Bhattacharyya, E. Fischer, H. Hatami, P. Hatami, S. Lovett, Every locally characterized affine-invariant property is testable, in: Proceedings of the 45th Annual ACM Symposium on Theory of Computing, STOC'13, ACM, New York, NY, USA, 2013, pp. 429–436.

[3] A. Bhattacharyya, E. Fischer, S. Lovett, Testing low complexity affine-invariant properties, in: Proc. 24th ACM-SIAM Symposium on Discrete Algorithms, 2013, pp. 1337–1355.

[4] H. Cohen, Number Theory: Vol. I: Tools and Diophantine Equations, Grad. Texts in Math., vol. 239, Springer Science & Business Media, 2008.

[5] W.T. Gowers, A new proof of Szemerédi's theorem, Geom. Funct. Anal. 11 (3) (2001) 465–588. MR1844079 (2002k:11014).

[6] W.T. Gowers, Decompositions, approximate structure, transference, and the Hahn–Banach theorem, Bull. Lond. Math. Soc. 42 (4) (2010) 573–606. MR2669681 (2011k:11015).

[7] W.T. Gowers, J. Wolf, The true complexity of a system of linear equations, Proc. Lond. Math. Soc. (3) 100 (1) (2010) 155–176. MR2578471 (2011a:11019).

[8] W.T. Gowers, J. Wolf, Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$, Geom. Funct. Anal. 21 (1) (2011) 36–69. MR2773103 (2012f:11024).

[9] B. Green, Montréal notes on quadratic Fourier analysis, in: Additive Combinatorics, in: CRM Proc. Lecture Notes, vol. 43, Amer. Math. Soc., Providence, RI, 2007, pp. 69–102. MR2359469 (2008m:11047).

[10] B. Green, T. Tao, The distribution of polynomials over finite fields, with applications to the Gowers norms, Contrib. Discrete Math. 4 (2) (2009) 1–36. MR2592422 (2011d:11022).

[11] B. Green, T. Tao, An arithmetic regularity lemma, an associated counting lemma, and applications, in: An Irregular Mind, in: Bolyai Soc. Math. Stud., vol. 21, János Bolyai Math. Soc., Budapest, 2010, pp. 261–334. MR2815606.

[12] B. Green, T. Tao, Linear equations in primes, Ann. of Math. (2) 171 (3) (2010) 1753–1850. MR2680398 (2011j:11177).

[13] B. Green, T. Tao, T. Ziegler, An inverse theorem for the Gowers $U^{s+1}[N]$-norm, Ann. of Math. (2) 176 (2) (2012) 1231–1372. MR2950773.

[14] H. Hatami, S. Lovett, Higher-order Fourier analysis of $\mathbb{F}_p^n$ and the complexity of systems of linear forms, Geom. Funct. Anal. 21 (6) (2011) 1331–1357. MR2860190 (2012m:11016).

[15] T. Kaufman, S. Lovett, Worst case to average case reductions for polynomials, in: IEEE Annual Symposium on Foundations of Computer Science, 2008, pp. 166–175.

[16] A. Samorodnitsky, Low-degree tests at large distances, in: STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 506–515. MR2402476 (2009f:68077).

[17] B. Szegedy, On higher order Fourier analysis, arXiv:1203.2260.

[18] T. Tao, Structure and randomness in combinatorics, in: FOCS'07. 48th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2007, pp. 3–15.

[19] T. Tao, T. Ziegler, The inverse conjecture for the Gowers norm over finite fields via the correspondence principle, Anal. PDE 3 (1) (2010) 1–20. MR2663409 (2011j:11018).

[20] T. Tao, T. Ziegler, The inverse conjecture for the Gowers norm over finite fields in low characteristic, Ann. Comb. 16 (1) (2012) 121–188. MR2948765.