

# Limits of Boolean Functions on $\mathbb{F}_p^n$

Hamed Hatami\*

McGill University  
Quebec, Canada

hatami@cs.mcgill.ca

Pooya Hatami

University of Chicago  
Illinois, U.S.A.

pooya@cs.uchicago.edu

James Hirst†

McGill University  
Quebec, Canada

james.hirst@mail.mcgill.ca

Submitted: Jun 9, 2014; Accepted: Sep 19, 2014; Published: Oct 2, 2014

Mathematics Subject Classifications: 11B30

## Abstract

We study sequences of functions of the form  $\mathbb{F}_p^n \rightarrow \{0,1\}$  for varying  $n$ , and define a notion of convergence based on the induced distributions from restricting the functions to a random affine subspace. Using a decomposition theorem and a recently proven equi-distribution theorem from higher-order Fourier analysis, we prove that the limits of such convergent sequences can be represented by certain measurable functions. We also show that every such limit object arises as the limit of some sequence of functions. These results are in the spirit of similar results which have been developed for limits of graph sequences. A more general, albeit substantially more sophisticated, limit object was recently constructed by Balázs Szegedy [Gowers norms, regularization and limits of functions on abelian groups, 2010. [arXiv:1010.6211](https://arxiv.org/abs/1010.6211)].

## 1 Introduction

In limit theories of discrete structures, one often studies a large object by studying its “local statistics”. More precisely, given a sampling rule that allows one to sample a random substructure, there is an induced probability measure on the set of possible small substructures. For example, given a graph  $G$  and a positive integer  $k$ , one can select  $k$  random vertices in  $G$  and look at the subgraph induced by  $G$  on these  $k$  vertices. This induces a probability distribution on  $k$ -vertex graphs. Every such sampling rule leads to a notion of convergence. Namely, a sequence of structures is called convergent if these probability distributions converge. So, in the above example, a sequence of graphs is

---

\*This author’s research was supported by an NSERC, and an FQRNT grant.

†This author’s research was partly supported by NSERC.

called convergent [LS06] if, for every  $k$ , the corresponding probability distributions on  $k$ -vertex graphs converges.

Let  $p$  be a fixed prime, and denote  $\mathbb{F} = \mathbb{F}_p$ . In this article we are interested in functions from vector spaces over the field  $\mathbb{F}$  to the set  $\{0, 1\}$ , or equivalently the subsets of such vector spaces. Given a subset of  $\mathbb{F}^n$ , the local information we would like to work with is the distribution of small linear structures, e.g. arithmetic progressions, contained within the set. To this end, we use the sampling rule that considers the restriction of a given function to a random affine subspace. We will describe a restriction of  $\mathbb{F}^n$  to an affine subspace of dimension  $k$  by an affine transformation  $A : \mathbb{F}^k \rightarrow \mathbb{F}^n$ . Recall that an affine transformation  $A$  is of the form  $L + c$  where  $L$  is linear and  $c$  is a constant vector in  $\mathbb{F}^n$ . More precisely, given a function  $f : \mathbb{F}^n \rightarrow \{0, 1\}$  and a positive integer  $k$ , we select a random affine transformation  $A : \mathbb{F}^k \rightarrow \mathbb{F}^n$  uniformly, and consider the random variable  $Af : \mathbb{F}^k \rightarrow \{0, 1\}$  defined as  $Af : x \mapsto f(Ax)$ . This induces a probability distribution on the set of functions  $\{\mathbb{F}^k \rightarrow \{0, 1\}\}$ . A sequence of functions  $\{f_i : \mathbb{F}^{n_i} \rightarrow \{0, 1\}\}_{i \in \mathbb{N}}$  is called *convergent* if these probability distributions converge for every  $k$ .

It is worth mentioning that, in the context of this paper, it is simply more natural to work with affine subspaces rather than linear ones. This is partly because most substructures of interest (e.g., arithmetic progressions) are translation invariant and, more generally, invariant under affine transformations. Consequently, it is desirable to require that a function  $f : \mathbb{F}^n \rightarrow \{0, 1\}$  and its translations  $f_a(x) := f(x + a)$  for  $a \in \mathbb{F}^n$  have the same local statistics. This would have not been the case had we defined convergence by restricting to random linear subspaces. Additionally, since linear subspaces always contain the origin, the corresponding notion of convergence is too sensitive to the values of the function at 0. For these reasons, we define convergence via restricting to affine subspaces. However, the techniques of this paper can be extended to other similar notions of convergence.

Now, the purpose of this article is to describe a limit object for such convergent sequences of functions. Before we can state our results we need to recall some results from higher-order Fourier analysis.

## 2 Basic background

Most of the material in this section is directly quoted from the full version of [BFH<sup>+</sup>13].

**Notation** For  $d \in \mathbb{N} \cup \{\infty\}$ , denote  $[d] := \{1, \dots, d\}$  if  $d < \infty$ , and  $[d] = \mathbb{N}$  otherwise. We shorthand  $\mathbb{F} = \mathbb{F}_p$  for a prime finite field. For  $f : \mathbb{F}^n \rightarrow \mathbb{C}$  we denote  $\|f\|_1 = \mathbf{E}[|f(x)|]$ ,  $\|f\|_2^2 = \mathbf{E}[|f(x)|^2]$  where  $x \in \mathbb{F}^n$  is chosen uniformly and  $\|f\|_\infty = \max |f(x)|$ . Note that  $\|f\|_1 \leq \|f\|_2 \leq \|f\|_\infty$ . The expression  $o_m(1)$  denotes quantities which approach zero as  $m$  grows. We shorthand  $x \pm \varepsilon$  for any quantity in  $[x - \varepsilon, x + \varepsilon]$ .

## 2.1 Uniformity norms and non-classical polynomials

**Definition 2.1** (Multiplicative Derivative). *Given a function  $f : \mathbb{F}^n \rightarrow \mathbb{C}$  and an element  $h \in \mathbb{F}^n$ , define the multiplicative derivative in direction  $h$  of  $f$  to be the function  $\Delta_h f : \mathbb{F}^n \rightarrow \mathbb{C}$  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  $f$  is determined by the expected multiplicative derivative of  $f$  in  $d$  random directions at a random point.

**Definition 2.2** (Gowers norm). *Given a function  $f : \mathbb{F}^n \rightarrow \mathbb{C}$  and an integer  $d \geq 1$ , the Gowers norm of order  $d$  for  $f$  is given by*

$$\|f\|_{U^d} = \left| \mathbf{E}_{h_1, \dots, h_d, x \in \mathbb{F}^n} [(\Delta_{h_1} \Delta_{h_2} \cdots \Delta_{h_d} f)(x)] \right|^{1/2^d}.$$

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  $P : \mathbb{F}^n \rightarrow \mathbb{F}$  is a polynomial of degree  $< d$ , then  $\|f\|_{U^d} = 1$ . If  $d < p$  and  $\|f\|_\infty \leq 1$ , then in fact, the converse holds, meaning that any function  $f : \mathbb{F}^n \rightarrow \mathbb{C}$  satisfying  $\|f\|_\infty \leq 1$  and  $\|f\|_{U^d} = 1$  is of this form. But when  $d \geq p$ , the converse is no longer true. In order to characterize functions  $f : \mathbb{F}^n \rightarrow \mathbb{C}$  with  $\|f\|_\infty \leq 1$  and  $\|f\|_{U^d} = 1$ , we need to define the notion of *non-classical polynomials*.

Non-classical polynomials might not necessarily be  $\mathbb{F}$ -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  $k \geq 0$ , let  $\mathbb{U}_k$  denote the subgroup  $\frac{1}{p^k}\mathbb{Z}/\mathbb{Z} \subseteq \mathbb{T}$ . Let  $\mathbf{e} : \mathbb{T} \rightarrow \mathbb{C}$  denote the character  $\mathbf{e}(x) = e^{2\pi i x}$ .

**Definition 2.3** (Additive Derivative). *Given a function<sup>1</sup>  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  and an element  $h \in \mathbb{F}^n$ , define the additive derivative in direction  $h$  of  $P$  to be the function  $D_h P : \mathbb{F}^n \rightarrow \mathbb{T}$  satisfying  $D_h P(x) = P(x+h) - P(x)$  for all  $x \in \mathbb{F}^n$ .*

**Definition 2.4** (Non-classical polynomials). *For an integer  $d \geq 0$ , a function  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  is said to be a non-classical polynomial of degree  $\leq d$  (or simply a polynomial of degree  $\leq d$ ) if for all  $h_1, \dots, h_{d+1}, x \in \mathbb{F}^n$ , it holds that*

$$(D_{h_1} \cdots D_{h_{d+1}} P)(x) = 0. \tag{1}$$

*The degree of  $P$  is the smallest  $d$  for which the above holds. A function  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  is said to be a classical polynomial of degree  $\leq d$  if it is a non-classical polynomial of degree  $\leq d$  whose image is contained in  $\mathbb{U}_1$ .*

It is a direct consequence of Definition 2.4 that a function  $f : \mathbb{F}^n \rightarrow \mathbb{C}$  with  $\|f\|_\infty \leq 1$  satisfies  $\|f\|_{U^{d+1}} = 1$  if and only if  $f = \mathbf{e}(P)$  for a (non-classical) polynomial  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  of degree  $\leq d$ .

---

<sup>1</sup>We try to adhere to the following convention: upper-case letters (e.g.  $F$  and  $P$ ) to denote functions mapping from  $\mathbb{F}^n$  to  $\mathbb{T}$  or to  $\mathbb{F}$ , lower-case letters (e.g.  $f$  and  $g$ ) to denote functions mapping from  $\mathbb{F}^n$  to  $\mathbb{C}$ , and upper-case Greek letters (e.g.  $\Gamma$  and  $\Sigma$ ) to denote functions mapping  $\mathbb{T}^C$  to  $\mathbb{T}$ .

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

**Lemma 2.5** (Lemma 1.7 in [TZ12]). *Let  $d \geq 1$  be an integer. A function  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  is a polynomial of degree  $\leq d$  if and only if  $P$  can be represented as*

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

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

Note that Lemma 2.5 immediately implies the following important observation:

**Remark 2.6.** *If  $Q : \mathbb{F}^n \rightarrow \mathbb{T}$  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  $\text{depth}(\lambda Q) = k$ .*

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  $\mathbb{U}_{k+1}$ . By a  $(d, k)$ -polynomial we mean a polynomial of degree  $d$  and depth  $k$ .

## 2.2 Rank

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 this: 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 2.7** (Rank of a polynomial). *Given a polynomial  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  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  $Q_1, \dots, Q_r : \mathbb{F}^n \rightarrow \mathbb{T}$  of degree  $\leq d - 1$  and a function  $\Gamma : \mathbb{T}^r \rightarrow \mathbb{T}$  satisfying  $P(x) = \Gamma(Q_1(x), \dots, Q_r(x))$ . If  $d = 1$ , then 1-rank is defined to be  $\infty$  if  $P$  is non-constant and 0 otherwise.*

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

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

**Definition 2.8** (Uniformity). Let  $\varepsilon > 0$  be a real. A degree  $d$  polynomial  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  is said to be  $\varepsilon$ -uniform if

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

The following theorem of Tao and Ziegler relates rank to uniformity by showing that high rank polynomials have small Gowers norms.

**Theorem 2.9** (Theorem 1.20 of [TZ12]). For any  $\varepsilon > 0$  and integer  $d > 0$ , there exists an integer  $r = r_{2.9}(d, \varepsilon)$  such that the following is true. For any polynomial  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  of degree  $\leq d$ , if  $\|\mathbf{e}(P)\|_{U^d} \geq \varepsilon$ , then  $\text{rank}_d(P) \leq r$ .

This immediately implies that a sufficiently regular polynomial will be uniform.

**Corollary 2.10.** Let  $\varepsilon$ ,  $d$ , and  $r = r_{2.9}(d, \varepsilon)$  be as in Theorem 2.9. Every  $r$ -regular polynomial of degree  $d$  is also  $\varepsilon$ -uniform.

Finally, we need to extend our definitions of rank and uniformity to a collection of polynomials. In particular, we will need these when the collection comes from a polynomial factor.

**Definition 2.11** (Rank and Regularity). A polynomial factor  $\mathcal{B}$  defined by a sequence of polynomials  $P_1, \dots, P_C : \mathbb{F}^n \rightarrow \mathbb{T}$  with respective depths  $k_1, \dots, k_C$  is said to have rank  $r$  if  $r$  is the least integer for which there exist weights  $(\lambda_1, \dots, \lambda_C) \in \mathbb{Z}^C$ , with  $(\lambda_1 \bmod p^{k_1+1}, \dots, \lambda_C \bmod p^{k_C+1}) \neq 0^C$ , such that  $\text{rank}_d\left(\sum_{i=1}^C \lambda_i P_i\right) \leq r$ , where  $d = \max_i \deg(\lambda_i P_i)$ .

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

**Definition 2.12** (Uniform Factor). Let  $\varepsilon > 0$  be a real. A polynomial factor  $\mathcal{B}$  defined by a sequence of polynomials  $P_1, \dots, P_C : \mathbb{F}^n \rightarrow \mathbb{T}$  with respective depths  $k_1, \dots, k_C$  is said to be  $\varepsilon$ -uniform if for every collection  $(\lambda_1, \dots, \lambda_C) \in \mathbb{Z}^C$ , with at least one  $\lambda_i \bmod p^{k_i+1}$  not 0, we have

$$\left\| \mathbf{e}\left(\sum_{i=1}^C \lambda_i P_i\right) \right\|_{U^d} < \varepsilon,$$

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

**Remark 2.13.** Let  $\varepsilon : \mathbb{N} \rightarrow \mathbb{R}_+$  be an arbitrary non-increasing function. Similar to Corollary 2.10, it also follows from Theorem 2.9 that an  $r$ -regular, degree  $d$  factor  $\mathcal{B}$  is also  $\varepsilon(|\mathcal{B}|)$ -uniform when  $r = r_{2.9}(d, \varepsilon(\cdot))$ .

The following decomposition theorem is one of the main tools in higher-order Fourier analysis.

**Theorem 2.14** (Strong Decomposition Theorem; Theorem 4.4 of [BFL13]). *Suppose  $\delta > 0$  and  $d \geq 1$  are integers. Let  $\eta : \mathbb{N} \rightarrow \mathbb{R}^+$  and  $r : \mathbb{N} \rightarrow \mathbb{N}$  be arbitrary non-decreasing functions. Then there exists  $C = C_{2.14}(\delta, \eta, r, d)$  such that the following holds.*

*Given  $f : \mathbb{F}^n \rightarrow [0, 1]$ , there exist three functions  $f_1, f_2, f_3 : \mathbb{F}^n \rightarrow \mathbb{R}$  and a polynomial factor  $\mathcal{B}$  of degree at most  $d$  and complexity at most  $C$  such that the following conditions hold:*

- (i)  $f = f_1 + f_2 + f_3$ .
- (ii)  $f_1 = \mathbf{E}[f|\mathcal{B}]$ .
- (iii)  $\|f_2\|_{U^{d+1}} \leq 1/\eta(|\mathcal{B}|)$ .
- (iv)  $\|f_3\|_2 \leq \delta$ .
- (v)  $f_1$  and  $f_1 + f_3$  have range  $[0, 1]$ ;  $f_2$  and  $f_3$  have range  $[-1, 1]$ .
- (vi)  $\mathcal{B}$  is  $r$ -regular.

**Remark 2.15.** *As in Remark 2.13, taking  $r = r_{2.9}(d, \varepsilon(\cdot))$ , we obtain a decomposition where now condition (vi) reads that the factor  $\mathcal{B}$  is  $\varepsilon(|\mathcal{B}|)$ -uniform, where  $\varepsilon : \mathbb{N} \rightarrow \mathbb{R}_+$  is an arbitrary non-increasing function.*

## 2.3 Complexity of systems of linear forms

In this article, a *linear form* in  $k$  variables is a vector  $L = (\lambda_1, \dots, \lambda_k) \in \mathbb{F}^k$  regarded as a linear function from  $V^k$  to  $V$  for any vector space  $V$  over  $\mathbb{F}$ : If  $\mathbf{x} = (x_1, \dots, x_k) \in V^k$ , then  $L(\mathbf{x}) := \lambda_1 x_1 + \dots + \lambda_k x_k$ . A linear form  $L = (\lambda_1, \lambda_2, \dots, \lambda_k)$  is said to be *affine* if  $\lambda_1 = 1$ . Affine linear forms are important in this work for the following reason. If  $\{L_1, \dots, L_m\} \subseteq \mathbb{F}^k$  and  $A : \mathbb{F}^k \rightarrow \mathbb{F}^n$  is a uniform random affine transformation, then  $(AL_1, \dots, AL_m)$  has the same distribution as  $(x_0 + L_1(x_1, \dots, x_k), \dots, x_0 + L_m(x_1, \dots, x_k))$  where  $x_0, \dots, x_k$  are uniform and independent random vectors in  $\mathbb{F}^n$ . The vectors  $(x_0 + L_1(x_1, \dots, x_k), \dots, x_0 + L_m(x_1, \dots, x_k))$  can be considered as the application of affine linear forms to the variables  $x_0, \dots, x_k$ . On the other hand if  $\{L_1, \dots, L_m\} \subseteq \mathbb{F}^k$  are all affine linear forms, and  $A : \mathbb{F}^k \rightarrow \mathbb{F}^n$  is a uniform random affine transformation, then  $(AL_1, \dots, AL_m)$  has the same distribution as  $(L_1(x_1, \dots, x_k), \dots, L_m(x_1, \dots, x_k))$  where  $x_1, \dots, x_k$  are uniform and independent random vectors in  $\mathbb{F}^n$ .

A *system of linear forms* in  $k$  variables is a finite set  $\mathcal{L} \subseteq \mathbb{F}^k$  of linear forms in  $k$  variables. A system of linear forms is called *affine* if it consists only of affine linear forms. For the reasons described in the previous paragraph we will mainly be interested in affine systems of linear forms.

Given a function  $f : \mathbb{F}^n \rightarrow \mathbb{C}$  and a system of linear forms  $\mathcal{L} = \{L_1, \dots, L_m\} \subseteq \mathbb{F}^\ell$ , define

$$t_{\mathcal{L}}(f) = \mathbf{E}_{x_1, \dots, x_\ell} \left[ \prod_{L \in \mathcal{L}} f(L(x_1, \dots, x_\ell)) \right]. \quad (2)$$

**Definition 2.16.** *A system of linear forms  $\mathcal{L} = \{L_1, \dots, L_m\} \subseteq \mathbb{F}^\ell$  is said to be of true*

complexity at most  $d$  if there exists a function  $\delta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $\lim_{\varepsilon \rightarrow 0} \delta(\varepsilon) = 0$  and

$$\left| \mathbf{E}_{x_1, \dots, x_\ell \in \mathbb{F}^n} \left[ \prod_{i=1}^m f_i(L_i(x_1, \dots, x_\ell)) \right] \right| \leq \min_i \delta(\|f_i\|_{U^{d+1}}),$$

for all  $f_1, \dots, f_m : \mathbb{F}^n \rightarrow [-1, 1]$ .

**Remark 2.17.** It is known [GW10] that the true complexity of an affine system linear forms  $\mathcal{L} = \{L_1, \dots, L_m\}$  is at most  $m$ .

The following lemma shows that for a proper decomposition  $f = f_1 + f_2 + f_3$ , the average  $t_{\mathcal{L}}(f)$  is a good approximation of  $t_{\mathcal{L}}(f_1)$ .

**Lemma 2.18.** Let  $\mathcal{L} = \{L_1, \dots, L_m\}$  be a system of linear forms of true complexity at most  $d$  and  $\varepsilon > 0$  be a constant. Let  $f : \mathbb{F}^n \rightarrow \{0, 1\}$  be decomposed to  $f = f_1 + f_2 + f_3$  according to Theorem 2.14. Then

$$|t_{\mathcal{L}}(f) - t_{\mathcal{L}}(f_1)| \leq \varepsilon,$$

provided that  $\delta$  is sufficiently small and  $\eta$  and  $r$  grow sufficiently fast.

*Proof.* We can expand  $t_{\mathcal{L}}(f) = t_{\mathcal{L}}(f_1 + f_2 + f_3)$  as

$$\sum_{(i_j)_{j \in [m]} \in [3]^m} \mathbf{E} \left[ \prod_{j=1}^m f_{i_j}(L_j(x_1, \dots, x_\ell)) \right].$$

Most of the terms in this sum are negligible: If any  $i_j = 2$ , then from the decomposition theorem and the definition of true complexity we get that the summand is at most  $\delta'(1/\eta(|\mathcal{B}|))$ , where  $\delta'$  is from Definition 2.16. Additionally, if some  $i_j = 3$ , then an application of Cauchy-Schwarz bounds the summand by  $\delta$ . The only other term is precisely  $t_{\mathcal{L}}(f_1)$ , so we get

$$|t_{\mathcal{L}}(f) - t_{\mathcal{L}}(f_1)| \leq 3^m \max\{\delta'(1/\eta(|\mathcal{B}|)), \delta\} = o_{\eta, r, \delta}(1).$$

□

### 3 Main results

We study the following notions of convergence. Given a function  $f : \mathbb{F}^n \rightarrow \{0, 1\}$  and an affine system of linear forms  $\mathcal{L} \subseteq \mathbb{F}^k$ , we can select a random affine transformation  $A : \mathbb{F}^k \rightarrow \mathbb{F}^n$  uniformly, and then consider the restriction of  $Af$  to the set  $\mathcal{L}$ . This induces a probability distribution  $\mu_f(\mathcal{L})$  over the set of functions  $\{\mathcal{L} \rightarrow \{0, 1\}\}$ .

**Definition 3.1.** A sequence of functions  $\{f_i : \mathbb{F}^{n_i} \rightarrow \{0, 1\}\}_{i \in \mathbb{N}}$  is called  $d$ -convergent if for every  $k$  and every  $\mathcal{L} \subseteq \mathbb{F}^k$  of true complexity at most  $d$ , the probability distributions  $\mu_{f_i}(\mathcal{L})$  converge.

It follows from Remark 2.17 that  $\{f_i : \mathbb{F}^{n_i} \rightarrow \{0, 1\}\}_{i \in \mathbb{N}}$  is convergent if it is  $d$ -convergent for every  $d$ . Thus in the sequel, the term  $\infty$ -convergence means convergence. Our first goal is to find a representation for the limit of a  $d$ -convergent sequence.

**Definition 3.2.** For every  $d \in \mathbb{N} \cup \{\infty\}$ , denote  $\mathbb{G}_d = \prod_{j=1}^d \prod_{k=0}^{\lfloor \frac{j-1}{p-1} \rfloor} \mathbb{U}_{k+1}^{\mathbb{N}}$ . So every element in this group is of the form  $a = (a_{j,k,i} : j \in [d], k \in [0, \lfloor \frac{j-1}{p-1} \rfloor], i \in \mathbb{N})$ , with each  $a_{j,k,i} \in \mathbb{U}_{k+1}$ .

A  $d$ -limit object is a measurable function  $\Gamma : \mathbb{G}_d \rightarrow [0, 1]$ .

To be precise,  $\mathbb{G}_d$  is a compact Abelian group with the topology generated by the open (clopen, even) sets which are the subsets of  $\mathbb{G}_d$  obtained by fixing finitely many coordinates. The unique Haar probability measure on  $\mathbb{G}_d$  can be obtained by specifying it on these sets (in the obvious way) and extending it to  $\mathbb{G}_d$  via Carathéodory's extension theorem.

In order to prove that  $d$ -limit objects correspond to the limits of  $d$ -convergent sequences, we need to define the probability distribution that a  $d$ -limit object induces on  $\{\mathcal{L} \rightarrow \{0, 1\}\}$ . First we need another definition.

**Definition 3.3** (Consistency). Let  $\mathcal{L} = \{L_1, \dots, L_m\}$  be a system of linear forms. A sequence of elements  $b_1, \dots, b_m \in \mathbb{T}$  is said to be  $(d, k)$ -consistent with  $\mathcal{L}$  if there exists a  $(d, k)$ -polynomial  $P$  and a point  $x$  such that  $P(L_i(x)) = b_i$  for every  $i \in [m]$ .

Given vectors  $\mathbf{d} = (d_1, \dots, d_C) \in \mathbb{Z}_{\geq 0}^C$  and  $\mathbf{k} = (k_1, \dots, k_C) \in \mathbb{Z}_{\geq 0}^C$ , a sequence of vectors  $\mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{T}^C$  is said to be  $(\mathbf{d}, \mathbf{k})$ -consistent with  $\mathcal{L}$  if for every  $i \in [C]$ , the elements  $\mathbf{b}_1(i), \dots, \mathbf{b}_m(i)$  are  $(d_i, k_i)$ -consistent with  $\mathcal{L}$ .

If  $\mathcal{B}$  is a polynomial factor, the term  $\mathcal{B}$ -consistent with  $\mathcal{L}$  is a synonym for  $(\mathbf{d}, \mathbf{k})$ -consistent with  $\mathcal{L}$  where  $\mathbf{d} = (d_1, \dots, d_C)$  and  $\mathbf{k} = (k_1, \dots, k_C)$  are respectively the degree and depth sequences of polynomials defining  $\mathcal{B}$ .

A sequence of elements  $\mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{G}_d$  is consistent with  $\mathcal{L}$  if for every  $j \in [d]$ ,  $k \in [0, \lfloor \frac{j-1}{p-1} \rfloor]$ , and  $i \in \mathbb{N}$ , the elements  $\mathbf{b}_1(j, k, i), \dots, \mathbf{b}_m(j, k, i)$  are  $(j, k)$ -consistent with  $\mathcal{L}$ .

Consider a highly uniform polynomial factor  $\mathcal{B}$  of degree  $d > 0$ , defined by the collection of polynomials  $P_1, \dots, P_C : \mathbb{F}^n \rightarrow \mathbb{T}$ , with respective degrees  $d_1, \dots, d_C$  and depths  $k_1, \dots, k_C$ . Let  $\mathcal{L} = \{L_1, \dots, L_m\}$  be an affine system of linear forms on  $\ell$  variables. We are interested in the distribution of the random matrix

$$\begin{pmatrix} P_1(L_1(X)) & P_2(L_1(X)) & \cdots & P_C(L_1(X)) \\ P_1(L_2(X)) & P_2(L_2(X)) & \cdots & P_C(L_2(X)) \\ \vdots & \vdots & & \vdots \\ P_1(L_m(X)) & P_2(L_m(X)) & \cdots & P_C(L_m(X)) \end{pmatrix}, \quad (3)$$

where  $X$  is a uniform random variable taking values in  $(\mathbb{F}^n)^\ell$ . Note that by the definition of consistency, for every  $i \in [C]$ , the  $i$ -th column of this matrix is always  $(d_i, k_i)$ -consistent with  $\mathcal{L}$ . To prove our main result we will use the following equi-distribution theorem from



[BFH<sup>+</sup>13], which shows that if the factor  $\mathcal{B}$  is sufficiently uniform, then the distribution of Eq. (3) can be made arbitrarily close to uniform over matrices satisfying the consistency condition above.

**Theorem 3.4** (See Theorem 3.10 of [BFH<sup>+</sup>13]). *Given  $\varepsilon > 0$ , let  $\mathcal{B}$  be an  $\varepsilon$ -uniform polynomial factor of degree  $d > 0$  and complexity  $C$ , that is defined by a tuple of polynomials  $P_1, \dots, P_C : \mathbb{F}^n \rightarrow \mathbb{T}$  having respective degrees  $d_1, \dots, d_C$  and respective depths  $k_1, \dots, k_C$ . Let  $\mathcal{L} = \{L_1, \dots, L_m\}$  be an affine system of linear forms on  $\ell$  variables.*

*Suppose  $\mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{T}^C$  are atoms of  $\mathcal{B}$  that are  $\mathcal{B}$ -consistent with  $\mathcal{L}$ . Then*

$$\Pr_{x_1, \dots, x_\ell} [\mathcal{B}(L_j(x_1, \dots, x_\ell)) = \mathbf{b}_j \ \forall j \in [m]] = \frac{1}{K} \pm \varepsilon,$$

where  $K$  is the number of tuples  $(\mathbf{b}_1, \dots, \mathbf{b}_m)$  that are  $\mathcal{B}$ -consistent with  $\mathcal{L}$ .

Consider a  $d$ -limit object  $\Gamma : \mathbb{G}_d \rightarrow [0, 1]$ . For an affine system of linear forms  $\mathcal{L} = \{L_1, \dots, L_m\}$ , select  $\mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{G}_d$  at random (according to the Haar measure on  $\mathbb{G}_d$ , defined following Definition 3.2), conditioned on being consistent with  $\mathcal{L}$ . Then define the random function  $g : \mathcal{L} \rightarrow \{0, 1\}$  by setting  $g(L_i) = 1$  with probability  $\Gamma(\mathbf{b}_i)$  and  $g(L_i) = 0$  with probability  $1 - \Gamma(\mathbf{b}_i)$  independently for every  $i \in [m]$ . This introduces a probability measure  $\mu_\Gamma(\mathcal{L})$  on the set of functions  $\{\mathcal{L} \rightarrow \{0, 1\}\}$ .

We say that a sequence of functions  $\{f_i : \mathbb{F}^{n_i} \rightarrow \{0, 1\}\}_{i \in \mathbb{N}}$   $d$ -converges to  $\Gamma$  if for every affine system of linear forms  $\mathcal{L}$  of true complexity at most  $d$ , the probability measures  $\mu_{f_i}(\mathcal{L})$  converge to  $\mu_\Gamma(\mathcal{L})$ .

**Theorem 3.5** (Main Theorem). *For every  $d \in \mathbb{N} \cup \{\infty\}$ , every  $d$ -convergent sequence  $d$ -converges to a  $d$ -limit object. On the other hand every  $d$ -limit object is the limit of a  $d$ -convergent sequence.*

## 4 Proof of the Main Theorem

For a measurable  $\Gamma : \mathbb{G}_d \rightarrow \mathbb{C}$ , similar to (2), we can define

$$t_{\mathcal{L}}(\Gamma) := \mathbf{E} \left[ \prod_{i \in [m]} \Gamma(\mathbf{b}_i) \right],$$

where  $\mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{G}_d$  are selected at random, conditioned on being consistent with  $\mathcal{L}$ .

We start with a simple and standard observation that allows us to work with the averages  $t_{\mathcal{L}}(\cdot)$  rather than the distributions  $\mu_{\mathcal{L}}$ .

**Observation 4.1.** *For every  $d \in \mathbb{N} \cup \{\infty\}$ , a sequence  $\{f_i : \mathbb{F}^{n_i} \rightarrow \{0, 1\}\}_{i \in \mathbb{N}}$   $d$ -converges to a  $d$ -limit object  $\Gamma$  if and only if for every affine system of linear forms  $\mathcal{L}$  of true complexity at most  $d$  we have  $\lim_{i \rightarrow \infty} t_{\mathcal{L}}(f_i) = t_{\mathcal{L}}(\Gamma)$ .*

*Proof.* The forward direction here is clear: The value of  $t_{\mathcal{L}}(f_i)$  is determined by the distribution  $\mu_{f_i}(\mathcal{L})$ .

For the other direction, write  $\mu := \mu_{f_i}(\mathcal{L})$  for some  $i$  and some affine  $\mathcal{L}$  of true complexity at most  $d$ . We can identify  $\mu$  with a function  $\mu : \{0, 1\}^{\mathcal{L}} \rightarrow [0, 1]$ . Then, applying the Fourier transform (for the group  $\mathbb{F}_2^{\mathcal{L}}$ ), we can write

$$\mu(x) = \sum_{S \subseteq \mathcal{L}} \hat{\mu}(S) \chi_S(x),$$

so that  $\mu$  is determined by the  $\hat{\mu}(S)$ . However, we have

$$\hat{\mu}(S) = \mathbf{E}_{y \in \{0,1\}^{\mathcal{L}}} [\mu(y) \chi_S(y)] = \mathbf{E}_y [\mu(y) (-1)^{\sum_{i \in S} y_i}] = \mathbf{E}_y \left[ \mu(y) \prod_{i \in S} (1 - 2y_i) \right].$$

Expanding this product, we see that  $\hat{\mu}(S)$  is a linear combination of terms of the form

$$\mathbf{E}_y \left[ \mu(y) \prod_{i \in S'} y_i \right] = t_{S'}(f_i),$$

where  $S' \subseteq S \subseteq \mathcal{L}$ . Thus the distribution  $\mu$  is determined by the values  $t_{\mathcal{L}'}(f_i)$ , where  $\mathcal{L}' \subseteq \mathcal{L}$  (and hence  $\mathcal{L}'$  has true complexity  $\leq d$ ).  $\square$

Theorem 3.5 follows from Lemma 4.2 and Lemma 4.4.

**Lemma 4.2.** *Let  $\{f_i : \mathbb{F}^{n_i} \rightarrow \{0, 1\}\}_{i \in \mathbb{N}}$  be a  $d$ -convergent sequence. There exists a  $d$ -limit object  $\Gamma$  such that  $\lim_{i \rightarrow \infty} t_{\mathcal{L}}(f_i) = t_{\mathcal{L}}(\Gamma)$  for every affine system of linear forms  $\mathcal{L}$  of true complexity at most  $d$ .*

*Proof.* Consider a decreasing sequence  $\{\varepsilon_i\}_{i \in \mathbb{N}}$  of positive reals tending to 0. Let the parameters  $\delta_i$ ,  $\eta_i$ , and  $r_i$  be chosen as required by Lemma 2.18 so that for every affine system of linear forms  $\mathcal{L} = \{L_1, \dots, L_m\}$  of true complexity at most  $d$ , if  $i$  is sufficiently large, then the following holds:

- (i)  $|t_{\mathcal{L}}(f_i) - t_{\mathcal{L}}(f_i^1)| \leq \varepsilon_i$  where  $f_i = f_i^1 + f_i^2 + f_i^3$  is decomposed according to Theorem 2.14 with the parameters  $\delta_i$ ,  $\eta_i$ , and  $r_i$ , and degree  $d_i$ , where  $d_i = d$  if  $d < \infty$ , and  $d_i = i$  if  $d = \infty$ .
- (ii) The assertion of Theorem 3.4 is true with  $\varepsilon = \varepsilon_i p^{-id_i C}$  when applied to the factor  $\mathcal{B}$  in the decomposition  $f_i = f_i^1 + f_i^2 + f_i^3$ . Here  $C$  is the complexity of  $\mathcal{B}$ .

Decompose each  $f_i$  as  $f_i = f_i^1 + f_i^2 + f_i^3$  according to Theorem 2.14 with the above mentioned parameters. We have  $f_i^1(x) = \tilde{\Gamma}_i(P_1^i(x), \dots, P_C^i(x))$  for some function  $\tilde{\Gamma}_i$  and polynomials  $P_1^i, \dots, P_C^i$ . Considering the degrees and the depths of the polynomials, the function  $\tilde{\Gamma}_i$  corresponds naturally to a  $d$ -limit object  $\Gamma_i$ : Indeed let  $\phi : [C] \rightarrow \mathbb{N} \times \mathbb{N} \times \mathbb{N}$  be any injective map that satisfies  $\phi(t) = (\deg(P_t^i), \text{depth}(P_t^i), t)$  for every  $t \in [C]$ . Define

$\pi : \mathbb{G}_d \rightarrow \mathbb{T}^C$  as  $\pi : \mathbf{b} \rightarrow (\mathbf{b}(\phi(1)), \dots, \mathbf{b}(\phi(C)))$ , and let  $\Gamma_i(\mathbf{b}) := \tilde{\Gamma}_i(\pi(\mathbf{b}))$  for every  $\mathbf{b} \in \mathbb{G}_d$ .

Let  $\mathcal{L} = \{L_1, \dots, L_m\}$  be an affine system of linear forms of true complexity at most  $d$ , and let  $i$  be sufficiently large. We will show that  $|t_{\mathcal{L}}(f_i^1) - t_{\mathcal{L}}(\Gamma_i)| \leq \varepsilon_i$ . Choose  $\mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{G}_d$  randomly, conditioned on being consistent with  $\mathcal{L}$ . Since consistency is defined coordinate-wise, it follows that  $(\pi(\mathbf{b}_1), \dots, \pi(\mathbf{b}_m))$  is distributed uniformly conditioned on being  $\mathcal{B}$ -consistent with  $\mathcal{L}$ , and hence that  $t_{\mathcal{L}}(\tilde{\Gamma}_i) = t_{\mathcal{L}}(\Gamma_i)$ .

Now we can write

$$t_{\mathcal{L}}(f_i^1) = \mathbf{E}_x \left[ \prod_{L_j \in \mathcal{L}} \tilde{\Gamma}_i(P_1^i(L_j(x)), \dots, P_C^i(L_j(x))) \right] = \mathbf{E} \left[ \prod_{L_j \in \mathcal{L}} \tilde{\Gamma}_i(y_j) \right],$$

where the  $y_j$  are distributed as  $(P_1(L_j(x)), \dots, P_C(L_j(x)))$ . The condition (ii) above shows that the distribution of  $(\mathbf{b}_1, \dots, \mathbf{b}_m)$ , where the  $\mathbf{b}_j \in \mathbb{T}^C$  are chosen uniformly and conditioned on being  $\mathcal{B}$ -consistent with  $\mathcal{L}$ , is  $\varepsilon$ -close in total variation distance to that of  $(P_1(L_j(x)), \dots, P_C(L_j(x)))_{j \in [m]}$  when  $x$  is chosen uniformly at random. This gives

$$\mathbf{E} \left[ \prod_{L_j \in \mathcal{L}} \tilde{\Gamma}_i(y_j) \right] \leq t_{\mathcal{L}}(\tilde{\Gamma}_i) + p^{id_i C} \varepsilon$$

since each  $P_t^i$ ,  $t \in [C]$ , has degree at most  $d_i$ , hence there are at most  $p^{id_i C}$  choices for the  $y_j$  when  $i \geq m$ . So we have the desired approximation.

So far we have established that for every system of affine linear forms  $\mathcal{L}$ , if  $i$  is sufficiently large, then

$$|t_{\mathcal{L}}(f_i) - t_{\mathcal{L}}(\Gamma_i)| \leq 2\varepsilon_i. \quad (4)$$

Next we construct the limit object  $\Gamma$ . For every  $t \in \mathbb{N}$  denote  $\mathbb{G}_d^t = \prod_{j=1}^d \prod_{k=0}^{\lfloor \frac{j-1}{p-1} \rfloor} \mathbb{U}_{k+1}^t$ . Note that  $\mathbb{G}_d^t$  corresponds to a partition of  $\mathbb{G}_d$ . For every measurable  $\Gamma : \mathbb{G}_d \rightarrow [0, 1]$  and  $t \in \mathbb{N}$ , define  $\mathcal{E}_t(\Gamma) = \mathbf{E}[\Gamma \mid \mathbb{G}_d^t]$ . Note that the set  $\{\Gamma_d^1 \rightarrow [0, 1]\}$  is a compact space, and thus one can find a subsequence of  $\{\Gamma_i\}_{i \in \mathbb{N}}$  such that  $\mathcal{E}_1(\Gamma_i)$  for  $i$  in this subsequence converges to a function  $\mu_1 : \mathbb{G}_d^1 \rightarrow [0, 1]$ . Now we restrict ourselves to this subsequence and consider  $\mathcal{E}_2$ . Again by compactness we can find a subsequence for which  $\mathcal{E}_2(\Gamma_i)$  converges to a function  $\mu_2 : \mathbb{G}_d^2 \rightarrow [0, 1]$ . Continuing in the same manner we define  $\mu_t : \mathbb{G}_d^t \rightarrow [0, 1]$  for every  $t$ . Note that since we restricted to a subsequence at every step, we have  $\mathbf{E}[\mu_t \mid \mathbb{G}_d^r] = \mu_r$  for every  $r < t$ . Furthermore, by picking the first element from the first subsequence, the second element from the second subsequence, and so on, we obtain a subsequence  $\Gamma'_1, \Gamma'_2, \dots$  of the original sequence that satisfies  $\lim \mathbf{E}[\Gamma'_i \mid \mathbb{G}_d^t] = \mu_t$  for every  $t \in \mathbb{N}$ .

The measure  $\mu_t$  is a  $\sigma$ -finite measure over the atoms  $\mathbb{G}_d^t$ , and thus by Carathéodory's extension theorem, there is a unique measure (also  $\sigma$ -finite)  $\mu$  on  $\mathbb{G}_d$  such that  $\mathbf{E}[\mu \mid \mathbb{G}_d^t] = \mu_t$  for every  $t$ . Now let  $\nu$  denote the Haar probability measure on  $\mathbb{G}_d$  (described after Definition 3.2), and note that for any  $t$  and any particular  $\Gamma_i$  we have  $\mathbf{E}[\Gamma_i \mid \mathbb{G}_d^t] \leq 1$ . Since  $\mu_t$  is a limit (over a subsequence) of these averages, we have  $\mu_t(A) \leq \nu(A)$  for every

$A \subseteq \mathbb{G}_d^t$ . It follows that  $\mu(A) \leq \nu(A)$  for any  $\mu$ -measurable  $A \subseteq \mathbb{G}_d$ . In particular,  $\mu$  is absolutely continuous with respect to  $\nu$ . Let  $\Gamma : \mathbb{G}_d \rightarrow [0, 1]$  be the Radon-Nikodym derivative of  $\mu$ .

Note that as  $\lim \mathbf{E}[\Gamma'_i \mid \mathbb{G}_d^t] = \mu_t$ , the sequence of  $\Gamma'_i$  converge to  $\Gamma$  in  $L_1$ , and consequently  $\lim t_{\mathcal{L}}(\Gamma'_i) = t_{\mathcal{L}}(\Gamma)$ . We showed in (Eq. (4)) that  $\lim t_{\mathcal{L}}(f_i) = \lim t_{\mathcal{L}}(\Gamma_i)$ , and since the former limit exists by assumption, it follows that  $t_{\mathcal{L}}(\Gamma) = \lim t_{\mathcal{L}}(f_i)$ .  $\square$

Before we can prove the second part of Theorem 3.5, we need a simple lemma which shows the existence of collections of high rank polynomials of arbitrary degree and depth sequence.

**Lemma 4.3.** *Let  $\mathbf{d} = (d_1, \dots, d_C) \in \mathbb{Z}_{>0}^C$  and  $\mathbf{k} = (k_1, \dots, k_C) \in \mathbb{Z}_{\geq 0}^C$  satisfy  $0 \leq k_i \leq \lfloor \frac{d_i-1}{p-1} \rfloor$  for every  $i$ , and let  $r$  be a positive integer. There exists a set of polynomials  $P_1, \dots, P_C$  of rank at least  $r$  such that  $P_i$  is of degree  $d_i$  and depth  $k_i$  for every  $i \in [C]$ .*

*Proof.* Let  $r'$  be an integer. For each  $i$ , let  $m_i$  satisfy  $d_i = m_i + (p-1)k_i$ . Allot variables  $x_1^i, \dots, x_{m_i r'}^i$  for exclusive use by  $P_i$ , and let  $P_i = (1/p^{k_i+1})(x_1^i \cdots x_{m_i}^i + \cdots + x_{m_i(r'-1)+1}^i \cdots x_{m_i r'}^i)$ . It is clear that  $P_1, \dots, P_C$  has the desired degree and depth sequence. For sufficiently large  $n$  we have enough variables to do this, and it is standard that for sufficiently large  $r'$ , these polynomials will have rank at least  $r$ .  $\square$

With this in hand, we can now complete the proof of Theorem 3.5.

**Lemma 4.4.** *Let  $d \in \mathbb{N} \cup \{\infty\}$ , and let  $\Gamma$  be a  $d$ -limit object. There exists a  $d$ -convergent sequence of functions  $\{f_i : \mathbb{F}^{n_i} \rightarrow \{0, 1\}\}_{i \in \mathbb{N}}$  whose limit is  $\Gamma$ .*

*Proof.* For every  $t \in \mathbb{N}$ , define the function  $\Gamma_t : \mathbb{G}_d \rightarrow [0, 1]$  to be the function obtained from  $\mathbf{E}[\Gamma \mid \mathbb{G}_d^t]$  (a map from  $\mathbb{G}_d^t$  to  $[0, 1]$ ) by extending it to a function on  $\mathbb{G}_d$ . The  $\Gamma_t$  converge to  $\Gamma$  in  $L_1$  norm, and each  $\Gamma_t$  depends on only a finite number of coordinates of  $\mathbb{G}_d$ . Let  $\mathbf{d}_t = (d_1^t, \dots, d_C^t)$  and  $\mathbf{k}_t = (k_1^t, \dots, k_C^t)$  be the degree and depth sequences corresponding to the coordinates of  $\mathbb{G}_d$  used by  $\Gamma_t$ .

For every  $r \in \mathbb{N}$ , we can apply Lemma 4.3 to get a collection of polynomials  $P_1, \dots, P_C$  of rank  $\geq r$  such that  $P_i$  has degree  $d_i^t$  and depth  $k_i^t$  for every  $i$ . Now define the function  $f_t^r : \mathbb{F}^{n_r} \rightarrow [0, 1]$  by letting  $f_t^r(x) = \Gamma(P_1(x), \dots, P_C(x))$ , where we treat  $(P_1(x), \dots, P_C(x))$  as an element of  $\mathbb{G}_d$ . It follows from Theorem 3.4 by the same argument used in the proof of Lemma 4.2 that we have  $t_{\mathcal{L}}(f_t^r) \rightarrow_r t_{\mathcal{L}}(\Gamma_t)$  for every affine  $\mathcal{L}$  of true complexity at most  $d$ . Taking a suitable diagonal subsequence of the  $f_t^r$ , we obtain a sequence of functions  $f_i : \mathbb{F}^{n_i} \rightarrow [0, 1]$  with  $t_{\mathcal{L}}(f_i) \rightarrow_i t_{\mathcal{L}}(\Gamma)$  for every affine  $\mathcal{L}$  of true complexity at most  $d$ .

To complete the proof, consider the random functions  $f'_i : \mathbb{F}^{n_i} \rightarrow \{0, 1\}$  where  $f'_i(x)$  takes value 1 with probability  $f_i(x)$ . It is well known that these  $d$ -converge to  $\Gamma$  with probability 1, and hence the existence of a  $d$ -convergent sequence converging to  $\Gamma$  is evinced.  $\square$

The use of Theorem 3.5 at this point is mainly one of convenience. Tools like Theorem 2.14 and Theorem 3.4 afford one the luxury to make simplifying assumptions about a function while incurring an arbitrarily small error. Working with  $d$ -limit objects here

is effectively allowing us to replace ‘arbitrarily small’ with 0. Similarly, we can substitute ‘arbitrarily high rank’ polynomials for ‘infinite rank’ ones. These kinds of assumptions can eliminate a significant amount of bookkeeping in proofs. Then, with a result regarding  $d$ -limit objects, the finite results are usually obtained for free (through Theorem 3.5). The state of affairs in the graph limit theory is mostly the same, although many conjecture that there should be some non-cosmetic benefit to the limit approach.

## 5 Necessary depths

The following inverse theorem characterizes functions with large  $U^{d+1}$  norm by showing that they must correlate highly with a degree  $\leq d$  (non-classical) polynomial.

**Theorem 5.1** (Theorem 1.11 of [TZ12]). *Suppose  $\delta > 0$  and  $d \geq 1$  is an integer. There exists an  $\varepsilon = \varepsilon_{5.1}(\delta, d)$  such that the following holds. For every function  $f : \mathbb{F}^n \rightarrow \mathbb{D}$  with  $\|f\|_{U^{d+1}} \geq \delta$ , there exists a polynomial  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  of degree  $\leq d$  that is  $\varepsilon$ -correlated with  $f$ , meaning*

$$\left| \mathbf{E}_{x \in \mathbb{F}^n} f(x) \mathbf{e}(-P(x)) \right| \geq \varepsilon.$$

Note that every polynomial that is not classical must have degree at least  $p$ . It is known that polynomials of degree  $d = p$  that are not classical are unnecessary in higher-order Fourier analysis. More precisely in Theorem 5.1, for the case  $d = p$ , one can assume that the polynomial  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  in the statement of the theorem is a *classical* polynomial of degree at most  $p$ . This can be carried further through Theorem 2.14 and then to the definition of the  $d$ -limit object. We will elaborate on this below, but first let us prove a generalization of this fact, which says that the polynomials of maximum possible depth are unnecessary in higher-order Fourier analysis.

**Lemma 5.2.** *Every  $(1 + k(p - 1), k)$ -polynomial  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  can be expressed as a function of a  $(1 + k(p - 1), k - 1)$ -polynomial, a  $(1 + (k - 1)(p - 1), k - 1)$ -polynomial, and a  $(k(p - 1), k - 1)$ -polynomial.*

*Proof.* By Lemma 2.5, we have  $P(x_1, \dots, x_n) = \frac{\sum c_i |x_i|}{p^{k+1}} + R(x_1, \dots, x_n) \pmod{1}$  for integers  $0 \leq c_i \leq p - 1$ , where  $R$  is a  $(1 + k(p - 1), k - 1)$ -polynomial. Let  $M := \sum c_i |x_i|$ , and let  $0 \leq a < p^k$  and  $b \in [p - 1]$  be the unique integers satisfying  $M \equiv a + bp^k \pmod{p^{k+1}}$ . The value of  $P$  is fixed by the three values  $a$ ,  $b$  and  $R$ . The value of  $a$  is determined by the value of the  $(1 + (k - 1)(p - 1), k - 1)$ -polynomial  $\frac{M}{p^k} \pmod{1}$ . Furthermore knowing  $a$ , the value of  $b$  is determined by the value of the  $\frac{M^p - M}{p^{k+1}} \pmod{1}$ . Indeed

$$bp^k \equiv (a + bp^k)^p - (a + bp^k) - (a^p - a) \pmod{p^{k+1}}. \quad (5)$$

It remains to show that  $\frac{M^p - M}{p^{k+1}} \pmod{1}$  is a  $(1 + k(p - 1), k - 1)$ -polynomial. Since degree and depth are invariant under affine transformations, it suffices to show that  $Q := \frac{|x_1|^p - |x_1|}{p^{k+1}} \pmod{1}$  is a  $(k(p - 1), k - 1)$ -polynomial. By Fermat’s little theorem  $p^k Q = 0$ , and thus  $Q$

is of depth  $k - 1$ . Furthermore, the identity  $|x_1|(|x_1| - 1) \dots (|x_1| - p + 1) = 0$  allows us to replace  $|x_1|^p$  with a polynomial of degree  $p - 1$ . This shows that  $Q$  is of degree at most  $(p - 1) + (p - 1)(k - 1) = k(p - 1)$ .  $\square$

It follows from Lemma 5.2 that in Theorem 2.14,  $(1 + k(p - 1), k)$ -polynomials can be avoided in the polynomials defining the factor  $\mathcal{B}$ . Consequently, every  $d$ -convergent sequence converges to a  $d$ -limit object  $\phi : \mathbb{G}_d \rightarrow [0, 1]$  such that  $\phi$  does not depend on the coordinates that correspond to  $(1 + k(p - 1), k)$ -polynomials. Next we will show that there are no other values of  $(d, k)$  that behave similarly. That is, for which every  $(d, k)$ -polynomial can be expressed as a function of a constant number of polynomials of either degree  $d$  and depth  $< k$ , or degree  $< d$ . We need the following theorem of [TZ12].

**Theorem 5.3** (See Theorem 4.1 of [TZ12]). *Let  $d > p$  be an integer, and  $\varepsilon > 0$ . There exists a  $\rho = \rho_{5.3}(\varepsilon, d)$  such that the following holds for sufficiently large  $n$ . If  $P : \mathbb{F}^n \rightarrow \mathbb{T}$  is a polynomial of degree  $d$  with  $\|\mathbf{e}(P)\|_{U^d} \geq \varepsilon$ , then  $pP : \mathbb{F}^n \rightarrow \mathbb{T}$  is a polynomial of degree  $\leq d - p + 1$  that satisfies*

$$\|\mathbf{e}(pP)\|_{U^{d-p+1}} \geq \rho.$$

The following lemma implies that unless  $d$  and  $k$  are as in Lemma 5.2, for every constant  $C$ , there exists a  $(d, k)$ -polynomial that cannot be expressed as a function of  $C$  polynomials, each of either degree  $d$  and depth  $< k$ , or of degree  $< d$ . The polynomial that we use in the proof of the lemma is reminiscent of the so called generalized inner product polynomial that was used by Babai, Nisan, and Szegedy [BNS92] for similar purposes (see also [VW08]).

**Lemma 5.4.** *Let  $2 \leq m \leq p - 1$  be an integer, and  $\varepsilon > 0$ . Then for every  $k \geq 0$ , defining  $d = m + k(p - 1)$ , there exists a degree  $(d, k)$ -polynomial  $Q$  such that*

$$|\langle \mathbf{e}(Q), \mathbf{e}(R_1 + R_2) \rangle| < \varepsilon, \tag{6}$$

for any polynomial  $R_1$  of degree at most  $d$  and depth less than  $k$ , and any polynomial  $R_2$  of degree at most  $d - 1$ .

*Proof.* Let

$$P = \sum_{i=0}^{\lfloor n/m \rfloor - 1} |x_{im+1}| \dots |x_{im+m}|.$$

Set  $\varepsilon_k = \varepsilon$ , and for every  $0 \leq i \leq k$ , let  $\varepsilon_i \in (0, \varepsilon)$  be constants satisfying

$$\varepsilon_i < \varepsilon_{5.1}(\rho_{5.3}(\varepsilon_{i+1}, d), d).$$

We show by induction on  $i$  that if  $n$  is sufficiently large, then the  $(m + i(p - 1), i)$ -polynomial  $Q = \frac{P}{p^{i+1}} \bmod 1$  satisfies the desired property with parameter  $\varepsilon_i$  in (6) instead of  $\varepsilon$ .

We first look at the classical case  $i = 0$ . Notice that in this case by taking  $n$  to be sufficiently large, we can guarantee that  $\|\mathbf{e}(Q)\|_{U^d}$  is sufficiently small, and this implies

that the correlation of  $Q$  with any polynomial of degree lower than  $m + i(p - 1) = m$  is smaller than  $\varepsilon_0$ .

Now let us consider the case  $i > 0$ . Assume for the sake of contradiction that  $|\langle \mathbf{e}(\frac{P}{p^{i+1}}), \mathbf{e}(R_1 + R_2) \rangle| \geq \varepsilon_i$  for a polynomial  $R_1$  of degree at most  $d_i = m + i(p - 1)$  and depth  $< k$ , and a polynomial  $R_2$  of degree  $\leq d_i - 1$ . This in particular implies  $\|\mathbf{e}(\frac{P}{p^{i+1}} - R_1 - R_2)\|_{U^{d_i}} \geq \varepsilon_i$ . Note that  $\frac{P}{p^{i+1}} - R_1 - R_2 \pmod{1}$  has degree  $d_i > p$ , and thus we can apply Theorem 5.3 to conclude that

$$\|\mathbf{e}(p(P/p^{i+1} - R_1 - R_2))\|_{U^{d_i-p+1}} \geq \rho(\|\mathbf{e}(P/p^{i+1} - R_1 - R_2)\|_{U^{d_i}}) \geq \rho_{5.3}(\varepsilon_i, d_i).$$

Therefore by Theorem 5.1 there exists a polynomial  $R'$  of degree at most  $d_i - p$  such that

$$|\langle \mathbf{e}(p(P/p^{i+1} - R_1 - R_2)), \mathbf{e}(R') \rangle| > \varepsilon_{5.1}(\rho_{5.3}(\varepsilon_i, d_i), d_i - p) \geq \varepsilon_{5.1}(\rho_{5.3}(\varepsilon_i, d_i), d_i) \geq \varepsilon_{i-1}.$$

It follows that  $|\langle \mathbf{e}(P/p^i), \mathbf{e}(pR_1 + pR_2 + R') \rangle| > \varepsilon_{i-1}$ , which contradicts our induction hypothesis.  $\square$

**Concluding remarks.** We conclude with an open question regarding the uniqueness of the limit object. Suppose that two limit objects  $\Gamma_1, \Gamma_2$  satisfy  $t_{\mathcal{L}}(\Gamma_1) = t_{\mathcal{L}}(\Gamma_2)$  for every affine system of linear forms  $\mathcal{L}$ . Then how do  $\Gamma_1$  and  $\Gamma_2$  relate to each other?

## References

- [BFH<sup>+</sup>13] Arnab Bhattacharyya, Eldar Fische, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every locally characterized affine-invariant property is testable. In *Proc. 45th Annual ACM Symposium on the Theory of Computing*, 2013. Full version at [arXiv:1212.3849](https://arxiv.org/abs/1212.3849). 2, 9
- [BFL13] Arnab Bhattacharyya, Eldar Fischer, and Shachar Lovett. Testing low complexity affine-invariant properties. In *Proc. 23rd ACM-SIAM Symposium on Discrete Algorithms*, 2013. 6
- [BNS92] László Babai, Noam Nisan, and Mária Szegedy. Multiparty protocols, pseudo-random generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. Twenty-first Symposium on the Theory of Computing (Seattle, WA, 1989). 14
- [GW10] William T. Gowers and Joseph A. Wolf. The true complexity of a system of linear equations. *Proc. Lond. Math. Soc. (3)*, 100(1):155–176, 2010. 7
- [LS06] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006. 2
- [TZ12] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *Ann. Comb.*, 16(1):121–188, 2012. 4, 5, 13, 14
- [VW08] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory Comput.*, 4:137–168, 2008. 14