Limits of Boolean Functions over Finite Fields

James Hirst

School of Computer Science
McGill University
Montreal, Canada

August 2014

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Masters in Computer Science

2014/08/11

Abstract

In this thesis, we study sequences of functions of the form $\mathbb{F}_p^n \to \{0, 1\}$ for varying n, and define a notion of convergence based on the induced distributions from restricting the functions to a random affine subspace. One of the key tools we use is the recently developed theory of 'higher order Fourier analysis', where the characters of standard Fourier analysis are replaced with exponentials of higher degree polynomials. This is not a trivial extension by any means, but when the polynomials are chosen with some care, the higher order decomposition can be taken to have properties analogous to those of the classical Fourier transform.

The result of applying higher order Fourier analysis in this setting is the necessity to determine the distribution of a collection of polynomials when they are composed with some additional linear structures. Here, we make use of a recently proven equidistribution theorem, relying on a near-orthogonality result showing that the higher order characters can be made orthogonal up to an arbitrarily small error term.

With these tools, we prove that the limit of every convergent sequence of functions can be represented by a limit object which takes the form of a certain measurable function on a group we construct. We also show that every such limit object arises as the limit of some sequence of functions. These results are in the spirit of analogous results which have been developed for limits of graph sequences. A more general, albeit substantially more sophisticated, limit object was recently constructed by Szegedy in [Sze10].

Abrégé

Cette thèse étudie les séquences de fonctions de la forme $\mathbb{F}_p^n \to \{0,1\}$ pour n variant, et définit une notion de convergence sur la base des distributions induits par la restriction des fonctions à un sous-espace affine statistique. Un des outils essentiels est la théorie de «l'ordre supérieur analyse de Fourier», où les caractères de l'analyse de Fourier sont remplacés par des exponentielles de polynômes de degré plus élevé. Ce n'est pas une extension triviale, mais lorsque les polynômes sont choisis avec soin, la décomposition d'ordre supérieur peut avoir des propriétés analogues à celles de la transformation classique de Fourier.

Le résultat de cet application de l'analyse de Fourier est la nécessité de déterminer la distribution d'un ensemble de polynômes quand ils sont composés avec les structures linéaires supplémentaires. Ici, un théorème d'équidistribution récemment prouvé est utilisé, en s'appuyant sur un résultat quasi-orthogonalité montrant que les caractères d'ordre supérieur peuvent être orthogonale à un terme d'erreur arbitrairement petit.

Avec ces outils, nous montrons que la limite des séquences de fonctions convergentes peut être représentée par un objet limite qui prend la forme d'une fonction mesurable sur un groupe que nous construisons. Nous montrons également que chaque objet de limite est le limite d'une séquence de fonctions. Ces résultats sont dans l'esprit des résultats analogues qui ont été développés pour les limites de séquences de graphes. Un objet de limite plus générale, quoique sensiblement plus sophistiqué, a été récemment construit par Szegedy dans [Sze10].

Acknowledgments

I would first like to thank both McGill University and NSERC for the funding they provided during the making of this thesis, allowing me to work unburdened by financial anxiety.

My thanks also go out to my supervisor Hamed Hatami for initiating and assisting the research that went into this thesis, as well as for always being available to suggest improvements and discuss details of proofs. I would also like to thank him for reading and giving constructive feedback on multiple iterations of this document.

Finally, I would like to thank Pooya Hatami, in his capacity as a coauthor, for his contributions to our main results.

Contents

Chapter 1

Introduction

A recently emerging theme in the theory of discrete structures is the interplay between local and global information. One reason for this advent is that modern applications necessitate the study of prohibitively large objects, in the sense that even traversing the entire structure is computationally intractable. The main technique we have for working with such objects is to randomly sample from them and work locally. Thus, there is some motivation to see what kind of global properties are discernible when looking only at a set of 'local statistics' which can be efficiently sampled. This has spurred activity in the field of property testing in recent years and, not surprisingly, this is a major area of application for much of the material that will be presented in this thesis.

Now, let us try to make this idea at least slightly more precise. Although 'local statistics' is a particularly vague expression, there is a fairly standard program at work here: Given a discrete structure and a corresponding notion of substructure, every rule for randomly sampling a (small) substructure induces a probability distribution over the set of all such substructures. Then one is interested in what kind of properties of the original structure can be inferred from this distribution. This is still rather abstract, so to dispense with generality, and also to further motivate what will follow, let us take the time to examine an example of this program applied in graph theory which has garnered some significant attention of late.

1.1 Subgraph sampling and graph limits

Given a graph G on n vertices and a positive integer $k < n$, our sampling rule is to independently choose k vertices from $V(G)$ and look at the subgraph of G induced by these vertices. For every graph H on k vertices, we denote the probability that a random map $\varphi : V(H) \to V(G)$ preserves both adjacency and non-adjacency by $t_{\text{ind}}(H, G)$. Here, if we were to condition on the map φ being injective, this would just be the probability that the subgraph induced by a k-subset of vertices is isomorphic to H. In some literature, this is even taken to be the definition of $t_{\text{ind}}(H, G)$. Note that, however, when $k = o(n)$, the difference between sampling independently or injectively is a quantity going to 0 as n grows.

The most optimistic question to ask at this point is what values $t_{\text{ind}}(H, G)$ are required to reconstruct G. Regrettably, this is not actually possible, and there is an easy counterexample: For any integer $m \geqslant 2$, define the m *blowup* of G, denoted by $G(m)$, to be the graph obtained from G by replacing each vertex with m copies, and connecting 2 vertices if and only if their originals are connected in G. Then note that G and $G(m)$ are indistinguishable by (independent) sampling, so we have that $t_{\text{ind}}(H, G) = t_{\text{ind}}(H, G(m))$ for every graph H. However, this is all that can go wrong. If G and G' are any two graphs such that $t_{\text{ind}}(H, G) = t_{\text{ind}}(H, G')$ for every graph H, then there is a third graph G'' such that both G and G' are (possibly different) blowups of G''. If G and G' are known to have the same number of vertices, then this implies that $G \cong G'$. For a graph G with n vertices, it is established that the values $t_{\text{ind}}(H, G)$ for all graphs H with $\leqslant n$ vertices determine G. It is conjectured that the values $t_{\text{ind}}(H, G)$ for H strictly smaller than G (in either vertex or edge sense) will suffice, but this remains a large open problem in the area.

Thinking of a graph as a list of subgraph densities gives an interesting perspective on things. This projects into a very well studied kind of space: real vector spaces. To actually work in this space, however, there are some hurdles to overcome. One immediate issue is that the set of graphs under this projection is sorely lacking in limit points. Consider a sequence of random graphs $G(n, 1/2)$. With high probability, every $t_{\text{ind}}(H, G)$ with $|V(H)| = k$ converges to $2^{-\binom{k}{2}}$, but there is no finite graph exhibiting these subgraph densities. So to complete this space, we will need to work with something more than graphs. A sequence of graphs $\{G_n\}_{n \in \mathbb{N}}$ is called *convergent* ([LS06]) if $t_{\text{ind}}(H, G_n)$ converges for every graph H. We would like to find a 'limit object', extending graphs, that correctly completes this notion of convergence.

A *graphon* is a symmetric, measurable function $W : [0,1]^2 \to [0,1]$. It is instructive to think of a graphon as a weighted, infinite adjacency matrix. Note that if G is a graph with adjacency matrix A_G then

$$t_{\text{ind}}(H,G) = \mathop{\mathbf{E}}_{v_1,\ldots,v_k \in V(G)} \left[\prod_{(i,j)\in E(H)} A_G(v_i,v_j) \prod_{(i,j)\notin E(H)} (1 - A_G(v_i,v_j)) \right]$$

for every graph H on k vertices, so it is not unnatural to define the density of H in a graphon W as

$$t_{\text{ind}}(H,W) = \mathop{\mathbf{E}}_{x_1,\ldots,x_k \in [0,1]} \left[\prod_{(i,j)\in E(H)} W(x_i,x_j) \prod_{(i,j)\notin E(H)} (1 - W(x_i,x_j)) \right].$$

For every graph G on n vertices, we can define a graphon W_G as follows. Partition $[0,1]$ into n intervals $I_1,\ldots,I_n$ of the same measure, and set W to 1 in every rectangle $I_i \times I_j$ where $i \sim j$ in G. Then the sequence of blowups $\{G(m)\}_{m\in\mathbb{N}}$ converges to W_G, in the sense that $t_{\text{ind}}(H,G_n) \to t_{\text{ind}}(H,W_G)$ for every graph H. In [LS06], the authors show that every convergent sequence of graphs converges to a graphon, so that graphons complete the space of graphs as desired. Moreover, they show that every graphon is the limit of some convergent sequence, indicating that graphons are precisely the right limit object to capture this notion of convergence.

1.2 Limits of boolean functions

In this thesis, we are interested in subsets of the vector space $\mathbb{F}_p^n$, where p is some fixed prime and n is a large positive integer. Equivalently, we will want to think of these as $\{0,1\}$-valued functions over $\mathbb{F}_p^n$, and so we will often conflate a set with its indicator function. Given a subset $A \subseteq \mathbb{F}_p^n$, the local information we would like to work with is the distribution of small linear structures, e.g. arithmetic progressions, contained within the set. Formally, we define a *linear form* in k variables as a vector $L = (\lambda_1,\ldots,\lambda_k) \in \mathbb{F}_p^k$, where we consider L to be a linear function from $(\mathbb{F}_p^n)^k$ to $\mathbb{F}_p^n$ by writing $L(X) = \sum_{i=1}^k \lambda_i x_i$ for every $X = (x_1,\ldots,x_k) \in (\mathbb{F}_p^n)^k$. We then define a *system of linear forms* to be a subset $\mathcal{L} \subseteq \mathbb{F}^k$ of linear forms. We are concerned with the distribution of $(A(L_1(X)), \dots, A(L_m(X)))$ where $X \in (\mathbb{F}_p^n)^k$ is taken uniformly at random. Let us denote this distribution by $\mu_A(\mathcal{L})$.

One quantity that is immediately accessible from this distribution is the *density* of $\mathcal{L}$ in A, denoted $t_{\mathcal{L}}(A)$, which is just the probability that all the $L_i(X)$ lie inside A. As an example, consider the system of linear forms $\{x, x+y, x+2y, \dots, x+(d-1)y\}$, where we have specified $\mathcal{L}$ by the action of each $L \in \mathcal{L}$ on a pair $(x, y) \in (\mathbb{F}_p^n)^2$. Here it is easy to see that $t_{\mathcal{L}}(A)$ is just the density of d term arithmetic progressions in A, so we can think of $\mathcal{L}$ as 'counting' this structure. Somewhat surprisingly, there is an observation to be made (see Observation 5.2) showing that the values $t_{\mathcal{L}}(A)$ for every $\mathcal{L}' \subseteq \mathcal{L}$ actually determine $\mu_A(\mathcal{L})$, so when it is convenient we may work with densities rather than the full distribution.

In fact, we can even further narrow our focus. A linear form $L = (\lambda_1, \dots, \lambda_k) \in \mathbb{F}_p^k$ is called an *affine linear form* if $\lambda_1 = 1$. Analogously, we define a *system of affine linear forms* as any system of linear forms consisting solely of affine linear forms. Affine systems of linear forms are, among other things, translation invariant, a property that makes them much more natural to work with than general systems. Notice that the system of linear forms counting arithmetic progressions is always affine, and indeed, all of the explicit systems we encounter will be affine.

For a function $f : \mathbb{F}_p^n \to \{0, 1\}$ and an affine system $\mathcal{L} = \{L_1, \dots, L_m\} \subseteq \mathbb{F}_p^k$, there is another interesting way to think of the distribution $\mu_f(\mathcal{L})$, which is that it is obtained by restricting f to a random affine subspace. Consider sampling a random affine transformation $T : \mathbb{F}_p^k \to \mathbb{F}_p^n$. Then the random variable $Tf : \mathbb{F}_p^k \to \{0, 1\}$ defined by $Tf : x \mapsto f(Tx)$ induces a probability distribution on the set of function $\{\mathbb{F}_p^k \to \{0, 1\}\}$. Let us denote this distribution by μ_f. Note that further restricting μ_f to $\mathcal{L}$, we obtain the distribution $\mu_f(\mathcal{L})$, since the distribution of $(TL_1, \dots, TL_m)$ over a random $T : \mathbb{F}_p^k \to \mathbb{F}_p^n$ is just that of $(L_1(X), \dots, L_m(X))$ over a random $X \in (\mathbb{F}_p^n)^k$.

Now we can define the notion of convergence whose investigation will become the main topic of this thesis. A sequence of functions $\{f_n : \mathbb{F}_p^n \to \{0, 1\}\}_{n \in \mathbb{N}}$ is called *convergent* if the distributions μ_{f_n} converge for every k, or equivalently, if the densities $t_{\mathcal{L}}(f_n)$ converge for every affine system of linear forms $\mathcal{L}$.

1.3 Fourier analysis and linear densities

For an arbitrary affine system $\mathcal{L}$ and a function $f : \mathbb{F}_p^n \to \{0,1\}$, it may be difficult to control the behavior of $t_{\mathcal{L}}(f)$ in terms of f. Developing the theory to do so will constitute a large portion of the work in this thesis, but there are certain types of systems that are amenable to a simple analysis. Unfortunately, arithmetic progressions will prove difficult to handle in full generality, but we will see that 3 term progressions (3-APs) behave quite nicely. Better yet, we have 3-parallelepipeds, which are given by the affine system $\mathcal{L} = \{x, x+y, x+z, x+y+z\}$. Larger parallelepipeds will play a major role in the general theory. The density of 3-parallelepipeds, i.e. $t_{\mathcal{L}}(f)$, is a key quantity, and it will be useful to distinguish it from other densities. We denote this density by $\|f\|_{U^2}^4$ and call $\|f\|_{U^2}$ the U^2 *norm* of f. An easy exercise shows that the density of 3-APs in a set is always bounded by its U^2 norm, so controlling the U^2 norm suffices to force the behavior of 3-APs as well.

There is a beautiful and powerful connection to be made between the U^2 norm and Fourier analysis. To do this, let us recall some of the basic theory. The space of functions $\{f : \mathbb{F}_p^n \to \mathbb{C}\}$ is a Hilbert space over $\mathbb{C}$ with the inner product of two functions $f, g : \mathbb{F}_p^n \to \mathbb{C}$ given by

$$\langle f, g \rangle = \mathop{\mathbf{E}}_x f(x)\overline{g(x)},$$

where $\overline{g(x)}$ denotes the complex conjugate of $g(x)$.

For a function $f : \mathbb{F}_p^n \to \mathbb{C}$, the Fourier transform expresses f in the basis of group characters (homomorphisms from $\mathbb{F}_p^n \to \mathbb{C}\backslash\{0\}$). On $\mathbb{F}_p^n$, these are precisely the functions $\{\chi_a\}_{a \in \mathbb{F}_p^n}$ where χ_a is defined by

$$\chi_a(x) = e^{\frac{2\pi i}{p} \sum_{i=1}^n a_i x_i}$$

for every $x \in \mathbb{F}_p^n$. The character χ_0 (which is just the constant 1) is called the principal character (of $\mathbb{F}_p^n$), and will usually exhibit different behavior from the non-principal characters.

For any character χ_a, we have that

$$\mathop{\mathbf{E}}_x \chi_a(x) = \mathop{\mathbf{E}}_x \chi_a(x+y) = \chi_a(y) \mathop{\mathbf{E}}_x \chi_a(x)$$

for any $y \in \mathbb{F}^n$, where the first equality makes use of the group structure of $\mathbb{F}^n$. So if

χ_a is non-principal, taking $\chi_a(y) \neq 1$ implies that $\mathbf{E}_x \chi_a(x) = 0$. Using this, for any two characters χ_a, χ_b, we have

$$\langle \chi_a, \chi_b \rangle = \mathop{\mathbf{E}}_x \chi_a(x)\overline{\chi_b(x)} = \mathop{\mathbf{E}}_x \chi_{a-b}(x)$$

which is 1 if and only if $a = b$. So distinct characters are orthogonal, and a dimension counting argument shows that they do in fact consist of a basis for the space of functions $\{\mathbb{F}_p^n \to \mathbb{C}\}$. This lets us write a function $f : \mathbb{F}_p^n \to \mathbb{C}$ as

$$f = \sum_{a \in \mathbb{F}_p^n} \hat{f}(a)\chi_a$$

for a unique choice of coefficients $\hat{f}(a) \in \mathbb{C}$. The values $\hat{f}(a)$ are called the *Fourier coefficients* of f, and are computed by $\mathbf{E}_x f(x)\chi_a(x)$. The Fourier coefficients of f satisfy the Parseval identity

$$\|f\|_2^2 = \langle f, f \rangle = \sum_{a \mathbb{F}_p^n} |\hat{f}(a)|^2,$$

which follows easily by replacing f with its Fourier transform: All the non-diagonal pairs will vanish because the characters are orthogonal.

The largest non-principal Fourier coefficient of a function, denoted by $\|\hat{f}\|_\infty$, is closely related to its U^2 norm. First, we have the identity

$$\|f\|_{U^2}^4 = \sum_{a \in \mathbb{F}_p^n} |\hat{f}(a)|^4, \tag{1.1}$$

whose proof is similar to that of Parseval's with only a bit more effort. Now note that for any $0 \neq a \in \mathbb{F}^n$, we have

$$\hat{f}(a) = \mathbf{E}\, f(x)\chi_a(x) \leqslant \|f\chi_a\|_{U^2} = \|f\|_{U^2},$$

where the non-trivial steps here are both easily implied by Eq. (1.1). Since $a \neq 0$ was arbitrary, this gives $\|f\|_{U^2} \geqslant \|\hat{f}\|_\infty$. This is an example of a *direct theorem*, asserting that functions with large maximum Fourier coefficient also have large U^2 norm. There is also an *inverse* theorem, which will complete the characterization of such functions. Using

Eq. (1.1) again, we have

$$\|f\|_{U^2}^4 = \sum_{a\in\mathbb{F}^n} |\hat{f}(a)|^4 \leqslant \max|\hat{f}(a)|^2 \sum_{a\in F_p^n} |\hat{f}(a)|^2,$$

so if f is bounded by 1 in L_2 norm (e.g. if f is {0,1}-valued), then Parseval implies that $\|f\|_{U^2} \leqslant (\|\hat{f}\|_\infty)^{1/2}$.

Now let us see, informally, how this could be used to determine the behavior of the densities of simple affine systems. The Fourier transform of f, along with the inverse theorem for the U^2 norm give us a way to decompose an arbitrary function $f : \mathbb{F}_p^n \to [0,1]$ into a 'structured' part and a 'quasirandom' part. Here, the structured part will hopefully depend only on a bounded number of characters, while the quasirandom part will be small in U^2 norm. Consider the following example of a simple decomposition theorem: For any $\varepsilon > 0$, take f_1 to be the function obtained from f by restricting to its Fourier coefficients larger than ε^2. Then trivially, $f_2 = f - f_1$ has $\|\hat{f}_2\|_\infty \leqslant \varepsilon^2$ and so the inverse theorem implies that $\|f\|_{U^2} \leqslant \varepsilon$. There are some significant shortcomings to this simple decomposition, but it will suffice for the purpose of this discussion.

For any affine system $\mathcal{L}$ with densities bounded by the U^2 norm (e.g. 3-APs), the study of $t_{\mathcal{L}}(f)$ can be reduced to that of $t_{\mathcal{L}}(f_1)$, since f_1 approximates f in the U^2 norm. Analyzing $t_{\mathcal{L}}(f_1)$ is not particularly difficult, since it is determined by the distribution of some characters composed with an affine system of linear forms. Using the orthogonality properties of characters, this distribution can be shown to be uniform over its support.

1.4 Main results and organization

The primary contribution of this thesis is the construction of a limit object for convergent sequences of functions $\{f_n : \mathbb{F}_p^n \to \{0,1\}\}$. The limit objects are measurable functions from a particular infinite group into $\mathbb{C}$ which encapsulates the idea of being a function of a (possibly infinite) collection of polynomials. In the special case where these are all linear polynomials, this group is just $\mathbb{F}_p^{\mathbb{N}}$, and the limit object can be thought of as specifying an infinite Fourier expansion for a function. Following along lines of work from the theory of graph limits, we prove that every convergent sequence of functions converges to some limit object. We further show that every limit object can be obtained as the limit of some convergent sequence.

The rest of the thesis will be organized as follows. For an abridged version of the results, we refer to the reader to our paper [HHH14], which contains only the main results from Chapter 3 and Chapter 4, along with the better part of Chapter 5.

Chapter 2 will be devoted to laying some necessary groundwork. First, we will define an extension of the U^2 norm (the U^d norm) and prove some of its basic properties. Next, before we can state the generalized direct/inverse theorem for the U^d norm, we will need introduce the notion of non-classical polynomials and develop some of their theory. Finally, we will give the previously mentioned theorem. The general inverse theorem is a highly non-trivial result, taking considerably more effort than the inverse theorem for the U^2 norm.

In Chapter 3, we will show how the inverse theorem Theorem 2.14 can be used to develop a theory of 'higher order Fourier analysis'. The key is to develop a decomposition theorem for the U^d norm, similar to the one we saw for the U^2 norm, which has properties similar to those of a Fourier decomposition. This will allow us to decompose an arbitrary function into a 'structured' part which is a function of some bounded degree polynomials, as well as a 'quasirandom' part which is small in the U^d norm.

Chapter 4 will introduce a notion of complexity for systems of linear forms such that bounded complexity systems have densities which can be controlled by some U^d norm. Then, after decomposing f according to Theorem 3.18, we will be able to discard all but the structured part of f. The chapter culminates with an equidistribution theorem which allows us to study the behavior of polynomials when composed with low complexity systems of linear forms using a near-orthogonality condition on the polynomials.

Finally, in Chapter 5 we will show how Theorem 3.18 and Theorem 4.10 can be used to find a limit object for convergent sequences of boolean functions. We also prove a semi-relevant result regarding non-classical polynomials which may be useful in other applications.

Chapter 2

Basic Background

Notation

The following is a list of the basic notation and conventions that will be used throughout.

For $d \in \mathbb{N} \cup \{\infty\}$, denote $[d] = \{1, \ldots, d\}$ if $d < \infty$, and $[d] = \mathbb{N}$ otherwise. We shorthand $\mathbb{F} = \mathbb{F}_p$ for a prime finite field p, and p is always implicitly assumed to be the characteristic of $\mathbb{F}$. We denote by $\mathbb{T}$ the circle group $\mathbb{R}/\mathbb{Z}$, and by $\mathbb{D}$ the unit disc in the complex plane. For a function $f : \mathbb{F}^n \to \mathbb{C}$, we use the normalized L_p norms $\|f\|_1 = \mathbf{E}|f(x)|$, $\|f\|_2^2 = \mathbf{E}|f(x)|^2$, and $\|f\|_\infty = \max|f(x)|$, as well as the inner product $\langle f, g\rangle = \mathbf{E}\, f(x)\overline{g(x)}$. Note that $\|f\|_1 \leqslant \|f\|_2 \leqslant \|f\|_\infty$. The expression $o_m(1)$, or just $o(1)$ when the limiting value m is implicit, denotes a quantity which approaches 0 as m grows. The limit notation $\to_r$ is also used occasionally to indicate convergence as the variable r is taken to infinity. We write $x \pm \varepsilon$ as shorthand for any quantity in the interval $[x - \varepsilon, x + \varepsilon]$.

The operator $\mathcal{C}$ will sometimes be used for complex conjugation. For a set $A \subseteq \mathbb{F}^n$, we will often also use A to denote its indicator function. On the other hand, if A is an event in some probability space, we will denote its indicator function by $\mathbf{1}(A)$. We will try to stick to the convention of using lower case letters like f or g to denote functions $\mathbb{F}^n \to \mathbb{C}$, while upper case letters like P or Q will be reserved for functions $\mathbb{F}^n \to \mathbb{T}$. For variables, lower case (possibly in boldface) will denote a vector, while upper case will be used for variables coming from a matrix group like $(\mathbb{F}^n)^k$, which we will usually shorthand as $\mathbb{F}^{nk}$.

2.1 Uniformity norms

In the paper [Gow01], Gowers introduced a family of norms $\{\| \cdot \|_{U^d}\}_{d\in\mathbb{N}}$, defined on functions $f : \mathbb{Z}_N \to \mathbb{C}$, and used them to give a new analytic proof of Szemeredi's famous theorem on arithmetic progressions. These norms would become known as Gowers uniformity norms, and have since been used extensively to tackle problems in additive combinatorics. Before we delve any deeper into this discussion, let us give the basic definition.

Definition 2.1 (Gowers uniformity norm)**.** *Let G be a finite abelian group, and $d \geqslant 1$ an integer. Given a function $f : G \to \mathbb{C}$, we define the* Gowers uniformity norm of order d *of f by*

$$\|f\|_{U^d} = \left| \mathop{\mathbf{E}}_{x,y_1,\ldots,y_d\in\mathbb{F}^n} \left[\prod_{S\subseteq[d]} \mathcal{C}^{|S|} f(x + \sum_{i\in S} y_i) \right] \right|^{1/2^d} . \tag{2.1}$$

From here on we will often refer to the Gowers uniformity norm of order d as simply the U^d norm. Also, while we have defined the norm on functions $f : G \to \mathbb{C}$, where G is an arbitrary finite abelian group, we will not need the full generality of this definition. We will only be interested in the case where $G = \mathbb{F}^n$, so in further instances we will make this distinction. A final note to make is that when working with U^d norms, it will often be instructive to consider the case when $f : \mathbb{F}^n \to \{0, 1\}$ is boolean-valued, so that f corresponds to a subset of $\mathbb{F}^n$. Here, when $d = 2$ we see that Definition 2.1 agrees with our previous definition of the U^2 norm for sets. Analogously, the U^d norm of a set has a clear meaning which is the U^d norm of its indicator function.

The U^d norm is, roughly, a measure of additive structure. For sets, we can see this from the definition: If $A \subseteq \mathbb{F}^n$ is a set, then $\|A\|_{U^d}^{2^d}$ is just the probability that a random d-parallelepiped lies in the set A. This may seem restrictive, but we will see (much) later that parallelepipeds are in fact a very general type of structure, in the sense that the study of arbitrary additive structures, e.g. arithmetic progressions, can be reduced to that of a suitably high dimensional parallelepiped. This should hopefully give some insight into why the U^d norms have become so prevalent in the field.

Now, there are many useful properties of the Gowers norms that are not immediate from Definition 2.1. Indeed, it is not even obvious whether $\|\cdot\|_{U^d}$ defines a norm at all, and in actuality, it is not completely true; there is a small caveat. In the case when $d = 1$, for any function $f : \mathbb{F}^n \to \mathbb{C}$ we have

$$\|f\|_{U^1} = \left(\mathop{\mathbf{E}}_{x,y} f(x)\overline{f(x+y)} \right)^{1/2} = |\mathop{\mathbf{E}}_{x} f(x)|.$$

Here, we have homogeneity and the triangle inequality, but as there are (many) non-trivial functions with expected value 0, the U^1 'norm' is in fact only a semi-norm. However, we will soon see that for $d \geqslant 2$, $\|\cdot\|_{U^d}$ does define a norm. The homogeneity requirement is trivial from the choice of normalization in Eq. (2.1). What remains is to check the triangle inequality, and to show that if $f : \mathbb{F}^n \to \mathbb{C}$ is a function, then $\|f\|_{U^d} = 0$ if and only if $f \equiv 0$.

These two facts, and many other basic properties of the U^d norms, can often be proven by repeated applications of the classic Cauchy-Schwarz inequality. This can be a rather arduous task, so we are fortunate to have the following lemma from [Gow01] which encapsulates several applications of Cauchy-Schwarz into a single inequality which can be applied in a multitude of ways.

Lemma 2.2 (Gowers-Cauchy-Schwarz inequality). *Let $d \geqslant 1$ be an integer and $\{f_S : \mathbb{F}^n \to \mathbb{C}\}_{S\subseteq[d]}$ a family of functions indexed by subsets of $[d]$. Then we have*

$$\left| \mathop{\mathbf{E}}_{x,y_1,\ldots,y_d\in\mathbb{F}^n} \left[\prod_{S\subseteq[d]} \mathcal{C}^{|S|} f_S(x + \sum_{i\in S} y_i) \right] \right| \leqslant \prod_{S} \|f_S\|_{U^d}. \tag{2.2}$$

To prove this, there is a convenient notation we will make use of.

Definition 2.3 (Gowers inner product). *Let $d \geqslant 1$ be an integer and $\{f_S : \mathbb{F}^n \to \mathbb{C}\}_{S\subseteq[d]}$ a family of of functions indexed by subsets of $[d]$. Then we denote*

$$\langle f_\emptyset, f_{\{1\}}, \ldots, f_{[d]} \rangle = \mathop{\mathbf{E}}_{x,y_1,\ldots,y_d\in\mathbb{F}^n} \left[\prod_{S\subseteq[d]} \mathcal{C}^{|S|} f_S(x + \sum_{i\in S} y_i) \right] \tag{2.3}$$

$$= \mathop{\mathbf{E}}_{z_1,\ldots,z_d,z'_1,\ldots,z'_d\in\mathbb{F}^n} \left[\prod_{S\subseteq[d]} \mathcal{C}^{|S|} f_S(\sum_{i\in S} z_i + \sum_{i\notin S} z'_i) \right].$$

The second equality in Eq. (2.3) is particularly useful, and comes from making the following change of variables. We set $x = z_1 + \cdots + z_d$ and $y_i = z'_i - z_i$ for $1 \leqslant i \leqslant d$;

the value of the expectation will remain fixed because of the group structure. As a matter of convenience, we will often identify subsets of $[d]$ with the integer set $[2^d]$ in the natural way, so that we may write

$$\langle f_1, \ldots, f_{2^d} \rangle$$

and have it be understood that we mean

$$\langle f_\emptyset, f_{\{1\}}, \ldots, f_{[d]} \rangle.$$

Finally, note that Eq. (2.3) also implies that for any function $f : \mathbb{F}^n \to \mathbb{C}$ we have

$$\|f\|_{U^d}^{2^d} = \langle f, \ldots, f \rangle = \mathop{\mathbf{E}}_{y_1,\ldots,y_d,y'_1,\ldots,y'_d \in \mathbb{F}^n} \left[\prod_{S \subseteq [d]} \mathcal{C}^{|S|} f(\sum_{i \in S} y_i + \sum_{i \notin S} y'_i) \right], \tag{2.4}$$

giving an alternate formulation of the U^d norm. Now let us prove Lemma 2.2.

Proof of Lemma 2.2. What we will show is that

$$\langle f_1, \ldots, f_{2^d} \rangle \leqslant \langle f_{2^{d-1}+1}, \ldots, f_{2^d}, f_{2^{d-1}+1}, \ldots f_{2^d} \rangle^{1/2} \langle f_1, \ldots, f_{2^{d-1}}, f_1, \ldots f_{2^{d-1}} \rangle^{1/2}.$$

From here, the result will follow by induction, using the fact that the inequality is trivial when all the f_S are the same (Eq. (2.4)).

To do this, let us begin with the alternate form of the Gowers inner product. From here, we can write this as an iterated expectation: First over $y_d, y'_d \in \mathbb{F}^n$, and then the outer expectation over the remaining $y_1, \ldots, y_{d-1}, y'_1, \ldots, y'_{d-1} \in \mathbb{F}^n$. We also split the product in two based on containing the element d. This looks like

$$\mathbf{E}\left[\mathbf{E} \left(\prod_{S \ni d} \mathcal{C}^{|S|} f_S(\sum_{i \in S} y_i + \sum_{i \notin S} y'_i) \prod_{S \subseteq [d-1]} \mathcal{C}^{|S|} f_S(\sum_{i \in S} y_i + \sum_{i \notin S} y'_i) \right) \right],$$

Now, all the terms in the first product use y_d, and all the terms in the second use y'_d, so they are independent over the inner expectation and we get

$$\mathbf{E}\left[\mathop{\mathbf{E}}_{y_d} \left(\prod_{S \ni d} \mathcal{C}^{|S|} f_S(\sum_{i \in S} y_i + \sum_{i \notin S} y'_i) \right) \mathop{\mathbf{E}}_{y'_d} \left(\prod_{S \ni d} \mathcal{C}^{|S|} f_S(\sum_{i \in S} y_i + \sum_{i \notin S} y'_i) \right) \right].$$

Applying Cauchy-Schwarz to this gives the upper bound

$$\left(\mathbf{E}\left[\mathop{\mathbf{E}}_{y_d}\prod_{S\ni d}\mathcal{C}^{|S|}f_S(\sum_{i\in S}y_i+\sum_{i\notin S}y_i')\right]^2\right)^{1/2}\left(\mathbf{E}\left[\mathop{\mathbf{E}}_{y_d'}\prod_{S\subseteq[d-1]}\mathcal{C}^{|S|}f_S(\sum_{i\in S}y_i+\sum_{i\notin S}y_i')\right]^2\right)^{1/2}.$$

Each of the terms here are exactly in the desired form: We can write

$$\left[\mathop{\mathbf{E}}_{y_d}\prod_{S\ni d}\mathcal{C}^{|S|}f_S(\sum_{i\in S}y_i+\sum_{i\notin S}y_i')\right]^2$$

as two independent expectations, yielding

$$\left(\mathbf{E}\left[\mathop{\mathbf{E}}_{y_d}\prod_{S\ni d}\mathcal{C}^{|S|}f_S(\sum_{i\in S}y_i+\sum_{i\notin S}y_i')\right]^2\right)^{1/2}=\langle f_{2^{d-1}+1},\ldots,f_{2^d},f_{2^{d-1}+1},\ldots f_{2^d}\rangle^{1/2},$$

and similarly for the other term. □

From here, we will be able to infer many facts about the U^d norm by simply choosing the family $\{f_S:\mathbb{F}^n\to\mathbb{C}\}_{S\subseteq[d]}$ suitably and applying the Gowers-Cauchy-Schwarz inequality.

Let us first complete the proof that $\|\cdot\|_{U^d}$, $d\geqslant 2$, does in fact define a norm. To prove the triangle inequality, take all the f_S to be $f+g$, where $f,g:\mathbb{F}^n\to\mathbb{C}$ are arbitrary. Then, an application of the standard triangle inequality gives

$$\|f+g\|_{U^d}^{2^d}=\langle f+g,\ldots,f+g\rangle\leqslant|\langle f,f+g,\ldots,f+g\rangle|+|\langle g,f+g,\ldots,f+g\rangle|.$$

Applying the Gowers-Cauchy-Schwarz inequality to both terms now concedes

$$\|f+g\|_{U^d}^{2^d}\leqslant\|f\|_{U^d}\|f+g\|_{U^d}^{2^d-1}+\|g\|_{U^d}\|f+g\|_{U^d}^{2^d-1},$$

and so we have the triangle inequality for $\|\cdot\|_{U^d}$. To see that $\|\cdot\|_{U^d}$ separates points, take $f_\emptyset=f_{[d]}=f$ for some $f:\mathbb{F}^n\to\mathbb{C}$, and the rest of the $f_S\equiv 1$. Since $d\geqslant 2$, there is a dependence relation in the system $\{x+\sum_{i\in S}y_i\}_{S\subseteq[d]}$, and so there is a change of variables which lets us write

$$\langle f,1,\ldots,1,f\rangle=\mathbf{E}\,f^2.$$

Applying Gowers-Cauchy-Schwarz here shows that

$$\mathbf{E}\, f^2 \leqslant \|f\|_{U^d}^2,$$

implying that $\|f\|_{U^d}= 0$ if and only if $f \equiv 0$.

Before moving on, there is one more useful fact we can get from Lemma 2.2. Taking $f_S = f$ whenever $S \subseteq [d-1]$ and $f_S = 1$ otherwise, Eq. (2.2) becomes

$$\|f\|_{U^{d-1}}^{2^{d-1}} \leqslant \|f\|_{U^d}^{2^{d-1}},$$

which shows that the U^d norms are increasing in d. As a special case of this, we have the identity

$$|\langle f, g\rangle| \leqslant \|f\overline{g}\|_{U^d}, \tag{2.5}$$

for every d, since $|\langle f, g\rangle|$ is in fact just $\|f\overline{g}\|_{U^1}$.

With the basic properties of Gowers norms we have established, we can begin to expose how they will become relevant to the work in this thesis. In particular, what we will show that there is a deep connection between Gowers norms and polynomials.

2.2 Derivatives and non-classical polynomials

To see how U^d can be related to polynomials, there is an important observation to make that will allow us to give a third formulation of the U^d norm. First, let us define the following differential operator.

Definition 2.4 (Multiplicative derivative)**.** *Let $f : \mathbb{F}^n \to \mathbb{C}$ be a function and fix an element $h \in \mathbb{F}^n$. The* multiplicative derivative in direction h of f *is defined to be the unique function $\Delta_h f : \mathbb{F}^n \to \mathbb{C}$ satisfying*

$$\Delta_h f(x) = f(x+h)\overline{f(x)}$$

for every $x \in \mathbb{F}^n$.

Now, observe that iterative applications of the multiplicative derivative have a familiar form. Indeed, for any $f : \mathbb{F}^n \to \mathbb{C}$ and directions $h_1, \ldots, h_d \in \mathbb{F}^n$, we have

$$(\Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_d} f)(x) = \prod_{S\subseteq[d]} \mathcal{C}^{|S|} f(x + \sum_{i\in S} h_i)$$

for every $x \in \mathbb{F}^n$. This is just the term appearing inside the expectation from Eq. (2.1), and so we can write

$$\|f\|_{U^d} = \left| \mathop{\mathbf{E}}_{h_1,\ldots,h_d,x\in\mathbb{F}^n} [(\Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_d} f)(x)] \right|^{1/2^d}. \tag{2.6}$$

In words, the U^d norm of f is given by taking the multiplicative derivative in d random directions and then computing the expected value. We will be able to infer a lot of useful information from Eq. (2.6). First, though, we need some notation.

For every $k \geqslant 1$, let $\mathbb{U}_k$ denote the subgroup $p^{-k}\mathbb{Z}/\mathbb{Z} \subseteq \mathbb{T}$, where $p^{-k}\mathbb{Z} = \{j/p^k \mid j \in \mathbb{Z}\} \subseteq \mathbb{R}$. For reasons that will soon become clear, it will be useful to identify $\mathbb{F}$ with its isomorphic subgroup $\mathbb{U}_1 \subseteq \mathbb{T}$ by the map $x \mapsto p^{-1}|x|$, where $|\cdot|$ denotes the standard map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\}$. In this way, if $P : \mathbb{F}^n \to \mathbb{F}$ is a polynomial in the usual sense, then we can also treat it as a function into $\mathbb{U}_1$, or more generally, $\mathbb{T}$.

Now, although polynomials do not behave particularly well under the multiplicative derivative, they *are* compatible with another differential operator.

Definition 2.5 (Additive derivative)**.** *Let $P : \mathbb{F}^n \to \mathbb{T}$ be a function and fix an element $h \in \mathbb{F}^n$. The* additive derivative in direction h of P *is defined be the unique function $D_h P : \mathbb{F}^n \to \mathbb{T}$ satisfying*

$$D_h P(x) = P(x+h) - P(x)$$

for every $x \in \mathbb{F}^n$.

It is not hard to see that the additive derivative of a polynomial of degree $\leqslant d$ is now a polynomial of degree $\leqslant d-1$, and hence such a polynomial vanishes under any $d+1$ additive derivatives.

The two derivatives we have defined so far are not unrelated. If $P : \mathbb{F}^n \to \mathbb{T}$ is a function and $h \in F^n$, then

$$(\Delta_h \mathsf{e}(P))(x) = \mathsf{e}((D_h P)(x)) \tag{2.7}$$

for every $x \in \mathbb{F}^n$, where $\mathsf{e} : \mathbb{T} \to \mathbb{C}$ denotes the character defined by $\mathsf{e}(x) = e^{2\pi i x}$ for every $x \in \mathbb{T}$.

Now, let us take a polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree $\leqslant d$, and consider the function $f : \mathbb{F}^n \to \mathbb{D}$ given by $f = \mathsf{e}(P)$. Then it follows from Eq. (2.7) that f becomes constant (in particular, 1) under any $d+1$ multiplicative derivatives, and it follows that $\|f\|_{U^{d+1}} = 1$.

Conversely, we would like to say something about functions with U^{d+1} norm 1. This is somewhat meaningless to ask if we allow our functions to be unbounded, so we will assume further that our function satisfies $\|f\|_\infty \leqslant 1$.

Unfortunately, there is still a subtle issue that arises. If $d < p$, then under the boundedness assumption we do get a full converse: Any function $g : \mathbb{F}^n \to \mathbb{D}$ satisfying $\|g\|_{U^{d+1}} = 1$ is of the form $\mathsf{e}(P)$, where $P : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leqslant d$. However, when $d \geqslant p$, we will see that there are different functions satisfying these properties. To deal with this, we will extend the collection of polynomials to a more general class of $\mathbb{T}$ valued functions.

Definition 2.6 (Non-classical polynomials)**.** *Let $d \geqslant 0$ be an integer. A function $P : \mathbb{F}^n \to \mathbb{T}$ is said to be a* non-classical polynomial of degree $\leqslant d$ *(or simply a* polynomial of degree $\leqslant d$*) if for all $h_1, \ldots, h_{d+1} \in \mathbb{F}^n$, it holds that*

$$(D_{h_1} D_{h_2} \cdots D_{h_{d+1}} P)(x) = 0 \tag{2.8}$$

for every $x \in \mathbb{F}^n$. The space of all such functions is denoted by $\mathrm{Poly}_d(\mathbb{F}^n)$.

A function $P : \mathbb{F}^n \to \mathbb{T}$ is said to be a classical polynomial *of degree $\leqslant d$ if $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ and additionally, the image of P is contained in $\mathbb{U}_1$.*

It follows directly from Definition 2.6 and Eq. (2.7) 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 (non-classical) polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$. It will be useful to give functions of this form a name.

Definition 2.7 (Phase polynomials)**.** *A function $f : \mathbb{F}^n \to \mathbb{D}$ is called a* phase polynomial of degree $\leqslant d$ *if*

$$f = \mathsf{e}(P)$$

for some $P \in \mathrm{Poly}_d(\mathbb{F}^n)$.

Now, the issue we stumbled upon regarding fields of low characteristic can be phrased as follows. When $d < p$, every polynomial is also classical, but for $d \geqslant p$, there are polynomials which are not classical: they take values in a larger subgroup of $\mathbb{T}$.

The following lemma of [TZ12] characterizing non-classical polynomials shows that they have a representation not unlike that of classical polynomials, i.e. as a suitably weighted combination of monomials.

Lemma 2.8. *Let $d \geqslant 1$ be an integer, and $P : \mathbb{F}^n \to \mathbb{T}$. Then $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ if and only if it has a representation of the form*

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

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 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.*

As a simple example, consider the function $P : \mathbb{F}_2 \to \mathbb{T}$ given by $P(x) = \frac{1}{4}|x|$. An easy computation of taking derivatives (there is only 1 non-trivial direction to check) shows that this is *not* a linear polynomial: it is in fact quadratic. Observe how we can read this off from Lemma 2.8 by noting that P is a depth 1 polynomial, which in turn implies that the degree is 2 rather than 1. Now, let us prove the lemma.

Proof. In the case that the image of P lies in a coset of $\mathbb{U}_1$ (including, e.g., when P is classical), it is easy to see that we must have $k = 0$ always, and a standard induction argument gives the result. For the general case, we will need to use the following claim.

Claim 2.9. *If $P \in \mathrm{Poly}_d(\mathbb{F}^n)$, then $pP \in \mathrm{Poly}_{d'}(\mathbb{F}^n)$, where $d' = \max(d - p + 1, 0)$.*

Proof. What we will show is that for any direction $h \in \mathbb{F}^n$, we have

$$\Delta_h^p P \equiv 0 \implies \Delta_h pP \equiv 0. \tag{2.9}$$

With this, if $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ with $d \geqslant p - 1$, then for any directions $h_1, \ldots, h_{d-p+2} \in \mathbb{F}^n$ we have

$$\Delta_{h_1}^p \Delta_{h_2} \cdots \Delta_{h_{d-p+2}} P \equiv 0 \implies \Delta_{h_1} \cdots \Delta_{h_{d-p+2}} pP \equiv 0,$$

which shows that $P \in \mathrm{Poly}_{d-p+1}(\mathbb{F}^n)$. When $d < p - 1$, essentially the same argument will show that $\deg(pP) = 0$.

Now, to prove Eq. (2.9), expand $(\Delta_h^p P)(x)$ as

$$(\Delta_h^p P)(x) = \sum_{S \subseteq [p]} (-1)^{|S|} P\left(x + \sum_S h\right) = \sum_{i=0}^{p} (-1)^i \binom{p}{i} P(x + ih).$$

Without loss of generality, suppose p is odd: The calculations will work when $p = 2$ with only minor differences. Then, using $ph = 0$, we can group terms to write this as

$$p\left[P(x-h) - P(x+h)\right] + O(p^2).$$

If this is 0 for every x, then it must be the case that $p\left[P(x-h) - P(x+h)\right]$ itself is as well, but an easy change of variables writes this as $(\Delta_h pP)(x)$, completing the proof of the claim. $\square$

Now, to prove the lemma, we will induct on the degree d using Claim 2.9. When $d < p-1$, $\deg(pP) = 0$, so P takes values in a coset of $\mathbb{U}_1$ and we are done. Otherwise, by induction $pP \in \mathrm{Poly}_{d-p+1}(\mathbb{F}^n)$ can be written in the form

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

Thus, we can write $P(x_1, \ldots, x_n)$ as

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

for some classical $Q \in \mathrm{Poly}_d(\mathbb{F}^n)$. Since the coefficients of Q are unique, and the remaining coefficients are unique by our induction hypothesis, the proof is finished. $\square$

For convenience of exposition, we will assume throughout this thesis that all polynomials have shift 0. This can be done without affecting any of the results we present. Hence, all polynomials of depth k take values in $\mathbb{U}_{k+1}$. For referring to a polynomial by its depth, we make the following definition.

Definition 2.10. *A polynomial $P : \mathbb{F}^n \to \mathbb{T}$ of degree exactly d and depth exactly k is called a* (d, k)-polynomial.

Now, Lemma 2.8 immediately implies some useful facts about non-classical polynomials.

Remark 2.11. *If $Q : \mathbb{F}^n \to \mathbb{T}$ is a (d, k)-polynomial, 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. If* $\lambda \in [1, p-1]$ *is an integer, then* $\deg(\lambda Q) = d$ *and* $\operatorname{depth}(\lambda Q) = k$*.*

Additionally, degree d *polynomials have depth at most* $k = \lfloor (d-1)/(p-1) \rfloor$*, and so such a polynomial takes at most* p^{k+1} *distinct values.*

We have seen that phase polynomials of degree $\leqslant d$ characterize the functions $f : \mathbb{F}^n \to \mathbb{D}$ that satisfy $\|f\|_{U^{d+1}} = 1$. Although this fact is interesting, it will only serve to motivate a stronger result. Indeed, we will need to be able to argue about functions which may only be bounded away from 0 by some small constant in the U^{d+1} norm.

2.3 Direct and inverse theorems

Recall from Chapter 1 the direct and inverse theorems for the U^2 norm which collectively show that having a large U^2 norm is roughly equivalent to having a large non-principal Fourier coefficient. Before we can give a more general version of this result for the U^{d+1} norm, we need to define a notion of similarity between two functions.

Definition 2.12 (Correlation)**.** *Let* $\varepsilon > 0$ *be fixed. If* $f, g : \mathbb{F}^n \to \mathbb{C}$ *are functions, then we say that* f *and* g *are* ε-correlated *if*

$$|\langle f, g\rangle| = \left| \mathop{\mathbf{E}}_{x \in \mathbb{F}^n} f(x)\overline{g(x)} \right| \geqslant \varepsilon.$$

For $d \geqslant 2$, we will not be able to show that a function $f : \mathbb{F}_p^n \to \mathbb{C}$ with large U^{d+1} norm has any relation to $\|\hat{f}\|_\infty$. Note, however, that $|\hat{f}(a)|$ is just the correlation of f with the character χ_a (a linear phase polynomial). So it seems natural that we should replace the notion of Fourier coefficients by correlation with degree d polynomials.

Let us assume then that f is ε-correlated with a phase polynomial of degree $\leqslant d$. Using Eq. (2.5), this implies that

$$\|f\mathsf{e}(-P)\|_{U^{d+1}} \geqslant |\langle f, e(-P)\rangle| \geqslant \varepsilon$$

for some $P \in \operatorname{Poly}_d(\mathbb{F}^n)$.

However we also have the easy fact that $\|f\mathsf{e}(P)\|_{U^{d+1}} = \|f\|_{U^{d+1}}$, which is immediate from Eq. (2.6). This in turn implies that $\|f\|_{U^{d+1}} \geqslant \varepsilon$. So $\|f\|_{U^{d+1}}$ is larger than the correlation of f with *any* phase polynomial. Thus, we have proved the following theorem.

Theorem 2.13 (Direct theorem for U^{d+1} norm)**.** *Let $f : \mathbb{F}^n \to \mathbb{C}$ be a function. Then*

$$\|f\|_{U^{d+1}} \geqslant \sup_{P \in \mathrm{Poly}_d(\mathbb{F}^n)} |\langle f, \mathsf{e}(-P) \rangle| .$$

Note that Theorem 2.13 also encompasses the equality case from earlier as 1-correlation is just equality when $f : \mathbb{F}^n \to \mathbb{D}$ is bounded.

We would now like an inverse theorem to complement this direct theorem. In particular, we would hope that bounded functions $f : \mathbb{F}^n \to \mathbb{D}$ with $\|f\|_{U^{d+1}} \geqslant \delta$ would δ-correlate with some degree $\leqslant d$ phase polynomial. This turns out to be too much to ask for, but we can salvage the idea. We will weaken the conclusion by allowing f to be only ε-correlated with a degree $\leqslant d$ phase polynomial, for some ε that is now allowed to depend on δ. This result is given formally by the following theorem from [TZ12], and is one of the most important results in the entire field. The proof will be omitted, as it is a highly non-trivial result, going beyond the scope of the thesis.

Theorem 2.14 (Inverse theorem for Gowers norms)**.** *Let $d \geqslant 1$ be an integer. For any $\delta > 0$, there exists an $\varepsilon = \varepsilon_{2.14}(\delta, d)$ such that the following holds. For every function $f : \mathbb{F}^n \to \mathbb{D}$ with $\|f\|_{U^{d+1}} \geqslant \delta$, there exists a polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ such that P is ε-correlated with f, so that*

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

There is an interesting history to Theorem 2.14. Barring the trivial cases for the U_1 and U_2 norms, the result for the U^3 norm was the first to be known ([GT08]). In this case, by a convenient technicality (see Lemma 5.12), f can be taken to correlate with a classical quadratic phase polynomial. In the same paper, the authors then conjectured for the general case that f could also correlate with a classical polynomial. However, the necessity of non-classical polynomial in fields of low characteristic (i.e. when $d \geqslant p$) was eventually discovered ([LMS08]). Following this, the theorem was proven in the high characteristic case ([BTZ10]) where non-classical polynomials can be avoided. Finally we have Theorem 2.14 which completes the picture.

Theorem 2.14 will be the starting point for all of the work in the section to follow.

Chapter 3

Higher Order Fourier Analysis

In Chapter 1, we saw, albeit somewhat informally, that the standard theory of Fourier analysis is a very useful tool for analyzing the density 'simple' systems of linear forms. For now, let us focus on the Fourier analysis part of that discussion. The key point is that there is an inverse theorem that gives a rough equivalence between having a large U^2 norm and having a large Fourier coefficient. This in turn implies a decomposition theorem which can write a function as a bounded combination of some characters as well as an error term which is small in U^2 norm. To analyze the structured part of a function, it remained only to use the basic properties of the Fourier expansion, namely that the characters span the entire space of functions $f : \mathbb{F}^n \to \mathbb{C}$ and more importantly that they form an orthonormal basis. This latter property allows us to significantly reduce formulae involving expectations.

Unfortunately, the U^2 norm is not a sufficiently strong error term to control densities of arbitrary linear structures, something we will see in more detail later on. Thus, we would like to develop a theory of 'higher order Fourier analysis'.

Lacking a true higher-order Fourier expansion, we will rather aim to prove a decomposition theorem not unlike the one for the U^2 norm, but with a stronger error condition. In the higher order setting, the error we are allowing ourselves is in the Gowers U^{d+1} norm. Theorem 2.14 gives an inverse theorem for the U^{d+1} norm, which says that functions with large U^{d+1} norm correlate with a phase polynomial of degree $\leqslant d$. When $d = 1$, note that correlating with a linear polynomial is the same as computing a Fourier coefficient, something we exploit in proving our decomposition theorem. So one should be optimistic that we can use the general inverse theorem to prove a decomposition into a bounded combination of degree $\leqslant d$ polynomials, admitting a small U^{d+1} error term.

Under such a decomposition, the 'structured' part of a function $f : \mathbb{F}^n \to [0,1]$ will be of the form $f_1(x) = \Gamma(P_1(x), \ldots, P_C(x))$, where $P_1, \ldots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$ and $\Gamma : \mathbb{T}^C \to [0,1]$ is a function. In fact, denoting by k_i the depth of each polynomial P_i, then since P_i takes values in $\mathbb{U}_{k_i+1}$, Γ is actually a function on $\prod_{i=1}^C \mathbb{U}_{k_i+1}$. Applying the usual Fourier transform to Γ and substituting in $P_1, \ldots, P_C$, this gives us

$$f_1(x) = \sum_\gamma \hat{\Gamma}(\gamma) \mathsf{e}\left(\sum_{i=1}^C \gamma(i) P_i(x)\right), \tag{3.1}$$

where we sum over all γ in the group $\prod_{i=1}^C \mathbb{Z}_{p^{k_i+1}}$. This can be thought of as the 'higher order Fourier expansion' of f, where the higher order characters are now phase polynomials of degree $\leqslant d$. In order to be able to use such an expansion, however, we still require that the higher-order characters be orthogonal. As it turns out, we do not even need the characters to be completely orthogonal. Since our decomposition theorem admits a certain amount of error, it is also acceptable if our characters are not completely orthogonal. Recall the orthogonality of linear characters follows from the fact that any non-principal character has expectation 0. Thus, all that we will require is that expectation of any non-trivial higher-order character is small. This will turn out to be possible by choosing our polynomials correctly.

We will now devote a fairly significant amount of effort to finding the decomposition we described above. We will in fact prove three decomposition theorems, each building on the last in order to obtain one with all the desired properties. Most of the work here follows a program which is fairly standard. Examples include the treatment from [Gre07] for the U^3 norm (before the inverse theorems for larger d were known), as well as the more recent [BFL13] which proves even more general decompositions than we will require. Since we will be working repeatedly with polynomials, our first order of business is to introduce a convenient notation that encapsulates a collection of polynomials with the structure it induces on $\mathbb{F}^n$.

3.1 Polynomial factors

Let $\{A_1, \dots, A_k\}$ be a finite partition of $\mathbb{F}^n$, and let $\mathcal{A}$ denote the σ-algebra on $\mathbb{F}^n$ generated by this partition. Recall that a function $f : \mathbb{F}^n \to \mathbb{C}$ is called $\mathcal{A}$-measurable if it is constant on each set A_i. Now consider a collection of polynomials $P_1, \dots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$. We can partition $\mathbb{F}^n$ according to the values that these polynomials take by writing $\mathbb{F}^n = \cup_{a \in \mathbb{T}^C} \{x \in \mathbb{F}^n : (P_i(x))_{j\in[c]} = a\}$. We shall give a special name to the σ-algebra generated by this partition.

Definition 3.1 (Polynomial factors)**.** *Let $P_1, \dots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$, and for every $a \in \mathbb{T}^C$ denote by B_a the set $\{x \in \mathbb{F}^n : (P_i(x))_{j\in[C]} = a\}$. The σ-algebra $\mathcal{B}$ on $\mathbb{F}^n$ generated by the partition $\{B_a\}_{a\in\mathbb{T}^C}$ is called a* polynomial factor of degree d and complexity at most C.

The sets B_a are called the atoms *of $\mathcal{B}$, and for any $x \in \mathbb{F}^n$, the notation $\mathcal{B}(x)$ is used to denote the unique atom of $\mathcal{B}$ containing x. The degree of $\mathcal{B}$ is denoted by* $\deg(\mathcal{B})$*, while the complexity is denoted by $|\mathcal{B}|$.*

Now let $\mathcal{B}$ be the polynomial factor defined by a collection of polynomials $P_1, \dots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$. In this notation, $\mathcal{B}$-measurable functions are of particular interest to us, since it follows from the definition of being $\mathcal{B}$-measurable that these functions depend only on the values of the polynomials $P_1, \dots, P_C$. In other words, they can be written in the form $\Gamma(P_1, \dots, P_C)$ for some $\Gamma : \mathbb{T}^C \to \mathbb{C}$.

Given an arbitrary function, the obvious way to obtain a $\mathcal{B}$-measurable function is to project it into the space of all $\mathcal{B}$-measurable functions by averaging it over each atom of the partition. More precisely, we have the following definition.

Definition 3.2 (Conditional expectation)**.** *Let $\mathcal{B}$ be the polynomial factor defined by a collection of polynomials $P_1, \dots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$. For any function $f : \mathbb{F}^n \to \mathbb{C}$, the* conditional expectation of f over $\mathcal{B}$*, denoted by $\mathbf{E}[f|\mathcal{B}]$, is the $\mathbb{C}$-valued function on $\mathbb{F}^n$ defined by writing*

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

for every $x \in \mathbb{F}^n$.

Remark 3.3. *Let $\mathcal{B}$ be a polynomial factor and $f : \mathbb{F}^n \to \mathbb{C}$ a function. To see that $\mathbf{E}[f|\mathcal{B}]$ is indeed a projection in the usual sense, note that, for any $\mathcal{B}$-measurable function*

$g : \mathbb{F}^n \to \mathbb{C}$, we have the identity

$$\langle f, g\rangle = \langle \mathbf{E}\left[f|\mathcal{B}\right], g\rangle, \tag{3.2}$$

which is easy to see from the definition of being $\mathcal{B}$-measurable. In fact, for general σ-algebras the conditional expectation is defined as the (unique) function satisfying Eq. (3.2) for every g.

When constructing a polynomial factor, we will often start with some initial factor (which may be the empty, or trivial, factor) and build it up iteratively. One way to extend a polynomial factor is simply to extend the collection of polynomials defining it. If the factor $\mathcal{B}$ is defined by the polynomials $P_1, \ldots, P_C$, and $\mathcal{B}'$ by the polynomials $P_1, \ldots, P_C, Q_1 \ldots, Q_{C'}$, then $\mathcal{B}'$ refines $\mathcal{B}$ as a σ-algebra. However, it is also possible for a polynomial factor to refine $\mathcal{B}$ without a relation between the polynomials. For this reason we will want to be able to differentiate between these types of refinement.

Definition 3.4 (Refinement of factors)**.** *Let $\mathcal{B}$ and $\mathcal{B}'$ be polynomial factors defined by the collections of polynomials $P_1, \ldots, P_C$ and $Q_1, \ldots, Q_{C'}$, respectively. If the collection $Q_1, \ldots, Q_{C'}$ extends $P_1, \ldots, P_C$ in the sense that $\{P_1, \ldots, P_C\} \subseteq \{Q_1, \ldots, Q_{C'}\}$, then $\mathcal{B}'$ is called a* syntactic refinement *of $\mathcal{B}$, denoted as $\mathcal{B}' \preceq_{syn} \mathcal{B}$. If $\mathcal{B}'$ refines $\mathcal{B}$ as a σ-algebra, in the sense that each atom of $\mathcal{B}'$ is contained in an atom of $\mathcal{B}$, then $\mathcal{B}'$ is called a* semantic refinement *of $\mathcal{B}$, denoted as $\mathcal{B}' \preceq_{sem} \mathcal{B}$. The bare notation $\preceq$ will be used synonymously with $\preceq_{sem}$.*

Remark 3.5. *It is clear from the definition that syntactic refinement implies semantic refinement. The converse of this is of course not true, but we can say something partial: If $\mathcal{B}$ is defined by $P_1, \ldots, P_C$ and $\mathcal{B}'$ by $Q_1, \ldots, Q_{C'}$ with $\mathcal{B}' \preceq \mathcal{B}$, then the factor $\mathcal{B}''$ defined by the polynomials in the union $\{P_1, \ldots, P_C\} \cup \{Q_1, \ldots, Q_{C'}\}$ has the same atoms as $\mathcal{B}'$, but now also satisfies $\mathcal{B}'' \preceq_{syn} \mathcal{B}$. The complexity of $\mathcal{B}''$ will in general increase, but it will at the very least satisfy $|\mathcal{B}''| \leqslant |\mathcal{B}'| + |\mathcal{B}|$.*

Now let us move on and give our first decomposition theorem.

3.2 Decomposition from inverse theorem

This first decomposition theorem will follow from the inverse theorem Theorem 2.14 with only a small amount of additional effort, much of which is notational. It should not, then, be too surprising that this will be a particularly weak decomposition theorem. However, it will be pivotal as a starting point for proving stronger decompositions.

Theorem 3.6 (Decomposition theorem I)**.** *Let $d \geqslant 1$ be an integer, and $\mathcal{B}_0$ be a polynomial factor of degree at most d. Given any $\delta > 0$, there exists a constant $C = C_{3.6}(\delta, d, |\mathcal{B}|)$ such that the following holds. For any function $f : \mathbb{F}^n \to [0,1]$, there exist two functions $f_1, f_2 : \mathbb{F}^n \to \mathbb{R}$ and a polynomial factor $\mathcal{B} \preceq \mathcal{B}_0$ of degree at most d and complexity at most C such that*

$$f = f_1 + f_2,$$

where

$$f_1 = \mathbf{E}\left[f|\mathcal{B}\right],$$

and

$$\|f_2\|_{U^{d+1}} \leqslant \delta.$$

Note that, since f is $[0,1]$-valued, it is clear that f_1 will also be $[0,1]$-valued as it just averages f over the atoms of $\mathcal{B}$. Then because $f_2 = f - f_1$, we see that f_2 will be $[-1,1]$-valued. An important thing to keep in mind here is that when f is $\{0,1\}$-valued, i.e. when f is the indicator function of a subset, f_1 will not be $\{0,1\}$-valued: it may take arbitrary values in $[0,1]$.

Now, before we can actually prove Theorem 3.6, we will need some setup, but the general idea is quite simple. If the conditional expectation $\mathbf{E}\left[f|\mathcal{B}_0\right]$ is far from f in U^{d+1} norm, then Theorem 2.14 implies that there exists a polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ that correlates highly with $f - \mathbf{E}\left[f|\mathcal{B}_0\right]$. Syntactically refining $\mathcal{B}_0$ by appending P to the underlying collection of polynomials, we obtain a new polynomial factor and repeat the procedure if necessary. We continue until the polynomial factor satisfies the conclusions of the theorem. We will show that, by our particular choice of the polynomial P, this process will terminate in a bounded number of steps. To make this arguement rigorous, we will need the following definition.

Definition 3.7 (Energy of a factor)**.** *Let $f : \mathbb{F}^n \to [-1, 1]$ be a function, and $\mathcal{B}$ a polynomial factor. The quantity*

$$\|\mathbf{E}[f|\mathcal{B}]\|_2^2 = \langle \mathbf{E}[f|\mathcal{B}], \mathbf{E}[f|\mathcal{B}]\rangle$$

is called the energy of f over $\mathcal{B}$*.*

The key observation to make regarding this definition is that if $f : \mathbb{F}^n \to [-1, 1]$ is a function and $\mathcal{B}$ is some polynomial factor, then refining $\mathcal{B}$ will never incur a decrease in energy. This is not enough, however, to guarantee that the process of repeatly refining a factor will eventually terminate. We need to further ensure that the energy increases by some amount that is bounded away from 0 at each step. A simple, yet important tool here is the next theorem, which is essentially a generalization of Pythagoras' theorem to σ-algebras. We will state the theorem for polynomial factors, but the result is true for arbitrary σ-algebras as well.

Theorem 3.8 (Pythagoras' theorem)**.** *Let $\mathcal{B}$ and $\mathcal{B}'$ be polynomial factors such that $\mathcal{B}' \preceq \mathcal{B}$. Then if $f : \mathbb{F}^n \to [-1, 1]$ is any function, we have*

$$\|\mathbf{E}[f|\mathcal{B}']\|_2^2 = \|\mathbf{E}[f|\mathcal{B}]\|_2^2 + \|\mathbf{E}[f|\mathcal{B}'] - \mathbf{E}[f|\mathcal{B}]\|_2^2. \tag{3.3}$$

Proof. We have

$$\|\mathbf{E}[f|\mathcal{B}'] - \mathbf{E}[f|\mathcal{B}]\|_2^2 = \langle \mathbf{E}[f|\mathcal{B}'] - \mathbf{E}[f|\mathcal{B}], \mathbf{E}[f|\mathcal{B}'] - \mathbf{E}[f|\mathcal{B}]\rangle$$

$$= \langle \mathbf{E}[f|\mathcal{B}'], \mathbf{E}[f|\mathcal{B}']\rangle - 2\langle \mathbf{E}[f|\mathcal{B}'], \mathbf{E}[f|\mathcal{B}]\rangle + \langle \mathbf{E}[f|\mathcal{B}], \mathbf{E}[f|\mathcal{B}]\rangle.$$

Now since $\mathcal{B}' \preceq \mathcal{B}$, it follows that $\mathbf{E}[\mathbf{E}[f|\mathcal{B}']|\mathcal{B}] = \mathbf{E}[f|\mathcal{B}]$, so using Eq. (3.2) from Remark 3.3 we get

$$\langle \mathbf{E}[f|\mathcal{B}'], \mathbf{E}[f|\mathcal{B}]\rangle = \langle \mathbf{E}[\mathbf{E}[f|\mathcal{B}']|\mathcal{B}], \mathbf{E}[f|\mathcal{B}]\rangle$$

$$= \langle \mathbf{E}[f|\mathcal{B}], \mathbf{E}[f|\mathcal{B}]\rangle,$$

which gives the desired result. ∎

With this, we are ready to prove Theorem 3.6.

Proof of Theorem 3.6. We will construct the polynomial factor $\mathcal{B}$ iteratively. Initially, take $\mathcal{B} = \mathcal{B}_0$. Now suppose we are at an arbitrary step, with $\mathcal{B}$ partially constructed; we do the following. Denote $g = f - \mathbf{E}[f|\mathcal{B}]$. If $\|g\|_{U^{d+1}} \leqslant \delta$, then we stop. Otherwise, by Theorem 2.14, there is a polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ that is $\varepsilon_{2.14}(\delta, d)$-correlated with g. Syntactically refine $\mathcal{B}$ by appending the polynomial P to obtain the factor $\mathcal{B}'$, and repeat this procedure with the factor $\mathcal{B}'$ in place of $\mathcal{B}$.

It is clear that if this process terminates in a bounded (depending only on δ and d) number of steps with a polynomial factor $\mathcal{B}$, then taking $f_1 := \mathbf{E}[f|\mathcal{B}]$ and $f_2 := f - f_1$ will satisfy the conclusions of the theorem. It remains to show that this process will indeed terminate. We will do this by showing that at each step, starting with the factor $\mathcal{B}$ and obtaining the new factor $\mathcal{B}'$, we have $\|\mathbf{E}[f|\mathcal{B}']\|_2^2 - \|\mathbf{E}[f|\mathcal{B}]\|_2^2 \geqslant \varepsilon^2$, where $\varepsilon = \varepsilon_{2.14}(\delta, d)$. Since energy is a quantity bounded by 1, this ensures that a maximum of $\lfloor 1/\varepsilon^2 \rfloor$ iterations can occur.

Applying Theorem 3.8 and using the fact that $\mathbf{E}[f|\mathcal{B}] = \mathbf{E}[\mathbf{E}[f|\mathcal{B}]|\mathcal{B}']$ whenever $\mathcal{B}' \preceq \mathcal{B}$, we have

$$\|\mathbf{E}[f|\mathcal{B}']\|_2^2 - \|\mathbf{E}[f|\mathcal{B}]\|_2^2 = \|\mathbf{E}[f|\mathcal{B}'] - \mathbf{E}[f|\mathcal{B}]\|_2^2 = \|\mathbf{E}[g|\mathcal{B}']\|_2^2.$$

Now, since P is ε-correlated with g, we get

$$\varepsilon \leqslant \left| \mathop{\mathbf{E}}_{x \in \mathbb{F}^n} g(x)\mathsf{e}(-P(x)) \right| = \left| \mathop{\mathbf{E}}_{x \in \mathbb{F}^n} \mathbf{E}[g|\mathcal{B}'](x)\mathsf{e}(-P(x)) \right|,$$

where we have used Eq. (3.2) with the fact that $\mathsf{e}(-P)$ is $\mathcal{B}'$-measurable (by definition). The triangle inequality and the fact that the L_p norms are increasing now yield

$$\varepsilon^2 \leqslant \|\mathbf{E}[g|\mathcal{B}']\|_1^2 \leqslant \|\mathbf{E}[g|\mathcal{B}']\|_2^2 = \|\mathbf{E}[f|\mathcal{B}']\|_2^2 - \|\mathbf{E}[f|\mathcal{B}]\|_2^2,$$

as desired. We complete the proof by noting that, since the complexity of our factor increases by 1 at each iteration, the final polynomial factor $\mathcal{B}$ we obtain has complexity at most $C = |\mathcal{B}_0| + \lfloor 1/\varepsilon^2 \rfloor$. □

We mentioned previously that the decomposition given by Theorem 3.6 will not suffice for our needs, and will only be used to prove stronger theorems. To see why this is so, note that we have made no attempt to impose any additional structure on the polynomials defining the factor $\mathcal{B}$. Recall that our goal was to obtain a decomposition with orthogonality properties similar to those in a Fourier decomposition. It turns out that, en route to a decomposition of this form, we will run into another problem, which is as follows. Although we can take the U^{d+1} error δ to be arbitrarily small, this in turn increases the constant C bounding the complexity of the polynomial factor $\mathcal{B}$. When we iteratively apply Theorem 3.6, in order for the error term to stay negligible, we will need it to be small as a function of the complexity C. This might initially seem like a difficult thing to ask for, and reasonably so, as in fact it turns out that such a decomposition theorem is simply not true. However, if we allow ourselves a reasonably small (not depending on C) global L_2 error, then we can ensure the U^{d+1} error is bounded as any function of the complexity.

If one is more familiar with graph theory, it is instructive to make an analogy to graph regularity. It is very reasonable to think of Theorem 3.6 as a weak regularity lemma. Continuing this line of thought, the decomposition we desire is a strong regularity lemma, which we will obtain by iteratively applying the weak lemma. Of course, as in the case of graph regularity, there is a downside to this process: the bounds we ultimately obtain will inevitably be of tower-type. Even worse, our third decomposition, which will iteratively apply the second, and will then end up with bounds which are towers of towers. This is to say nothing of the implicit reliance on the constant appearing in the inverse theorem Theorem 2.14. The full version of this theorem which we are using is proved using a limit approach which does not even give any bounds. In applications where one cares about particular bounds, one will start with a quantitative (albeit weaker) version of the inverse theorem, and then prove something similar to Theorem 3.6, e.g. in [GT09].

Let us now state and prove the strong decomposition theorem.

Theorem 3.9 (Decomposition theorem II)**.** *Let $d \geqslant 1$ be an integer, and $\mathcal{B}_0$ be a polynomial factor of degree at most d and complexity at most C_0. Given any $\delta > 0$ and a non-decreasing function $\omega : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ (where ω may depend on δ), there exists a constant $C = C_{3.9}(\delta, \omega, d, |\mathcal{B}_o|)$ such that the following holds. For any function $f : \mathbb{F}^n \to [0, 1]$, there exist three functions $f_1, f_2, f_3 : \mathbb{F}^n \to \mathbb{R}$ and a polynomial factor $\mathcal{B} \preceq \mathcal{B}_0$ of degree at most d and complexity at most C such that*

$$f = f_1 + f_2 + f_3,$$

where

$$f_1 = \mathbf{E}\left[f|\mathcal{B}\right],$$

$$\|f_2\|_{U^{d+1}} \leqslant 1/\omega(C),$$

and

$$\|f_3\|_2 \leqslant \delta.$$

Proof. We will apply Theorem 3.6 iteratively with parameters δ_i, $i \geqslant 1$, to obtain a sequence of polynomial factors $\mathcal{B}_i$ of respective complexities at most C_i such that for every i, we have

1. $\mathcal{B}_i \preceq \mathcal{B}_{i-1}$;
2. $\|f - \mathbf{E}[f|\mathcal{B}_i]\|_{U^{d+1}} \leqslant \delta_i$;
3. C_i depends only on C_{i-1} and δ_i;
4. $\delta_{i+1} \leqslant 1/\omega(C_i)$.

To do this, we start by taking $0 < \delta_1 < 1/\omega(C_0)$. Applying Theorem 3.6 with parameters δ_1 and $\mathcal{B}_0$, we obtain a polynomial factor $\mathcal{B}_1 \preceq \mathcal{B}_0$ of complexity at most C_1 satisfying conditions $(1)-(3)$ above. Taking $0 \leqslant \delta_2 \leqslant 1/\omega(C_1)$ to satisfy condition (4), we iterate this procedure by applying Theorem 3.6 with parameters δ_2 and $\mathcal{B}_1$.

Now, since the sequence of energies $\|\mathbf{E}[f|\mathcal{B}_i]\|_2^2$ is non-decreasing and bounded by 1, the pigeonhole principle implies that there is some $i \leqslant \lceil 1/\delta^2 \rceil$ such that

$$\|\mathbf{E}[f|\mathcal{B}_{i+1}]\|_2^2 - \|\mathbf{E}[f|\mathcal{B}_i]\|_2^2 \leqslant \delta^2.$$

For such an i, take

$$f_1 = \mathbf{E}[f|\mathcal{B}_i],$$

$$f_2 = \mathbf{E}[f|\mathcal{B}_{i+1}] - \mathbf{E}[f|\mathcal{B}_i],$$

and

$$f_3 = f - \mathbf{E}[f|\mathcal{B}_{i+1}],$$

so that $f = f_1 + f_2 + f_3$. Taking $C = C_i$, which clearly depends only on δ, ω and d, our choice of the δ_i ensures that f_2 will satisfy $\|f_2\|_{U^{d+1}} \leqslant 1/\omega(C)$. Finally, it follows directly from Theorem 3.8 that $\|f_2\|_2 \leqslant \delta$, and so we have obtained the requisite decomposition. $\square$

Again, it follows from the conclusion that f_1 will be $[0,1]$-valued, while it is clear from the proof that f_2 and f_3 will be $[-1,1]$-valued.

With Theorem 3.9, we will now be able work towards imposing the promised orthogonality conditions on our polynomial factor.

3.3 Rank

To motivate some of the definitions to follow, recall Eq. (3.1) which gives something like a higher order Fourier decomposition. The characters of this decomposition are now phase polynomials of degree $\leqslant d$. Ideally, the characters would be completely orthogonal, but we will be satisfied with an approximate orthogonality, such that the expectation of any non-trivial character is small. In fact, we can ask for even more: That the U^d norm of any character is small. It will be useful to give a name to polynomials with this property.

Definition 3.10 (Uniform polynomials). *Fix some $\varepsilon > 0$. A polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ is said to be ε-uniform if*

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

To see what kind of structure uniform polynomials have, consider the following. If a polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ satisfies $\|\mathsf{e}\,(P)\|_{U^d} = 1$, then we have seen that in fact we must have $P \in \mathrm{Poly}_{d-1}(\mathbb{F}^n)$. It is reasonable to suspect, then, that if such a P now satisfies $\|\mathsf{e}\,(P)\|_{U^d} \geqslant \varepsilon$, for some ε, then there should be an algebraic explanation for this in terms of degree $d-1$ polynomials. This leads us to the following definition.

Definition 3.11 (Rank of a polynomial). *For a polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ and an integer $d \geqslant 1$, the d-*rank* of P, denoted $\mathsf{rank}_d(P)$ is defined to be the smallest integer r such that there exist polynomials $Q_1, \ldots, Q_r \in \mathrm{Poly}_{d-1}(\mathbb{F}^n)$ and a function $\Gamma : \mathbb{T}^r \to \mathbb{T}$ satisfying $P(x) = \Gamma(Q_1(x), \ldots, Q_r(x))$ for every $x \in \mathbb{F}^n$. When $d = 1$, we define $\mathsf{rank}_1(P)$ to be ∞ if P is non-constant and 0 otherwise.*

The rank *of a polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ is its $\deg(P)$-rank, and is denoted as just $\mathsf{rank}(P)$. We say that P is r-regular if $\mathsf{rank}(P) \geqslant r$.*

It may be difficult to see how this corresponds to a notion of rank. It is instructive to consider the case when $d = 2$ where P is a (classical) quadratic polynomial. Here, we can write the quadratic part of P in the form $x^T M x$, where M is a matrix and x is the vector of variables. In this case, $\mathsf{rank}(P)$ simply corresponds to the rank of the matrix M.

We now seek a proper correspondence between uniformity and rank that justifies our heuristic argument. This turns out to be difficult, and we call upon the following theorem from [TZ12].

Theorem 3.12 (Inverse theorem for polynomials)**.** *For any $\varepsilon > 0$ and integer $d \geqslant 1$, there exists an integer $r = r_{3.12}(\varepsilon, d)$ such that the following is true. For any polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$, if P is ε-uniform, then $\mathsf{rank}_d(P) \leqslant r$.*

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

Corollary 3.13. *Let ε, d, and $r(d, \varepsilon)$ be as in Theorem 3.12. Then every r-regular polynomial is also ε-uniform.*

As with the inverse theorem for the Gowers norms, Theorem 3.12 is a deep result which is beyond this thesis to prove. In fact, in [TZ12], Theorem 2.14 is deduced from Theorem 3.12 along with a weaker version of the inverse theorem that was already known. Hence, one could say that Theorem 3.12 underpins almost all of the work in this thesis.

At this point, we would like 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. However, it is not immediately obvious what the correct definition should be. It will not be sufficient to simply require a rank (uniformity) condition on each individual polynomial in the collection. To see why this is so, we look again to Eq. (3.1). The characters of this decomposition are phase polynomials, where the phase is an arbitrary linear combination of polynomials. In order to obtain the approximate orthogonality that we desire, we will have to require that all linear combinations of our polynomials be high rank (uniform).

Definition 3.14 (Rank of a polynomial factor)**.** *A polynomial factor $\mathcal{B}$ defined by a sequence of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective depths $k_1, \ldots, k_C$ is said to have rank r if r is the least integer for which there exist $(\lambda_1, \ldots, \lambda_C) \in \mathbb{Z}^C$ with $\lambda_i \bmod p^{k_1+1} \neq 0$ for all $i \in [C]$, such that $\mathsf{rank}_d\left(\sum_{i=1}^{C} \lambda_i P_i\right) \leqslant r$, where $d = \max_i \deg(\lambda_i P_i)$.*

Given a polynomial factor $\mathcal{B}$ of complexity at most C 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 $r(C)$.

Definition 3.15 (Uniform Factors)**.** *Let $\varepsilon > 0$ be a real. A polynomial factor $\mathcal{B}$ defined by a sequence of polynomials $P_1, \ldots, P_C : \mathbb{F}^n \to \mathbb{T}$ with respective depths $k_1, \ldots, k_C$ is said to* *be ε-uniform if for every collection $(\lambda_1, \ldots, \lambda_C) \in \mathbb{Z}^C$ with $\lambda_i \bmod p^{k_1+1} \neq 0$ for all $i \in [C]$, we have*

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

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

Remark 3.16. *Let $\varepsilon : \mathbb{N} \to \mathbb{R}_+$ be an arbitrary non-increasing function. Similar to Corollary 3.13, it also follows from Theorem 3.12 that an r-regular, degree d factor $\mathcal{B}$ is also $\varepsilon(|\mathcal{B}|)$-uniform, where $r = r_{3.12}(d, \varepsilon(\cdot))$.*

From Remark 3.16, we see that it is not particularly important whether we work with rank or uniformity. As such, we will generally use whichever is most convenient. Rank is particularly useful when trying to prove things about collections of polynomials, since the definition gives us some algebraic structure we can try to exploit. Uniformity, on the other hand, is more useful when we are actually using these results, since it gives a quantitative bound on the U^d norm of the polynomials. In what remains of this section, we will be working with rank.

Now, we would like to obtain a decomposition theorem akin to Theorem 3.9 but where the factor $\mathcal{B}$ is now r-regular for some growth function $r : \mathbb{N} \to \mathbb{N}$. To do this, we will need the following lemma of [BFL13], which shows that any polynomial factor $\mathcal{B}$ can be refined into a rank factor with complexity bounded by some universal function depending only on the desired rank and the degree.

Lemma 3.17 (Polynomial regularity lemma)**.** *Let $d \geqslant 1$ be an integer, and $r : \mathbb{N} \to \mathbb{N}$ an arbitrary non-decreasing function. Then there exists another non-decreasing function $\tau : \mathbb{N} \to \mathbb{N}$ with the following property. If $\mathcal{B}$ is any polynomial factor of degree at most d and complexity at most C, then there is another polynomial factor $\mathcal{B}'$ (semantically) refining $\mathcal{B}$, also of degree d, of complexity at most $\tau(C)$ and rank at least $r(C')$.*

The idea here is quite simple. If the factor is not sufficiently high rank, then there is an explanation in terms of a bounded number of low degree polynomials. Thus we can refine $\mathcal{B}$ by removing a high degree polynomial and replacing it with lower degree ones. Iterating if necessary, we will argue that the process terminates by using the fact that the degree of the factor $\mathcal{B}$ decreases.

Proof. Suppose $\mathcal{B}$ is defined by the polynomials $P_1, \ldots, P_C$ of respective depths $k_1, \ldots, k_C$. If $\mathcal{B}$ does not have rank at least $r(C)$, then there exist $\lambda_i \in \mathbb{Z}_{p^{k_i+1}}$ not all 0 such that

$$\sum_{i=1}^{C} \lambda_i P_i = \Gamma(Q_1, \ldots, Q_{r(C)})$$

for a function $\Gamma : \mathbb{T}^{r(C)} \to \mathbb{T}$ and polynomials $Q_1, \ldots, Q_{r(C)} \in \text{Poly}_{d'-1}(\mathbb{F}^n)$, where $d' = \max_i\{\deg(\lambda_i P_i) : \lambda_i \neq 0\}$. Suppose without loss of generality that $d' = \deg(\lambda_1 P_1)$. The value of P_1 is determined by the values of $P_2, \ldots, P_C$ along with the values of $Q_1, \ldots, Q_{r(c)}$; hence the factor $\mathcal{B}'$ obtained by removing P_1 and adding $Q_1, \ldots, Q_{r(C)}$ to the polynomials underlying $\mathcal{B}$ semantically refines $\mathcal{B}$.

If $\mathcal{B}'$ still does not have the desired rank, then we can iterate this process. Because we always choose to remove a polynomial of maximum degree, this will eventually reduce the degree of $\mathcal{B}$, in a number of iterations depending only on d, r and C. If eventually we are left with only linear polynomials, then we are done, and so this defines the function τ. $\square$

Finally, we have all the tools we need to give our third and final decomposition theorem.

3.4 A stronger decomposition theorem

First, let us give the statement of the theorem.

Theorem 3.18 (Decomposition theorem III)**.** *Let $d \geqslant 1$ be an integer. Given any $\delta > 0$ and two arbitrary non-decreasing functions $\eta : \mathbb{N} \to \mathbb{R}_{>0}$, $r : \mathbb{N} \to \mathbb{N}$, there exists a constant $C = C_{3.18}(\delta, \eta, r, d)$ such that the following holds. For any function $f : \mathbb{F}^n \to [0, 1]$, there exist three functions $f_1, f_2, f_3 : \mathbb{F}^n \to \mathbb{R}$ and a polynomial factor $\mathcal{B}$ of degree at most d and complexity at most C so that the following conditions are satisfied:*

1. $f = f_1 + f_2 + f_3$.
2. $f_1 = \mathbf{E}[f|\mathcal{B}]$.
3. $\|f_2\|_{U^{d+1}} \leqslant 1/\eta(C)$.
4. $\|f_3\|_2 \leqslant \delta$.
5. *f_1 has range $[0, 1]$; f_2 and f_3 have range $[-1, 1]$.*

6. $\mathcal{B}$ is r-regular.

Remark 3.19. *As in Remark 3.16, taking $r = r_{3.12}(d, \varepsilon(\cdot))$, we obtain a decomposition where now condition (vi) reads that the factor $\mathcal{B}$ is $\varepsilon(C)$-regular, where $\varepsilon : \mathbb{N} \to \mathbb{R}_{>0}$ is an arbitrary non-increasing function.*

Before we try to prove Theorem 3.18, let us see what we have achieved. We appear to have incorporated everything we desired into a single theorem. Let $f : \mathbb{F}^n \to [0,1]$ be a function, and decompose f as $f_1 + f_2 + f_3$ according to Theorem 3.18. Let $\mathcal{B}$ denote the resulting polynomial factor defined by the polynomials $P_1, \ldots, P_C$ and write $C = |\mathcal{B}|$. As in Remark 3.19, let us take $\mathcal{B}$ to be $\varepsilon(C)$-regular for some function $\varepsilon : \mathbb{N} \to \mathbb{R}_{>0}$. This gives us bounds of the form

$$\mathop{\mathbf{E}}_{x} \mathsf{e}\left(\sum_{i=1}^{C} \gamma_i P_i(x)\right) \leqslant \varepsilon(C)$$

for any γ from a suitable product group (depending on the depths of the P_i). It is not clear yet why this is so useful, but expectations of this form will become prevalent in Chapter 4.

Proving Theorem 3.18 will not be too much effort at this point. As with the previous two theorems, we proceed iteratively. We will apply Theorem 3.9 with some carefully chosen parameters and then use Lemma 3.17 to refine the factor into one of high rank. The energy increment argument from the proof of Theorem 3.6 will again show that the process must terminate in a bounded number of steps. Note that we are not passing an initial factor $\mathcal{B}_0$ as an input in Theorem 3.18; we will start iterating from the trivial factor. It is not that this presents a problem for us, rather that we have no need of it, and so it is omitted to clean up the statement marginally. If one were interested in proving an even stronger decomposition theorem, it would be necessary to retain this additional input. After the proof of Theorem 3.18, we will wrap up the section with a short discussion regarding such stronger decompositions.

Proof of Theorem 3.18. Apply Theorem 3.9 with $\mathcal{B}_0$ as the trivial factor, and the remaining parameters chosen so that the decomposition $f = f_1 + f_2 + f_3$ and the resulting polynomial factor $\mathcal{B}$ (with $C = |\mathcal{B}|$) satisfies

$$\|f_2\|_{U^{d+1}} \leqslant 1/\eta(\tau(C))$$

and

$$\|f_3\|_2 \leqslant \delta/2,$$

where τ is the function obtained when applying Lemma 3.17 to r. Applying this lemma to $\mathcal{B}$ gives another polynomial factor $\mathcal{B}' \preceq \mathcal{B}$ of complexity $C' \leqslant \tau(C)$ and rank at least $r(C')$. Now write

$$f_1' = \mathbf{E}\left[f|\mathcal{B}'\right]$$

$$f_2' = f_2$$

and

$$f_3' = f_3 + \mathbf{E}\left[f|\mathcal{B}\right] - \mathbf{E}\left[f|\mathcal{B}'\right],$$

so that $f = f_1' + f_2' + f_3'$. In this decomposition, by our choice of parameters, f_1' and f_2' are as desired, so if the decomposition fails it is because $\|f_3'\|_2 > \delta$. This implies, by our choice of f_3, that

$$\|\mathbf{E}\left[f|\mathcal{B}\right] - \mathbf{E}\left[f|\mathcal{B}'\right]\|_2 > \delta/2.$$

By Theorem 3.8, this leads to the energy increment

$$\|\mathbf{E}\left[f|\mathcal{B}'\right]\|_2^2 - \|\mathbf{E}\left[f|\mathcal{B}\right]\|_2^2 > \delta^2/4.$$

Thus, if $f = f_1' + f_2' + f_3'$ is not the desired decomposition, we may iterate this procedure, and the process will ultimately terminate, after at most $\lceil 4/\delta^2 \rceil$ steps. □

Now, Theorem 3.18 will suffice for our needs, but before moving on, it is worth mentioning that there are cases when an even stronger decomposition is required. In particular, one would like the L_2 error δ to also be able to decrease as a function of the complexity of the factor $\mathcal{B}$. For a function $f : \mathbb{F}^n \to [0,1]$, consider such a decomposition $f = f_1 + f_2 + f_3$, where f_2 is small in U^{d+1} norm as a function of $|\mathcal{B}|$, and f_3 is small in L_2 norm as a function of $|\mathcal{B}|$. Since the U^{d+1} norm is always bounded by the L_2 norm, by combining f_2 and f_3 this would imply a decomposition of the form $f = g_1 + g_2$, where g_2 is small in U^{d+1} norm as a function of $|\mathcal{B}|$. This is essentially equivalent to the decomposition we initially proposed following Theorem 3.6, which we have already mentioned is impossible.

What turns out to be possible is that the L_2 error decreases as a function of the complexity of a different polynomial factor $\mathcal{B}' \preceq \mathcal{B}$. The factor $\mathcal{B}'$ is close to $\mathcal{B}$ in a sense that we will not go into. If one is so inclined, the factor $\mathcal{B}'$ can also be taken to be of high rank. The details of this decomposition are contained in [BFL13], and is aptly named the super decomposition theorem.

Chapter 4

Equidistibution of Regular Factors

The decomposition given by Theorem 3.18 is a powerful tool which reduces the study of general functions to ones of the form $\Gamma(P_1, \ldots, P_C)$, where $P_1, \ldots, P_C$ is a uniform collection of polynomials (in the sense that they define a uniform polynomial factor). What remains is to actually be able to study these objects. As a warm-up, let us see what can be said about the distribution of $(P_1(x), \ldots, P_C(x))$, where $x \in \mathbb{F}^n$ is taken uniformly at random. The following theorem, which appears in [BFL13], but whose proof implicitly goes back to [Gre07] and possibly further, shows that this distribution can be made arbitrarily close to the uniform distribution by taking $P_1, \ldots, P_C$ to be sufficiently uniform.

Theorem 4.1 (Polynomial Equidistribution)**.** *Let $\mathcal{B}$ be a polynomial factor defined by the collection of (classical) polynomials $P_1, \ldots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$. Suppose that $\mathcal{B}$ is ε-uniform for some $\varepsilon > 0$, then, identifying the atoms of $\mathcal{B}$ with elements in $\mathbb{F}^C$, we have that for every $b \in \mathbb{F}^C$,*

$$\mathbf{Pr}\left[\mathcal{B}(x) = b\right] = p^{-C} \pm \varepsilon. \tag{4.1}$$

This will follow almost directly from the definition of uniform factors.

Proof. We will exploit the fact that $P_i(x) = b_i$ if and only if $p^{-1} \sum_{\lambda \in \mathbb{F}} \mathsf{e}\left(\lambda(P_i(x) - b(i)\right) = 0$. This is because the latter formula is just the expectation of a character which is 0 if and only if that character is principal. Now we can write

$$\mathbf{Pr}\left[\mathcal{B}(x) = b\right] = \mathop{\mathbf{E}}_{x \in \mathbb{F}^n}\left[\prod_{i=1}^{C} \frac{1}{p} \sum_{\lambda_i \in \mathbb{F}} \mathsf{e}\left(\lambda_i(P_i(x) - b(i)\right)\right].$$

Interchanging the expectation and the sum, this reduces to

$$p^{-C} \sum_{\lambda_1,\ldots,\lambda_C \in \mathbb{F}^C} \left[\mathop{\mathbf{E}}_{x\in\mathbb{F}^n} \mathsf{e}\left(\sum_{i=1}^{C} \lambda_i (P_i(x) - b(i)) \right) \right] = p^{-C}(1 \pm p^C \varepsilon),$$

where we have used the fact that $\mathcal{B}$ is ε-uniform to bound the inner expectation by ε whenever the λ_i are not all 0. □

We gave this theorem for a polynomial factor defined by classical polynomials, but with a bit of care one can give an analogous result for arbitrary polynomials. The result will be the same: The distribution will be uniform over a product group depending on the depths of the polynomials. We do not give it here because it is more effort than we are willing to spend on a warm-up, and also because it will follow from a much stronger theorem that we will give later on. Indeed, we will need to understand more than just the distribution of $(P_1(x), \ldots, P_C(x))$, even if we allow the polynomials to be non-classical. To see why this is the case, we need to introduce the notion of complexity.

4.1 Complexity of linear forms

A central quantity associated with a system of linear forms, which we discussed in Chapter 1, is a measure of density. Given a set $A \subseteq \mathbb{F}^n$ and a system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^\ell$, the probability, taken over a uniform $X \in \mathbb{F}^{n\ell}$, that $L_i(X) \in A$ for every $i \in [m]$ is a density of sorts, and will be particularly important when we come to Chapter 5. There is a natural functional analogue of this, which is given below.

Definition 4.2 (Density)**.** *Let $f : \mathbb{F}^n \to \mathbb{C}$ be a function and $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^k$ a system of linear forms. We define*

$$t_{\mathcal{L}}(f) := \mathop{\mathbf{E}}_{X\in\mathbb{F}^n} \left[\prod_{i=1}^{m} f(L_i(X)) \right] \tag{4.2}$$

Recall that we are primarily interested in affine systems of linear forms, as they satisfy a certain homogeneity condition (translation invariance). We will not actually need the following definition, but include it for the sake of completeness.

Definition 4.3 (Homogeneous linear forms)**.** *A system of linear forms* $\mathcal{L} = \{L_1, \ldots, L_m\}$ *in* k *variables is called* homogeneous *if the distribution of* $(L_1(X), \ldots, L_m(X))$ *over a uniform* $X \in \mathbb{F}^{nk}$ *is the same as that of* $(c + L_1(X), \ldots, c + L_m(X))$ *for every* $c \in \mathbb{F}^n$*.*

A problem we have alluded to at least twice is that for any fixed $d \geqslant 1$, there are systems of linear forms whose behavior is not controlled by the Gowers U^{d+1} norm. More precisely, there are systems of linear forms $\mathcal{L}$ such that we cannot in general bound $t_{\mathcal{L}}(f)$ in terms of $\|f\|_{U^{d+1}}$, where $f : \mathbb{F}^n \to [0, 1]$ is a function. Thus, we would like to define a notion of complexity so that systems of linear forms with bounded complexity have well behaved densities. The obvious definition to make here is the following one from [GW10].

Definition 4.4 (True complexity)**.** *Let* $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^k$ *be a system of linear forms. The* true complexity *of* $\mathcal{L}$ *is defined to be the smallest integer* d *such that there exists a function* $\delta : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ *with* $\lim_{\varepsilon \to 0} \delta(\varepsilon) = 0$ *and*

$$\left| \mathop{\mathbf{E}}_{x_1, \ldots, x_k} \left[\prod_{i=1}^{m} f_i(L_i(x_1, \ldots, x_k)) \right] \right| \leqslant \min_i \delta(\|f_i\|_{U^{d+1}})$$

for all $f_1, \ldots, f_m : \mathbb{F}^n \to [-1, 1]$*.*

This is clearly the correct definition, but as we have done little more than restate what we wanted, it does not lend itself to actually describing what kind of systems of linear forms have bounded true complexity. A characterization of such systems does exist: See [HHL14], which resolves a conjecture of [GW10] on this matter. However, there is a different notion of complexity, introduced in [GT10], which immediately gives some structure to systems with bounded complexity.

Definition 4.5 (Cauchy-Schwarz complexity)**.** *Let* $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^k$ *be a system of linear forms. The* Cauchy-Schwarz complexity *of* $\mathcal{L}$ *is defined to be the smallest integer* s *such that for every* $1 \leqslant i \leqslant m$ *we can partition the set* $\mathcal{L} \backslash \{L_i\}$ *into* $s + 1$ *subsets so that* L_i *does not lie in the linear span of any subset.*

An immediate consequence of this definition is that an affine system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^k$ has Cauchy-Schwarz complexity at most $m - 2$. This is because affine systems of linear forms are pairwise independent, so we can always split $\mathcal{L} \backslash \{L_i\}$ into the $m - 1$ singleton subsets.

Cauchy-Schwarz complexity is related to true complexity via the following lemma from [GT10], whose proof is via some clever applications of the eponymous Cauchy-Schwarz inequality.

Lemma 4.6 (Counting lemma)**.** *Let $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^k$ be a system of linear forms of Cauchy-Schwarz complexity s. Then for any functions $f_1, \ldots, f_m : \mathbb{F}^n \to [-1, 1]$, we have*

$$\left| \mathop{\mathbf{E}}_{x_1,\ldots,x_k \in \mathbb{F}^n} \left[\prod_{i=1}^{m} f_i(L_i(x_1, \ldots, x_k)) \right] \right| \leqslant \min_i \|f_i\|_{U^{s+1}}$$

Proof. We will only prove the inequality in the case when $s = 2$. The same argument will extend to the general case, except that the notation becomes somewhat toxic.

Without loss of generality, suppose $\|f_1\|_{U^3} = \min_i \|f_i\|_{U^3}$. We may also assume (through a trick from [GT10]) that L_1 is supported on its first 3 entries and that it is the only linear form in $\mathcal{L}$ with this property. Thus, we can write

$$\mathop{\mathbf{E}}_{x_1,\ldots,x_k} \left[\prod_{i=1}^{m} f_i(L_i(x_1, \ldots, x_k)) \right] = \mathbf{E} \mathop{\mathbf{E}}_{x_1,x_2,x_3 \in \mathbb{F}^n} f_1(x_1, x_2, x_3) g_1(x_1, x_2) g_2(x_1, x_3) g_3(x_2, x_3),$$

where the g_i are bounded functions that may additionally depend on any of the variables $x_4, \ldots, x_k$ in the scope of the outer expectation.

To bound the inner expectation, a first application of Cauchy-Schwarz gives

$$\mathop{\mathbf{E}}_{x_1,x_2,x_3 \in \mathbb{F}^n} g = f_1(x_1, x_2, x_3) g_1(x_1, x_2) g_2(x_1, x_3) g_3(x_2, x_3)$$

$$\leqslant \left(\mathop{\mathbf{E}}_{x_1,x_2} g_1(x_1, x_2)^2 \left(\mathop{\mathbf{E}}_{x_3} f_1(x_1, x_2, x_3) g_2(x_1, x_3) g_3(x_2, x_3) \right)^2 \right)^{1/2}.$$

Bounding $g_1(x_1, x_2)^2$ by 1 and expanding the the squared expectation as two independent expectations, this becomes

$$\leqslant \left(\mathop{\mathbf{E}}_{x_1,x_2,x_3,x_3'} f_1(x_1, x_2, x_3) f_1(x_1, x_2, x_3') g_2(x_1, x_3) g_3(x_2, x_3) g_2(x_1, x_3') g_3(x_2, x_3') \right)^{1/2}$$

$$= \left(\mathop{\mathbf{E}}_{x_3,x_3'} \mathop{\mathbf{E}}_{x_1,x_2} [f_1(x_1, x_2, x_3) f_1(x_1, x_2, x_3')][g_2(x_1, x_3) g_2(x_1, x_3')][g_3(x_2, x_3) g_3(x_2, x_3')] \right)^{1/2}.$$

Here, the inner expectation has exactly the form we started with except with one fewer variables (consider the functions indicated by the brackets and suppress the dependence on the variables x_3, x_3'). Thus, after two more iterations of Cauchy-Schwarz, we will obtain the upper bound

$$(\mathbf{E}\, f_1(x_1,x_2,x_3)f_1(x_1',x_2,x_3)f_1(x_1,x_2',x_3)f_1(x_1',x_2',x_3)\cdot \\ \cdot f_1(x_1,x_2,x_3')f_1(x_1',x_2,x_3')f_1(x_1,x_2',x_3')f_1(x_1',x_2',x_3'))^{1/8},$$

which is exactly $\|f_1\|_{U^3}$ (from Eq. (2.4)). Taking the expectation over the remaining variables $x_4,\ldots,x_k$ completes the proof. □

Remark 4.7. *It follows directly from Lemma 4.6 that Cauchy-Schwarz complexity is an upper bound for true complexity. In particular, we have that the true complexity of an affine system of linear forms is bounded by m.*

Now let $d \geqslant 1$ be an integer, and take $\mathcal{L}$ to be a system of linear forms of true complexity at most d. By suitably decomposing a function $f : \mathbb{F}^n \to [0,1]$ as $f_1+f_2+f_3$ according to Theorem 3.18, it follows that we can make both $t_{\mathcal{L}}(f_2)$ and $t_{\mathcal{L}}(f_3)$ arbitrarily small. The culmination of our discussion regarding complexity is the following lemma, which shows that, in fact, $t_{\mathcal{L}}(f_1)$ can be made to arbitrarily approximate $t_{\mathcal{L}}(f)$.

Lemma 4.8. *Let $\mathcal{L} = \{L_1,\ldots,L_m\} \subseteq \mathbb{F}^k$ be a system of linear forms of true complexity at most d and $\varepsilon > 0$ a constant. Decompose $f : \mathbb{F}^n \to [0,1]$ as $f = f_1+f_2+f_3$ according to Theorem 3.18 with parameters δ, η, r. Then*

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

provided that δ is sufficiently small and η and r grow sufficiently fast.

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

$$\sum_{(i_j)_{j\in[m]}\in[3]^m} \mathop{\mathbf{E}}_{X\in\mathbb{F}^{nk}} \left[\prod_{j=1}^{m} f_{i_j}(L_j(X))\right].$$

Most of the terms in this sum are negligible: If any $i_j = 2$, then we get that the summand is at most $\delta'(1/\eta(|\mathcal{B}|))$, where δ' is from Definition 4.4 and $\mathcal{B}$ is the polynomial factor from

Theorem 3.18. Additionally, if some $i_j = 3$, then an application of Cauchy-Schwarz yields

$$\mathop{\mathbf{E}}_{X \in \mathbb{F}^{nk}} \left[\prod_{j=1}^{m} f_{i_j}(L_j(X)) \right] \leqslant \left(\mathop{\mathbf{E}}_{X \in \mathbb{F}^{nk}} f_3(X)^2 \right)^{1/2},$$

and so we can bound the summand by δ. The only other term, when all the $i_j = 1$, is precisely $t_{\mathcal{L}}(f_1)$, so we get

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

□

It is worth noting that the proof of Lemma 4.8 does not use the full power of Theorem 3.18. It would suffice to decompose f as $f_1 + f_2$ according to Theorem 3.6. We have stated the lemma this way, however, because we will only be applying it as a first step, after which the additional conclusions of Theorem 3.18 will be required.

With these tools, we will be able to complete the discussion we started at the beginning of the section.

4.2 Consistency

Consider a function $f : \mathbb{F}^n \to \{0, 1\}$, and a system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^k$ of true complexity at most d. Lemma 4.8 above shows that we can arbitrarily approximate $t_{\mathcal{L}}(f)$ by $t_{\mathcal{L}}(f_1)$, where $f = f_1 + f_2 + f_3$ is decomposed according to Theorem 3.18 with suitably chosen parameters. Suppose this decomposition writes $f_1 = \mathbf{E}[f|\mathcal{B}]$ for some highly uniform polynomial factor $\mathcal{B}$ defined by the polynomials $P_1, \ldots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$ with respective degrees $d_1, \ldots, d_C$ and depths $k_1, \ldots, k_C$. Then we can write $f_1 = \Gamma(P_1, \ldots, P_C)$ for some function $\Gamma : \mathbb{T}^C \to [0, 1]$, and this gives

$$t_{\mathcal{L}}(f_1) = t_{\mathcal{L}}(\Gamma(P_1, \ldots, P_C)) = \mathop{\mathbf{E}}_{X \in \mathbb{F}^{nk}} \left[\prod_{i=1}^{m} \Gamma(P_1(L_i(X)), \ldots, P_C(L_i(X))) \right].$$

Thus, rather than the distribution of $(P_1(x), \ldots, P_C(x))$ over a uniform $x \in \mathbb{F}^n$, we would like to be able to understand the distribution of the random matrix

$$\begin{pmatrix} P_1(L_1(X)) & P_2(L_1(X)) & \cdots & P_C(L_1(X)) \\ P_1(L_2(X)) & P_2(L_2(X)) & \cdots & P_C(L_2(X)) \\ \vdots & \vdots & & \vdots \\ P_1(L_m(X)) & P_2(L_m(X)) & \cdots & P_C(L_m(X)) \end{pmatrix}, \tag{4.3}$$

where X is a uniform random variable taking values in $\mathbb{F}^{nk}$. Each column of this matrix takes values in $\mathbb{U}_{k_i+1}^m$ for some i. When $m = 1$ and (without loss of generality) $L_1 = (1)$, the non-classical analogue of Theorem 4.1 we mentioned earlier says that the distribution can be made as close as desired to uniform over the product group $\prod_{i=1}^C \mathbb{U}_{k_i+1}$ by taking $\mathcal{B}$ to be sufficiently uniform. However, when $m \geqslant 2$, more complicated situations can arise.

The most obvious issue comes from the fact that for a degree d polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$, we have the derivative identity Eq. (2.8). Hence, if we take $\mathcal{L} \subseteq \mathbb{F}^k$ to be the system of linear forms indexed by subsets $S \subseteq [d+1]$, where $L_S = (1_S(i))_{i \in S}$, the rows of the matrix will sum to 0. In this case, the i-th column will not even be fully supported on $\mathbb{U}_{k_i+1}^m$, since the (say) last entry will be determined by the previous ones. Before we say anything more, let us give a name to the columns on which the distribution of Eq. (4.3) is supported.

Definition 4.9 (Consistency)**.** *Let $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^\ell$ be a system of linear forms. A sequence of elements $b_1, \ldots, b_m \in \mathbb{T}$ is said to be (d,k)-*consistent with $\mathcal{L}$ *if there exists a (d,k)-polynomial $P \in \mathrm{Poly}_{d,k}(\mathbb{F}^n)$ and a point $X \in \mathbb{F}^{n\ell}$ such that $P(L_i(X)) = b_i$ for every $i \in [m]$.*

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

*If $\mathcal{B}$ is a polynomial factor, the term $\mathcal{B}$-*consistent *is synonymous with $(\mathbf{d}, \mathbf{k})$-consistent, where $\mathbf{d}$ and $\mathbf{k}$ are, respectively, the degree and depth sequences of the polynomials defining $\mathcal{B}$.*

In this language, the rows of Eq. (4.3) are $\mathcal{B}$-consistent with $\mathcal{L}$, which is equivalent to saying that each (i-th) column is (d_i, k_i)-consistent with $\mathcal{L}$. Fortunately, consistency turns out to be the only obstruction to getting equidistribution. The following theorem shows that, once we condition on the consistency requirement, the distribution of Eq. (4.3) can be made arbitrarily close to uniform. There is a homogeneity requirement here, which can either be on the polynomials or the linear forms. Since we are mainly interested in affine systems of linear forms, the form of the theorem we will use is from [BFH+13], but more recently in [HHL14] the result was proved for general systems under the assumption that all the polynomials P_i are homogeneous. This is a much stronger result, as with a bit of effort, one can give an analogue of Theorem 3.18 where the polynomial factor is completely defined by homogeneous polynomials. This is nearly pointless effort for us, however, so we will stick with affine systems of linear forms.

Theorem 4.10 (Equidistribution over linear forms)**.** *For some $\varepsilon > 0$, let $\mathcal{B}$ be an ε-uniform polynomial factor of degree $d > 0$ and complexity C, that is defined by the polynomials $P_1, \ldots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$ having respective degrees $d_1, \ldots, d_C$ and depths $k_1, \ldots, k_C$. Let $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^\ell$ be an affine system of linear forms.*

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

$$\mathop{\mathbf{Pr}}_{X \in \mathbb{F}^{nk}} [\mathcal{B}(L_j(X)) = \mathbf{b}_j \ \forall j \in [m]] = \frac{1}{|K|} \pm \varepsilon,$$

where K denotes the subgroup of tuples $(\mathbf{b}_1, \ldots, \mathbf{b}_m)$ that are $\mathcal{B}$ consistent with $\mathcal{L}$.

To prove Theorem 4.10, we would like to proceed as in the proof of Theorem 4.1. Before we can do this, however, we need to prove that the uniformity condition on the polynomials is somehow preserved even when they are composed with linear forms. This result is often referred to as 'strong near-orthogonality', and will be most of the effort towards proving the equidistribution theorem.

4.3 Strong near-orthogonality

Let $\mathcal{B}$ be a uniform polynomial factor defined by the polynomials $P_1, \ldots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$. The uniformity condition here implies that if P is any non-trivial linear combination of the P_i, then $\|\mathsf{e}(P)\|_{U^{d+1}}$ can be made arbitrarily small by increasing the uniformity. Unfortunately, if $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^\ell$ is a system of linear forms, then there are non-trivial linear combinations of the $P_i(L_j)$ which are identically 0. There is an obvious source of these: If any tuple $(\lambda_1, \ldots, \lambda_m)$ satisfies $\sum_{j=1}^m \lambda_j b_j$ for every (d, k)-consistent sequence $b_1, \ldots, b_m$, then clearly $\sum_{j=1}^{m} \lambda_j P(L_j) \equiv 0$ for every (d,k)-polynomial P. Let us give a name to this set.

Definition 4.11 (Dependency)**.** *Let $d \geqslant 1$ and $k \geqslant 0$ be integers. For a system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^\ell$, the* (d,k)-dependency *of $\mathcal{L}$ is defined to be the set of tuples $(\lambda_1, \ldots, \lambda_m) \in \mathbb{Z}^m$ such that $\sum_{j=1}^{m} \lambda_j b_j = 0$ for every sequence $b_1, \ldots, b_m$ that is (d,k)-consistent with $\mathcal{L}$.*

Equivalently, the (d,k)-dependency of $\mathcal{L}$ is the orthogonal complement of the subgroup of (d,k)-consistent sequences.

The question now is whether there are more non-trivial solutions to $\sum \lambda_{i,j} P_i(L_j) \equiv 0$ than those where each tuple $(\lambda_{i,j})_{j \in [m]}$ lies in the (d_i,k_i)-dependency of $\mathcal{L}$. It turns out that, when the polynomials are sufficiently uniform, the answer is that there are not. This leads us to the following dichotomy theorem from [BFH+13].

Theorem 4.12 (Near-orthogonality over linear forms)**.** *Given $\varepsilon > 0$, suppose $\mathcal{B}$ is an ε-uniform polynomial factor defined by the polynomials $P_1, \ldots, P_C \in \mathrm{Poly}_d(\mathbb{F}^n)$. Let $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^\ell$ be an affine system of linear forms, and for every tuple Λ of integers $(\lambda_{i,j})_{i \in [C], j \in [m]}$ define*

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

for every $X \in \mathbb{F}^{n\ell}$. Then, one of the two statements below is true.

1. *$P_{\mathcal{L},\mathcal{B},\Lambda} \equiv 0$.*

2. *$P_{\mathcal{L},\mathcal{B},\Lambda}$ is non-constant and $|\mathbf{E}_{X \in \mathbb{F}^{n\ell}} \mathsf{e}(P_{\mathcal{L},\mathcal{B},\Lambda}(X))| \leqslant \varepsilon$.*

Furthermore, $P_{\mathcal{L},\mathcal{B},\Lambda} \equiv 0$ if and only if for every $i \in [C]$, the tuple $(\lambda_{i,j})_{j \in [m]}$ lies in the (d_i,k_i)-dependency of $\mathcal{L}$.

Proof. First, we will need to make some modifications to the system of linear forms $\mathcal{L}$. For any affine linear form $L = (\lambda_1, \ldots, \lambda_\ell) \in \mathbb{F}^\ell$, we denote $|L| = \sum_{i=2}^{\ell} |\lambda_i|$. Now suppose that $|L| > d$. Then if $P \in \mathrm{Poly}_d(\mathbb{F}^n)$ is any polynomial, for any $y_1, \ldots, y_k$, where $k = |L|$, Eq. (2.8) implies that

$$\sum_{S \subseteq [k]} (-1)^{|S|} P(x + \sum_{i \in S} y_i) = 0.$$

for every $x \in \mathbb{F}^n$. Take the first $|\lambda_2|$ of the y_i as x_2, the next $|\lambda_3|$ of the y_i as x_3, and so on. Then the $S = [k]$ term above will be exactly $P(L(X))$, where $X = (x, x_2, \dots, x_\ell) \in \mathbb{F}^{n\ell}$. All the remaining terms are of the form $P(L'(X))$ where $|L'| < |L|$, and so we can write $P(L(X))$ as a linear combination of these. Iterating this process, we can write $P(L(X))$ as a linear combination of $P(L'(X))$ where every $|L'| \leqslant d$. Doing this every polynomial P_i underlying $\mathcal{B}$, we can write $P_{\mathcal{L},\mathcal{B},\Lambda}$ as $P_{\mathcal{L}',\mathcal{B},\Lambda'}$, where every $L'_j \in \mathcal{L}'$ satisfies $|L'_j| \leqslant \deg \lambda'_{i,j} P_i$ for every $i \in [C]$.

Now, if $P_{\mathcal{L},\mathcal{B},\Lambda} \equiv 0$, then $P_{\mathcal{L}',\mathcal{B},\Lambda'} \equiv 0$ as well. However, since the transformation $\Lambda \to \Lambda'$ depends only on the degree and depth sequences of $\mathcal{B}$, $P_{\mathcal{L},\mathcal{B}',\Lambda} \equiv 0$ for any factor $\mathcal{B}'$ generated by polynomials with the same degree and depth sequence as $\mathcal{B}$. This is precisely the requirement for Λ to have marginals in the dependency of $\mathcal{L}$.

On the other hand, if $P_{\mathcal{L}',\mathcal{B},\Lambda'} \not\equiv 0$, we will be able to show that $|\mathbf{E}\, \mathsf{e}(P_{\mathcal{L}',\mathcal{B},\Lambda'})| \leqslant \varepsilon$. To do this, we will 'derive' $P_{\mathcal{L}',\mathcal{B},\Lambda'}$ until only a maximal term remains. More precisely, we will suppose without loss of generality that L'_1 satisfies

1. $\lambda'_{i,1} \neq 0$ for some $i \in [C]$.

2. L'_1 is maximal in the sense that for every $j \neq 1$, either $\lambda'_{i,j} = 0$ for all $i \in [C]$ or it is the case that $|\lambda_{j,t}| < |\lambda_{1,t}|$ for some $t \in [\ell]$.

Given a vector $a = (\alpha_1, \dots, \alpha_\ell) \in \mathbb{F}^\ell$, a direction $y \in \mathbb{F}^n$, and a function $P : \mathbb{F}^{n\ell} \to \mathbb{T}$, we define the differential operator $D_{a.y}$ by

$$D_{a,y} P(x_1, \dots, x_\ell) = P(x_1 + \alpha_1 y, \dots, x_\ell + \alpha_\ell y) - P(x_1, \dots, x_\ell).$$

Note that the operator $D_{a,y}$ behaves particularly well when P is composed with a linear form $L \in \mathbb{F}^\ell$:

$$D_{a,y} P_i(L(x_1, \dots, x_\ell)) = P_i(L(x_1, \dots, x_\ell) + \langle L, a \rangle y) - P_i(L(x_1, \dots, x_\ell))$$

$$= (D_{\langle L, a \rangle, y} P_i)(L(x_1, \dots, x_\ell)).$$

Hence, if a and L are orthogonal, then $D_{a,y} P_i(L) \equiv 0$ for every y.

Let $d = |L'_1|$, and define the vectors $a_1, \dots, a_d \in \mathbb{F}^\ell$ by taking, for every $i \in [2, \ell]$ and $0 \leqslant \lambda \leqslant |\lambda_{1,i}| - 1$, the vector $(-\lambda, 0, \dots, 0) + e_i$. Since L'_1 is maximal as defined above,

$\langle L_1', a_k\rangle \neq 0$ for every $k \in [d]$, but for any $j > 1$, there exists some $k \in [d]$ such that $\langle L_j', a_k\rangle = 0$. This implies that

$$D_{a_d,y_d}\cdots D_{a_1,y_1}P_{\mathcal{L}',\mathcal{B},\Lambda'} \equiv D_{a_d,y_d}\cdots D_{a_1,y_1}\left(\sum_{i=1}^{C}\lambda'_{i,1}P_i(L_1')\right)$$

$$\equiv \left(D_{\langle L_1',a_d\rangle,y_d}\cdots D_{\langle L_1',a_1\rangle,y_1}\sum_{i=1}^{C}\lambda'_{i,1}P_i\right)(L_1')$$

for every choice of directions $y_1,\ldots,y_d \in \mathbb{F}^n$. Taking expectations and making a change of variables sending L_1' to e_1, we have shown that

$$\mathop{\mathbf{E}}_{\substack{y_1,\ldots,y_d,\\ x_1,\ldots,x_\ell\in\mathbb{F}^n}} \mathsf{e}\left((D_{a_d,y_d}\cdots D_{a_1,y_1}P_{\mathcal{L}',\mathcal{B},\Lambda'})(x_1,\ldots,x_\ell)\right) = \|\sum_{i=1}^{C}\lambda'_{i,1}P_i\|_{U^d}^{2^d}. \tag{4.4}$$

To complete the proof we will use the following claim.

Claim 4.13.

$$\mathop{\mathbf{E}}_{\substack{y_1,\ldots,y_d,\\ x_1,\ldots,x_\ell\in\mathbb{F}^n}} \mathsf{e}\left((D_{a_d,y_d}\cdots D_{a_1,y_1}P_{\mathcal{L}',\mathcal{B},\Lambda'})(x_1,\ldots,x_\ell)\right) \geqslant \left(\left|\mathop{\mathbf{E}}_{x_1,\ldots,x_\ell\in\mathbb{F}^n}\mathsf{e}\left(P_{\mathcal{L}',\mathcal{B},\Lambda'}(x_1,\ldots,x_\ell)\right)\right|\right)^{2^d}.$$

Proof. It suffices to show that for any function $Q:\mathbb{F}^{n\ell}\to\mathbb{T}$ and non-zero $a\in\mathbb{F}^\ell$ we have

$$\left|\mathop{\mathbf{E}}_{y,x_1,\ldots,x_\ell\in\mathbb{F}^n}\mathsf{e}\left((D_{a,y}Q)(x_1,\ldots,x_\ell)\right)\right| \geqslant \left|\mathop{\mathbf{E}}_{x_1,\ldots,x_\ell\in\mathbb{F}^n}\mathsf{e}\left(Q(x_1,\ldots,x_\ell)\right)\right|^2.$$

The claim will then follow by iteratively applying this inequality.

Without loss of generality we can assume that $a = e_1$. Then

$$\left|\mathop{\mathbf{E}}_{y,x_1,\ldots,x_\ell}\mathsf{e}\left((D_{e_1,y}Q)(x_1,\ldots,x_\ell)\right)\right| = \left|\mathop{\mathbf{E}}_{y,x_1,\ldots,x_\ell}\mathsf{e}\left(Q'(x_1+y,x_2,\ldots,x_\ell)-Q'(x_1,\ldots,x_\ell)\right)\right|.$$

Since x_1 and x_1+y are independent, we can write this as

$$\mathop{\mathbf{E}}_{x_2,\ldots,x_\ell}\left|\mathop{\mathbf{E}}_{x_1}\mathsf{e}\left(Q(x_1,\ldots,x_\ell)\right)\right|^2 \geqslant \left|\mathop{\mathbf{E}}_{x_1,\ldots,x_\ell}\mathsf{e}\left(Q(x_1,\ldots,x_\ell)\right)\right|^2.$$

□

Combining Eq. (4.4) with Claim 4.13, the result follows directly from our uniformity assumption on the factor $\mathcal{B}$. □

From here, it will be easy to push the proof of Theorem 4.10 through in a manner analogous to that of Theorem 4.1: We replace the use of the uniformity of the polynomials by an application of Theorem 4.12. Of course, we will also need to treat with non-classical polynomials, but that will only require taking care with our notation.

Proof of Theorem 4.10. As before, we will want to rewrite the indicator function of the event $P_i(L_j(x)) = \mathbf{b}_j(i)$ as the expectation of a character. The polynomial P_i takes values in the subgroup $\mathbb{U}_{k_i+1}$, so the correct statement to make is that

$$\mathbf{1}(P_i(L_j(X)) = \mathbf{b}_j(i)) = p^{-(k_i+1)} \sum_{\lambda \in \mathbb{Z}_{p^{k_i+1}}} \mathsf{e}\left(\lambda(P_i(L_j(X)) - \mathbf{b}_j(i))\right)$$

for any $X \in \mathbb{F}^{nk}$. This then implies that

$$\mathbf{Pr}\left[\mathcal{B}(L_j(X)) = \mathbf{b}_j \ \forall j \in [m]\right] = \mathbf{E}\left[\prod_{i,j}\left(p^{-(k_i+1)} \sum_{\lambda_{i,j} \in \mathbb{Z}_{p^{k_i+1}}} \mathsf{e}\left(\lambda_{i,j}(P_i(L_j(X)) - \mathbf{b}_j(i))\right)\right)\right].$$

Interchanging expectation with sum, this reduces to

$$p^{-m(K+1)} \sum_{\lambda_{i,j} \in \mathbb{Z}_{p^{k_i+1}}} \mathsf{e}\left(-\sum_{i,j} \lambda_{i,j}\mathbf{b}_j(i)\right) \mathbf{E}\left[\mathsf{e}\left(\sum_{i,j} \lambda_{i,j} P_i(L_j(X))\right)\right],$$

where $K = \sum_{i=1}^{C} k_i$. Now for every $i \in [C]$, let Λ_i denote the (d_i, k_i)-dependency of $\mathcal{L}$. Whenever $(\lambda_{i,j})_{j\in[m]} \in \Lambda_i$ for every $i \in [C]$, it follows from the fact that the $\mathbf{b}_j$ are consistent with $\mathcal{L}$ and Theorem 4.12 that the corresponding summand is equal to 1. Otherwise, Theorem 4.12 shows that the summand is bounded by ε. Putting things together, this gives us

$$\mathbf{Pr}\left[\mathcal{B}(L_j(X)) = \mathbf{b}_j \ \forall j \in [m]\right] = p^{-m(K+1)}\left(\prod_{i=1}^{C}|\Lambda_i| \pm \varepsilon p^{mK}\right).$$

Using the fact that Λ_i is the orthogonal complement to the subgroup of (d_i, k_i)-consistent atoms, this immediately implies the result. $\square$

The equidistribution theorem Theorem 4.10, along with Theorem 3.18, will be our main tools for finding a limit object for convergent sequences.

Chapter 5

Main Results

5.1 Convergence and limit objects

Let us first recall the sampling rule that we will use to define convergence. Given a function $f : \mathbb{F}^n \to \{0,1\}$ and an affine system of linear forms, $\mathcal{L} \subseteq \mathbb{F}^k$, we select a random affine transformation $A : \mathbb{F}^k \to \mathbb{F}^n$ uniformly. By looking at the function $Af : x \mapsto f(Ax)$, this induces a probability distribution μ_f over the set of functions $\{\mathbb{F}^k \to \{0,1\}\}$. Then the distribution $\mu_f(\mathcal{L})$ is obtained by restricting μ_f to the set of functions $\{\mathcal{L} \to \{0,1\}\}$. We defined a sequence of functions $\{f_i : \mathbb{F}^{n_i} \to \{0,1\}\}_{i\in\mathbb{N}}$ to be convergent if the distributions μ_f converge for every k. It will be easier to work with the following notion of convergence in terms of the distributions $\mu_f(\mathcal{L})$.

Definition 5.1 (d-convergence)**.** *Let $d \geqslant 1$ be an integer. A sequence of functions $\{f_i : \mathbb{F}^{n_i} \to \{0,1\}\}_{i\in\mathbb{N}}$ is called* d-convergent *if for every integer $k \geqslant 1$ and every affine system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\} \subseteq \mathbb{F}^k$ of true complexity at most d, the probability distributions $\mu_{f_i}(\mathcal{L})$ converge.*

It follows from Remark 4.7 that a sequence $\{f_i : \mathbb{F}^{n_i} \to \{0,1\}\}_{i\in\mathbb{N}}$ is convergent if and only if it is d-convergent for every d. Thus, as a slight abuse of notation we will often talk about ∞-convergence, which is understood to mean convergence.

To study convergent sequences, we would like to be able work with the averages $t_{\mathcal{L}}(f_i)$ rather than the distributions $\mu_{f_i}(\mathcal{L})$. Recall that $\mu_{f_i}(\mathcal{L}) \sim (f(L_1(X)), \ldots, f(L_m(X)))$, where $X \in \mathbb{F}^{nk}$ is chosen uniformly. This clearly determines the value $t_{\mathcal{L}}(f_i)$. The following observation shows us that, in fact, the distribution $\mu_{f_i}(\mathcal{L})$ is determined by the values $t_{\mathcal{L}'}(f_i)$, where $\mathcal{L}' \subseteq \mathcal{L}$.

Observation 5.2. *For every $d \in \mathbb{N} \cup \{\infty\}$, a sequence of functions $\{f_i : \mathbb{F}^{n_i} \to \{0,1\}\}_{i\in\mathbb{N}}$ is d-convergent if and only if for every affine system of linear forms $\mathcal{L}$ of true complexity at most d, the values $t_{\mathcal{L}}(f_i)$ converge.*

Proof. To simplify the notation, let us write $\nu := \mu_{f_i}(\mathcal{L})$ for some i and affine $\mathcal{L}$ of true complexity at most d. We can identify the distribution ν with a function $\nu : \{0,1\}^{\mathcal{L}} \to [0,1]$. Then, applying the Fourier transform, we can write

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

for every $x \in \{0,1\}^{\mathcal{L}}$, so that ν is determined by the $\hat{\nu}(S)$. However, we have

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

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

$$\mathop{\mathbf{E}}_{y}\left[\nu(y)\prod_{i\in S'} y_i\right],$$

where $S' \subseteq S \subseteq \mathcal{L}$. By definition, $\nu(y)$ is the probability that $(f(L_1(X)), \ldots, f(L_m(X))) = y$ over a uniform $X \in \mathbb{F}^{nk}$, which shows that

$$\mathop{\mathbf{E}}_{y}\left[\nu(y)\prod_{i\in S'} y_i\right] = t_{S'}(f_i).$$

It follows that the values $\hat{\nu}(S)$, and hence the distribution ν, are determined by the values $t_{\mathcal{L}'}(f_i)$, where $\mathcal{L}' \subseteq \mathcal{L}$ (and so has true complexity at most d). $\square$

Now we would like to find a representation for the limit of a d-convergent sequence. At this point, we hope it seems reasonably natural to consider the following definition. We will elaborate on our choice more shortly.

Definition 5.3 (d-limit objects)**.** *For every $d \in \mathbb{N} \cup \{\infty\}$, denote*

$$\mathbb{G}_d := \prod_{j=1}^{d} \prod_{k=1}^{\lfloor \frac{j-1}{p-1} \rfloor} \mathbb{U}_{k+1}^{\mathbb{N}}.$$

So every element in the group $\mathbb{G}_d$ 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 $a_{j,k,i} \in U_{k+1}$.

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

A d-limit object $\Gamma : \mathbb{G}_d \to [0,1]$ can be thought of as being a function of infinitely many polynomials: For every degree $j \in [d]$ and then for every possible depth $k \in [0, \lfloor \frac{j-1}{p-1} \rfloor]$ (recall Lemma 2.8), $\mathbb{G}_d$ has a factor of $\mathbb{U}_{k+1}^{\mathbb{N}}$, which is the group a countable collection of (j,k)-polynomials takes values in.

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 the set of functions $\{\mathcal{L} \to \{0,1\}\}$. First, we need another definition, extending the notion of consistency to elements of $\mathbb{G}_d$.

Definition 5.4 (Consistency II)**.** *Let $\mathcal{L} = \{L_1, \dots, L_m\} \subseteq \mathbb{F}^\ell$ be a system of linear forms. 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 $\mathcal{L}$.*

From Theorem 4.10, we are led to suggest the following sampling rule. Consider a d-limit object $\Gamma : \mathbb{G}_d \to [0,1]$. For any affine system of linear forms $\mathcal{L} = \{L_1, \dots, L_m\} \subseteq \mathbb{F}^\ell$, select $\mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{G}_d$ uniformly at random, conditioned on being consistent with $\mathcal{L}$. Then define the random function $g : \mathcal{L} \to \{0,1\}$ 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 induces a probability measure $\mu_\Gamma(\mathcal{L})$ on the set of function $\{\mathcal{L} \to \{0,1\}\}$.

Definition 5.5. *For every $d \in \mathbb{N} \cup \{\infty\}$, we say that a sequence of functions $\{f_i : \mathbb{F}^{n_i} \to \{0,1\}\}_{i \in \mathbb{N}}$* d-converges to Γ *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 $\mu_\Gamma(\mathcal{L})$.*

To continue working with the values $t_{\mathcal{L}}(f_i)$ as we would like, we need an analogue of Definition 4.2 for d-limit objects.

Definition 5.6. *For a d-limit object $\Gamma : \mathbb{G}_d \to [0,1]$, we define*

$$t_{\mathcal{L}}(\Gamma) = \mathbf{E}\left[\prod_{i=1}^{m} \Gamma(\mathbf{b}_i)\right],$$

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

Observation 5.2 can be extended in light of Definition 5.6.

Observation 5.7. *For every $d \in \mathbb{N} \cup \{\infty\}$, a sequence of functions $\{f_i : \mathbb{F}^{n_i} \to \{0,1\}\}_{i\in\mathbb{N}}$ d-converges to a d-limit object $\Gamma : \mathbb{G}_d \to [0,1]$ if and only if for every affine system of linear forms $\mathcal{L}$ of true complexity at most d, we have $\lim_{i\to\infty} t_{\mathcal{L}}(f_i) = t_{\mathcal{L}}(\Gamma)$.*

For any notion of convergence, there are two properties basic of a limit object that should be satisfied if we would like it to properly represent convergent sequences. First, every convergent sequence should converge to a limit object, so that the space of limit objects is complete. Second, every limit object should be obtainable as the limit of a convergent sequence. This second property can be thought of as the space of limit objects being dense in the original structures.

The main theorem of this thesis below shows that d-limit objects have both of these properties with respect to d-convergence. This holds even when we allow $d = \infty$, in which case d-convergence corresponds to our original notion of convergence.

Theorem 5.8 (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.*

5.2 Proof of the main theorem

We will prove our main theorem via two lemma which correspond to the two statements that make up Theorem 5.8. First, we have the completeness lemma.

Lemma 5.9 (Completeness)**.** *Let $\{f_i : \mathbb{F}^{n_i} \to \{0,1\}\}_{i\in\mathbb{N}}$ be a d-convergent sequence. There exists a d-limit object Γ such that $\lim_{i\to\infty} t_{\mathcal{L}}(f_i) = t_{\mathcal{L}}(\Gamma)$ for every affine system of linear forms $\mathcal{L}$ of true complexity at most d.*

The idea of the proof is quite straightforward. If we decompose a function $f : \mathbb{F}^n \to \{0,1\}$ as $f = f_1 + f_2 + f_3$ according to Theorem 3.18, then there is a natural way to treat f_1 as a d-limit object $\Gamma : \mathbb{G}_d \to [0,1]$ (using only finitely many of the coordinates in $\mathbb{G}_d$). By choosing the parameters in the decomposition correctly, we can not only ensure that $t_{\mathcal{L}}(f_1)$ well approximates $t_{\mathcal{L}}(f)$ (Lemma 4.8), but additionally that $t_{\mathcal{L}}(\Gamma)$ well approximates $t_{\mathcal{L}}(f_1)$ (Theorem 4.10). Doing this for each function in our sequence, we will be able to construct a d-limit object which is the limit of the sequence.

Proof. Consider a decreasing sequence $\{\varepsilon_i\}_{i\in\mathbb{N}}$ of positive reals tending to 0. Let the parameters δ_i, η_i, and r_i be chosen as required by Lemma 4.8 so that for every affine system of linear forms $\mathcal{L} = \{L_1, \ldots, L_m\}$ of true complexity at most d, if i is sufficiently large, then the following holds:

(i) $|t_{\mathcal{L}}(f_i) - t_{\mathcal{L}}(f_i^1)| \leqslant \varepsilon_i$ where $f_i = f_i^1 + f_i^2 + f_i^3$ is decomposed according to Theorem 3.18 with the parameters δ_i, η_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 4.10 is true with $\varepsilon = \varepsilon_i p^{-id_iC}$ 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 3.18 with the above mentioned parameters. We have $f_i^1(x) = \tilde{\Gamma}_i(P_1^i(x), \ldots, P_C^i(x))$ for some function $\tilde{\Gamma}_i : \mathbb{T}^C \to [0,1]$ and polynomials $P_1^i, \ldots, P_C^i \in \mathrm{Poly}_{d_i}(\mathbb{F}^n)$. Considering the degrees and the depths of the polynomials, the function $\tilde{\Gamma}_i$ corresponds naturally to a d-limit object Γ_i: Indeed, let $\phi : [C] \to \mathbb{N} \times \mathbb{N} \times \mathbb{N}$ be any injective map that satisfies $\phi(t) = (\deg(P_t^i), \mathrm{depth}(P_t^i), \cdot)$ for every $t \in [C]$. Define $\pi : \mathbb{G}_d \to \mathbb{T}^C$ as $\pi : \mathbf{b} \to (\mathbf{b}(\phi(1)), \ldots, \mathbf{b}(\phi(C)))$, and let $\Gamma_i(\mathbf{b}) := \tilde{\Gamma}_i(\pi(\mathbf{b}))$ for every $\mathbf{b} \in \mathbb{G}_d$.

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

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

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

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

since each P_t^i, $t \in [C]$, has degree at most d_i, hence there are at most p^{id_iC} choices for the y_j when $i \geqslant m$. 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

$$|t_{\mathcal{L}}(f_i) - t_{\mathcal{L}}(\Gamma_i)| \leqslant 2\varepsilon_i. \tag{5.1}$$

Next we construct the limit object Γ. For every $t \in \mathbb{N}$ denote $\mathbb{G}_d^t = \prod_{j=1}^{t} \prod_{k=0}^{\lfloor \frac{j-1}{p-1} \rfloor} \mathbb{U}_{k+1}^t$. Note that $\mathbb{G}_d^t$ corresponds to a partition of $\mathbb{G}_d$. For every measurable $\Gamma : \mathbb{G}_d \to [0,1]$ and $t \in \mathbb{N}$, define $\mathcal{E}_t(\Gamma) = \mathbf{E}\left[\Gamma \mid \mathbb{G}_d^t\right]$. Note that the set $\{\mathbb{G}_d^1 \to [0,1]\}$ 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 $\mu_1 : \mathbb{G}_d^1 \to [0,1]$. 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 $\mu_2 : \mathbb{G}_d^2 \to [0,1]$. Continuing in the same manner we define $\mu_t : \mathbb{G}_d^t \to [0,1]$ for every t. Note that since we restricted to a subsequence at every step, we have $\mathbf{E}[\mu_t|\mathbb{G}_d^r] = \mu_r$ 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 $\Gamma_1', \Gamma_2', \ldots$ of the original sequence that satisfies $\lim \mathbf{E}\left[\Gamma_i' \mid \mathbb{G}_d^t\right] = \mu_t$ for every $t \in \mathbb{N}$.

The measure μ_t is a σ-finite measure over the atoms $\mathbb{G}_d^t$, and thus by Carathéodory's extension theorem, there is a unique measure (also σ-finite) μ on $\mathbb{G}_d$ such that $\mathbf{E}[\mu|\mathbb{G}_d^t] = \mu_t$ for every t. Now let ν denote the uniform measure, and note that for any t and any particular Γ_i we have $\mathbf{E}[\Gamma_i \,|\mathbb{G}_d^t] \leqslant 1$. Since μ_t is a limit (over a subsequence) of these averages, we have $\mu_t(A) \leqslant \nu(A)$ for every $A \subseteq \mathbb{G}_d^t$. It follows that $\mu(A) \leqslant \nu(A)$ for any μ-measurable $A \subseteq \mathbb{G}_d$. In particular, μ is absolutely continuous with respect to ν. Let $\Gamma : \mathbb{G}_d \to [0, 1]$ be the Radon-Nikodym derivative of μ.

Note that as $\lim \mathbf{E}\left[\Gamma'_i \mid \mathbb{G}_d^t\right] = \mu_t$, the sequence of Γ'_i converge to Γ in L_1, and consequently $\lim t_{\mathcal{L}}(\Gamma'_i) = t_{\mathcal{L}}(\Gamma)$. We showed in (Eq. (5.1)) that $\lim t_{\mathcal{L}}(f_i) = \lim t_{\mathcal{L}}(\Gamma_i)$, and since the former limit exists by assumption, it follows that $t_{\mathcal{L}}(\Gamma) = \lim t_{\mathcal{L}}(f_i)$. ☐

Before we can prove the second part of Theorem 5.8, we will need an additional lemma which shows the existence of collections of uniform polynomials with arbitrary degree and depth sequences.

Lemma 5.10. *Let $\mathbf{d} = (d_1, \ldots, d_C) \in \mathbb{Z}_{>0}^C$ and $\mathbf{k} = (k_1, \ldots, k_C) \in \mathbb{Z}_{\geqslant 0}^C$ satisfy $0 \leqslant k_i \leqslant \lfloor \frac{d_i - 1}{p-1} \rfloor$ for every i, and let $\varepsilon > 0$ be a constant. There exists an ε-uniform collection of polynomials $P_1, \ldots, P_C$ such that P_i is of degree d_i and depth k_i for every $i \in [C]$.*

The obvious way to generate polynomials with few low-degree dependencies is to simply have each monomial use a disjoint set of variables. The sum of r monomials of this type will clearly have rank that is bounded as a function of r.

Proof. Let r' be an integer. For each i, let m_i satisfy $d_i = m_i + (p-1)k_i$. Allot variables $x_1^i, \ldots, x_{m_i r'}^i$ for exclusive use by P_i, and let $P_i = p^{-k_i - 1}(x_1^i \cdots x_{m_i}^i + \cdots + x_{m_i(r'-1)+1}^i \cdots x_{m_i r'}^i)$. It is clear that $P_1, \ldots, P_C$ has the desired degree and depth sequence. For sufficiently large n we have enough variables to do this, and for sufficiently large choice of r', these polynomials will have rank at least $r_{3.12}(d, \varepsilon)$, where $d = \max_i d_i$. ☐

With Lemma 5.10, we can now complete the proof of Theorem 5.8. The technique should not be surprising: We will restrict $\mathbb{G}_d$ to the finite subgroup $\mathbb{G}_d^t$ and generate a uniform collection of polynomials with degrees and depths corresponding to the entries of $\mathbb{G}_d^t$. We will then use these polynomials to define a sequence of functions that can easily be shown to d-converge to the desired d-limit object.

Lemma 5.11 (Denseness)**.** *Let $d \in \mathbb{N} \cup \{\infty\}$, and let Γ be a d-limit object. There exists a d-convergent sequence of functions $\{f_i : \mathbb{F}^{n_i} \to \{0,1\}\}_{i \in \mathbb{N}}$ whose limit is Γ.*

Proof. For every $t \in \mathbb{N}$, define the function $\Gamma_t : \mathbb{G}_d \to [0, 1]$ to be the function obtained from $\mathbf{E}\left[\Gamma \mid \mathbb{G}_d^t\right]$ (a map from $\mathbb{G}_d^t$ to $[0, 1]$) by extending it to a function on $\mathbb{G}_d$. The Γ_t converge to Γ in L_1 norm, and each Γ_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 Γ_t.

For every $r \in \mathbb{N}$, we can apply Lemma 5.10 to get a collection of polynomials $P_1, \dots, P_C$ of rank $\geqslant r$ such that P_i has degree d_i^t and depth k_i^t for every i. Now define the function $f_t^r : \mathbb{F}^{n_r} \to [0,1]$ 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 4.10 by the same argument used in the proof of Lemma 5.9 that we have $t_{\mathcal{L}}(f_t^r) \to_r t_{\mathcal{L}}(\Gamma_t)$ 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 $f_i : \mathbb{F}^{n_i} \to [0,1]$ with $t_{\mathcal{L}}(f_i) \to_i t_{\mathcal{L}}(\Gamma)$ for every affine $\mathcal{L}$ of true complexity at most d.

To complete the proof, consider the random functions $f_i' : \mathbb{F}^{n_i} \to \{0,1\}$ where $f_i'(x)$ takes value 1 with probability $f_i(x)$. It is not hard to see that these d-converge to Γ with probability 1, and hence the existence of a d-convergent sequence converging to Γ is evinced. $\square$

This concludes the proof of Theorem 5.8. This answers the two most important questions regarding the veracity of d-limit objects. Now, before we conclude, there is an interesting problem to consider that is relevant to the discussion.

5.3 Necessary depths

Recall the inverse theorem for the Gowers U^{d+1} norm Theorem 2.14, which shows that functions with large U^{d+1} norm must correlate with a polynomial $P \in \mathrm{Poly}_d(\mathbb{F}^n)$. From Lemma 2.8, it is clear that any P that is non-classical must have degree at least p. However, it is further known that polynomials of degree $d = p$ that are not-classical are unnecessary in higher order Fourier analysis (hence why non-classical polynomials can be avoided in quadratic Fourier analysis). More precisely, in Theorem 2.14, taking $d = p$, one can assume that the polynomial P is in fact a classical polynomial of degree at most p. This can be carried through the decomposition theorems to Theorem 3.18, and on to the definition of a 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 5.12. *Every $(1 + k(p-1), k)$-polynomial $P : \mathbb{F}^n \to \mathbb{T}$ 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 2.8, we have $P(x_1, \dots, x_n) = \frac{\sum c_i |x_i|}{p^{k+1}} + R(x_1, \dots, x_n) \mod 1$ for integers $0 \leqslant c_i \leqslant p-1$, where R is a $(1+k(p-1), k-1)$-polynomial. Let $M := \sum c_i |x_i|$, and let $0 \leqslant a < p^k$ and $b \in [p-1]$ be the unique integers satisfying $M \equiv a + bp^k \mod p^{k+1}$. 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 $\frac{M}{p^k} \mod 1$. Furthermore knowing a, the value of b is determined by the value of the $\frac{M^p - M}{p^{k+1}} \mod 1$. Indeed

$$bp^k \equiv (a + bp^k)^p - (a + bp^k) - (a^p - a) \mod p^{k+1}. \tag{5.2}$$

It remains to show that $\frac{M^p - M}{p^{k+1}} \mod 1$ is a $(1+k(p-1), k-1)$-polynomial. Since degree and depth are invariant under affine transformations, it suffices to show that $Q := \frac{|x_1|^p - |x_1|}{p^{k+1}} \mod 1$ 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 $(p-1) + (p-1)(k-1) = k(p-1)$. ☐

It follows from Lemma 5.12 that in Theorem 3.18, $(1 + k(p-1), k)$-polynomials can be avoided in the polynomials defining the factor $\mathcal{B}$. Consequently, every d-convergent sequence converges to a d-limit object $\phi : \mathbb{G}_d \to [0, 1]$ such that ϕ 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 $< d$. To do this, we need the following theorem of [TZ12], whose proof we omit.

Theorem 5.13. *Let $d > p$ be an integer, and $\varepsilon > 0$. There exists a $\rho = \rho_{5.13}(\varepsilon, d)$ such that the following holds for sufficiently large n. If $P : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree d with $\|\mathsf{e}(P)\|_{U^d} \geqslant \varepsilon$, then $pP : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leqslant d - p + 1$ that satisfies*

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

Now, the next lemma implies that unless d and k are as in Lemma 5.12, the following holds. 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 $< d$.

Lemma 5.14. *Let $m \geqslant 2$ be an integer, and $\varepsilon > 0$. Then for every $k \geqslant 0$, defining $d = m + k(p-1)$, there exists a degree (d,k)-polynomial Q such that*

$$|\langle \mathsf{e}(Q), \mathsf{e}(R_1 + R_2)\rangle| < \varepsilon, \tag{5.3}$$

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

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

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

$$\varepsilon_i < \varepsilon_{2.14}(\rho_{5.13}(\varepsilon_{i+1}, d), d).$$

We show by induction on i that if n is sufficiently large, then the $(m+i(p-1), i)$-polynomial $Q = \frac{P}{p^{i+1}} \mod 1$ satisfies the desired property with parameter ε_i in (5.3) instead of ε.

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 $\|\mathsf{e}(Q)\|_{U^d}$ 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 ε_0.

Now let us consider the case $i > 0$. Assume for the sake of a contradiction that $|\langle \mathsf{e}\left(\frac{P}{p^{i+1}}\right), \mathsf{e}(R_1 + R_2)\rangle| \geqslant \varepsilon_i$ 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 $\leqslant d_i - 1$. This in particular implies $\|\mathsf{e}\left(\frac{P}{p^{i+1}} - R_1 - R_2\right)\|_{U^{d_i}} \geqslant \varepsilon_i$. Note that $\frac{P}{p^{i+1}} - R_1 - R_2 \mod 1$ has degree $d_i > p$, and thus we can apply Theorem 5.13 to conclude that

$$\|\mathsf{e}\left(p(P/p^{i+1} - R_1 - R_2)\right)\|_{U^{d_i - p + 1}} \geqslant \rho\left(\|\mathsf{e}\left(P/p^{i+1} - R_1 - R_2\right)\|_{U^{d_i}}\right) \geqslant \rho_{5.13}(\varepsilon_i, d_i).$$

Therefore by Theorem 2.14 there exists a polynomial R' of degree at most $d_i - p$ such that

$$\left|\left\langle \mathsf{e}\left(p(P/p^i - R_1 - R_2)\right), \mathsf{e}\left(R'\right)\right\rangle\right| > \varepsilon_{2.14}(\rho_{5.13}(\varepsilon_i, d_i), d_i - p) \geqslant \varepsilon_{2.14}(\rho_{5.13}(\varepsilon_i, d_i), d_i) \geqslant \varepsilon_{i-1}.$$

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

Although we have seen that these lemmas do have implications for d-limit objects, Lemma 5.12 and Lemma 5.14 are results that are interesting in their own right, and may be useful even outside the scope of this thesis.

Chapter 6

Concluding Remarks

We would like to wrap up by looking at an interesting direction in which to continue this line of work. So far we have seen that d-limit objects correctly capture the notion of d-convergence in the sense that every d-convergent sequence converges to a d-limit object and every d-limit object is the limit of some d-convergent sequence. What is still missing from this picture is to study in what sense d-limit objects are *unique*. To see how powerful a characterization of uniqueness can be, let us again appeal to the graph limits example.

Two graphons $U, W : [0,1]^2 \to [0,1]$ are called *weakly isomorphic* if $t_{\text{ind}}(F, U) = t_{\text{ind}}(F, W)$ for every simple graph F. Recall that a transformation $\sigma : [0,1] \to [0,1]$ is called measure preserving if $\lambda(\sigma^{-1}(A)) = \lambda(A)$ for every measurable $A \subseteq [0,1]$. If W is a graphon, and σ is measure preserving, then we can define another graphon W^σ by writing $W^\sigma(x,y) = W(\sigma(x), \sigma(y))$. It follows from the definition of being measure preserving that W is always weakly isomorphic to W^σ.

The direct converse to this is false: There exist weakly isomorphic graphons W_1, W_2 such that neither can be obtained from the other by composing with a measure preserving map. It is not far from being true, however. In [BCL10], the authors prove that if W_1, W_2 are weakly isomorphic graphons, then there exists a third graphon W and two measure preserving maps σ_1, σ_2 such that $W = W_1^{\sigma_1}$ and $W = W_2^{\sigma_2}$ almost everywhere. With a bit of effort, this characterization can be used to show that the space of graphons modulo weak isomorphism is in fact a *compact* metric space. This is an extremely useful fact which has been (and still is) used to prove many new results in the field.

In our setting, we would like to see what can be said about d-limit objects satisfying

$t_{\mathcal{L}}(\Gamma_1) = t_{\mathcal{L}}(\Gamma_2)$ for every affine system of linear forms $\mathcal{L}$ of true complexity at most d. Let us borrow terminology and call Γ_1 and Γ_2 *d-weakly isomorphic* (or just weakly isomorphic when this is not ambiguous) in such a case. This is an interesting problem because, together Theorem 5.8, it would imply something about the types of functions from $\{\mathbb{F}^n \to \{0,1\}\}$ which be distinguished by sampling densities. As before, there is an easy way to generate weakly isomorphic limit objects: If $f = \Gamma(P_1, \ldots, P_C)$ where each $\deg(P_i) \leqslant d$, then by taking the rank of the collection $\{P_1, \ldots, P_C\}$ to infinity, we obtain a d-limit object in the obvious way (recall the proof of Lemma 5.11). Replacing each P_i by any Q_i of the same degree and depth and letting the rank of $\{Q_1, \ldots, Q_C\}$ go to infinity, we obtain a new d-limit object which will be weakly isomorphic to the original. This follows directly from Theorem 4.12.

There is more that can go wrong, however. It may be that in the subspace spanned by P_1 and P_2, Γ is only supported on, say, $P_1 + P_2$. In this case, we can write f as $\Gamma'(P_1 + P_2, P_3, \ldots, P_C)$ for some Γ'. As the rank goes to infinity, we will again obtain two weakly isomorphic limit objects, where here we have used Theorem 4.10. For 1-convergence, where all the polynomials are linear, this appears to be a complete characterization. We can prove that if $f = \Gamma_1(P_1, \ldots, P_{C_1})$ and $g = \Gamma_2(Q_1, \ldots, Q_{C_2})$ are weakly isomorphic, where the P_i and Q_i are all non-constant and linear (hence, classical as well), then there exists a third function Γ and two linear transformations $T_1 : \mathbb{T}^{C_1} \to \mathbb{T}^C$ and $T_2 : \mathbb{T}^{C_2} \to \mathbb{T}^C$ such that

$$\Gamma_1(P_1, \ldots, P_{C_1}) \equiv \Gamma(T_1(P_1, \ldots, P_{C_1}))$$

and

$$\Gamma_2(Q_1, \ldots, Q_{C_2}) \equiv \Gamma(T_2(Q_1, \ldots, Q_{C_2})).$$

Regrettably, this result does not appear to generalize to the higher degree cases. Even for 2-convergence, there are more subtle issues than those incurred by linear transformations. Indeed, consider taking $\deg(Q) = 2$ and $\deg(P_1) = \deg(P_2) = 1$. Then suppose $f = \Gamma(Q, P_1, P_2)$ is such that we can write f as $\Gamma'(Q + P_1 \cdot P_2)$ for some Γ'. Letting the rank of $\{Q, P_1, P_2\}$ go to infinity, we obtain two limit objects (call them Γ and Γ' in a minor abuse of notation) that are clearly not linearly related.

Now, the rank condition implies that $t_{\mathcal{L}}(\Gamma) = t_{\mathcal{L}}(f)$ for every $\mathcal{L}$ of complexity at most 2. For Γ', note that since the rank of $P_1 \cdot P_2$ is low (it is rank 1), the rank of $Q + P_1 \cdot P_2$ is essentially that of Q. So it follows that $t_{\mathcal{L}}(\Gamma') = t_{\mathcal{L}}(f)$ as well, and we have weak isomorphism.

We can push this example further, and show that d-weak isomorphism cannot be characterized by lower degree relations. Let $f = \Gamma(P_1, \dots, P_C)$ for some high rank collection $\{P_1, \dots, P_C\}$ of degree $\leqslant d$. Then suppose we can find a Γ' and degree $\leqslant d$ polynomials $Q_1, \dots, Q_{C'}$ satisfying $\Gamma(P_1, \dots, P_C) \equiv \Gamma'(R(Q_1, \dots, Q_{C'}))$, where $R : \mathbb{T}^{C'} \to \mathbb{T}^C$ is a degree $\leqslant d$ polynomial in each coordinate (of $\mathbb{T}^C$) such that the degree and depth sequences of $P_1, \dots, P_C$ and $R(Q_1, \dots, Q_{C'})$ are the same. Passing to the limit we will obtain weakly isomorphic d-limit objects. It seems possible that we can characterize weakly isomorphic d-limit objects this way, i.e., by the existence of a third limit object and two degree $\leqslant d$ polynomials mapping to the same degree and depth sequence, but there may yet be other ways to generate weakly isormophic limit objects that we do not know of.

References

[BCL10] Christian Borgs, Jennifer Chayes, and László Lovász. Moments of two-variable functions and the uniqueness of graph limits. *Geom. Funct. Anal.*, 19(6):1597–1619, 2010. 61

[BFH+13] Arnab Bhattacharyya, Eldar Fische, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every locally characterized afine-invariant property is testable. In *Proc. 45th Annual ACM Symposium on the Theory of Computing*, 2013. Full version at arXiv:1212.3849. 44, 45

[BFL13] Arnab Bhattacharyya, Eldar Fischer, and Shachar Lovett. Testing low complexity affine-invariant properties. In *Proc. 23rd ACM-SIAM Symposium on Discrete Algorithms*, 2013. 22, 32, 36, 37

[BTZ10] Vitaly Bergelson, Terence Tao, and Tamar Ziegler. An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}^\omega$. *Geom. Funct. Anal.*, 19(6):1539–1596, 2010. 20

[Gow01] William T. Gowers. A new proof of Szeméredi's theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001. 10, 11

[Gre07] Ben Green. Montréal notes on quadratic Fourier analysis. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 69–102. Amer. Math. Soc., Providence, RI, 2007. 22, 37

[GT08] Ben Green and Terence Tao. An inverse theorem for the Gowers U^3-norm. *Proc. Edin. Math. Soc.*, 51:73–153, 2008. 20

[GT09] Ben Green and Terence Tao. New bounds for Szemerédi's theorem. II. A new bound for $r_4(N)$. In *Analytic number theory*, pages 180–204. Cambridge Univ. Press, Cambridge, 2009. 28

[GT10] Benjamin Green and Terence Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010. 39, 40

[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. 39

[HHH14] Hamed Hatami, Pooya Hatami, and James Hirst. Limits of boolean functions on $\mathbb{F}_p^n$. 2014. http://arxiv.org/abs/1308.4108. 8

[HHL14] Hamed Hatami, Pooya Hatami, and Shachar Lovett. General systems of linear forms: equidistribution and true complexity. May 2014. http://arxiv.org/abs/1403.7703. 39, 44

[LMS08] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 547–556, New York, NY, USA, 2008. ACM. 20

[LS06] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006. 2, 3

[Sze10] Balázs Szegedy. Gowers norms, regularization and limits of functions on abelian groups. 2010. Full version at arXiv:1010.6211. i, ii

[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. 16, 20, 31, 58