Combinatorial and dynamical properties of polynomial progressions

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Science and Engineering

2021

Borys J. Kuca

School of Natural Sciences
Department of Mathematics

Blank Page

Contents

Word Count: 36944

Blank Page

Abstract

In this thesis, we study combinatorial and dynamical questions on polynomial progressions that are connected with the polynomial Szemerédi theorem of Bergelson and Leibman. Some of these questions stem from additive combinatorics whereas others have ergodic-theoretic flavour.

Firstly, we prove upper bounds for the size of subsets of finite fields lacking certain polynomial progressions. Specifically, we look at two single-dimensional families of progressions and one multidimensional family. In doing so, we obtain asymptotics for the number of certain progressions in subsets of finite fields with quantitative error terms; we also get a quantitative control of certain polynomial configurations by low-degree Gowers norms. These results are obtained using discrete Fourier analysis, a basic theory of Gowers norms, the degree-lowering argument of Peluse, and variations of the PET induction scheme of Bergelson and Leibman.

Secondly, we qualitatively study several notions of complexity of polynomial progressions. One of them comes from additive combinatorics and describes the smallest-degree Gowers norm controlling a given configuration. Another two originate in ergodic theory and refer to the smallest characteristic factor for the convergence of the multiple ergodic averages associated with the progression. The last one is purely algebraic, concerning the algebraic relations between terms of the progression. We conjecture that these four notions agree for all polynomial progressions. We show this for all homogeneous progressions, a large class of progressions that includes most of the progressions for which complexity results have previously been obtained, and many more. We also prove a number of smaller results: the equivalence of true and algebraic complexity for a certain family of inhomogeneous polynomial progressions, asymptotics for the count of progressions of complexity 1 or multiple recurrence results for these configurations. Our proofs apply techniques from higher order Fourier analysis and ergodic theory. In the process of deriving our results, we give new equidistribution results on nilmanifolds for certain types of polynomial sequences.

Blank Page

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Blank Page

Copyright Statement

Blank Page

Outline

This thesis is written in a journal format style in accordance with the university guidelines. It consists of the introductory Chapter 1, followed by three paper chapters, each of which is our own work, and concluded by a summary section. Each chapter has its own bibliography, and the paper chapters also have their own abstracts. In this section, we outline the structure of the thesis together with the relations of the particular chapters to our papers and preprints [Kuc21a; Kuc21b; Kuc21c; Kuc21d].

Chapter 1 summarises our results in the context of earlier works and gives the taste of the most important ideas behind our results. In Section 1.1, we discuss celebrated theorems of van der Waerden, Szemerédi, Bergelson, Leibman, Host, Kra and Walsh in additive combinatorics and ergodic theory. We also list questions inspired by these theorems that have motivated our work presented in this thesis. We then proceed to summarise the state of the art for specific questions; this includes our own contributions. Where possible, we give a brief explanation of the underlying ideas behind our proofs to aid the reading of the main body of the thesis. In Section 1.2, we discuss known bounds in the polynomial Szemerédi theorem together with our results proved in [Kuc21a; Kuc21b]. In Section 1.3, we define various notions of complexity of polynomial progressions, present a conjecture relating these notions, and discuss the state of the art for this conjecture. In particular, we summarise our contributions from [Kuc21c; Kuc21d]. Section 1.4 contains the basic definitions related to nilmanifolds, and it provides a glimpse into our results from [Kuc21c; Kuc21d] on the equidistribution of polynomial sequences on nilmanifolds. Finally, Section 1.5 briefly discusses the related problems of determining which progressions have good lower bounds for multiple recurrence or contain many popular common differences. This section includes some results of ours that have been proved in [Kuc21b; Kuc21c].

Chapter 2 contains the material from our paper "Further bounds in the polynomial Szemerédi theorem over finite fields" [Kuc21a] that has been published by Acta Arithmetica. It contains new upper bounds for certain cases of the polynomial Szemerédi theorem over finite fields. This chapter is very close to the version published in Acta Arithmetica. Apart from minor changes in notation, two deviations from the Acta Arithmetica version include a more succinct proof of Lemma 2.4.6 and the deletion of a short concluding chapter.

Chapter 3 contains our paper "Multidimensional polynomial Szemerédi theorem in finite fields for polynomials of distinct degrees" [Kuc21b] and proves some upper bounds in the multidimensional version of the polynomial Szemerédi theorem. The only noteworthy deviation from the up-to-date arXiv version is a more concise reformulation of Section 3.6.

Chapter 4 contains our paper "On several notions of complexity of polynomial progressions" [Kuc21c]. It has been extended by several results from our earlier paper "True complexity of polynomial progressions in finite fields" [Kuc21d], published in the Proceedings of the Edinburgh Mathematical Society. We have decided to merge these two papers into one thesis chapter to make the thesis more concise and facilitate the examiners' job. Chapter 4 retains the structure of [Kuc21c], and a far majority of the content of this Chapter comes from this paper; all the additions from [Kuc21d] are clearly designated as such in the body of the chapter.

Chapter 5 lists potential future directions that follow from our work, together with a brief summary of obstacles that one has to overcome to make further progress.

Acknowledgements

PhD was not just an intellectual challenge that forced me to get out of my comfort zone and surpass my limitations; it was not merely a fascinating learning experience during which I managed to make original contributions to the bottomless pool of human knowledge, something that I have always dreamt of. It was first and foremost an exhilarating journey, with many twists and turns, from which I have emerged wiser and more mature - or at least so I hope. Over the course of my doctoral degree, I continually felt the support of many people, to whom I owe my deepest gratitude.

I would like to first thank my three supervisors - Sean Prendiville, Tuomas Sahlsten and Donald Robertson. Thank you for all the trust, help and support that I have received from you in both professional and non-professional matters. In particular, thank you for setting me on research track, guiding me in my exploration of additive combinatorics and ergodic theory, patiently answering my countless questions, keeping me informed about conferences and other educational opportunities, reading my work, and helping me obtain postdoctoral positions. Thank you for our numerous mathematical and non-mathematical conversations in the last three years, during which I learnt more than I can express. I would also like to thank all the other academics in Manchester who supported me at various stages of my degree, and the university itself for providing generous financial support, good working space and friendly environment.

During the PhD, I was incredibly lucky to be supported by a number of amazing friends, without whom the everyday experience of doing research would have lost much of its charm. It is the time spent with you that made my PhD unforgettable, and I cannot express how thankful I am for that. I would like to thank my academic brothers - Jon Chapman, Ioannis Tsokanos, Joe Thomas and Connor Stevens - for all the good time that we had together, and for all our fascinating mathematical and non-mathematical discussions. I wish to express my gratitude to my officemates - Vicky Torega, Phil Richardson, Luke Webb, Alex Hiles, Xiaoxi Pang and Gabriele Incorvaia - for providing an excellent office atmosphere and great company before the pandemic forced us out of the Alan Turing Building. Lastly, I would like to thank all my other PhD friends from Yale and Manchester who accompanied and supported me on this thrilling journey - Horia Teodorescu, Dan Kluger, Ioanna Nikolopoulou, Anja Meyer, Ricardo Palomino, Konstantin Siroki, Alex Batsis, Rudradip Biswas, Marco Monti and many, many others.

Lastly, I would like to thank my entire extended family: my parents Agata and Jerzy, my brother Kacper, my cousins Ania and Maciek, my aunts Bożena and Madzia, and my uncles Marek and Robert. You may not have had a clear idea of what is this research that I am wasting taxpayers' money on, yet you supported me nonetheless, in too many important ways to enumerate.

Notation

This section explains the basic notation used in this thesis. However, particular chapters may introduce further notational conventions that are useful in the discussion of the results derived in these chapters.

The labels $\mathbb{N}$, $\mathbb{N}_+$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ denote the sets of natural numbers, positive integers, integers, rational numbers, real numbers, and complex numbers respectively.

The letter p always denotes a prime. The set $\mathbb{F}_p$ is the finite field with p elements. $\mathbb{Z}/N\mathbb{Z}$ is the cyclic group with N elements.

$D \in \mathbb{N}_+$ is a fixed integer, and we denote elements of $\mathbb{Z}^D$ or $\mathbb{F}_p^D$ using bolded letters, e.g. $\mathbf{x} = (x_1, ..., x_D) \in \mathbb{Z}^D$.

$\vec{P}$ is a polynomial progression, defined in (1.2).

Constants are denoted by $0 < c < 1 < C$, registering dependence on parameters in subscripts when necessary. For instance, c_k is a constant between 0 and 1 depending on a parameter k. Moreover, $f \ll g$, $g \gg f$, $f = O(g)$ or $g = \Omega(f)$ means that $|f(p)| \leqslant C|g(p)|$ for sufficiently large p, or other variable when appropriate. Similarly, $f \ll_k g$, $g \gg_k f$, $f = O_k(g)$ or $g = \Omega_k(f)$ means that the implied constant depends on the parameter k. We also let $f = \Theta(g)$ if both $f \ll g$ and $g \ll f$ are satisfied. All the quantitative results that we prove hold for a fixed progression $\vec{P}$, and so constants are allowed to depend on $\vec{P}$ without this dependence being recorded explicitly.

We also say that $f = o(g)$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

For a finite, nonempty set X, the label $|X|$ and 1_X denote its cardinality and indicator function respectively. For a function $f : X \to \mathbb{C}$, the expression $\mathop{\mathbb{E}}_{x\in X} f(x) = \frac{1}{|X|} \sum_{x\in X} f(x)$ is the average of the function f on X. If X is clear from the context, we may abbreviate $\mathop{\mathbb{E}}_{x\in X} f(x)$ as $\mathbb{E} f$.

On a probability space $(X, \mathcal{X}, \mu)$, we define the inner product $\langle f, g\rangle = \int_X f\overline{g} d\mu$, the norms $\|f\|_{L^q(\mu)} = (\int_X |f|^q d\mu)^{\frac{1}{q}}$ for $1 \leqslant q < \infty$, as well as $\|f\|_{L^\infty(\mu)} = \operatorname{ess\,sup}\{|f(x)| : x \in X\}$. If $X = \mathbb{F}_p^D$, then μ is the uniform probability measure, and we write $\|f\|_{L^q} = (\mathbb{E}_{\mathbf{x}\in X} |f(\mathbf{x})|^q)^{\frac{1}{q}}$ whenever $1 \leqslant q < \infty$ and $\|f\|_\infty = \max\{|f(\mathbf{x})| : \mathbf{x} \in X\}$.

For a sub-σ-algebra $\mathcal{Y}$ of a probability space $(X, \mathcal{X}, \mu)$ and a function $f \in L^\infty(\mu)$, we denote $\mathbb{E}(f|\mathcal{Y})$ to be the conditional expectation of f with respect to $\mathcal{Y}$. In particular, $\mathcal{Z}_s = \mathcal{Z}_s(X)$ denotes the Host-Kra factor of X of degree s and is defined in Section 1.4.

We let $e(x) = e^{2\pi i x}$ be the exponential function, and

$$e_p(x) = e(x/p) = e^{2\pi i x/p}.$$

The Fourier transform of $f : \mathbb{F}_p \to \mathbb{C}$ is given by

$$\hat{f}(\alpha) = \mathop{\mathbb{E}}_{x \in \mathbb{F}_p} f(x) e_p(-\alpha x)$$

for any $\alpha \in \mathbb{Z}$.

We denote the multiplicative derivative of $f : \mathbb{F}_p \to \mathbb{C}$ to be $\Delta_h f(x) = f(x)\overline{f(x+h)}$.

$\mathcal{C} : z \mapsto \overline{z}$ is the conjugation operator.

For any integer vector $w \in \mathbb{Z}^s$, we set $|w| = |w_1| + ... + |w_s|$.

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is a respective binomial coefficient.

Inside a group G, we use 1 to denote both its identity and the trivial subgroup.

1 Background and summary of results

1.1 Introduction

Additive combinatorics studies arithmetic patterns in subsets of natural numbers, abelian groups, finite fields, and other algebraically structured sets. One of the key results in the field is the classical theorem of Szemerédi on arithmetic progressions in subsets of natural numbers.

Theorem 1.1.1 (Szemerédi theorem, [Sze75]). *Let $t \geqslant 3$ be an integer and $A \subseteq \mathbb{N}$ be a subset of positive upper density*[1]*. Then A contains a nontrivial arithmetic progression of length t, i.e. a configuration of the form*

$$(x,\ x+y,\ ...,\ x+(t-1)y) \tag{1.1}$$

for some $y \neq 0$.

Szemerédi theorem has been described as a "Rosetta stone" [Tao06c] in that its many proofs contain insights from various branches of mathematics, including graph theory [Sze75], hypergraph theory [RS04; NRS06; Gow07; Tao06b], ergodic theory [Fur77; FKO82; HK05b; Tao06a; Aus10], higher order Fourier analysis [Gow01; GT10a; GT10b] or model theory [Tow10]. As remarked in [Ara15], its appeal to mathematicians can be judged by the fact that four Fields medalists (Klaus Roth, Jean Bourgain, Timothy Gowers and Terence Tao) and two Abel prize winners (Endre Szemerédi and Hillel Furstenberg) have made important contributions to the theorem. Szemerédi theorem has also been used in discussions of philosophical flavour as an example of "good" [Tao07] or "deep" mathematics [Ara15].

Itself being a cornerstone of additive combinatorics, Szemerédi theorem generalises an earlier theorem of van der Waerden which to this day is one of

[1]The upper density of $A \subseteq [N]$ is $\overline{d}(A) = \limsup\limits_{N\to\infty} \frac{|A\cap[N]|}{N}$. Furstenberg showed in [Fur77] that the statement holds more generally for sets having positive upper Banach density, which is defined as $d^*(A) = \limsup\limits_{N-M\to\infty} \dfrac{|A\cap[M,N)|}{N-M}$.

the major results of Ramsey theory.

Theorem 1.1.2 (Van der Waerden's theorem, [Van27]). *For any $t, k \in \mathbb{N}_+$ there exists $N \in \mathbb{N}_+$ such that any k-colouring of $[N]$ contains a nontrivial arithmetic progression of length t.*

Szemerédi theorem has been generalized in many directions. One of its most remarkable extensions is the celebrated theorem of Green and Tao on arithmetic progressions in primes.

Theorem 1.1.3 (Green-Tao theorem, [GT08b]). *Let $t \geqslant 3$ be an integer. Then the set of primes contains a nontrivial arithmetic progression of length t.*

Most of our work has been connected with the following polynomial generalization of Szemerédi theorem by Bergelson and Leibman. We call a polynomial $Q \in \mathbb{R}[\mathbf{x}]$ *integral* if $Q(\mathbb{Z}^D) \subseteq \mathbb{Z}$ and $Q(0, ..., 0) = 0$. For integral polynomials $P_1, ..., P_t \in \mathbb{R}[y]$, a *(single-dimensional) polynomial progression* of length $t+1$ is a configuration of the form

$$\vec{P}(x, y) = (x,\ x + P_1(y),\ ...,\ x + P_t(y)) \tag{1.2}$$

in $\mathbb{R}[x, y]^{t+1}$. Throughout, we shall use the expression "polynomial progression" somewhat ambiguously to denote both an element of $\mathbb{R}[x, y]^{t+1}$, and an element $\vec{P}(x, y)$ of $\mathbb{Z}^{t+1}$ or $\mathbb{F}_p^{t+1}$ for some x and nonzero y. It will always be clear from the context which meaning we use at a given moment. We shall also say that a set A contains $\vec{P}(x, y)$ if $\vec{P}(x, y) \in A^{t+1}$.

Theorem 1.1.4 (Polynomial Szemerédi theorem, [BL96]). *Let $t \in \mathbb{N}_+$, $A \subseteq \mathbb{N}$ be a subset of positive upper density, and $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ be a polynomial progression. Then A contains $\vec{P}(x, y)$ for some $x \in \mathbb{Z}$ and a nonzero $y \in \mathbb{Z}$.*

Theorem 1.1.4 follows from the following result in ergodic theory. For us, a *system* is an abbreviation for an invertible measure-preserving dynamical system.

Theorem 1.1.5. *[BL96; HK05a] Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system, $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ be a polynomial progression. If $\mu(A) > 0$ for $A \in \mathcal{X}$, then*

$$\lim_{N \to \infty} \mathop{\mathbb{E}}_{n \in [N]} \mu(A \cap T^{P_1(n)} A \cap \cdots \cap T^{P_t(n)} A) > 0. \tag{1.3}$$

Theorem 1.1.4 can be derived from 1.1.5 using the following tools that allow one to transfer results from ergodic theory into combinatorics.

Theorem 1.1.6 (Furstenberg correspondence principle, [Fur77])**.** *Let $A \subseteq \mathbb{N}$, and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a polynomial progression. There exists an ergodic system $(X, \mathcal{X}, \mu, T)$ and a set $B \in \mathcal{X}$ for which $\overline{d}(A) = \mu(B)$ and*

$$\overline{d}(A \cap (A+n_1) \cap \cdots \cap (A+n_t)) \geqslant \mu(A \cap T^{n_1}A \cap \cdots \cap T^{n_t}A)$$

for all $t \in \mathbb{N}_+$ and $n_1, ..., n_t \in \mathbb{Z}$.

Theorems 1.1.4 and 1.1.5 have given rise to two lines of investigation, a finitary (combinatorial) and an infinitary (ergodic) one. Since the infinitary methods of ergodic theory yield no bounds for the size of subsets of $\mathbb{N}$ lacking (1.2), one of the outstanding problems has been to find a finitary proof for Theorem 1.1.4 that would provide an answer to the following question.

Question 1.1.7 (Bounds over integers)**.** *How big can a subset $A \subseteq [N]$ be if it lacks $\vec{P}(x,y)$ for $y \neq 0$?*

One can ask a related question over finite fields. Theorem 1.1.4 immediately implies the following finite-field version.

Theorem 1.1.8 (Polynomial Szemerédi theorem in finite fields)**.** *Let $t \in \mathbb{N}_+$, $0 < \alpha < 1$, and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a polynomial progression. There exists $p_0(\alpha) > 0$ such that for all primes $p > p_0$, each subset $A \subseteq \mathbb{F}_p$ of size at least αp contains $\vec{P}(x,y)$ for some $y \neq 0$.*

For nonlinear configurations $\vec{P}$, the question of finding bounds for subsets lacking $\vec{P}$ over finite fields is quite different from the same question over integers. This is because $\mathbb{F}_p$ contains $\Theta(p)$ images of an integral polynomial $Q \in \mathbb{R}[y]$ of degree d while $[N]$ contains only $\Theta(N^{\frac{1}{d}})$ images of Q. Thus, there are many more configurations $\vec{P}(x,y)$ in $\mathbb{F}_p$ than there are in $[N]$ for $N = p$. In particular, upper bounds for sets lacking polynomial progressions that work for $[N]$ will automatically hold for $\mathbb{F}_p$; but the converse need not be true, and some of the upper bounds in $\mathbb{F}_p$ that we have are stronger than the respective bounds in $[N]$. For instance, all subsets of $[N]$ lacking $(x,\ x+y,\ x+y^2)$ for $y \neq 0$ are known to be of size $O(N/(\log N)^c)$ for some $c > 0$ [PP20]; in $\mathbb{F}_p$, a different argument gives a bound of the form $O(p^{11/12})$ [DLS20].

Not only are bounds for nonlinear configurations likely to be different in $\mathbb{F}_p$ and $[N]$, but also $\mathbb{F}_p$ is a much nicer space to work with than $[N]$ due to its field properties. Finite fields are closed under arithmetic operations, and therefore tools from Fourier analysis and higher order Fourier analysis such as

Fourier transforms or nilsequences take simpler forms in $\mathbb{F}_p$. Moreover, some of the technical issues that show up in integers do not appear in finite fields, increasing the clarity of the arguments in $\mathbb{F}_p$. Finite fields serve therefore as a good testing ground for arguments and techniques that may later be extended to integers. This has indeed been the case with the work of Peluse [Pel19] on the polynomial Szemerédi theorem in $\mathbb{F}_p$ for linearly independent polynomials that was successfully adapted to integers [PP19; PP20; Pel20]. Working in finite fields allows us to distil the essential aspects of the problem from technicalities, and to understand what is going on behind the scenes.

One of our goals has therefore been to derive bounds in the polynomial Szemerédi theorem in finite fields. Letting $\vec{P}$ be a polynomial progression, this problem breaks down into three interconnected questions.

Question 1.1.9 (Bounds over $\mathbb{F}_p$)**.** *How big can a subset $A \subseteq \mathbb{F}_p$ be if it lacks $\vec{P}(x, y)$ for $y \neq 0$?*

Question 1.1.10 (Asymptotic count)**.** *How many configurations $\vec{P}(x, y)$ does A contain?*

The notion of Gowers norms mentioned in the following question is defined in (1.7).

Question 1.1.11 (True complexity)**.** *What is the smallest-degree Gowers norm controlling $\vec{P}$?*

The relationship between these three questions is as follows. To obtain bounds for subsets $A \subseteq \mathbb{F}_p$ lacking $\vec{P}(x, y)$ for $y \neq 0$, we first want to obtain an asymptotic count for the number of configurations $\vec{P}(x, y)$ in A. That is, we want to estimate the quantity

$$|\{(x, y) \in \mathbb{F}_p^2 : \vec{P}(x, y) \in A^{t+1}\}|. \tag{1.4}$$

If we can show that (1.4) is a sum of the main term $M(\alpha)p^2$, where $\alpha = |A|/p$, and an error term of size at most $E(p) = o(p^2)$, then on assuming that A lacks $\vec{P}(x, y)$ for $y \neq 0$, we deduce that $\alpha \leqslant s(p)$ for some nonnegative, nonincreasing function s satisfying $s(p) \to 0$ as $p \to \infty$. The quantity $s(p)p$ is therefore an upper bound for the size of subsets of $\mathbb{F}_p$ lacking $\vec{P}$. Question 1.1.9 thus reduces to Question 1.1.10.

Due to the availability of tools such as Fourier analysis, we prefer to work with the analytic expression

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} 1_A(x)1_A(x+P_1(y))\cdots 1_A(x+P_t(y)), \tag{1.5}$$

where 1_A is the indicator function of A, rather than with the combinatorial quantity (1.4). More generally, we need to understand expressions of the form

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x)f_1(x+P_1(y))\cdots f_t(x+P_t(y)) \tag{1.6}$$

for any functions $f_0,...,f_t:\mathbb{F}_p\to\mathbb{C}$ that are *1-bounded*, i.e. satisfy $\|f_i\|_\infty\leqslant 1$. The expressions like (1.6), which we shall refer to as the *counting operator* for $\vec{P}$, can be understood with the help of so-called Gowers norms. The *Gowers norm* of $f:\mathbb{F}_p\to\mathbb{C}$ of degree $s\in\mathbb{N}_+$ is defined by the formula

$$\|f\|_{U^s}=\left(\mathop{\mathbb{E}}_{x,h_1,...,h_s\in\mathbb{F}_p}\prod_{w\in\{0,1\}^s}\mathcal{C}^{|w|}f(x+w_1h_1+...+w_sh_s)\right)^{\frac{1}{2^s}}, \tag{1.7}$$

where $\mathcal{C}:z\mapsto\overline{z}$ is the conjugation operator and $|w|=w_1+...+w_s$. It was proved in [Gow01] that Gowers norms control arithmetic progressions, in the sense that

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x)f_1(x+y)\cdots f_s(x+sy)\right|\leqslant\min_{0\leqslant i\leqslant s}\|f_i\|_{U^s} \tag{1.8}$$

for all 1-bounded $f_0,...,f_s:\mathbb{F}_p\to\mathbb{C}$. Gowers norms are also known to control polynomial progressions in the following way.

Proposition 1.1.12 (Gowers norms control polynomial progressions, Proposition 2.2 of [Pel19])**.** *Let $\vec{P}\in\mathbb{R}[x,y]^t$ be a polynomial progression. For all $0\leqslant i\leqslant t$, there exist $s\in\mathbb{N}_+$ and $c>0$ such that*

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x)f_1(x+P_1(y))\cdots f_t(x+P_t(y))\right|\leqslant\|f_i\|_{U^s}^c+O(p^{-c})$$

for all 1-bounded functions $f_0,...,f_t:\mathbb{F}_p\to\mathbb{C}$.

In the light of the monotonicity property of Gowers norms

$$\|f\|_{U^1}\leqslant\|f\|_{U^2}\leqslant\|f\|_{U^3}\leqslant..., \tag{1.9}$$

derived e.g. in Section 1 of [GT08a], it is natural to ask what is the smallest-degree Gowers norm controlling a given configuration. This is useful because lower-degree Gowers norms are easier to understand than higher-degree Gowers norms. In particular, the first two Gowers norms are related to classical quantities. The U^1 seminorm is just the absolute value of the average, i.e. $\|f\|_{U^1} = |\mathbb{E}_{x\in\mathbb{F}_p} f(x)|$, and it is therefore a seminorm rather than a true norm. The U^2 norm is a genuine norm, and is related to Fourier analysis via

$$\|\hat{f}\|_\infty \leqslant \|f\|_{U^2} \leqslant \|f\|_\infty^{\frac{1}{2}} \|\hat{f}\|_\infty^{\frac{1}{2}},$$

implying that having a large U^2 norm is equivalent to having a large Fourier coefficient for all 1-bounded functions. More generally, having a large U^s norm amounts to correlating with an object called a *nilsequence* of degree $s-1$ [GT08a; GTZ11; GTZ12; Man18], which are more complicated and harder to work with for $s \geqslant 3$ than for smaller s.

We will illustrate the utility of having these simple expressions for lower-degree Gowers norms with the following example. In [Kuc21a], we show that the last term of

$$(x,\ x+y,\ x+2y,\ x+y^3) \tag{1.10}$$

is controlled by the U^1 norm. More precisely, there exists $c>0$ such that

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x)f_1(x+y)f_2(x+2y)f_3(x+y^3)\right| \leqslant \|f_3\|_{U^1}^c + O(p^{-c})$$

for all 1-bounded functions $f_0, f_1, f_2, f_3 : \mathbb{F}_p \to \mathbb{C}$. Decomposing

$$1_A = \alpha + (1_A - \alpha),$$

where once again $\alpha = |A|/p$, and using the fact that

$$\|1_A - \alpha\|_{U^1} = \left|\mathop{\mathbb{E}}_{x\in\mathbb{F}_p} 1_A(x) - \alpha\right| = 0,$$

we deduce that

$$\begin{aligned}&\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} 1_A(x)1_A(x+y)1_A(x+2y)1_A(x+y^3)\\ =&\alpha \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} 1_A(x)1_A(x+y)1_A(x+2y) + O(p^{-c}),\end{aligned}$$

relating the number of configurations (1.10) in A to the number of 3-term arithmetic progressions, which is well-understood.

Thus, Question 1.1.9 reduces to Question 1.1.10, which in many cases can be deduced from Question 1.1.11. In Section 1.2, we shall outline our results in the direction of Question 1.1.9 that have been proved in [Kuc21a], together with related results for Questions 1.1.10 and 1.1.11. In Section 1.3, we in turn discuss our results in the direction of Question 1.1.11 that we proved in [Kuc21d; Kuc21c], together with some of their consequences for Question 1.1.10.

The second line of investigation stemming from Theorem 1.1.4 is of an ergodic-theoretic nature and focuses on understanding the limiting behaviour of averages of the form

$$\mathop{\mathbb{E}}_{n\in[N]} T^{P_1(n)}f_1\cdots T^{P_t(n)}f_t \tag{1.11}$$

for $f_1, ..., f_t \in L^\infty(\mu)$ defined on a system $(X, \mathcal{X}, \mu, T)$. The strategy in dealing with such averages, initiated by Furstenberg in [Fur77], has been to find a *factor* $\mathcal{Z}$ (i.e. a T-invariant sub-σ-algebra of $\mathcal{X}$), which is complex enough so that the L^2 limit of (1.11) remains unchanged if the functions $f_1, ..., f_t$ are projected onto $\mathcal{Z}$, but has an extra algebraic structure that facilitates the study of the convergence of (1.11). We say that a factor $\mathcal{Z}\subseteq\mathcal{X}$ is *characteristic* for the L^2 convergence[2] of $\vec{P}$ at $i\in[t]$ for the system $(X, \mathcal{X}, \mu, T)$ if for all functions $f_1, ..., f_t \in L^\infty(\mu)$, we have the convergence

$$\lim_{N\to\infty}\left\|\mathop{\mathbb{E}}_{n\in[N]} T^{P_1(n)}f_1\cdots T^{P_t(n)}f_t\right\|_{L^2(\mu)} = 0$$

whenever $\mathbb{E}(f_i|\mathcal{Z}) = 0$. In [HK05a; HK05b], Host and Kra constructed a family of *factors* $(\mathcal{Z}_s)_{s\in\mathbb{N}_+}$, called henceforth *Host-Kra factors*, that are characteristic for the convergence of polynomial progressions. These factors will be discussed in-depth in Section 1.3.

Theorem 1.1.13 ([HK05a; HK05b])**.** *Let $\vec{P}\in\mathbb{R}[x,y]^{t+1}$ be a polynomial progression. For every $i\in[t]$ there exists $s\in\mathbb{N}$ such that for every system $(X, \mathcal{X}, \mu, T)$, the factor $\mathcal{Z}_s$ is characteristic for the L^2 convergence of $\vec{P}$ at i. Moreover, the L^2 limit of (1.11) exists.*

If $\mathcal{Z}_s$ satisfies the conditions of Theorem 1.1.13, then we say that $\mathcal{Z}_s$ is characteristic for the L^2 convergence of $\vec{P}$ at i (without reference to a specific system). It turns out that $\mathcal{Z}_s$ is a factor of $\mathcal{Z}_{s+1}$ for each $s\in\mathbb{N}_+$, and so one wants to find the smallest $s\in\mathbb{N}_+$ that is characteristic for the convergence of a given progression.

[2] We can analogously define characteristic factors for weak or pointwise convergence.

Question 1.1.14 (Host-Kra complexity)**.** *For a polynomial progression $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ and $i \in [t]$, what is the smallest $s \in \mathbb{N}_+$ such that $\mathcal{Z}_s$ is characteristic for the L^2 convergence of $\vec{P}$?*

Host-Kra factors turn out to be useful in the study of convergence of (1.11) because they are inverse limits of nilsystems [HK05a; HK05b], a class of systems with rich algebraic structure that we define and discuss in Section 1.3. Consequently, the average (1.11) can be approximated arbitrarily well by an expression

$$\mathop{\mathbb{E}}_{n \in [N]} T_a^{P_1(n)} \tilde{f}_1 \cdots T_a^{P_t(n)} \tilde{f}_t, \tag{1.12}$$

where $\tilde{f}_1, ..., \tilde{f}_t : G/\Gamma \to \mathbb{C}$ are continuous functions on nilmanifolds G/Γ, and $T_a x = ax$ is a *nilrotation* on G/Γ. Understanding (1.12) therefore comes down to understanding the distribution of orbits

$$(a^{P_1(n)}x, ..., a^{P_t(n)}x)_{n \in \mathbb{N}} \tag{1.13}$$

inside G^t/Γ^t. Incidentally, understanding the distribution of the orbit (1.13) inside G^t/Γ^t is also crucial for resolving Question 1.1.11, which shows another deep connection between the infinitary and finitary approaches to the polynomial Szemerédi theorem. The connection between these questions will be explained in more detail in Sections 1.3 and 1.4.

We now move on to discuss another type of problems investigated with regards to polynomial progressions. It has been proved in [BHK05] that 3- and 4-term arithmetic progressions satisfy the following property.

Theorem 1.1.15 (Lower bounds for multiple recurrence, [BHK05])**.** *Let the system $(X, \mathcal{X}, \mu, T)$ be ergodic, and $A \in \mathcal{X}$ be a set of positive measure. For every $\varepsilon > 0$, the sets*

$$\{n \in \mathbb{Z} : \mu(A \cap T^n A \cap T^{2n} A) \geqslant \mu(A)^3 - \varepsilon\}$$

and

$$\{n \in \mathbb{Z} : \mu(A \cap T^n A \cap T^{2n} A \cap T^{3n} A) \geqslant \mu(A)^4 - \varepsilon\}$$

are syndetic[3]*.*

[3] A set $A \subseteq \mathbb{Z}$ is *syndetic* if is has bounded gaps, i.e. if there exists $L > 0$ such that $|[n, n+L) \cap A| > 0$ for every $n \in \mathbb{Z}$.

Curiously, the property fails for arithmetic progression of length $t \geqslant 5$ by a construction of Ruzsa presented in an appendix to [BHK05]. In the light of Theorem 1.1.15, it is natural to ask the following.

Question 1.1.16 (Lower bounds for multiple recurrence)**.** *For what polynomial progressions $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ is it true that for every ergodic system $(X, \mathcal{X}, \mu, T)$, every set $A \in \mathcal{X}$ of positive measure, and every $\varepsilon > 0$, the set*

$$\{n \in \mathbb{Z} : \mu(A \cap T^{P_1(n)}A \cap T^{P_2(n)}A \cap \cdots \cap T^{P_t(n)}) \geqslant \mu(A)^{t+1} - \varepsilon\}$$

is syndetic?

Question 1.1.16 has a natural finite-field analogue.

Question 1.1.17 (Popular common differences)**.** *For what polynomial progressions $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ is the following true: for each $\alpha, \varepsilon > 0$ and $A \subseteq \mathbb{F}_p$ of size $|A| \geqslant \alpha p$, the set*

$$\{y \in \mathbb{F}_p : |A \cap (A + P_1(y)) \cap (A + P_2(y)) \cap \cdots \cap (A + P_t(y))| \geqslant (\alpha^{t+1} - \varepsilon)p\} \tag{1.14}$$

has $\Omega_{\alpha,\varepsilon}(p)$ elements?

An element $y \in \mathbb{F}_p$ belonging to the set (1.14) is often referred to as a *popular common difference* for $\vec{P}$.

Questions 1.1.16 and 1.1.17 are discussed in Section 1.5. The relationship between these two question is similar to the connection between Questions 1.1.11 and 1.1.14: both pairs of questions come down to resolving similar problems on nilmanifolds.

The questions that we have discussed so far can be generalized to multidimensional progressions. For $D \in \mathbb{N}_+$ and nonzero vectors $\mathbf{v}_1, ..., \mathbf{v}_t \in \mathbb{Z}^D$ and integral polynomials $P_1, ..., P_t \in \mathbb{R}[y]$, a *multidimensional polynomial progression* of length $t + 1$ is a configuration of the form

$$\vec{P}(\mathbf{x}, y) = (\mathbf{x},\ \mathbf{x} + \mathbf{v}_1 P_1(y),\ ...,\ \mathbf{x} + \mathbf{v}_t P_t(y)) \tag{1.15}$$

in $\mathbb{R}[\mathbf{x}, y]^{t+1}$, where $\mathbf{v}_1 P_1(y), ..., \mathbf{v}_t P_t(y)$ are all distinct. If $D = 1$, then the vectors $\mathbf{v}_1, ..., \mathbf{v}_t$ are nonzero scalars, and so (1.15) reduces to (1.2). If $\mathbf{v}_1, ..., \mathbf{v}_t$ are all linearly dependent, then we can similarly reduce the study of (1.15) to the study of (1.2). A nontrivial example that does not reduce to the one-dimensional case is

$$\vec{P}(x_1, x_2, y) = ((x_1,\ x_2),\ (x_1 + y,\ x_2),\ (x_1,\ x_2 + y^2)). \tag{1.16}$$

Theorem 1.1.4 has originally been proved for all multidimensional polynomial progressions.

Theorem 1.1.18 (Multidimensional polynomial Szemerédi theorem, [BL96])**.** *Let $t, D \in \mathbb{N}_+$, $A \subseteq \mathbb{Z}^D$ be a subset of positive upper density, and $\vec{P} \in \mathbb{R}[\boldsymbol{x}, y]^{t+1}$ be a polynomial progression. Then A contains $\vec{P}(\boldsymbol{x}, y)$ for some $\boldsymbol{x} \in \mathbb{Z}^D$ and a nonzero $y \in \mathbb{Z}$.*

Questions 1.1.7, 1.1.9, 1.1.10 can be generalized to multidimensional polynomial progressions as follows.

Question 1.1.19. *How big can a subset A of $[N]^D$ or $\mathbb{F}_p^D$ be if it lacks $\vec{P}(\boldsymbol{x}, y)$ for $y \neq 0$?*

Question 1.1.20. *How many configurations $\vec{P}(\boldsymbol{x}, y)$ does A contain?*

One can also generalize Question 1.1.11.

Question 1.1.21. *What norm $\|\cdot\|$ "controls" $\vec{P}$ at $i \in [t]$ in the sense that for every $\varepsilon > 0$ there exists $\delta > 0$ and $p_0 \in \mathbb{N}$ such that for all $p > p_0$ and all 1-bounded functions $f_0, ..., f_t : \mathbb{F}_p^D \to \mathbb{C}$, we have the bound*

$$\left| \mathop{\mathbb{E}}_{x \in \mathbb{F}_p^D, y \in \mathbb{F}_p} f_0(\boldsymbol{x}) f_1(\boldsymbol{x} + \boldsymbol{v}_1 P_1(y)) \cdots f_t(\boldsymbol{x} + \boldsymbol{v}_t P_t(y)) \right| < \varepsilon$$

whenever $\|f_i\| < \delta$?

The multidimensional analogue of (1.11) is an expression of the form

$$\mathop{\mathbb{E}}_{n \in [N]} T_1^{P_1(n)} f_1 \cdots T_t^{P_t(n)} f_t, \tag{1.17}$$

where $T_1, ..., T_t$ are distinct measure-preserving transformations on a probability space $(X, \mathcal{X}, \mu)$. A major result on the L^2 convergence of averages (1.17) comes from Walsh.

Theorem 1.1.22 (Walsh's theorem, [Wal12])**.** *Let $(X, \mathcal{X}, \mu)$ be a probability space, $T_1, ..., T_t$ be measure-preserving transformations on X and $P_1, ..., P_t \in \mathbb{R}[y]$ be integral polynomials. Suppose that the group generated by $T_1, ..., T_t$ is nilpotent. Then for every $f_1, ..., f_t \in L^\infty(\mu)$, the L^2 limit of (1.17) exists.*

If the transformations $T_1, ..., T_t$ do not form a nilpotent group, the limit of (1.17) need not exist [BL02].

While Theorem 1.1.22 ensures the convergence of (1.17) for a large class of transformations $T_1, ..., T_t$, it does not give much information about the nature of the limit. Even when $P_1, ..., P_t$ are distinct linear polynomials, the limit of (1.17) need not be invariant when $f_1, ..., f_t$ are projected onto some Host-Kra factors $\mathcal{Z}_s(T_1), ..., \mathcal{Z}_s(T_t)$ corresponding to the transformations $T_1, ..., T_t$. For instance, if $t = 2$, $T_1 = T^2$, $T_2 = T$ for an irrational rotation T on the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, and $P_i(n) = in$, then

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} T_1^{P_1(n)} f_1 \cdot T_2^{P_2(n)} f_2 = \lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} T^{2n}(f_1 f_2) = \mathbb{E}(f_1 f_2|\mathcal{I}(T)) = \int_{\mathbb{T}} f_1 f_2,$$

and the right hand side cannot be expressed in terms of the conditional expectations $\mathbb{E}(f_1|\mathbb{Z}_s(T_1)) = \int_{\mathbb{T}} f_1$ and $\mathbb{E}(f_2|\mathbb{Z}_s(T_2)) == \int_{\mathbb{T}} f_2$ for a general choice of $f_1, f_2 \in L^\infty(\mathbb{T})$. One therefore faces the following rather general question, which can be viewed as an ergodic analogue of Question 1.1.21 or as a multidimensional analogue of Question 1.1.14.

Question 1.1.23. *Does there exist a naturally-defined factor $\mathcal{Z}$ of $\mathcal{X}$ for $i \in [t]$ and measure-preserving transformations $T_1, ..., T_t$ on a probability space $(X, \mathcal{X}, \mu)$ such that for all functions $f_1, ..., f_t \in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty} \left\| \mathop{\mathbb{E}}_{n\in[N]} T_1^{P_1(n)} f_1 \cdots T_t^{P_t(n)} f_t \right\|_{L^2(\mu)} = 0$$

whenever $\mathbb{E}(f_i|\mathcal{Z}) = 0$?

1.2 Upper bounds in the polynomial Szemerédi theorem in finite fields

Having presented a general context for our research, we now delve deeper into Questions 1.1.7, 1.1.9 and 1.1.19 on finding bounds in single- and multidimensional variants of the Szemerédi and polynomial Szemerédi theorem over integers and finite fields.

For arithmetic progressions, such bounds exist due to the Fourier analytic work of Gowers [Gow01]. In the special case of three- and four-term arithmetic progressions, these bounds have been improved several times [Rot53; Hea87; Sze90; Bou99; Bou08; San11; San12; Blo16; Blo19; Sch20; GT09]. The current best bounds are of the following form.

Theorem 1.2.1 (Szemerédi theorem, quantitative version)**.** *Let $t \geqslant 3$ be an integer, and suppose $A \subseteq [N]$ lacks arithmetic progressions of length t. There exist constants $0 < c_t < 1$ such that*

$$|A| \ll_t \begin{cases} N/(\log N)^{1+c_3}, & t = 3 \text{ [BS20]}, \\ N/(\log N)^{c_4}, & t = 4 \text{ [GT17]}, \\ N/(\log\log N)^{c_t}, & t > 4 \text{ [Gow01]}. \end{cases}$$

For $t > 4$, we can take $c_t = 2^{-2^{t+9}}$.

For polynomial progressions containing nonlinear terms, bounds only exist in several special cases where appropriate analytic arguments have been found. A lot of work has been done for configurations of length 2 [Sár78a; Sár78b; Bal+94; Sli03; Luc06; Ric19], and the best bound comes from Rice, who showed that subsets $A \subseteq [N]$ lacking $\vec{P}(x, y) = (x,\ x + P(y))$ for any integral $P \in \mathbb{Q}[y]$ have size at most $O(N/(\log N)^{c \log\log\log\log N})$ [Ric19]. For $P(y) = y^2$, this bound has been improved to $O(N/(\log N)^{c \log\log\log N})$ by Bloom and Maynard [BM20]. For longer progressions, the first bound has been obtained by Prendiville; he proved that for any $k \in \mathbb{N}_+$, subsets $A \subseteq [N]$ lacking

$$(x,\ x + y^k,\ \ldots,\ x + (t-1)y^k) \tag{1.18}$$

for $y \neq 0$ have size at most $O_{t,k}\left(N(\log\log N)^{-c_{t,k}}\right)$ for some $c_{t,k} > 0$ [Pre17]. Prendiville and Peluse further showed that subsets $A \subseteq [N]$ lacking

$$(x,\ x + y,\ x + y^2)$$

for $y \neq 0$ have size at most $O\left(N/(\log N)^c\right)$ for some $c > 0$ [PP20], improving on an earlier result [PP19]. Finally, Peluse has derived an upper bound of the form $O(N/(\log\log N)^c)$ for subsets of $[N]$ lacking $\vec{P}$ with distinct-degree polynomials $P_1, \ldots, P_t$ [Pel20].

All the aforementioned upper bounds proved in the integer setting also hold in $\mathbb{F}_p$. However, there have been some further results in $\mathbb{F}_p$ for

$$(x,\ x + y,\ x + y^2,\ \ldots,\ x + y^t). \tag{1.19}$$

that give strictly better bounds than what is known over integers. These bounds have been obtained using analytic number theory [BC17], algebraic geometry [DLS20] and Fourier analysis [Pel18], with the most general result being the Fourier-analytic work of Peluse [Pel19].

Whereas upper bounds are obtained by Fourier analytic arguments, lower bounds come from explicit constructions of large, structured sets lacking given configurations. For 3-term arithmetic progressions, the main lower bound comes from Behrend [Beh46] and has later been slightly improved by Elkin [Elk11] to $\Omega((\log N)^{\frac{1}{4}} N e^{-C\sqrt{\log N}})$ (see also [GW10b] for a more succinct exposition). For nonlinear configurations, known bounds are of a much worse shape. Ruzsa constructed subsets of $[N]$ of size $\Omega(N^{0.733...})$ lacking $x,\ x+y^2$ with $y \neq 0$ [Ruz84]. This has later been improved by Younis, who proved the existence of subsets of $[N]$ of size $\Omega(N^{0.768...})$ lacking $x,\ x+y,\ x+y^2$ with $y \neq 0$ [You19].

In Section 1.2.1, we discuss the main result of [Pel19] and present our generalization thereof. In Section 1.2.2, we present our bounds for (1.18). Both results are contained in [Kuc21a]. Finally, in Section 1.2.3, we state our results in the multidimensional case from [Kuc21b].

1.2.1 Extending a result of Peluse

Using a Fourier-analytic argument, Peluse obtained the following bounds for subsets of $\mathbb{F}_p$ lacking linearly independent progressions.

Theorem 1.2.2 ([Pel19])**.** *Let $P_1, ..., P_t \in \mathbb{Z}[y]$ be linearly independent integral polynomials. There exists $c > 0$ such that if $A \subseteq \mathbb{F}_p$ lacks*

$$(x,\ x+P_1(y),\ ...,\ x+P_t(y)),$$

with $y \neq 0$, then $|A| = O(p^{1-c})$.

In particular, Theorem 1.2.2 gives the first bounds for the size of subsets of $\mathbb{F}_p$ lacking shifted geometric progressions (1.19). It can be deduced from the following counting result, in which we set $P_0(y) = 0$.

Theorem 1.2.3 ([Pel19])**.** *Let $P_1, ..., P_t \in \mathbb{Z}[y]$ be linearly independent integral polynomials. There exists $c > 0$ with the property that for any 1-bounded functions $f_0, ..., f_t : \mathbb{F}_p \to \mathbb{C}$, we have*

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \prod_{i=0}^{t} f_i(x+P_i(y)) = \prod_{i=0}^{t} \mathop{\mathbb{E}}_{x\in\mathbb{F}_p} f_i(x) + O(p^{-c}).$$

It is worth noting that Theorem 1.2.2 precedes any similar results for integers [PP19; PP20; Pel20]. Moreover, Theorem 1.2.2 is quantitatively stronger than the results in [PP19; PP20; Pel20] in that the upper bound

$O(p^{1-c})$ in Theorem 1.2.2 is better than the bounds $O(N/(\log N)^c)$ [PP20] and $O(N/(\log\log N)^c))$ [PP19; Pel20] in integers; finally, Theorem 1.2.2 covers all linearly independent progressions while the analogous result in integers requires the polynomials $P_1, ..., P_t$ to have distinct degrees [Pel20].

We extend Theorem 1.2.2 to a larger family of progressions.

Theorem 1.2.4 (Theorem 1 of [Kuc21a]). *Let $t, k \in \mathbb{N}_+$ and $t \geqslant 3$. There exist constants $0 < c_t < 1$ such that if $A \subseteq \mathbb{F}_p$ lacks the progression*

$$(x,\ x+y,\ ...,\ x+(t-1)y,\ x+y^t,\ ...,\ x+y^{t+k-1}) \tag{1.20}$$

for $y \neq 0$, then

$$|A| \ll_{t,k} \begin{cases} p/(\log p)^{1+c_3}, & t=3 \\ p/(\log p)^{c_4}, & t=4 \\ p/(\log\log p)^{c_t}, & t>4. \end{cases}$$

For $t > 4$, we can take $c_t = 2^{-2^{t+9}}$.

The progression (1.20) is the union of an arithmetic progression and a geometric progression, and it is the first polynomial configuration with known upper bounds where polynomials in y are neither homogeneous of the same degree, as is the case with arithmetic progression or arithmetic progression with higher-power differences, nor linearly independent. Theorem 1.2.4 follows from a more general result.

Theorem 1.2.5 (Theorem 2 of [Kuc21a]). *Let $t, k \in \mathbb{N}_+$, $t \geqslant 3$, and $p \cdot s_t(p)$ be the size of the largest subset of $\mathbb{F}_p$ lacking t-term arithmetic progressions (1.1). Suppose P_t, ..., P_{t+k-1} are integral polynomials such that*

$$a_t P_t + ... + a_{t+k-1} P_{t+k-1}$$

is a polynomial of degree at least t unless $a_t = ... = a_{t+k-1} = 0$. If $A \subseteq \mathbb{F}_p$ lacks

$$(x,\ x+y,\ ...,\ x+(t-1)y,\ x+P_t(y),\ ...,\ x+P_{t+k-1}(y)) \tag{1.21}$$

with $y \neq 0$, then

$$|A| \ll p \cdot s_t(cp^c)$$

where c and the implied constant are positive and depend on t, k, and P_t, ..., P_{t+k-1}, but not on A or p.

As of right now, the best bounds for s_t are of the form

$$s_t(p) \ll_t \begin{cases} (\log p)^{-1-c_3}, \ t = 3 \\ (\log p)^{-c_4}, \ t = 4 \\ (\log\log p)^{-c_t}, \ t > 4. \end{cases}$$

for some constants $0 < c_t < 1$, where $c_t = 2^{-2^{t+9}}$ for $t > 4$. From this general result follow the bounds in Theorem 1.2.4. What Theorem 1.2.5 is saying is that up to the values of implicit constants, our bounds are optimal in the sense that they are of the same shape as the bounds in Szemerédi theorem.

We should note that in Theorem 1 of [Kuc21a], our bound in the case $t = 3$ is different from one given in Theorem 1.2.4. This is because the bound for 3-term arithmetic progressions has been improved twice since the release of [Kuc21a] - first by Schoen [Sch20], then by Bloom and Sisask [BS20].

We prove Theorem 1.2.5 by first proving an analogue of Theorem 1.2.3, from which Theorem 1.2.5 can be deduced easily.

Theorem 1.2.6 (Theorem 3 of [Kuc21a])**.** *Let $t, k \in \mathbb{N}_+$ and $t \geqslant 3$. Suppose P_t, ..., P_{t+k-1} are integral polynomials such that*

$$a_t P_t + ... + a_{t+k-1} P_{t+k-1}$$

is a polynomial of degree at least t unless $a_t = ... = a_{t+k-1} = 0$. There exists $c > 0$ depending on $t, k, P_t, ..., P_{t+k-1}$ with the following property: if $f_0, ..., f_{t+k-1} : \mathbb{F}_p \to \mathbb{C}$ are 1-bounded functions, then

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \left(\prod_{j=0}^{t-1} f_j(x+jy)\right)\left(\prod_{j=t}^{t+k-1} f_j(x+P_j(y))\right) \tag{1.22}$$
$$= \left(\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \prod_{j=0}^{t-1} f_j(x+jy)\right)\left(\prod_{j=t}^{t+k-1} \mathop{\mathbb{E}}_{x\in\mathbb{F}_p} f_j(x)\right) + O(p^{-c}).$$

We encourage the reader to think of (1.22) probabilistically, as a "discorrelation" result: the occurrence of the polynomials P_t, ..., P_{t+k-1} takes place independently from the occurrence of t-term arithmetic progressions up to an error $O(p^{-c})$. This result generalizes Theorem 2.1 of [Pel19], which covers the case $t = 1$. In particular, we obtain the following corollary upon taking all the functions $f_0, \ldots, f_{t+k-1}$ to be the indicator function of some set $A \subseteq \mathbb{F}_p$.

Corollary 1.2.7. *Let $t, k \in \mathbb{N}_+$ and $t \geqslant 3$. Suppose $P_t, \ldots, P_{t+k-1}$ are integral polynomials such that*

$$a_t P_t + \ldots + a_{t+k-1} P_{t+k-1}$$

is a polynomial of degree at least t unless $a_t = \ldots = a_{t+k-1} = 0$. There exists $c > 0$ depending on $t, k, P_t, \ldots, P_{t+k-1}$ such that for every $A \subseteq \mathbb{F}_p$, we have

$$\left|\left\{(x,y) \in \mathbb{F}_p^2 : (x,\ x+y,\ \ldots,\ x+(t-1)y,\ x+P_t(y),\ \ldots,\ x+P_{t+k-1}(y)) \in A^{t+k}\right\}\right|$$
$$= \left|\left\{(x,y) \in \mathbb{F}_p^2 : (x,\ x+y,\ \ldots,\ x+(t-1)y) \in A^t\right\}\right| \cdot \left(\frac{|A|}{p}\right)^k + O(p^{2-c}).$$

The algebraic condition imposed on the polynomials $P_t, \ldots, P_{t+k-1}$ turns out to be necessary, as evidenced by the progression $(x,\ x+y,\ x+2y,\ x+y^2)$. The polynomial y^2 has degree 2, which is less than the length of the arithmetic progression, therefore y^2 is a linear combination of $x^2, (x+y)^2, (x+2y)^2$. Thus, the terms of this progression satisfy an algebraic relation

$$(x^2+2x) - 2(x+y)^2 + (x+2y)^2 - 2(x+y^2) = 0$$

We can use this relation to construct a counterexample to (1.22) for this configuration. Taking $f_j(u) = e_p(Q_j(u))$ for $0 \leqslant j \leqslant 3$ with

$$Q_0(u) = u^2 + 2u, \quad Q_1(u) = -2u^2, \quad Q_2(u) = u^2, \quad Q_3(u) = -2u$$

gives

$$\mathop{\mathbb{E}}_{x,y \in \mathbb{F}_p} f_0(x) f_1(x+y) f_2(x+2y) f_3(x+y^2) = 1$$

while

$$\mathop{\mathbb{E}}_{x,y \in \mathbb{F}_p} f_0(x) f_1(x+y) f_2(x+2y) \mathop{\mathbb{E}}_{x \in \mathbb{F}_p} f_3(x) = 0$$

because of the orthogonality of additive characters. It happens more generally that whenever a nonzero linear combination of $P_t,\ P_{t+1},\ \ldots,\ P_{t+k-1}$ has degree $d < t$, there exists a nontrivial polynomial relation of degree d between the linear terms $x,\ x+y,\ \ldots,\ x+(t-1)y$ and the nonlinear terms $x+P_t(y), \ldots, x+P_{t+k-1}(y)$. This relation prevents discorrelation from happening in an analogous way to what we have seen for $(x,\ x+y,\ x+2y,\ x+y^2)$. We will elaborate on the importance of algebraic relations like this in Section 1.3.

The proof of Theorem 1.2.6 involves Gowers norms. Indeed, Theorem 1.2.6 is equivalent to the following result.

Proposition 1.2.8. *Let $t, k \in \mathbb{N}$, $t \geqslant 3$ and $P_t,\ \ldots,\ P_{t+k-1}$ be integral polynomials such that*

$$a_t P_t + \ldots + a_{t+k-1} P_{t+k-1}$$

is a polynomial of degree at least t unless $a_t = \ldots = a_{t+k-1} = 0$. There exists a constant $c > 0$ for which the following holds: if $f_0, \ldots, f_{t+K-1} : \mathbb{F}_p \to \mathbb{C}$ are 1-bounded functions, then

$$\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \left(\prod_{j=0}^{t-1} f_j(x+jy) \right) \left(\prod_{j=t}^{t+k-1} f_j(x+P_j(y)) \right) \right| \leqslant \min_{t\leqslant j\leqslant t+k-1} \|f_j\|_{U^1} + O(p^{-c}).$$

Theorem 1.2.6 follows from Proposition 1.2.8 by splitting each f_j with $t \leqslant j \leqslant t+k-1$ into $f_j = \mathbb{E}\, f_j + (f_j - \mathbb{E}\, f_j)$, applying multilinearity, and observing that the U^1 norm of $f_j - \mathbb{E}\, f_j$ is 0.

Proposition 1.2.8 is proved by induction on the pairs (t,k), ordered $(t,k) < (t',k')$ when $t < t'$ or when $t = t'$ and $k < k'$. The case $(t,1)$ follows by applying the Cauchy-Schwarz inequality and a change of variables so that

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \left(\prod_{j=0}^{t-1} f_j(x+jy) \right) \left(\prod_{j=t}^{t+k-1} f_j(x+P_j(y)) \right) \right|^2 \\
&\leqslant \mathop{\mathbb{E}}_{x,y,h\in\mathbb{F}_p} \left(\prod_{j=1}^{t-1} \Delta_{jh} f_j(x+jy) \right) f_t(x+P_j(y)) \overline{f_t(x+P_j(y+h))} \\
&\leqslant \mathop{\mathbb{E}}_{h\in\mathbb{F}_p} \left| \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \left(\prod_{j=1}^{t-1} \Delta_{jh} f_j(x+(j-1)y) \right) f_t(x+P_j(y)-y) \overline{f_t(x+P_j(y+h)-y)} \right|.
\end{aligned}$$

In doing so, we have reduced the configuration

$$(x,\ x+y,\ \ldots,\ x+(t-1)y,\ x+P_t(y))$$

to the configuration

$$(x,\ x+y,\ \ldots,\ x+(t-2)y,\ x+P_t(y)-y,\ x+P_t(y+h)-y),$$

which satisfies the condition of Proposition 1.2.8 for every $h \neq 0$ (the polynomial $P_t(y+h) - y$ may not be integral because its constant term $P_t(h)$ might not be 0, but this nuisance is easy to overcome). Applying the inductive hypothesis for the case $(t-1, 2)$ therefore gives us the result.

This argument no longer works for $k > 1$, where a much longer strategy, similar to one used by Peluse in [Pel19], is needed. We illustrate the key steps that one takes in deriving (t, k) for $k > 1$. We want to show that there exists $c > 0$ such that

$$\left| \mathop{\mathbb{E}}_{x,y \in \mathbb{F}_p} \left(\prod_{j=0}^{t-1} f_j(x+jy) \right) \left(\prod_{j=t}^{t+k-1} f_j(x+P_j(y)) \right) \right| \leqslant \min_{t \leqslant j \leqslant t+k-1} \|f_j\|_{U^1} + O(p^{-c})$$

for any choice of 1-bounded functions $f_0, ..., f_{t+k-1} : \mathbb{F}_p \to \mathbb{C}$. To prove this, we examine properties of the dual function

$$F(x) = \mathop{\mathbb{E}}_{y \in \mathbb{F}_p} \prod_{j=0}^{t+k-2} f_j(x + P_j(y) - P_{t+k-1}(y)),$$

where we set $P_j(y) = jy$ for $0 \leqslant j \leqslant t-1$. The dual function is named so because

$$\mathop{\mathbb{E}}_{x,y \in \mathbb{F}_p} \prod_{j=0}^{t+k-1} f_j(x+P_j(y)) = \langle F, \overline{f_{t+k-1}} \rangle. \tag{1.23}$$

By (1.23) and an application of the Cauchy-Schwarz inequality, we can bound

$$\left| \mathop{\mathbb{E}}_{x,y \in \mathbb{F}_p} \prod_{j=0}^{t+k-1} f_j(x+P_j(y)) \right| \leqslant \|F\|_{L^2} = \mathop{\mathbb{E}}_{x,y \in \mathbb{F}_p} \prod_{j=0}^{t+k-2} f_j(x+P_j(y)) \overline{F(x+P_{t+k-1})}.$$

Together with Proposition 1.1.12, this implies that the counting operator (1.2.6) for $\vec{P}$ is controlled by some Gowers norm of F; that is, there exists $s \in \mathbb{N}_+$ and $c > 0$ independent of $f_0, ..., f_{t+k-1}$, for which

$$\left| \mathop{\mathbb{E}}_{x,y \in \mathbb{F}_p} \prod_{j=0}^{t+k-1} f_j(x+P_j(y)) \right| \leqslant \|F\|_{U^s}^c + O(p^{-c}). \tag{1.24}$$

Normally, Gowers norms satisfy the monotonicity property (1.9). For the dual norms, a partial converse is also true. The following lemma is a version of a degree-lowering result, first proved in [Pel19], adapted to our dual function F.

Lemma 1.2.9 (Degree lowering, Lemma 8 of [Kuc21a]). *For each $s' \in \mathbb{N}+$,* *there exists $c > 0$ independent of $f_0, ..., f_{t+k-1}$ such that*

$$\|F\|_{U^{s'+1}} \ll \|F\|_{U^{s'}}^c + p^{-c}.$$

Inducting on Lemma 1.2.9 and using (1.24), one can therefore show that

$$\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \prod_{j=0}^{t+k-1} f_j(x + P_j(y)) \right| \ll \|F\|_{U^1}^c + p^{-c}$$

for some $c > 0$ independent of the choice of $f_0, ..., f_{t+k-1}$. Finally, the definition of the U^1 norm of F and inductive hypothesis in the case $(t, k-1)$ implies the bound

$$\|F\|_{U^1} \leqslant \min_{t\leqslant j\leqslant t+k-2} \|f_j\|_{U^1} + O(p^{-c}). \tag{1.25}$$

Splitting $f_j = \mathbb{E}\, f_j + (f_j - \mathbb{E}\, f_j)$, using $\|f_j - \mathbb{E}\, f_j\|_{U^1}$, and applying the inductive hypothesis in the case $(t, 1)$ gives the (t, k) case of Proposition 1.2.8.

The proof of Lemma 1.2.9 uses the combinatorial tool known as *popularity principle*, to which we shall refer frequently.

Lemma 1.2.10 (Popularity principle, Exercise 1.1.4 of [TV06])**.** *Let X be a nonempty finite set and $f : X \to [0, 1]$. If $\mathbb{E}_{x\in X}\, f(x) = \varepsilon$ then the set $H = \{x \in X : f(x) \geqslant \frac{\varepsilon}{2}\}$ has at least $\frac{\varepsilon}{2}|X|$ elements.*

1.2.2 Extending a result of Prendiville

Our second result in [Kuc21a] covers progressions (1.18), i.e. arithmetic progressions of length t whose common difference is a k-th power. As mentioned earlier, it has been proved by Prendiville in [Pre17] that all subsets of $[N]$ lacking (1.18) have size at most $O_{t,k}(N/(\log\log N)^{c_{t,k}})$ for some $c_{t,k} > 0$. The constant $c_{t,k}$ is however not explicitly given. Prendiville's bound holds in the finite field setting; however, we have found a different argument that gives superior bounds in this context.

Theorem 1.2.11 (Theorem 4 of [Kuc21a])**.** *Let $t, k \in \mathbb{N}_+$ and $t \geqslant 3$. There exist constants $0 < c_t < 1$ such that if $A \subseteq \mathbb{F}_p$ lacks*

$$(x, x + y^k, ..., x + (t-1)y^k)$$

for $y \neq 0$, then

$$|A| \ll_{t,k} \begin{cases} p/(\log p)^{1+c_3}, \ t=3 \\ p/(\log p)^{c_4}, \ t=4 \\ p/(\log\log p)^{c_t}, \ t>4. \end{cases}$$

For $t>4$, we can take $c_t = 2^{-2^{t+9}}$.

The bounds in Theorem 1.2.11 are once again of the same shape as in Theorem 1.2.1. This is no coincidence, since the count of (1.18) bears a close connection to the count of t-term arithmetic progressions. The result below is a special case of Theorem 5 of [Kuc21a].

Theorem 1.2.12. *Let $t, k \in \mathbb{N}_+$ and $t \geqslant 3$. There exists $c = c_{t,k} > 0$ with the following property: if $f_0, ..., f_{t-1} : \mathbb{F}_p \to \mathbb{C}$ are 1-bounded functions, then*

$$\begin{aligned} &\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x) f_1(x+y^k)\cdots f_{t-1}(x+(t-1)y^k) \\ =&\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x) f_1(x+y)\cdots f_{t-1}(x+(t-1)y) + O(p^{-c}). \end{aligned}$$

Taking $f_0 = ... = f_{t-1} = 1_A$ for a set $A \subseteq \mathbb{F}_p$, we obtain the following estimate.

Corollary 1.2.13. *Let $t, k \in \mathbb{N}_+$, $t \geqslant 3$. There exists $c = c_{t,k} > 0$ such that for any $A \subseteq \mathbb{F}_p$, we have*

$$\begin{aligned} &\left|\left\{(x,y)\in\mathbb{F}_p^2 : (x,\ x+y^k,\ ...,\ x+(t-1)y^k) \in A^t\right\}\right| \\ =&\left|\left\{(x,y)\in\mathbb{F}_p^2 : (x,\ x+y,\ ...,\ x+(t-1)y) \in A^t\right\}\right| + O(p^{2-c}). \end{aligned}$$

The argument used to prove Theorem 1.2.12 is quite different from, and much simpler than, the argument leading to Theorem 1.2.6. For $k \in \mathbb{N}_+$, we define

$$Q_k = \{x \in \mathbb{F}_p : x = y^k \text{ for some } y \in \mathbb{F}_p\}$$

to be the set of k-th power residues. We use the fact that each nonzero element $x \in Q_k$ has $\gcd(k, p-1)$ representations of the form $x = y^k$ to rewrite

$$\begin{aligned} &\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x) f_1(x+y^k)\cdots f_{t-1}(x+(t-1)y^k) \qquad (1.26)\\ =&\gcd(k,p-1)\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x) f_1(x+y)\cdots f_{t-1}(x+(t-1)y) 1_{Q_k}(y) + O(p^{-1}). \end{aligned}$$

By a straightforward application of Cauchy-Schwarz t times, we bound

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x)f_1(x+y)\cdots f_{t-1}(x+(t-1)y)1_{Q_k}(y)\right| \leqslant \|1_{Q_k}\|_{U^t}.$$

If χ is a nontrivial multiplicative character on $\mathbb{F}_p$, we decompose

$$1_{Q_k}(y) = \frac{1}{k}(1+\chi(y)+\chi(y)^2+...+\chi(y)^{k-1})$$

using the orthogonality of characters. An algebraic argument using Weil's bound (Corollary 11.24 of [IK04]) implies that $\|\chi^i\|_{U^t} = O(p^{-c})$ for some $c>0$ unless $\gcd(k,p-1)|i$, from which it follows that

$$\begin{aligned}&\gcd(k,p-1)\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x)f_1(x+y)\cdots f_{t-1}(x+(t-1)y)1_{Q_k(y)} \\ &= \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x)f_1(x+y)\cdots f_{t-1}(x+(t-1)y) + O(p^{-c}).\end{aligned} \tag{1.27}$$

Theorem 1.2.12 can be inferred from combining (1.26) and (1.27).

1.2.3 Bounds in the multidimensional polynomial Szemerédi theorem

Compared to the polynomial Szemerédi theorem in single dimension (Theorem 1.1.4), few bounds are known in the multidimensional version (Theorem 1.1.18). The earliest works have focused on *corners*, i.e. the configuration

$$((x_1,x_2),\ (x_1+y,x_2),\ (x_1,x_2+y)). \tag{1.28}$$

Shkredov showed in [Shk06a] that subsets of $[N]\times[N]$ lacking (1.28) with $y\neq 0$ have size $O(N^2/(\log\log N)^c)$ for some $c>0$, improving on his earlier triple logarithmic bound [Shk06b]. For nonlinear configurations, the first bound is due to Han, Lacey and Yang, who showed that for distinct-degree integral polynomials $P_1,P_2\in\mathbb{Z}[y]$, subsets of $\mathbb{F}_p^2$ lacking

$$((x_1,x_2),\ (x_1+P_1(y),x_2),\ (x_1,x_2+P_2(y)))$$

with $y\neq 0$ have size at most $O(p^{2-1/16})$ [HLY21]. In [Kuc21b], we have used methods from [Pel19; Kuc21a; CFH11] to prove the following quantitative version of the multidimensional polynomial Szemerédi theorem in finite fields for distinct-degree polynomials, thus extending the aforementioned results of [HLY21].

Theorem 1.2.14 (Theorem 1.1 of [Kuc21b])**.** *Let $D,t\in\mathbb{N}_+$, $\boldsymbol{v}_1,...,\boldsymbol{v}_t\in\mathbb{Z}^D$ be* *nonzero vectors and* $P_1, ..., P_t \in \mathbb{Z}[y]$ *be integral polynomials satisfying* $\deg P_1 < ... < \deg P_t$. *There exists* $c > 0$ *and a threshold* $p_0 \in \mathbb{N}$ *such that for all primes* $p > p_0$, *each subset* $A \subseteq \mathbb{F}_p^D$ *of size at least* $\Omega(p^{D-c})$ *contains*

$$(\boldsymbol{x},\ \boldsymbol{x} + \boldsymbol{v}_1 P_1(y),\ ...,\ \boldsymbol{x} + \boldsymbol{v}_t P_t(y)) \tag{1.29}$$

for some $\boldsymbol{x} \in \mathbb{F}_p^D$ *and nonzero* $y \in \mathbb{F}_p$.

Once again, Theorem 1.2.14 can be deduced from a counting result in much the same way as Theorem 1.2.5 has been derived from Theorem 1.2.6. In the result that follows, we let $\mathbf{v}_0 = \mathbf{0}$, $P_0(y) = 0$, $V_i = \mathrm{Span}_{\mathbb{F}_p}\{\mathbf{v}_i\}$ and $\mathbb{E}(f_i|V_i)(\mathbf{x}) = \mathbb{E}_{n\in\mathbb{F}_p}(\mathbf{x} + n\mathbf{v}_i)$ be the average of f_i on the coset $\mathbf{x} + V_i$.

Theorem 1.2.15. *Let* $D, t \in \mathbb{N}_+$, $\boldsymbol{v}_1, ..., \boldsymbol{v}_t \in \mathbb{Z}^D$ *be nonzero vectors and* $P_1, ..., P_t \in \mathbb{Z}[y]$ *be integral polynomials satisfying* $0 < \deg P_1 < ... < \deg P_t$. *There exists* $c > 0$ *and a threshold* $p_0 \in \mathbb{N}$ *such that for all primes* $p > p_0$ *and all 1-bounded functions* $f_0, ..., f_t : \mathbb{F}_p^D \to \mathbb{C}$, *we have*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{F}_p^D,\, y\in\mathbb{F}_p} \prod_{i=0}^{t} f_i(\boldsymbol{x} + \boldsymbol{v}_i P_i(y)) = \mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{F}_p^D} \prod_{i=0}^{t} \mathbb{E}(f_i|V_i)(\boldsymbol{x}) + O(p^{-c}).$$

In particular, Theorem 1.2.15 implies that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_1,x_2,y\in\mathbb{F}_p} f_0(x_1,x_2)f_1(x_1+y,x_2)f_2(x_1,x_2+y^2)\\
&= \mathop{\mathbb{E}}_{\substack{x_1,x_2,\\ x_1',x_2'\in\mathbb{F}_p}} f_0(x_1,x_2)f_1(x_1',x_2)f_2(x_1,x_2') + O(p^{-c})
\end{aligned}$$

for some $c > 0$ uniformly in all 1-bounded functions $f_0, f_1, f_2 : \mathbb{F}_p^2 \to \mathbb{C}$, which has been proved in [HLY21] with an explicit constant $c = \frac{1}{8}$. It also implies a novel estimate

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\substack{x_1,x_2,x_3,\\ y\in\mathbb{F}_p}} f_0(x_1,x_2,x_3)f_1(x_1+y,x_2,x_3)f_2(x_1,x_2+y^2,x_3)f_3(x_1,x_2,x_3+y^3)\\
&= \mathop{\mathbb{E}}_{\substack{x_1,x_2,x_3,\\ x_1',x_2',x_3'\in\mathbb{F}_p}} f_0(x_1,x_2,x_3)f_1(x_1',x_2,x_3)f_2(x_1,x_2',x_3)f_3(x_1,x_2,x_3') + O(p^{-c}).
\end{aligned}$$

The proof of Theorem 1.2.15 bears strong resemblance to the proof of Theorem 1.2.6, which itself has been inspired by the proof of Theorem 1.2.3 by Peluse. One important difference is that we now use a "directional" version of Gowers norms. For $f : \mathbb{F}_p^D \to \mathbb{C}$ and $\mathbf{v} \in \mathbb{F}_p^D$, we define the *Gowers norm* of

f of *degree* s *along* $\boldsymbol{v}$ to be

$$\|f\|_{U^s(\mathbf{v})} = \left(\mathop{\mathbb{E}}_{\substack{\mathbf{x}\in\mathbb{F}_p^D,\\ h_1,\dots,h_s\in\mathbb{F}_p}} \prod_{w\in\{0,1\}^s} \mathcal{C}^{|w|} f(\mathbf{x}+\mathbf{v}(w_1h_1+\dots+w_sh_s)) \right)^{\frac{1}{2^s}}.$$

This definition is an adaptation of Host-Kra seminorms from ergodic theory corresponding to the transformation $T\mathbf{x} = \mathbf{x}+\mathbf{v}$ on $\mathbb{F}_p^D$. For example, if $D = 2$ and $\mathbf{v} = (1, 0)$, then

$$\|f\|_{U^2(\mathbf{v})} = \left(\mathop{\mathbb{E}}_{\substack{x_1,x_2,\\ h_1,h_2\in\mathbb{F}_p}} f(x_1,x_2)\overline{f(x_1+h_1,x_2)f(x_1+h_2,x_2)}f(x_1+h_1+h_2,x_2) \right)^{\frac{1}{4}}.$$

Directional Gowers norms can be used to control (1.29) in a way that should be reminiscent of Proposition 1.1.12. One difference is that for certain technical reasons, we need to be able to control the L^2 norm of

$$G_t(\mathbf{x}) = \mathop{\mathbb{E}}_{y\in\mathbb{F}_p} \prod_{i=1}^{t} f_i(\mathbf{x}+\mathbf{v}_i P_i(y)).$$

Proposition 1.2.16 (Proposition 4.1 of [Kuc21b])**.** *Let $D, t \in \mathbb{N}_+$, $\boldsymbol{v}_1, \dots, \boldsymbol{v}_t \in \mathbb{Z}^D$ be nonzero vectors and $P_1, \dots, P_t \in \mathbb{Z}[y]$ be integral polynomials satisfying $0 < \deg P_1 < \dots < \deg P_t$. There exist $s \in \mathbb{N}_+$ and $c > 0$ such that for any 1-bounded functions $f_0, \dots, f_t : \mathbb{F}_p \to \mathbb{C}$, we have the bound*

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{F}_p^D,\, y\in\mathbb{F}_p} \prod_{i=0}^{t} f_i(\boldsymbol{x}+\boldsymbol{v}_i P_i(y)) \right| \leqslant \|G_t\|_{L^2} \leqslant \|f_t\|_{U^s(\boldsymbol{v}_t)}^c + O(p^{-c}).$$

The rest of the argument goes very similarly to the proof of Theorem 1.2.6. By considering an appropriate dual function, we show that

$$\|G_t\|_{L^2} \leqslant \|f_t\|_{U^1(\mathbf{v}_t)} + O(p^{-c})$$

for some constant $c > 0$. Decomposing $f_t = \mathbb{E}(f_t|V_t) + (f_t - \mathbb{E}(f_t|V_t))$ and using the identities $\|f_t - \mathbb{E}(f_t|V_t)\|_{U^1(\mathbf{v}_t)} = 0$ and $\mathbb{E}(f_t|V_t)(\mathbf{x}+n\mathbf{v}_t) = \mathbb{E}(f_t|V_t)(\mathbf{x})$ for any $n \in \mathbb{F}_p$, we get

$$\begin{aligned}&\mathop{\mathbb{E}}_{\mathbf{x}\in\mathbb{F}_p^D,\, y\in\mathbb{F}_p} \prod_{i=0}^{t} f_i(\mathbf{x}+\mathbf{v}_i P_i(y))\\ &= \mathop{\mathbb{E}}_{\mathbf{x}\in\mathbb{F}_p^D,\, y\in\mathbb{F}_p} \left(\prod_{i=0}^{t-1} f_i(\mathbf{x}+\mathbf{v}_i P_i(y))\right) \mathbb{E}(f_t|V_t)(\mathbf{x}) + O(p^{-c}),\end{aligned}$$

from which Theorem 1.2.15 follows by induction on t.

One final point we want to make is that the assumption of distinct degrees in Proposition 1.2.16 is essential. Without this assumption, one can only get a control of the counting operator of (1.29) in terms of some box norm of f_t, and these norms are much less understood than Gowers norms. For the same reason, Proposition 1.2.16 only gives control in terms of a directional Gowers norm of f_t rather than all the functions $f_0, ..., f_t$, which is another difference compared to what one faces when dealing with single dimensional configurations (Proposition 1.1.12).

1.3 Complexity of polynomial progressions

The results that we have stated in Section 1.2 show the utility of Gowers norms in finding bounds in the polynomial Szemerédi theorem. For instance, Theorem 1.2.6 follows from Proposition 1.2.8, which asserts that progressions like (1.20) can be controlled by the U^1 norm of some of the functions. Because of the particularly simple form of the U^1 norm, being able to control a counting operator by this norm is very helpful. However, this is rarely the case, and we usually have to resort to higher-degree Gowers norms. The monotonicity property (1.9) of Gowers norms makes it desirable to find the smallest-degree Gowers norm controlling a given progression, leading naturally to the following definition, originally introduced in the works of Gowers and Wolf on systems of linear forms [GW10a; GW11a; GW11b; GW11c].

Definition 1.3.1 (True complexity). *Let $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a polynomial progression. We say that $\vec{P}$ has* true complexity s at an index $0 \leqslant i \leqslant t$, *denoted $\mathcal{T}_i(\vec{P})$, if s is the smallest natural number with the following property: for every $\varepsilon > 0$, there exist $\delta > 0$ and $p_0 \in \mathbb{N}$ such that for all primes $p > p_0$ and all 1-bounded functions $f_0, ..., f_t : \mathbb{F}_p \to \mathbb{C}$, we have*

$$\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x) f_1(x+P_1(y)) \cdots f_t(x+P_t(y)) \right| < \varepsilon$$

whenever $\|f_i\|_{U^{s+1}} < \delta$.

Proposition 1.1.12 guarantees that true complexity is well-defined for each polynomial progression.

True complexity is one of several notions of complexity of polynomial progressions that have been studied. Another two come from ergodic theory and have to do with understanding the convergence of ergodic averages (1.11). We recall that a *factor* of a system $(X, \mathcal{X}, \mu, T)$ is a T-invariant sub-σ-algebra. Equivalently, it is a system $(Y, \mathcal{Y}, \nu, S)$ equipped with a surjective measurable map $\pi : X' \to Y'$, called *factor map*, where $X' \subseteq X, Y' \subseteq Y$ are respectively $T-$ and $S-$invariant subsets of full measure and the map π satisfies $\pi \circ T = S \circ \pi$ on X' as well as $\mu \circ \pi^{-1} = \nu$. Lastly, a factor can be specified by choosing a T-invariant subalgebra of $L^\infty(\mu)$.

Definition 1.3.2 (Characteristic factors)**.** *Let* $(X, \mathcal{X}, \mu, T)$ *be a system,* $t \in \mathbb{N}_+$ *and* $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ *be a polynomial progression.*

Suppose that $1 \leqslant i \leqslant t$*. A factor* $\mathcal{Y}$ *of* $\mathcal{X}$ *is* characteristic for the L^2-convergence of $\vec{P}$ at i *if for all choices of* $f_1, ..., f_t \in L^\infty(\mu)$*, we have*

$$\lim_{N\to\infty} \left\| \mathop{\mathbb{E}}_{n\in[N]} T^{P_1(n)} f_1 \cdots T^{P_t(n)} f_t \right\|_{L^2(\mu)} = 0$$

whenever $\mathbb{E}(f_i|\mathcal{Y}) = 0$.

Similarly, suppose that $0 \leqslant i \leqslant t$*. A factor* $\mathcal{Y}$ *of* $\mathcal{X}$ *is* characteristic for the weak convergence of $\vec{P}$ at i *if for all choices of* $f_0, ..., f_t \in L^\infty(\mu)$*, we have*

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_X f_0 \cdot T^{P_1(n)} f_1 \cdots T^{P_t(n)} f_t d\mu = 0 \tag{1.30}$$

whenever $\mathbb{E}(f_i|\mathcal{Y}) = 0$.

We recall from Section 1.1 that for each system, there exists a family $(\mathcal{Z}_s)_{s\in\mathbb{N}}$ of naturally-defined factors called Host-Kra factors introduced in [HK05a; HK05b]. While we work with properties of Host-Kra factors rather than their definition, we give the definition for the sake of completeness, following the presentation in [Hos06]. We start by inductively constructing a system $(X^{[s]}, \mathcal{X}^{[s]}, \mu^{[s]}, T^{[s]})$. For $s = 0$, this is just $(X, \mathcal{X}, \mu, T)$, and so we assume that $(X^{[s]}, \mathcal{X}^{[s]}, \mu^{[s]}, T^{[s]})$ is defined for some $s \geqslant 0$. We let $X^{[s+1]} = X^{[s]} \times X^{[s]}$ be the product of 2^{s+1} copies of X, and $\mathcal{X}^{[s+1]} = \mathcal{X}^{[s]} \otimes \mathcal{X}^{[s]}$ be the product σ-algebra. We then define $(X^{[s+1]}, \mathcal{X}^{[s+1]}, \mu^{[s+1]}, T^{[s+1]})$ to be the relatively independent joining of two copies of $(X^{[s]}, \mathcal{X}^{[s]}, \mu^{[s]}, T^{[s]})$ over $\mathcal{I}^{[s]}$, the σ-algebra of $T^{[s]}$ invariant subsets of $X^{[s]}$. That is, the measure $\mu^{[s+1]}$ is characterised by the fact that for any $F_1, F_2 \in L^\infty(\mu^{[s]})$, we have

$$\int_{X^{[s+1]}} F_1(x_1)F_2(x_2)d\mu^{[s+1]}(x_1,x_2) = \int_{X^{[s]}} \mathbb{E}(F_1|\mathcal{I}^{[s]})(x)\,\mathbb{E}(F_2|\mathcal{I}^{[s]})(x)d\mu^{[s]}(x).$$

The definition of the system $(X^{[s+1]}, \mathcal{X}^{[s+1]}, \mu^{[s+1]}, T^{[s+1]})$ looks rather formidable, so it may help to unpack it in several special cases.

1. If $T^{[s]}$ is ergodic, i.e. $\mathcal{I}^{[s]}$ is trivial, then $\mu^{[s+1]} = \mu^{[s]} \times \mu^{[s]}$ is just the product measure of two copies of $\mu^{[s]}$.

2. In particular, if T is weak-mixing, then every product $T \times \cdots \times T$ is ergodic, and so $(X^{[s]}, \mathcal{X}^{[s]}, \mu^{[s]}, T^{[s]})$ is just the product system composed of 2^s copies of $(X, \mathcal{X}, \mu, T)$.

3. Let T be an irrational translation on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Then $\mu^{[s]}$ is the Haar measure on the subtorus of $\mathbb{T}^{[s]}$ that is the image of the $(s+1)$-dimensional space

   $$G^{[s]} = \mathrm{Span}_{\mathbb{R}}\{(x_w)_{w\in\{0,1\}^s} : x_w = x + w_1h_1 + ... + w_sh_s,\ x, h_1, ..., h_s \in \mathbb{R}\}$$

   inside $\mathbb{T}^{[s]} = \mathbb{T}^{\{0,1\}^s}$. Thus, for instance, $X^{[2]}$ is the image of

   $$G^{[2]} = \mathrm{Span}_{\mathbb{R}}\{(x,\ x+h_1,\ x+h_2,\ x+h_1+h_2 :\ x, h_1, h_2 \in \mathbb{R}\}$$

   in $\mathbb{T}^{[2]} = \mathbb{T}^{\{0,1\}^2}$, and

   $$\int_{X^{[2]}} F d\mu^{[2]} = \int_{\mathbb{T}^3} F(x,\ x+h_1,\ x+h_2,\ x+h_1+h_2)dxdh_1dh_2 \qquad (1.31)$$

   for any $F \in L^\infty(\mu^{[2]})$.

A *side* of $\{0,1\}^s$ is a subset of the form $\{w \in \{0,1\}^s : w_i = 0\}$ or $\{w \in \{0,1\}^s : w_i = 1\}$ for some $i \in [s]$. The measure $\mu^{[s]}$ is invariant under the *side transformations*

$$(T_\alpha^{[s]}x)_w = \begin{cases} Tx_w, & w \in \alpha \\ x_w, & w \notin \alpha, \end{cases}$$

where α is a side of $\{0,1\}^s$. We note that there are s side transformations which translate the coordinate indexed by $0 = (0, ..., 0)$, and s transformations that leave this coordinate invariant.

For instance, the integral (1.31) is invariant under the side transformations $T\times T\times Id_X\times Id_X$, $Id_X\times Id_X\times T\times T$, $T\times Id_X\times T\times Id_X$ and $Id_X\times T\times Id_X\times T$, which correspond to the sides $\{w\in\{0,1\}^2 : w_1=0\}$, $\{w\in\{0,1\}^2 : w_1=1\}$, $\{w\in\{0,1\}^2 : w_2=0\}$ and $\{w\in\{0,1\}^2 : w_2=1\}$ respectively. Of these side transformations, the second and the fourth leave the coordinate indexed by 0 invariant. We let $X^{[s]*}=X^{2^s-1}$, so that each point $x\in X^{[s]}$ is written as $x=(x_0,\tilde{x})$ for $x_0\in X$ and $\tilde{x}\in X^{[s]*}$. Letting $\mathcal{X}^{[s]*}$ be the product σ-algebra, $\mu^{[s]*}$ be the projection of $\mu^{[s]}$ onto $X^{[s]*}$ and $T^{[s]*}$ be the product of 2^s-1 copies of T, we get that $(X^{[s]},\mathcal{X}^{[s]},\mu^{[s]},T^{[s]})$ is the joining of the systems $(X,\mathcal{X},\mu,T)$ and $(X^{[s]*},\mathcal{X}^{[s]*},\mu^{[s]*},T^{[s]*})$. Denoting the s transformations of $X^{[s]}$ that leave the coordinate indexed by 0 invariant as $T_1^{[s]}$, ..., $T_s^{[s]}$, we observe that we can write them as $T_i^{[s]}=Id_X\times T_i^{[s]*}$ for some transformation $T_i^{[s]*}$ on $X^{[s]*}$. We let $\mathcal{I}^{[s]*}$ be the sub-σ-algebra of $\mathcal{X}^{[s]*}$ that is invariant under all the transformations $T_1^{[s]*},...,T_s^{[s]*}$. As proved in [HK05b], there is a bijection (up to null sets) between the sub-σ-algebra $\mathcal{I}^{[s]*}$ of $\mathcal{X}^{[s]*}$ and some sub-σ-algebra $\mathcal{Z}_s$ of X - and this is how the Host-Kra factor $\mathcal{Z}_s$ is defined.

Some of Host-Kra factors take up familiar forms: the factor $\mathcal{Z}_0$ is the σ-algebra of T-invariant sets while the factor $\mathcal{Z}_1$ is the *Kronecker factor*, the largest factor of X isomorphic to rotation on a compact abelian Lie group. For $s>1$, however, the factors $\mathcal{Z}_s$ get ever more complicated as s increases. This has to do with the important result of Host and Kra that $\mathcal{Z}_s$ is an inverse limit of *s-step nilsystems* [HK05b], and these are harder to understand and work with for $s>1$. The fact that $\mathcal{Z}_s$ is an inverse limit of nilsystems implies that a $\mathcal{Z}_s$-measurable function can be approximated arbitrarily well by continuous functions on nilmanifolds. We will explain the role of nilsystems in Section 1.4.

We have mentioned in Section 1.1 that the $\mathcal{Z}_s$ factor is a factor of $\mathcal{Z}_{s+1}$ for each $s\in\mathbb{N}_+$, and that for each polynomial progression $\vec{P}\in\mathbb{R}[x,y]^{t+1}$, some Host-Kra factor is characteristic for the L^2 convergence (Theorem 1.1.13). In particular, if $\mathcal{Z}_s$ is characteristic for the L^2 convergence of $\vec{P}$ at some index, then so is $\mathcal{Z}_{s+1}$, but the converse need not hold. This motivates the following definition, variants of which have previously appeared and been studied in [BLL07; Fra08; Lei09; Fra16].

Definition 1.3.3 (Host-Kra complexity)**.** *Let $t\in\mathbb{N}_+, 0\leqslant i\leqslant t$ and $\vec{P}\in\mathbb{R}[x,y]^{t+1}$ be a polynomial progression. The progression $\vec{P}$ has* Host-Kra complexity *s at i, denoted $\mathcal{HK}_i(\vec{P})$, if s is the smallest natural number such that the factor $\mathcal{Z}_s$ is characteristic for the weak convergence of $\vec{P}$ at i for all totally* *ergodic*[4] *systems* $(X, \mathcal{X}, \mu, T)$.

Although we have defined Host-Kra complexity in terms of weak convergence, the existence of L^2 limit (Theorem 1.1.13) and an application of the Cauchy-Schwarz inequality imply that the weak and L^2 limits coincide. One could also define Host-Kra complexity in terms of ergodic systems. However, as shown in the proof of Corollary 4.1.14, passing from the totally ergodic to the ergodic setting requires that the Host-Kra complexity of $\vec{P}$ at every index is the same as the Host-Kra complexity of a related progression

$$\vec{\tilde{P}}(x,y) = (x,\ x+\tilde{P}_1(y),\ ...,\ x+\tilde{P}_t(y))$$

for every $r \in \mathbb{N}_+$ and $0 \leqslant j < r$, where $\tilde{P}_i(y) = (P_i(r(y-1)+j) - P_i(j))/r$. We have been unable to prove this statement, therefore we restrict our definition of Host-Kra complexity to totally ergodic systems.

There is a special class of systems, called *Weyl systems* for which studying complexity has been of particular interest. This is because in the early works on Host-Kra complexity, the authors first reduced the question of finding the smallest characteristic factor to the case of Weyl systems [FK05; FK06; Fra08] or explicitly focused on studying complexity for these systems [BLL07; Lei09].

Definition 1.3.4. *A* Weyl system *is an ergodic system* $(X, \mathcal{X}, \mu, T)$ *defined by a unipotent affine transformation on a compact abelian Lie group* X. *That is,* T *is given by* $Tx = \phi(x) + a$ *for* $a \in X$ *and an automorphism* ϕ *of* X *satisfying* $(\phi - \mathrm{Id_X})^{\mathrm{s}} = 0$ *for some* $s \in \mathbb{N}_+$.

Studying the convergence of averages (1.11) over Weyl systems leads to another notion of complexity.

Definition 1.3.5 (Weyl complexity). *Let* $t \in \mathbb{N}_+$, $0 \leqslant i \leqslant t$ *and* $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ *be a polynomial progression. The progression* $\vec{P}$ *has* Weyl complexity s *at* i, *denoted* $\mathcal{W}_i(\vec{P})$, *if* s *is the smallest natural number such that the factor* $\mathcal{Z}_s$ *is characteristic for the weak convergence of* $\vec{P}$ *at* i *for all Weyl systems* $(X, \mathcal{X}, \mu, T)$.

Like Host-Kra complexity, the notion of Weyl complexity has previously appeared in several versions and under various names in a number of papers, including [BLL07; Lei09; Fra08; Fra16].

[4] A system $(X, \mathcal{X}, \mu, T)$ is *totally ergodic* if T^r is ergodic for every $r \in \mathbb{N}_+$.

For $f \in L^\infty(\mu)$, we define the *Gowers-Host-Kra seminorm* of degree s to be

$$\begin{aligned}|||f|||_s &= \left(\int_{X^{[s]}} \prod_{w\in\{0,1\}^s} \mathcal{C}^{|w|} f(x_w) d\mu^{[s]}(x)\right)^{\frac{1}{2^s}} \\ &= \left(\lim_{N_1,\ldots,N_s\to\infty} \mathop{\mathbb{E}}_{n_1\in[N_1]} \cdots \mathop{\mathbb{E}}_{n_s\in[N_s]} \int_X \prod_{w\in\{0,1\}^s} \mathcal{C}^{|w|} T^{w_1 n_1+\ldots+w_s n_s} f d\mu\right)^{\frac{1}{2^s}}.\end{aligned}$$

As the name suggests, Gowers-Host-Kra seminorms are seminorms on $L^\infty(\mu)$. However, the seminorm $|||\cdot|||_s$ is a norm when restricted to the algebra of $\mathcal{Z}_{s-1}$-measurable functions (cf. Proposition 4.2 of [Hos06]). The last statement, together with the fact $|||f|||_s = |||\,\mathbb{E}(f|\mathcal{Z}_{s-1})|||_s$, can be reformulated as

$$|||f|||_s = 0 \iff \mathbb{E}(f|\mathcal{Z}_{s-1}) = 0. \tag{1.32}$$

For the transformation $Tx = x+1$ on $X = \mathbb{F}_p$, the weak limit (1.30) becomes

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x) f_1(x+P_1(y)) \cdots f_t(x+P_t(y)), \tag{1.33}$$

and so $|||f|||_s = \|f\|_{U^s}$. Gowers-Host-Kra seminorms thus extend the notion of Gowers norms to more general probability spaces. Like Gowers norms, they satisfy the monotonicity property

$$|||f|||_1 \leqslant |||f|||_2 \leqslant |||f|||_3 \leqslant \ldots \tag{1.34}$$

for every $f \in L^\infty(\mu)$.

Finally, we define one more notion of complexity, this time a purely algebraic one.

Definition 1.3.6 (Algebraic relations and algebraic complexity)**.** *Let $t \in \mathbb{N}_+, 0 \leqslant i \leqslant t$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a polynomial progression. An* algebraic relation *of degree d satisfied by $\vec{P}$ is a tuple $(Q_0, \ldots, Q_t) \in \mathbb{Q}[u]^{t+1}$ such that*

$$Q_0(x) + Q_1(x+P_1(y)) + \ldots + Q_t(x+P_t(y)) = 0 \tag{1.35}$$

and $\max_i \deg Q_i = d$. The progression $\vec{P}$ has algebraic complexity *s at i, denoted $\mathcal{A}_i(\vec{P})$, if s is the smallest natural number with the property that for any algebraic relation $(Q_0, \ldots, Q_t)$ satisfied by $\vec{P}$, the degree of Q_i is at most s.*

For instance, the progression $(x,\ x+y,\ x+y^2,\ x+y+y^2)$ satisfies one algebraic relation (up to scaling): the relation $(u, -u, -u, u)$ of degree 1.

Conjecture 1.3.7 (The four notions of complexity are the same, Conjecture 1.9 of [Kuc21c]). *Let $t \in \mathbb{N}_+, 0 \leqslant i \leqslant t$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a polynomial progression. Then*

$$\mathcal{HK}_i(\vec{P}) = \mathcal{W}_i(\vec{P}) = \mathcal{T}_i(\vec{P}) = \mathcal{A}_i(\vec{P}) \leqslant t-1.$$

Conjecture 1.3.7 in fact consists of two statements. First, all the four notions of complexity defined so far are the same; consequently, the smallest characteristic factor or the smallest degree Gowers norm controlling a given progression depend purely on the algebraic relations between the terms of the progression. Second, complexity is bounded from above in terms of the length of the progression. More precisely, a progression of length $t+1$ cannot satisfy an algebraic relation of degree t or higher.

Various subparts of Conjecture 1.3.7 have been posed previously, and those that have not are a part of the folklore in one form or another. The question of finding the upper bound for Host-Kra complexity together with the conjecture $\mathcal{HK}_i(\vec{P}) \leqslant t-1$ for $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ has been posed in [BLL07; Lei09; Fra08; Fra16]; some of these papers also state the trivial fact that Weyl complexity is no greater than the Host-Kra complexity and surmise that they might be equal. Additionally, some of the aforementioned papers contain the observation that progressions with algebraic relations of high degree cannot have low Host-Kra complexity, however the equivalence of Host-Kra complexity and algebraic complexity is not explicitly conjectured in any of them, nor is the equivalence of Weyl and algebraic complexity. By contrast, the conjecture relating true and algebraic complexity has been explicitly posed and examined for systems of linear forms in the works of Gowers and Wolf [GW10a; GW11a; GW11b; GW11c].

It it not hard to show that $\mathcal{T}_i(\vec{P}) \geqslant \mathcal{A}_i(\vec{P})$ for any progression $\vec{P}$ and index i. Indeed, if $\mathcal{A}_i(\vec{P}) = s$, then we have an algebraic relation (1.35) with $\deg Q_i = s$, and so setting $f_j(u) = e_p(Q_j(u))$ for $0 \leqslant j \leqslant t$ gives an example of functions satisfying

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \prod_{j=0}^{t} f_j(x+P_j(y)) = \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} e_p\left(\sum_{j=0}^{t} Q_j(x+P_j(y))\right) = 1,$$

yet $\|f_j\|_{U^s} = O(p^{-c})$ for some $c > 0$. Hence $\mathcal{T}_i(\vec{P}) \geqslant s$. An adaptation of this argument to the ergodic setting gives the inequality $\mathcal{HK}_i(\vec{P}) \geqslant \mathcal{A}_i(\vec{P})$. The difficult part is to show that true and Host-Kra complexity are no greater than algebraic complexity.

In Section 11 of [Kuc21c], we prove the equivalence of Weyl and algebraic complexity.

Theorem 1.3.8 (Theorem 1.16 of [Kuc21c])**.** *Let $t \in \mathbb{N}_+, 0 \leqslant i \leqslant t$ and $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ be a polynomial progression. Then $\mathcal{W}_i(\vec{P}) = \mathcal{A}_i(\vec{P})$.*

Although the argument leading to Theorem 1.3.8 comes down to unpacking the definition of Weyl complexity accompanied by a certain amount of algebraic manipulations, we are not aware of this result being stated before.

Conjecture 1.3.7 is known to be true for arithmetic progression. The inequality (1.8) tells us that t-term arithmetic progressions have true complexity at most $t-2$. We cannot do better than that, and the counterexample comes precisely from analysing the algebraic relations between the terms of arithmetic progressions. Let $f_j(u) = e_p\left((-1)^j \binom{t-1}{j} u^{t-2}\right)$. A direct computation and standard estimates of exponential sums show that $\|f_j\|_{U^{t-2}} \ll p^{-c}$ for some $c > 0$. At the same time, we have

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x) f_1(x+y)\cdots f_{t-1}(x+(t-1)y) = 1$$

due to the algebraic relation

$$\binom{t-1}{0} x^{t-2} - \binom{t-1}{1}(x+y)^{t-2} + \ldots + (-1)^{t-1}\binom{t-1}{t-1}(x+(t-1)y)^{t-2} = 0, \tag{1.36}$$

examples of which is the relation $x - 2(x+y) + (x+2y) = 0$ for 3-term arithmetic progressions and the relation $x^2 - 3(x+y)^2 + 3(x+2y)^2 - (x+3y)^2 = 0$ satisfied by 4-term arithemtic progressions. The norm U^{t-2} cannot therefore be used to control t-term arithmetic progressions, implying that t-term arithmetic progressions have true complexity $t-2$. Straightforward adaptations of these arguments to the ergodic setting show that Host-Kra and Weyl complexities of t-term arithmetic progressions are also $t-2$. The example of arithmetic progressions thus shows that the bound in Conjecture 1.3.7 is sharp.

For more general systems of linear forms, true complexity has been studied by Gowers and Wolf [GW10a; GW11a; GW11b; GW11c], Green and Tao [GT10a], Altman [Alt21] and Manners [Man18; Man21]. The strongest results so far come from recent preprints of Altman and Manners. While Altman first proved that true and algebraic complexity coincide for all systems of linear forms [Alt21], improving on an earlier work of Green and Tao [GT10a], Manners has recently provided a quantitative proof of this statement.

Theorem 1.3.9 (Theorem 1.1.5 of [Man21])**.** *Let*

$$\Psi(\boldsymbol{n}) = (\psi_1(\boldsymbol{n}),\ \ldots,\ \psi_t(\boldsymbol{n})) \in \mathbb{Z}[\boldsymbol{n}]^t$$

be a system of linear forms in $\boldsymbol{n} \in \mathbb{Z}^D$*. There exist* $s \in \mathbb{N}_+$ *and* $c > 0$ *such that for all 1-bounded functions* $f_1, \ldots, f_t : \mathbb{F}_p \to \mathbb{C}$*, we have the bound*

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{n}\in\mathbb{F}_p^D} f_1(\psi_1(\boldsymbol{n}))\cdots f_t(\psi_t(\boldsymbol{n}))\right| \leqslant \min_i \|f_i\|_{U^s}^c.$$

A number of results has also been proved before with regards to Host-Kra complexity. The fact that for any integral polynomial P and integers $0 < a_1 < \ldots < a_t$, the progression

$$\vec{P}(x,y) = (x,\ x + a_1 P(y),\ \ldots,\ x + a_t P(y)) \tag{1.37}$$

has Host-Kra, Weyl and algebraic complexity equal to $t-1$ has been proved in [Fra08]. A straightforward adaptation of this argument to the combinatorial setting would also show that the true complexity of such progressions equals $t-1$, although this has never been explicitly carried out. If $P(y) = y^k$ for some $k \in \mathbb{N}_+$, then the fact that $\vec{P}$ has true complexity $t-1$ follows from Theorem 1.2.12 with a quantitative error term.

Conjecture 1.3.7 has been proved for linearly independent progressions, i.e. progressions $\vec{P}$ with $P_1, \ldots, P_t$ being linearly independent. The equivalence of Host-Kra, Weyl and algebraic complexity in this case follows from the works of Frantzikinakis and Kra [FK05; FK06]. The equivalence of true and algebraic complexity is a consequence of the previously cited work of Peluse [Pel19].

For progressions of the form (1.20), such as

$$\vec{P}(x,y) = (x,\ x+y,\ \ldots,\ x+(t-1)y,\ x+y^t,\ \ldots,\ x+y^{t+k-1}),$$

the equivalence of true and algebraic complexities is implied by Theorem 1.2.6 as well as the fact that the only relations satisfied by terms of such progressions involve the linear terms.

The examples that we have discussed so far - arithmetic progressions with polynomial differences, linearly independent progressions and configurations (1.20) - all fall into the class of *homogeneous progressions* defined below.

Definition 1.3.10 (Homogeneity and inhomogeneity)**.** *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a polynomial progression. An* algebraic relation $(Q_0, ..., Q_t) \in \mathbb{R}[u]^{t+1}$ *is* homogeneous *of degree d if it is of the form*

$$(Q_0(u), ..., Q_t(u)) = (a_0 u^d, ..., a_t u^d)$$

for $a_0, ..., a_t \in \mathbb{R}$, some but not all of which may be zero. Otherwise, we call it inhomogeneous*. The progression $\vec{P}$ is* homogeneous *if all the algebraic relations that it satisfies are linear combinations of its homogeneous algebraic relations, and it is called* inhomogeneous *otherwise.*

Thus, arithmetic progressions with polynomial differences, linearly independent progressions and configurations (1.20) are all examples of homogeneous progressions. For instance, linearly independent progressions are homogeneous for trivial reasons because they satisfy no algebraic relation, whereas $(x,\ x+y,\ x+2y,\ x+y^3)$ only satisfies a homogeneous relation

$$x - 2(x+y) + (x+2y) = 0. \tag{1.38}$$

An example of an inhomogeneous progression is

$$(x,\ x+y,\ x+2y,\ x+y^2); \tag{1.39}$$

it satisfies both the homogeneous relation (1.38) and the inhomogeneous relation

$$(x^2+2x) - 2(x+y)^2 + (x+2y)^2 - 2(x+y^2) = 0$$

that cannot be written down as a sum of homogeneous relations.

Our main result in [Kuc21c] is the following.

Theorem 1.3.11 (Conjecture 1.3.7 holds for homogeneous progressions, Theorem 1.11 of [Kuc21c])**.** *Let $t \in \mathbb{N}+$. If $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ is a homogeneous progression, then it satisfies Conjecture 1.3.7. That is, the notions of Host-Kra, Weyl, true and algebraic complexity coincide for homogeneous progressions, and they are all at most $t-1$.*

Theorem 1.3.11 extends our Theorems 1.2-1.5 of [Kuc21d], where we prove the equivalence of true and algebraic complexities for certain special families of homogeneous progressions. Theorem 1.3.11 also partially extends the results of [Fra08], where Host-Kra complexity has been determined for all progressions of length four, both homogeneous and inhomogeneous. In Section 1.4, we illustrate the proof of Theorem 1.3.11. Now, we briefly explain where the upper bound $t-1$ comes from. The homogeneity of $\vec{P}$ implies that if $\vec{P}$ satisfies a nontrivial algebraic relation of degree t, then it must satisfy a nontrivial homogeneous relation of this degree:

$$a_0x^t + a_1(x+P_1(y))^t + ... + a_t(x+P_t(y))^t = 0.$$

Expanding each $(x+P_i(y))^t$ with the help of the binomial formula, we get a relation $a_0 + ... + a_t = 0$ by looking at the coefficient of x^t, as well as t relations of the form

$$0 = \sum_{i=1}^t a_i \binom{t}{j} x^{t-j} P_i(y)^j = \binom{t}{j} x^{t-j} \left(a_1 P_1(y)^j + ... + a_t P_t(y)^j\right)$$

for $1 \leqslant j \leqslant t$. This gives us t equations

$$\begin{aligned}
a_1P_1(y) + ...+\ a_tP_t(y) &= 0\\
a_1P_1(y)^2 + ...+a_tP_t(y)^2 &= 0\\
\vdots \qquad\qquad\qquad &\vdots\\
a_1P_1(y)^t + ...+a_tP_t(y)^t &= 0.
\end{aligned}$$

The invertibility of the Vandermonde matrix implies that these t equations can only be satisfied if $a_1 = ... = a_t = 0$, which also implies $a_0 = 0$. Therefore, $\vec{P}$ satisfies no nontrivial homogeneous relation of degree t, which together with its homogeneity implies that $\max_i \mathcal{A}_i(\vec{P}) \leqslant t-1$.

In particular, Theorem 1.3.11 resolves Conjecture 1.3.7 for progressions of complexity 1, i.e. those progressions whose terms only satisfy linear relations. As a corollary, we can estimate the number of such progressions in subsets of finite fields.

Theorem 1.3.12 (Theorem 1.14(i) of [Kuc21c])**.** *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a polynomial progression of algebraic complexity at most 1. Suppose that $Q_1, ..., Q_d \in \mathbb{R}[y]$ are integral polynomials such that $P_i(y) = \sum_{j=1}^d a_{ij} Q_j(y)$ for $a_{ij} \in \mathbb{Z}$ for each $0 \leqslant i \leqslant t$ and $1 \leqslant j \leqslant d$. Let $L_i(y_1, ..., y_d) = \sum_{j=1}^d a_{ij} y_j$. For any 1-bounded functions $f_0, ..., f_t : \mathbb{F}_p \to \mathbb{C}$, we have*

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \prod_{i=0}^t f_i(x+P_i(y)) = \mathop{\mathbb{E}}_{x,y_1,...,y_d\in\mathbb{F}_p} \prod_{i=0}^t f_i(x+L_i(y_1,...,y_d)) + o(1),$$

where the error term $o(1)$ is taken as $p \to \infty$ over primes and does not depend on the choice of $f_0, ..., f_t$. In particular,

$$\left\{(x, y) \in \mathbb{F}_p^2 : (x,\ x + P_1(y),\ ...,\ x + P_t(y)) \in A^{t+1}\right\}$$
$$= \frac{1}{p^{d-1}} \left\{(x, y_1, ..., y_d) \in \mathbb{F}_p^{d+1} : (x,\ x + L_1(y_1, ..., y_d),\ ...,\ x + L_t(y_1, ..., y_d)) \in A^{t+1}\right\} + o(p^2)$$

for each $A \subseteq \mathbb{F}_p$, where the error term is uniform in all choices of A.

One important remark is that the error term in Theorem 1.3.12 is qualitative as opposed to Theorems 1.2.6 or 1.2.12, where we know the error to be at most polynomial in p.

In particular, Theorem 1.3.12 relates the number of progressions

$$(x,\ x + y,\ x + y^2,\ x + y + y^2)$$

to the number of linear configurations

$$(x,\ x + y,\ x + z,\ x + y + z)$$

via the equation

$$\left\{(x, y) \in \mathbb{F}_p^2 : (x,\ x + y,\ x + y^2,\ x + y + y^2) \in A^4\right\}$$
$$= \frac{1}{p} \left\{(x, y, z) \in \mathbb{F}_p^3 : (x,\ x + y,\ x + z,\ x + y + z) \in A^4\right\} + o(p^2),$$

or the number of progressions

$$(x,\ x + y^2,\ x + 2y^2,\ x + y^3,\ x + 2y^3)$$

to the number of linear configurations

$$(x,\ x + y,\ x + 2y,\ x + z,\ x + 2z)$$

by

$$\left\{(x, y) \in \mathbb{F}_p^2 : (x,\ x + y^2,\ x + 2y^2,\ x + y^3,\ x + 2y^3) \in A^5\right\}$$
$$= \frac{1}{p} \left\{(x, y, z) \in \mathbb{F}_p^3 : (x,\ x + y,\ x + 2y,\ x + z,\ x + 2z) \in A^5\right\} + o(p^2).$$

We obtain a similar result in the setting of totally ergodic systems.

Theorem 1.3.13 (Theorem 1.14(ii) of [Kuc21c]). *Let $\vec{P}$, $L_1, \ldots, L_t$ be as in Theorem 1.3.12. For any totally ergodic system $(X, \mathcal{X}, \mu, T)$ and $f_0, \ldots, f_t \in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_X \prod_{i=0}^t T^{P_i(n)} f_i d\mu = \lim_{N\to\infty} \mathop{\mathbb{E}}_{n_1,\ldots,n_d\in[N]} \int_X \prod_{i=0}^t T^{L_i(n_1,\ldots,n_d)} f_i d\mu.$$

In particular, for any $A \in \mathcal{X}$, we have

$$\begin{aligned}&\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \mu\left(A \cap T^{P_1(n)}A \cap \cdots \cap T^{P_t(n)}A\right)\\ &= \lim_{N\to\infty} \mathop{\mathbb{E}}_{n_1,\ldots,n_d\in[N]} \mu\left(A \cap T^{L_1(n_1,\ldots,n_d)}A \cap \cdots \cap T^{L_t(n_1,\ldots,n_d)}A\right).\end{aligned}$$

Inhomogeneous progressions are much harder to deal with for reasons that shall be explained in the next section. We therefore have not been able to resolve Conjecture 1.3.7 for all inhomogeneous progressions. We have, however, proved the equivalence of true and algebraic complexity for a special family that includes (1.39).

Theorem 1.3.14 (Theorem 1.10 of [Kuc21d]). *Let $t, d \in \mathbb{N}_+$ satisfy $t \geqslant 3$ and $2 \leqslant d \leqslant t-1$, and*

$$\vec{P}(x,y) = (x,\ x+y,\ \ldots,\ x+(t-1)y,\ x+y^d). \tag{1.40}$$

1. *If $d | t-1$, then (1.40) has true complexity $t-1$ at $0 \leqslant i \leqslant t-1$ and $\frac{t-1}{d}$ at t.*

2. *If $d \nmid t-1$, then (1.40) has true complexity $t-2$ at $0 \leqslant i \leqslant t-1$ and $\left\lfloor \frac{t-1}{d} \right\rfloor$ at t.*

We note that Theorem 1.3.14 holds for any $t, d \in \mathbb{N}_+$ with $d \geqslant 2$, however the configuration (1.40) is inhomogeneous precisely when $t \geqslant 3$ and $2 \leqslant d \leqslant t-1$. The reason for the inhomogeneity of (1.40) for these values of t and d is that the monomials

$$x^{kd},\ (x+y)^{kd},\ \ldots,\ (x+(t-1)y)^{kd} \tag{1.41}$$

span the space of homogeneous polynomials in x and y of degree kd whenever $kd \leqslant t-1$. Hence $y^{dk} = (y^d)^k$ with $kd \leqslant t-1$ is a linear combination of the elements (1.41). It can be inferred from this that for every $k \in \mathbb{N}_+$ satisfying

$kd \leqslant t-1$, there exists an algebraic relation

$$Q_0(x) + Q_1(x+y) + ... + Q_{t-1}(x+(t-1)y) + (x+y^d)^k = 0,$$

with $\deg Q_0 = ... = \deg Q_{t-1} = kd$, from which it follows that $\mathcal{A}_t(\vec{P}) \geqslant \left\lfloor \frac{t-1}{d} \right\rfloor$ and $\mathcal{A}_i(\vec{P}) \geqslant d\left\lfloor \frac{t-1}{d} \right\rfloor$ for $0 \leqslant i \leqslant t-1$. Due to the algebraic relation (1.36), we also have $\mathcal{A}_i(\vec{P}) \geqslant t-2$ for $0 \leqslant i \leqslant t-1$. Combining all these bounds with Theorem 1.3.14 and the observation $\mathcal{A}_i(\vec{P}) \leqslant \mathcal{T}_i(\vec{P})$ mentioned before, we deduce that $\mathcal{A}_i(\vec{P}) = \mathcal{T}_i(\vec{P})$ for all $0 \leqslant i \leqslant t$. One should be able to adapt the proof of Theorem 1.3.14 to the ergodic setting to deduce that progressions (1.40) satisfy Conjecture 1.3.7, although we have not attempted it.

We have found three arguments to prove Theorem 1.3.14 in the special case of the progression (1.39). The simplest one is based on a simple application of the Cauchy-Schwarz inequality and a change of variables that lead to the inequality

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x) f_1(x+y) f_2(x+2y) f_3(x+y^2) \right|^2 \\
&\leqslant \mathop{\mathbb{E}}_{x,y,h\in\mathbb{F}_p} \Delta_h f_1(x) \Delta_{2h} f_2(x+y) f_3(x+y^2-y) f_3(x+(y+h)^2-y) \\
&\leqslant \mathop{\mathbb{E}}_{h\in\mathbb{F}_p} \left| \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} \Delta_h f_1(x) \Delta_{2h} f_2(x+y) f_3(x+y^2-y) f_3(x+(y+h)^2-y) \right|
\end{aligned}$$

For all $h \neq 0$, the configuration

$$(x,\ x+y,\ x+y^2-y,\ x+(y+h)^2-y)$$

is homogeneous and has complexity 1, and so fixing such h and using Theorem 1.3.11, we deduce that the true complexity of (1.39) at $i = 3$ is 1. Proving that the true complexity (1.39) at $i = 0, 1, 2$ is 2 then follows from a Fourier analytic argument described in details in Section 12 of [Kuc21d]. A similar argument has been used by Frantzikinakis in [Fra08] to prove that Host-Kra complexity equals algebraic complexity for this progression.

The idea behind the argument that we have just described is that by an appropriate use of the Cauchy-Schwarz inequality and change of variables, we want to replace an inhomogeneous progression by a homogeneous one. The simple trick we used for (1.39) no longer works for $t > 3$; however, we can adapt the method just described to handle these progressions. Instead of applying the Cauchy-Schwarz inequality once, we apply it t times, getting rid of f_0 in the first application, f_1 in the second application, etc. Applying this new method to (1.39), we bound

$$\begin{aligned}&\Big|\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x)f_1(x+y)f_2(x+2y)f_3(x+y^2)\Big|^8\\ &\leqslant \mathop{\mathbb{E}}_{x,y,h_1,h_2,h_3}\prod_{w\in\{0,1\}^3}\mathcal{C}^{|w|}f_3(x+\eta_w(y,h_1,h_2,h_3)),\end{aligned}$$

where

$$\eta_w(y,h_1,h_2,h_3)=(y+w_1h_1+w_2h_2+w_3h_3)^2-w_2h_2-2w_3h_3.$$

The progression $(x+\eta_w(y,h_1,h_2,h_3))_{w\in\{0,1\}^3}$ is homogeneous, and it can be shown to have true complexity 1 at each index by adapting the proof of Theorem 1.3.11, or alternatively by a direct argument presented in Section 12 of [Kuc21d].

In Chapter 4.10, we present an alternative method of proving Theorem 1.3.14 for (1.39). This method resembles more directly the proof of Theorem 1.3.11, and it seems to have a greater potential to be extended to cover all inhomogeneous progressions. We have not been able to carry out this generalisation, though. This method is already quite complicated for (1.39), and having it extended to larger classes of inhomogeneous progressions seems to require significant new insights.

Lastly, it is perhaps worth mentioning that Gowers norms need not be the smallest naturally-defined norms controlling a given progression. For instance, the first three terms of (1.39) are controlled by the u^3 norm given by the formula

$$\|f\|_{u^s}=\max\left\{\Big|\mathop{\mathbb{E}}_{x\in\mathbb{F}_p} f(x)e_p(Q(x))\Big|:\ \deg Q=s\right\}.$$

More precisely, for each $\varepsilon>0$ there exists $\delta>0$ and a threshold $p_0\in\mathbb{N}$ such that for all primes $p>p_0$ and all 1-bounded functions $f_0,f_1,f_2:\mathbb{F}_p\to\mathbb{C}$, we have the bound

$$\Big|\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_p} f_0(x)f_1(x+y)f_2(x+2y)f_3(x+y^2)\Big|\leqslant\varepsilon$$

whenever $\|f_i\|_{u^3}\leqslant\delta$. The u^s norms, sometimes called *polynomial bias norms*, satisfy $\|f\|_{u^s}\leqslant\|f\|_{U^s}$ [GT08a], and for $s\geqslant 3$, they are not equivalent to the U^s norms over $\mathbb{F}_p$, hence having control by the u^3 norm is stronger than having a control by the U^3 norm. A similar phenomenon takes place on the ergodic theoretic side; there exists an affine factor $\mathcal{A}_2$, finer than the Host-Kra factor $\mathcal{Z}_2$, which is characteristic for (1.39) at $i = 0, 1, 2$ [Fra08].

1.4 Equidistribution on nilmanifolds

We have indicated before that the proof of Theorem 1.3.11 comes down to solving an equidistribution problem on nilmanifolds. We elaborate on this connection in this section. The theoretical framework for studying such questions has been developed in many works, including but not limited to [Fur77; HK05a; HK05b; HK18; Zie07; Lei02; Lei05a; Lei05b; Lei07; Lei09; FK05; FK06; Fra08; GT08a; GT10a; GT12; GTZ11; GTZ12; CS12]. We start by defining the most important concepts related to nilmanifolds.

Definition 1.4.1 (Filtrations). *Let G be a group. A* filtration *on G of degree s is a chain of subgroups*

$$G = G_0 = G_1 \geqslant G_2 \geqslant ... \geqslant G_s \geqslant G_{s+1} = G_{s+2} = ... = 1$$

satisfying $[G_i, G_j] \leqslant G_{i+j}$ for each $i, j \in \mathbb{N}$. We denote it as $G_\bullet = (G_i)_{i=0}^\infty$.

A natural example of filtration comes from the lower central series, given by $G_{k+1} = [G, G_k]$ for each $k > 1$, where the *commutator* of two elements $a, b \in G$ is defined as $[a, b] = a^{-1}b^{-1}ab$, and $[A, B]$ is the subgroup of G generated by all the commutators $[a, b]$ with $a \in A, b \in B$.

Definition 1.4.2 (Nilpotent groups). *A group G is s-*step nilpotent *if the lower central series filtration has degree s.*

We like to think of nilpotent groups as generalisations of abelian groups since 1-step nilpotent groups are precisely abelian groups. An example of a 2-step nilpotent group is the Heisenberg group

$$G = \begin{pmatrix} 0 & \mathbb{R} & \mathbb{R} \\ 0 & 0 & \mathbb{R} \\ 0 & 0 & 0 \end{pmatrix},$$

whose lower central series is given by $G_1 = G$, $G_2 = \begin{pmatrix} 0 & 0 & \mathbb{R} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, and $G_3 = G_4 = ... = 1$.

Definition 1.4.3 (Nilmanifolds)**.** *Let G be a nilpotent Lie group and $\Gamma \leqslant G$ be a cocompact lattice. The compact manifold $X = G/\Gamma$ is called a* nilmanifold*. We additionally say that X is a* filtered nilmanifold of degree s *if G comes equipped with a filtration $G_\bullet$ of degree s.*

From now on, we assume that G is a nilpotent Lie group and $\Gamma \leqslant G$ is a cocompact lattice. By replacing G with its universal cover if necessary, we also assume that G is simply connected[5]. There is a special class of subgroups of nilpotent groups that we care about.

Definition 1.4.4 (Rational subgroups and filtrations)**.** *A subgroup $H \leqslant G$ is* rational *if $H/(H \cap \Gamma)$ is closed in G/Γ. A filtration $G_\bullet$ is rational if G_i is a rational subgroup for each $i \in \mathbb{N}_+$.*

We assume that each filtration on a nilmanifold that we discuss is rational, from which it follows that G_i/Γ_i is a subnilmanifold of G/Γ for each $i \in \mathbb{N}_+$, where $\Gamma_i = G_i \cap \Gamma$. The assumption of the rationality of filtrations can be made for two reasons. First, the nilmanifolds that we work with in the study of true complexity come from the inverse theorem for Gowers norms [GTZ12], and the nilmanifolds in this result are endowed with rational filtrations. Second, the filtration $G_\bullet^o$ defined below that we work with in the study of Host-Kra complexity is rational as a consequence of the standard fact that the lower central series filtration is rational (see e.g. Section 1 of [GT12]).

Nilmanifolds naturally give rise to a class of dynamical systems that we use extensively.

Definition 1.4.5 (Nilsystems and nilrotations)**.** *A nilsystem is a system $(X, \mathcal{X}, \mu, T_a)$, where $X = G/\Gamma$ is a nilmanifold, $\mathcal{X}$ is the Borel σ-algebra, μ is the Haar measure, and for each $x = b\Gamma \in X$ and $a \in G$, the map T_a is given by $T_a x = ax = (ab)\Gamma$. We call T_a a* nilrotation *by the element a, and we call a ergodic if T_a is ergodic.*

The utility of nilsystems is that for each $s \in \mathbb{N}$ and each ergodic system $(X, \mathcal{X}, \mu, T)$, the Host-Kra factor $\mathcal{Z}_s$ is a uniform limit of ergodic s-step nilsystems [HK05b]. For each $\varepsilon > 0$, this gives rise to a decomposition

$$f = f_{str} + f_{sml} + f_{unf} \tag{1.42}$$

[5] For us, simple connectedness does not imply connectedness; rather, it means that each connected component is simply connected. This is consistent with how the term 'simple connectedness' is used in the literature on the ergodic theory on nilmanifolds, e.g. in [HK18].

of a function $f \in L^\infty(\mu)$. The functions $f_{str}, f_{sml}, f_{unf} \in L^\infty(\mu)$ satisfy $\mathbb{E}(f_{unf}|\mathcal{Z}_s) = 0$, $\|f_{sml}\|_{L^1(\mu)} \leqslant \varepsilon$, and $f_{str}(x) = F(a^n\pi(x))$, where $\pi : X \to Y$ is a factor map onto an ergodic s-step nilsystem $(Y, \mathcal{Y}, \nu, S)$ and F is a continuous function on Y. If T is ergodic, then so is S. Sequences of the form $\psi(n) = F(a^n y)$ are often called *nilsequences* or *basic nilsequences*, although we refrain from using this term in this section since there are several related but not equivalent definitions of nilsequences in the literature. For an ergodic system $(X, \mathcal{X}, \mu, T)$ and an integral progression $\vec{P} \in \mathbb{R}[x, y]^{t+1}$, we can use the decomposition (1.42) to approximate limits

$$\begin{aligned}&\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_X f_0(x) f_1(T^{P_1(n)}x) \cdots f_t(T^{P_t(n)}x)\, d\mu(x)\\ &= \lim_{N\to\infty} \mathop{\mathbb{E}}_{n,m\in[N]} \int_X f_0(T^m x) f_1(T^{m+P_1(n)}x) \cdots f_t(T^{m+P_t(n)}x)\, d\mu(x)\end{aligned}$$

arbitrarily well by limits of the form

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n,m\in[N]} \int_Y \tilde{f}_0(a^m b\Gamma) \tilde{f}_1(a^{m+P_1(n)} b\Gamma) \cdots \tilde{f}_t(a^{m+P_t(n)} b\Gamma)\, d\nu(b\Gamma) \tag{1.43}$$

for continuous functions $\tilde{f}_i : Y \to \mathbb{C}$ on an ergodic nilsystem $(Y, \mathcal{Y}, \nu, S)$. Each ergodic nilsystem is a disjoint union of finitely many totally ergodic nilsystems $Y_i = a^i Y_0$ (see e.g. Section 0.6 of [Lei09]), and by splitting the domain of the integration accordingly, we can assume without loss of generality that Y itself is totally ergodic. For ergodic nilsystems, total ergodicity is equivalent to connectedness (Corollary 7 and 8 in Chapter 11 of [HK18]); we therefore assume that $Y = G/\Gamma$ is a connected, ergodic nilsystem. We caution the reader that the connectedness of Y is not equivalent to the connectedness of G. For instance, the disconnected group

$$G = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix} \tag{1.44}$$

induces a connected nilmanifold $G/\Gamma \cong \mathbb{T}^2$ when quotiented by $\Gamma = G \cap GL_3(\mathbb{Z})$.

Definition 1.4.6 (Polynomial sequences)**.** *Let $D \in \mathbb{N}_+$ and G be a nilpotent Lie group with a filtration $G_\bullet$ of degree s. A* polynomial sequence $g : \mathbb{Z}^D \to G$

adapted to $G_\bullet$ is a sequence

$$g(\boldsymbol{n}) = \prod_{j=0}^{s} \prod_{|\boldsymbol{i}|=j} g_{\boldsymbol{i}}^{\binom{\boldsymbol{n}}{\boldsymbol{i}}}, \tag{1.45}$$

with the property that $g_{\boldsymbol{i}} \in G_{|\boldsymbol{i}|}$ for each $\boldsymbol{i}$, where $\binom{\boldsymbol{n}}{\boldsymbol{i}} = \binom{n_1}{i_1} \cdots \binom{n_D}{i_D}$ and $|\boldsymbol{i}| = i_1 + \ldots + i_D$. We denote the set of polynomial sequences from $\mathbb{Z}^D$ to G adapted to the filtration $G_\bullet$ as $\mathrm{poly}(\mathbb{Z}^D, G_\bullet)$.

Proposition 1.4.7 (Proposition 6.2 of [GT12])**.** *Let G be a connected, simply connected, nilpotent Lie group with a filtration $G_\bullet$. The set $\mathrm{poly}(\mathbb{Z}^D, G_\bullet)$is a group with the group operation being the multiplication of polynomial sequences.*

It is important for us to understand the equidistribution properties of polynomial sequences, by which we mean the following.

Definition 1.4.8 (Equidistribution of polynomial sequences)**.** *Let $D \in \mathbb{N}_+$ and G/Γ be a filtered nilmanifold with a filtration $G_\bullet$. A sequence $g \in \mathrm{poly}(\mathbb{Z}^D, G_\bullet)$ is equidistributed if*

$$\lim_{N\to\infty} \mathbb{E}_{\boldsymbol{n}\in[N]^D} F(g(\boldsymbol{n})\Gamma) = \int_{G/\Gamma} F$$

for any continuous function $F : G/\Gamma \to \mathbb{C}$.

By [Lei05b], understanding the equidistribution of a polynomial sequence comes down to finding its closure, which itself has a nilmanifold structure.

Proposition 1.4.9 ([Lei05b])**.** *Let $D \in \mathbb{N}_+$, G/Γ be a filtered nilmanifold with a filtration $G_\bullet$ and $g \in \mathrm{poly}(\mathbb{Z}^D, G_\bullet)$. There exists a finite group W, a rational subgroup $H \leqslant G$, points $\{x_w \in G/\Gamma : w \in W\}$ and a homomorphism $\eta : \mathbb{Z}^D \to W$, with the property that for every $w \in W$, the sequence $(g(\boldsymbol{n}))_{\boldsymbol{n}\in\eta^{-1}(w)}$ is equidistributed on the connected nilmanifold $Hx_w = \overline{(g(\boldsymbol{n})\Gamma)_{\boldsymbol{n}\in\eta^{-1}(w)}}$.*

In particular, if $\overline{(g(\boldsymbol{n})\Gamma)_{\boldsymbol{n}\in\mathbb{Z}^D}}$ is connected, then g is equidistributed on $\overline{(g(\boldsymbol{n})\Gamma)_{\boldsymbol{n}\in\mathbb{Z}^D}} = g_{\boldsymbol{0}}H/(H \cap \Gamma)$.

As an illustration, consider the sequence $g(n) = an^2$ on $G/\Gamma = \mathbb{R}/\mathbb{Z}$. If a is irrational, then $\overline{\{g(n)\Gamma\}_{n\in\mathbb{Z}}} = \mathbb{R}/\mathbb{Z}$, and Weyl's equidistribution theorem guarantees that g is equidistributed on $\mathbb{R}/\mathbb{Z}$. If $a = \frac{1}{3}$, then $g(n) = 0 \mod \mathbb{Z}$ when $n \in 3\mathbb{Z}$ and $g(n) = \frac{1}{3} \mod \mathbb{Z}$ otherwise, and so $\{g(n)\Gamma\}_{n\in\mathbb{Z}} = \{0, \frac{1}{3}\}$ inside $\mathbb{R}/\mathbb{Z}$. In this case, Proposition 1.4.9 holds with $W = \mathbb{Z}/3\mathbb{Z}$, $H = \{0\}$, $\eta(n) = n \mod 3$, $x_0 = 0$ and $x_1 = x_2 = \frac{1}{3}$.

An important class of polynomial sequences is that of irrational polynomial sequences.

Definition 1.4.10 (Irrational sequences)**.** *Let G be a connected, simply connected, nilpotent Lie group with a filtration $G_\bullet$ of degree s and a cocompact lattice Γ. For $i \in \mathbb{N}_+$, let*

$$G_i^\nabla = \langle G_{i+1}, [G_j, G_{i-j}] : 1 \leqslant j < i \rangle.$$

*An i-*th level character *is a continuous group homomorphism $\eta_i : G_i \to \mathbb{R}$ that vanishes on G_i^∇ and satisfies $\eta_i(\Gamma_i) \subseteq \mathbb{Z}$.*

An element $g_i \in G_i$ is irrational *if $\eta_i(g_i) \notin \mathbb{Z}$ unless η_i is trivial.*

A polynomial sequence $g(n) = \prod_{i=0}^{s} g_i^{\binom{n}{i}} \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ is irrational *if g_i is irrational for every $1 \leqslant i \leqslant s$.*

Irrationality is a sufficient condition for a sequence to be equidistributed by Lemma 3.7 of [GT10a]. For instance, the sequence $g(n) = a_0 + a_1 n + ... + a_s n^s$ on $\mathbb{R}$ adapted to the filtration $G_0 = ... = G_s = \mathbb{R}$, $G_{s+1} = G_{s+2} = ... = \{0\}$ is irrational if and only if $a_s \notin \mathbb{Q}$. However, it is equidistributed if and only if at least one coefficient $a_1, ..., a_s$ is irrational.

To simplify the problem further, we want to express each sequence $n \mapsto a^n b$ on G as a polynomial sequence on the connected component G^o of G adapted to the filtration $G^o_\bullet = (G^o_i)_{i\in\mathbb{N}}$ given by $G^o_i = G_i \cap G^o$. The connectedness of $X = G/\Gamma$ implies that the quotients of G and G^o give rise to the same nilmanifold; that is, $G/\Gamma = G^o/\Gamma^o$, where $\Gamma^o = \Gamma \cap G^o$. The group G^o is a connected, simply connected nilpotent Lie group, and as a result, it can be described more explicitly. In particular, there is a diffeomorphism $\psi : G^o \to \mathbb{R}^m$ for some $m \in \mathbb{N}$, called *Mal'cev coordinate map*, that satisfies $\psi(\Gamma) = \mathbb{Z}^m$ and $\psi(G^o_i) = \{0\}^{m-m_i} \times \mathbb{R}^{m_i}$ for some $m_i \in \mathbb{N}_+$. The connectedness of $X = G/\Gamma$ further implies that $G = G^o\Gamma$ (Corollary 7 of Chapter 11 of [HK18]), and so each element $x \in X$ can be realised as $x = b\Gamma$ for some $b \in G^o$. Passing from G to G^o is accomplished by the following proposition.

Proposition 1.4.11 (Proposition 3.2 of [Kuc21c])**.** *Let G be a simply connected, nilpotent Lie group, G/Γ be a connected nilmanifold, $a \in G$ be ergodic and $b \in G^o$. There exists an irrational sequence $g_b \in \mathrm{poly}(\mathbb{Z}, G^o_\bullet)$ such that $a^n b\Gamma = g_b(n)\Gamma^o$. Moreover, for a.e. $b\Gamma \in G/\Gamma$, the sequence g_b is irrational.*

To illustrate Proposition 1.4.11, let G be as in (1.44) with the lower central series filtration. We write the group in a more compact notation, as $G =$

$\mathbb{Z}\times\mathbb{R}\times\mathbb{R}$ with the group operation given by $(a,b,c)*(a',b',c')=(a+a',b+b',c+c'+ab')$. Thus, $\Gamma=\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}$, $G^o=\{0\}\times\mathbb{R}\times\mathbb{R}$ and $G_2=\{0\}\times\{0\}\times\mathbb{R}$. If $a=(k,b,c)$ and $x=(0,y,z)$, then the product a^nx equals

$$\begin{aligned}a^nx=(k,b,c)^n*(0,y,z)&=(kn,bn+y,kb\binom{n}{2}+(c+ky)n+z)\\&=(0,bn+y,kb\binom{n}{2}+(c+ky)n+z)\mod\Gamma.\end{aligned}$$

We deduce that the element a induces an ergodic nilrotation if and only if $b\notin\mathbb{Q}$. The sequence

$$g_x(n)=(bn+y,kb\binom{n}{2}+(c+ky)n+z)=(y,z)+(b,c+ky)n+(0,kb)\binom{n}{2}$$

in $G=\mathbb{R}\times\mathbb{R}$ is irrational with respect to the filtration $G_1=\mathbb{R}\times\mathbb{R}$, $G_2=\{0\}\times\mathbb{R}$, $G_3=G_4=...=\{0\}\times\{0\}$ if and only if b and kb are irrational, which - in the face of the fact that $k\in\mathbb{Z}$ - is equivalent to the irrationality of b. Thus, the ergodicity of a indeed implies the irrationality of g_x for every $x\in G^o$, confirming Proposition 1.4.11 in this simple case.

The expression (1.43) therefore becomes

$$\lim_{N\to\infty}\mathop{\mathbb{E}}_{n,m\in[N]}\int_{G^o/\Gamma^o}\tilde{f}_0(g_b(m)\Gamma^o)\tilde{f}_1(g_b(m+P_1(n))\Gamma^o)\cdots\tilde{f}_t(g_b(m+P_t(n))\Gamma^o)d\nu(b\Gamma^o).\tag{1.46}$$

Through the sequence of simplifications described above, we have arrived at the following problem.

Question 1.4.12. *Let $\vec{P}\in\mathbb{R}[x,y]^{t+1}$ be a polynomial progression and G be a nilpotent, connected, simply connected Lie group with a rational filtration $G_\bullet$ and a cocompact lattice Γ. Given an irrational polynomial sequence $g\in\mathrm{poly}(\mathbb{Z},G_\bullet)$, what is the limit of*

$$\mathop{\mathbb{E}}_{m,n\in[N]}F(g(m)\Gamma,g(m+P_1(n))\Gamma,...,g(m+P_t(n))\Gamma)$$

for a continuous function $F:G^{t+1}/\Gamma^{t+1}\to\mathbb{C}$?

For the rest of the section, we assume that G is connected in addition to the earlier assumptions that it is a simply connected, nilpotent Lie group.

By Proposition 1.4.9, Question 1.4.12 comes down to understanding the distribution of

$$g^P(m,n) = (g(m)\Gamma, g(m+P_1(n))\Gamma, ..., g(m+P_t(n))\Gamma)$$

inside G^{t+1}/Γ^{t+1}. Thus, we want to find a rational subgroup $\tilde{G} \leqslant G^{t+1}$ such that the closure of g^P is a finite union of translates of $\tilde{G}/\tilde{\Gamma}$, where $\tilde{\Gamma} = \tilde{G} \cap \Gamma$.

It turns out that as long as $\vec{P}$ is homogeneous, the structure of the group $\tilde{G}$ depends only on $\vec{P}$ and the filtration $G_\bullet$. We define a family of vector spaces

$$\mathcal{P}_k = \operatorname{Span}_{\mathbb{R}}\left\{(x^k, (x+P_1(y))^k, ..., (x+P_t(y))^k) : x, y \in \mathbb{R}\right\}$$

as well as groups

$$G^P = \langle h_i^{\vec{v}_i} : \; h_i \in G_i, \; \vec{v}_i \in \mathcal{P}_i : 1 \leqslant i \leqslant s\rangle$$

and $\Gamma^P = \Gamma^{t+1} \cap G^P$. Our main technical result in [Kuc21c] is the following equidistribution theorem.

Theorem 1.4.13 (Dichotomy between homogeneous and inhomogeneous progressions, Theorem 1.15 of [Kuc21c])**.** *Let $t \in \mathbb{N}_+$, $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a polynomial progression and G is a connected, simply-connected, nilpotent Lie group with a rational filtration $G_\bullet$ and a cocompact lattice Γ.*

1. *If $\vec{P}$ is homogeneous, then for every irrational polynomial sequence $g \in \operatorname{poly}(\mathbb{Z}, G_\bullet)$, the sequence g^P is equidistributed on G^P/Γ^P.*

2. *If $\vec{P}$ is inhomogeneous, then for every irrational polynomial sequence $g \in \operatorname{poly}(\mathbb{Z}, G_\bullet)$, the closure of g^P is a union of finitely many translates of a subnilmanifold of G^P/Γ^P. For every $\vec{P}$, we can moreover find a filtered nilmanifold G/Γ and an irrational polynomial sequence $g: \mathbb{Z} \to G$ such that g^P is equidistributed on a proper subnilmanifold of G^P/Γ^P.*

We illustrate Theorem 1.4.13 with specific examples similar to examples in Section 11 of [Kuc21d] or Section 9 of [Kuc21c]. Let $G = G_1 = \mathbb{R}^2$, $G_2 = \{0\} \times \mathbb{R}$, $G_3 = G_4 = ... = \{0\} \times \{0\}$. The sequence $g(n) = (an, bn^2)$ is adapted to the filtration $G_\bullet$, and it is irrational if and only if a and b are irrational. We identify G^4 with $\mathbb{R}^8$ through the map

$$\begin{gathered} G^4 \to \mathbb{R}^8 \\ ((x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4)) \mapsto (x_1,x_2,x_3,x_4,y_1,y_2,y_3,y_4). \end{gathered}$$

Letting $\vec{e}_1, ..., \vec{e}_8$ be the standard coordinate vectors in $\mathbb{R}^8$ and setting

$$\vec{v}_{11} = \vec{e}_1 + \vec{e}_2 + \vec{e}_3 + \vec{e}_4, \quad \vec{v}_{12} = \vec{e}_2 + 2\vec{e}_3, \quad \vec{v}_{13} = \vec{e}_4,$$
$$\vec{v}_{21} = \vec{e}_5 + \vec{e}_6 + \vec{e}_7 + \vec{e}_8, \quad \vec{v}_{22} = \vec{e}_6 + 2\vec{e}_7, \quad \vec{v}_{23} = \vec{e}_8, \quad \vec{v}_{24} = \vec{e}_6 + 4\vec{e}_7,$$

we observe that $G^P = \mathrm{Span}\{\vec{v}_{11}, \vec{v}_{12}, \vec{v}_{13}, \vec{v}_{21}, \vec{v}_{22}, \vec{v}_{23}, \vec{v}_{24}\}$ is a 7-dimensional subspace of $\mathbb{R}^8$. If

$$\vec{P}(x,y) = (x,\ x+y,\ x+2y,\ x+y^3), \tag{1.47}$$

then the sequence g^P is given by

$$g^P(x,y) = a\vec{v}_{11}x + a\vec{v}_{12}y + a\vec{v}_{13}y^3 + b\vec{v}_{21}x^2 + 2b\vec{v}_{22}xy + b\vec{v}_{23}(2xy^3 + y^6) + b\vec{v}_{24}y^4,$$

and the irrationality of a, b, linear independence of the polynomials

$$x,\ y,\ y^3,\ x^2,\ xy,\ 2xy^3 + y^6,\ y^4,$$

and Weyl's equidistribution theorem imply that g^P is equidistributed on G^P/Γ^P.

Now, we let

$$\vec{P}(x,y) = (x,\ x+y,\ x+2y,\ x+y^2). \tag{1.48}$$

This time, g^P is given by

$$g^P(x,y) = a\vec{v}_{11}x + a\vec{v}_{12}y + (a\vec{v}_{13} + b\vec{v}_{24})y^2 + b\vec{v}_{21}x^2 + 2b\vec{v}_{22}xy + b\vec{v}_{23}(2xy^2 + y^4).$$

We observe that g^P is now a linear combination of 6 linearly independent monomials

$$x,\ y,\ y^2,\ x^2,\ xy,\ 2xy^2 + y^4.$$

In particular, the coefficient of y^2 is a sum of two terms which would have been coefficients of two separate monomials for the progression (1.47). This is a consequence of the inhomogeneous algebraic relation

$$(x^2 + 2x) - 2(x+y)^2 + (x+2y)^2 - 2(x+y^2) = 0. \tag{1.49}$$

Thus, the closure of the orbit of g^P depends on the interactions between a and b. If a, b and 1 are rationally independent, then g^P is equidistributed on

G^P. If a and b are rationally dependent, then g^P is equidistributed on the 6-dimensional subspace

$$\tilde{G} = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_{11}, \vec{v}_{12}, a\vec{v}_{13} + b\vec{v}_{24}, \vec{v}_{21}, \vec{v}_{22}, \vec{v}_{23}\}. \tag{1.50}$$

Finally, if some rational linear combination of a and b is a rational number q/r in its lower terms with $r > 1$, then the closure of g^P is a union of at most r translates of a 6-dimensional subtorus of G^P/Γ^P. For instance, if $a = \sqrt{2}$ and $b = \sqrt{2} + \frac{1}{3}$, then we define

$$\tilde{G} = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_{11}, \vec{v}_{12}, \vec{v}_{13} + \vec{v}_{24}, \vec{v}_{21}, \vec{v}_{22}, \vec{v}_{23}\}, \tag{1.51}$$

and observe that the sequences g_0^P, g_1^P, g_2^P defined by $g_i^P(x,y) = g^P(x, 3y+i)$ are equidistributed on $\tilde{G}/\tilde{\Gamma}$, $\frac{1}{3}\vec{v}_{24}+\tilde{G}/\tilde{\Gamma}$ and $\frac{1}{3}\vec{v}_{24}+\tilde{G}/\tilde{\Gamma}$ respectively. In particular, for inhomogeneous progressions, the group $\tilde{G}$ depends not only on the filtration $G_\bullet$ and the progression $\vec{P}$, but also on the interactions between the coefficients of g.

Theorem 1.4.13 has the following consequence which plays an important role in the derivation of Theorem 1.3.11.

Corollary 1.4.14. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a homogeneous polynomial progression. Suppose that G is a connected, simply-connected, nilpotent Lie group with a rational filtration $G_\bullet$ and a cocompact lattice Γ. Let $0 \leqslant i \leqslant t$ be an integer and $\mathcal{A}_i(\vec{P}) = s$. If $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ is irrational and $f_0, ..., f_t : G/\Gamma \to \mathbb{C}$ are continuous functions such that f_i vanishes on almost every*[6] *coset of G_{s+1}/Γ_{s+1}, then*

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{m,n\in[N]} f_0(g(m)\Gamma) f_1(g(m+P_1(n))\Gamma) \cdots f_t(g(m+P_t(n))\Gamma) = 0.$$

Corollary 1.4.14 follows from the observation that if $\mathcal{A}_i(\vec{P}) = s$, then the basis vector $\vec{e}_i$ belongs to the subspace $\mathcal{P}_{s+1}$. Consequently, the group $G_{s+1}^{\vec{e}_i} = \{1\}^i \times G_{s+1} \times \{1\}^{t-i}$ is a subgroup of G^P. We let $\hat{G} = G^P/G_{s+1}^{\vec{e}_i}$ and $\hat{\Gamma} = \Gamma^P/(G_{s+1}^{\vec{e}_i} \cap \Gamma^P)$. By the measure disintegration theorem and the assumption

[6] With respect to the Haar measure on $G/(G_{s+1}\Gamma)$.

that f_i vanishes on almost every coset of G_{s+1}/Γ_{s+1}, we have

$$\left|\int_{G^P/\Gamma^P} f_0 \otimes \cdots \otimes f_t((u_0, ..., u_t)\Gamma^P)\ d(u_0, ..., u_t)\Gamma^P\right|$$
$$\ll \max_j \|f_j\|_{L^\infty(\nu)}^{t-1} \int_{\hat{G}/\hat{\Gamma}} \left|\int_{G_{s+1}/\Gamma_{s+1}} f_i(u_i v\Gamma_{s+1})\ dv\Gamma_{s+1}\right| d(u_0, ..., u_t)\hat{\Gamma} = 0.$$

Corollary 1.4.14 follows upon combining the argument above with Theorem 1.4.13.

We now outline how Theorem 1.4.13 can be used to derive Theorem 1.3.11 through a sequence of reductions. Using the decomposition (1.42), we approximate an arbitrary totally ergodic system with a connected ergodic nilsystem. Proposition 1.4.11 then allows us to replace a dense orbit $n \mapsto a^n b\Gamma$ on a connected nilmanifold G/Γ by an irrational sequence g_b on G^o/Γ^o. The final step is to use the fact proved by Ziegler [Zie07] that Host-Kra factors of G/Γ take a particularly simple form

$$Z_s = \frac{G}{G_{s+1}\Gamma} = \frac{G^o}{G^o_{s+1}\Gamma^o}, \tag{1.52}$$

and so the assumption $\mathbb{E}(f_i|\mathcal{Z}_s) = 0$ implies that f_i has zero average over almost every coset of G^o_{s+1}/Γ^o_{s+1}. Theorem 1.3.11 is then a straightforward consequence of Corollary 1.4.14. This argument can be modified to works for ergodic systems that are not totally ergodic; in the case $s = 0$, however, a minor technical complication forces us to replace the factor $\mathcal{Z}_0$ with the rational Kronecker factor $\mathcal{K}_{rat}$.

The argument sketched above is used to prove that the notions of Host-Kra and algebraic complexity agree for homogeneous progression. An argument showing that true and algebraic complexities also agree is very similar, except that it is carried out in a finitary setting, with quantitative errors of which one needs to take care. In the finitary argument, the Host-Kra decomposition (1.42) is replaced by the arithmetic regularity lemma (Theorem 1.2 of [GT10a]), and the statement that g^P equidistributes on G^P/Γ^P for a homogeneous progression $\vec{P}$ is replaced by its finitary analogue. In this section, we have outlined the infinitary argument because it is notationally cleaner and can be described without falling into too many technicalities; in [Kuc21c], however, we prove a finitary version of Theorem 1.4.13(i), from which we deduce the infinitary statement given in Theorem 1.4.13.

1.5 Lower bounds for multiple recurrence and popular common differences

Results from the previous sections can be used to deduce two closely connected families of results: lower bounds for multiple recurrence and popular common differences, which fall into Questions 1.1.16 and 1.1.17 respectively. We start by quoting once again the result from [BHK05] that for every ergodic system $(X, \mathcal{X}, \mu, T)$, $A \in \mathcal{X}$ of positive measure and $\varepsilon > 0$, the sets

$$\{n \in \mathbb{Z} : \mu(A \cap T^n A \cap T^{2n} A) \geqslant \mu(A)^3 - \varepsilon\} \tag{1.53}$$

and

$$\{n \in \mathbb{Z} : \mu(A \cap T^n A \cap T^{2n} A \cap T^{3n} A) \geqslant \mu(A)^4 - \varepsilon\} \tag{1.54}$$

are syndetic, i.e. have bounded gaps. Using a variant of the Furstenberg correspondence principle (Theorem 1.1.6), one can then show that for every set $A \subseteq \mathbb{Z}$ of positive upper Banach density and every $\varepsilon > 0$, the sets

$$\{n \in \mathbb{Z} : d^*(A \cap (A+n) \cap (A+2n)) \geqslant \mu(A)^3 - \varepsilon\} \tag{1.55}$$

and

$$\{n \in \mathbb{Z} : d^*(A \cap (A+n) \cap (A+2n) \cap (A+3n)) \geqslant \mu(A)^4 - \varepsilon\} \tag{1.56}$$

are syndetic.

The ergodicity assumption is necessary; Bergelson, Host and Kra present in [BHK05] an example of a nonergodic system $(X, \mathcal{X}, \mu, T)$ with the property that for every $l \geqslant 1$, there exists a set $A \in \mathcal{X}$ of positive measure for which

$$\mu(A \cap T^n A \cap T^{2n} A) \leqslant \frac{1}{2}\mu(A)^l.$$

A construction by Ruzsa attached as an appendix to [BHK05] gives an example of an ergodic system $(X, \mathcal{X}, \mu, T)$ with the property that for every $l \geqslant 1$, there exists a set $A \in \mathcal{X}$ of positive measure for which

$$\mu(A \cap T^n A \cap T^{2n} A \cap T^{3n} A \cap T^{4n} A) \leqslant \frac{1}{2}\mu(A)^l. \tag{1.57}$$

Results (1.53), (1.54), (1.55), (1.56), (1.57) have natural analogues in the finite-field setting. It has been shown in [GT10a] that for every $\alpha, \varepsilon > 0$, there exists $c > 0$ with the property that for every subset $A \subseteq \mathbb{F}_p$ with at least αp elements, the sets

$$\{y \in \mathbb{F}_p : |A \cap (A+y) \cap (A+2y)| \geqslant (\alpha^3 - \varepsilon)p\} \tag{1.58}$$

and

$$\{y \in \mathbb{F}_p : |A \cap (A+y) \cap (A+2y) \cap (A+3y)| \geqslant (\alpha^4 - \varepsilon)p\} \tag{1.59}$$

have at least cp elements. Once again, an analogous result fails for 5-term progressions.

We say that a polynomial progression $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ *has good bounds for multiple recurrence* if for every ergodic system $(X, \mathcal{X}, \mu, T)$, $A \in \mathcal{X}$ of positive measure and $\varepsilon > 0$, the set

$$\left\{n \in \mathbb{Z} : \mu(A \cap T^{P_1(n)}A \cap \cdots \cap T^{P_t(n)}A) \geqslant \mu(A)^{t+1} - \varepsilon\right\}$$

is syndetic. We say that $\vec{P}$ *has a popular common difference* if for every $\alpha, \varepsilon > 0$, there exists a threshold p_0 such that for all primes $p > p_0$ and each subset $A \subseteq \mathbb{F}_p$ with at least αp elements, there exists $y \neq 0$ satisfying

$$|A \cap (A + P_1(y)) \cap \cdots \cap (A + P_t(y))| \geqslant (\alpha^{t+1} - \varepsilon)p.$$

Finally, $\vec{P}$ *has many popular common differences* if for every $\alpha, \varepsilon > 0$, there exists a constant $c > 0$ such that for every prime p and every subset $A \subseteq \mathbb{F}_p$ with at least αp elements, the set

$$\left\{y \in \mathbb{F}_p : |A \cap (A + P_1(y)) \cap \cdots \cap (A + P_t(y))| \geqslant (\alpha^{t+1} - \varepsilon)p\right\}$$

has at least cp elements.

It is straightforward to see that having many popular common differences implies having a popular common difference. In all the cases that we analyse in this section, however, the two concepts are equivalent. In fact, we believe the following to be true.

Conjecture 1.5.1. *Let $t \geqslant 2$ be an integer and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral progression. Then $\vec{P}$ has good bounds for multiple recurrence if and only if it has a popular common difference if and only if it has many popular common differences.*

The plausibility of Conjecture 1.5.1 stems from the fact that in the known cases, the methods that one uses to prove that a given progression satisfies any of the three properties mentioned in Conjecture 1.5.1 are very similar, as are the counterexamples in the cases of progressions that fail to satisfy substatements of Conjecture 1.5.1. This is yet another example of a deep connection between questions in additive combinatorics and ergodic theory, the common ground for which is the equidistribution theory on nilmanifolds.

We start with linear configurations, where the question of classifying progressions having many popular common differences has been resolved in [BHK05; Gre05; GT10a; SSZ20].

Theorem 1.5.2. *Let $t \geqslant 2$ and $0 < a_1 < ... < a_t$ be integers. The progression*

$$\vec{P}(x,y) = (x,\ x + a_1 y,\ ...,\ x + a_t y) \tag{1.60}$$

has a popular common difference if and only if it has many popular common differences, which happens precisely in the following cases:

1. *for all a_1, a_2 when $t = 2$;*
2. *if and only if $a_1 + a_2 = a_3$ when $t = 3$.*

The proof of Theorem 1.5.2 in the case $t = 3$ carried out in [SSZ20] is computationally involved and has entailed a good deal of computer assistance.

An analogue of Theorem 1.5.2 for good bounds for multiple recurrence has not been proved, although some special cases have been resolved in [BHK05; Fra08].

We now proceed to nonlinear progressions. In [Fra08], Frantzikinakis showed that for every integral polynomial Q, the progression

$$\vec{P}(x,y) = (x,\ x + a_1 Q(y),\ ...,\ x + a_t Q(y)) \tag{1.61}$$

with $0 < a_1 < ... < a_t$ has good bounds for multiple recurrence if $t = 2$ or if $a_1 + a_2 = a_3$ in the case $t = 3$. It seems very likely that (1.61) has good bounds for multiple recurrence if and only if (1.60) does.

It follows from results in [FK05; FK06; Pel19] that all linearly independent progressions have good bounds for multiple recurrence and many popular common differences. It has been further proved in [Fra08] that all polynomial progressions of length 4 and complexity 1 have good lower bounds for multiple recurrence. We extend this result to longer configurations of complexity 1.

Theorem 1.5.3 (Theorem 1.14 of [Kuc21c])**.** *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a polynomial progression of algebraic complexity 1, satisfying the following property: there exist linearly independent integral polynomials $Q_1, ..., Q_k$ such that*

$$\{a_1 Q_1 + ... + a_k Q_k : \ a_1, ..., a_k \in \mathbb{Z}\} = \{b_1 P_1 + ... + b_t P_t : \ b_1, ..., b_t \in \mathbb{Z}\}. \tag{1.62}$$

Then $\vec{P}$ has good bounds for multiple recurrence and many popular common differences.

Theorem 1.5.3 does not cover all the cases of polynomial progressions of length 4 and algebraic complexity 1 resolved by Frantzikinakis in [Fra08] since not all such progressions satisfy the condition (1.62). Partly for this reason, we believe the condition (1.62) to be merely an artefact of the proof that could be disposed of with more careful arguments, and that Theorem 1.5.3 holds for every progression of complexity 1. Some examples of progressions that satisfy the conditions of (1.5.3) include

$$(x, \ x+y, \ x+y^2, \ x+y+y^2) \quad \text{and} \quad (x, \ x+2y^2, \ x+3y^2, \ x+y^3, \ x+2y^3).$$

Finally, we mention some results in the multidimensional case. Chu proved in [Chu11] that if $(X_1, \mathcal{X}_1, \mu_1, S_1)$ and $(X_1, \mathcal{X}_1, \mu_1, S_1)$ are ergodic systems and $X = X_1 \times X_2$ is the product system with $T_1 = S_1 \times \mathrm{Id}_{\mathrm{X}_2}$ and $T_2 = \mathrm{Id}_{\mathrm{X}_1} \times \mathrm{S}_2$, then the set

$$\{n \in \mathbb{Z} : \mu(A \cap T_1^{n_1} A \cap T_2^{n_2} A \geqslant \mu(A)^3 - \varepsilon\}$$

is syndetic for every $A \in \mathcal{X} = \mathcal{X}_1 \otimes \mathcal{X}_2$ with positive measure $\mu = \mu_1 \times \mu_2$ and every $\varepsilon > 0$. However, this is no longer the case for arbitrary commuting ergodic transformations T_1, T_2 on a probability space $(X, \mathcal{X}, \mu)$. In this case, we have to allow a worse exponent, as shown by the result below.

Theorem 1.5.4 ([Chu11])**.** *Let $(X, \mathcal{X}, \mu)$ be a probability space and T_1, T_2 be two commuting ergodic transformations. For every $A \in \mathcal{X}$ of positive measure and every $\varepsilon > 0$, the set*

$$\{n \in \mathbb{Z} : \mu(A \cap T_1^n A \cap T_2^n A) \geqslant \mu(A)^4 - \varepsilon\}$$

is syndetic.

However, for every $0 < c < 1$ there exists a probability space $(X, \mathcal{X}, \mu)$, commuting ergodic transformations T_1, T_2, and a set $A \in \mathcal{X}$ of positive measure for which

$$\{n \in \mathbb{Z} : \mu(A \cap T_1^n A \cap T_2^n A) \leqslant c\mu(A)^3\}.$$

Theorem 1.2.15 together with Lemma 1.6 of [Chu11] imply the following.

Corollary 1.5.5. *Let $t \in \mathbb{N}_+$, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_t \in \mathbb{Z}^D$ be nonzero vectors and $P_1, \ldots, P_t \in \mathbb{Z}[y]$ be integral polynomials with $0 < \deg P_1 < \ldots < \deg P_t$. Then*

$$\vec{P}(\boldsymbol{x}, y) = (\boldsymbol{x},\ \boldsymbol{x} + \boldsymbol{v}_1 P_1(y),\ \ldots,\ \boldsymbol{x} + \boldsymbol{v}_t P_t(y))$$

has many popular common differences.

Under the extra assumption that $y^{\deg P_i + 1} | P_{i+1}(y)$, it has been proved in [CFH11] that the progression from Corollary 1.5.5 has good bounds for multiple recurrence. The authors of [CFH11] conjecture that this extra assumption is not necessary.

References

[Alt21] D. Altman. "On a conjecture of Gowers and Wolf". In: *ArXiv e-prints* (2021). arXiv: `2106.15437`.

[Ara15] A. Arana. "On the depth of Szemerédi's theorem". In: *Philos. Math.* **23** (2 2015), pp. 163–176.

[Aus10] T. Austin. "Deducing the multidimensional Szemerédi theorem from an infinitary removal lemma". In: *J. Anal. Math.* **111**.1 (2010), pp. 131–150.

[Bal+94] A. Balog et al. "Difference sets without kth powers". In: *Acta Math. Hungar.* **65**.2 (1994), pp. 165–187.

[BC17] J. Bourgain and M.-C. Chang. "Nonlinear Roth type theorems in finite fields". In: *Israel J. Math.* **221** (2017), pp. 853–867.

[Beh46] F. Behrend. "On sets of integers which contain no three terms in arithmetical progression". In: *Proc. Natl. Acad. Sci. USA* **32** (12 1946), pp. 331–2.

[BHK05] V. Bergelson, B. Host, and B. Kra. "Multiple recurrence and nilsequences". In: *Invent. Math.* **160** (2005). With an appendix by I. Ruzsa, pp. 261–303.

[BL02] V. Bergelson and A. Leibman. "A nilpotent Roth theorem". In: *Invent. Math.* **147** (2002), pp. 429–470.

[BL96] V. Bergelson and A. Leibman. "Polynomial extensions of van der Waerden's and Szemerédi's theorems". In: *J. Amer. Math. Soc.* **9** (1996), pp. 725–753.

[BLL07] V. Bergelson, A. Leibman, and E. Lesigne. "Complexities of finite families of polynomials, Weyl systems, and constructions in combinatorial number theory". In: *J. Anal. Math.* **103** (2007), pp. 47–92.

[Blo16] T. Bloom. "A quantitative improvement for Roth's theorem on arithmetic progressions". In: *J. Lond. Math. Soc.* **93** (2016), pp. 643–663.

[Blo19] T. Bloom. "Logarithmic bounds for Roth's theorem via almost-periodicity". In: *Discrete Anal.* **4** (2019), 20 pp.

[BM20] T. Bloom and J. Maynard. "A new upper bound for sets with no square differences". In: *ArXiv e-prints* (2020). arXiv: `2011.13266`.

[Bou08] J. Bourgain. "Roth's theorem on progressions revisited". In: *J. Anal. Math.* **56** (104 2008), pp. 155–192.

[Bou99] J. Bourgain. "On triples in arithmetic progression". In: *Geom. Funct. Anal.* **56** (9 1999), pp. 968–984.

[BS20] T. Bloom and O. Sisask. "Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions". In: *ArXiv e-prints* (2020). arXiv: `0902.0885`.

[CFH11] Q. Chu, N. Frantzikinakis, and B. Host. "Ergodic averages of commuting transformations with distinct degree polynomial iterates". In: *Proc. Lond. Math. Soc.* **102** (2011), pp. 801–842.

[Chu11] Q. Chu. "Multiple recurrence for two commuting transformations". In: *Ergodic Theory Dynam. Systems* **31** (2011), pp. 771–792.

[CS12] P. Candela and O. Sisask. "Convergence results for systems of linear forms on cyclic groups and periodic nilsequences". In: *SIAM J. Discrete Math.* **28** (2012), pp. 786–810.

[DLS20] D. Dong, X. Li, and W. Sawin. "Improved estimates for polynomial Roth type theorems in finite fields". In: *J. Anal. Math.* **141** (2020), pp. 689–705.

[Elk11] M. Elkin. “An improved construction of progression-free sets”. In: *Israel J. Math.* **184**.1 (2011), p. 93.

[FK05] N. Frantzikinakis and B. Kra. “Polynomial averages converge to the product of integrals”. In: *Israel J. Math.* **148**.1 (2005), pp. 267–276.

[FK06] N. Frantzikinakis and B. Kra. “Ergodic averages for independent polynomials and applications”. In: *J. Lond. Math. Soc.* **74** (2006), pp. 131–142.

[FKO82] H. Furstenberg, Y. Katznelson, and D. Ornstein. “The ergodic theoretical proof of Szemerédi’s theorem”. In: *Bull. Amer. Math. Soc.* **7**.3 (1982), pp. 527–552.

[Fra08] N. Frantzikinakis. “Multiple ergodic averages for three polynomials and applications”. In: *Trans. Amer. Math. Soc.* **360**.10 (2008), pp. 5435–5475.

[Fra16] N. Frantzikinakis. “Some open problems on multiple ergodic averages”. In: *Bull. Hellenic Math. Soc.* **60** (2016), pp. 41–90.

[Fur77] H. Furstenberg. “Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions”. In: *J. Anal. Math.* **71** (1977), pp. 204–256.

[Gow01] W. T. Gowers. “A new proof of Szemerédi’s theorem”. In: *Geom. Funct. Anal.* **11**.3 (2001), pp. 465–588.

[Gow07] W. T. Gowers. “Hypergraph regularity and the multidimensional Szemerédi theorem”. In: *Ann. of Math.* **166**.3 (2007), pp. 897–946.

[Gre05] B. Green. “A Szemerédi-type regularity lemma in abelian groups, with applications”. In: *Geom. Funct. Anal.* **15** (2005), pp. 340–376.

[GT08a] B. Green and T. Tao. “An inverse theorem for the Gowers $U^3(G)$ norm”. In: *Proc. Edinb. Math. Soc.* **51**.1 (2008), pp. 73–153.

[GT08b] B. Green and T. Tao. “The primes contain arbitrarily long arithmetic progressions”. In: *Ann. of Math.* **167**.2 (2008), pp. 481–547.

[GT09] B. Green and T. Tao. “New bounds for Szemerédi’s theorem, II: a new bound for $r_4(N)$”. In: *Analytic number theory: essays in honour of Klaus Roth.* Cambridge UP, 2009, pp. 180–204.

[GT10a] B. Green and T. Tao. "An arithmetic regularity lemma, an associated counting lemma, and applications". In: *An irregular mind. Szemerédi is 70.* Vol. **21**. Bolyai Soc. Math. Stud., 2010, pp. 261–334.

[GT10b] B. Green and T. Tao. "Yet another proof of Szemerédi's theorem". In: *An irregular mind. Szemerédi is 70.* Vol. **21**. Bolyai Soc. Math. Stud., 2010, pp. 335–342.

[GT12] B. Green and T. Tao. "The quantitative behaviour of polynomial orbits on nilmanifolds". In: *Ann. of Math.* **175** (2 2012), pp. 465–540.

[GT17] B. Green and T. Tao. "New bounds for Szemerédi's theorem, III: a polylogarithmic bound for $r_4(N)$". In: *Mathematika* **63**.3 (2017), pp. 944–1040.

[GTZ11] B. Green, T. Tao, and T. Ziegler. "An inverse theorem for the Gowers U^4 norm". In: *Glasg. Math. J.* **53**.1 (2011), pp. 1–50.

[GTZ12] B. Green, T. Tao, and T. Ziegler. "An inverse theorem for the Gowers $U^{s+1}[N]$-norm". In: *Ann. of Math.* **176**.2 (2012), pp. 1231–1372.

[GW10a] W. T. Gowers and J. Wolf. "The true complexity of a system of linear equations". In: *Proc. Lond. Math. Soc.* **100**.1 (2010), pp. 155–176.

[GW10b] B. Green and J. Wolf. "A note on Elkin's improvement of Behrend's construction". In: *Additive number theory: Festschrift in honor of the sixtieth birthday of Melvyn B. Nathanson.* Springer New York, 2010, pp. 141–144.

[GW11a] W. T. Gowers and J. Wolf. "Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$". In: *Geom. Funct. Anal.* **21** (2011), pp. 36–69.

[GW11b] W. T. Gowers and J. Wolf. "Linear forms and quadratic uniformity for functions on $\mathbb{F}_p^n$". In: *Mathematika* **57** (2 2011), pp. 215–237.

[GW11c] W. T. Gowers and J. Wolf. "Linear forms and quadratic uniformity for functions on $\mathbb{Z}_N$". In: *J. Anal. Math.* **115**.1 (2011), pp. 121–186.

[Hea87] D. R. Heath-Brown. "Integer sets containing no arithmetic progressions". In: *J. Lond. Math. Soc.* **35**.3 (1987), pp. 385–394.

[HK05a] B. Host and B. Kra. "Convergence of polynomial ergodic averages". In: *Israel J. Math.* **149**.1 (2005), pp. 1–19.

[HK05b] B. Host and B. Kra. "Nonconventional ergodic averages and nilmanifolds". In: *Ann. of Math.* **161**.1 (2005), pp. 397–488.

[HK18] B. Host and B. Kra. *Nilpotent structures in ergodic theory*. AMS, 2018.

[HLY21] R. Han, M. T. Lacey, and F. Yang. "A polynomial Roth theorem for corners in finite fields". In: *ArXiv e-prints* (2021). arXiv: `2012.11686`.

[Hos06] B. Host. "Convergence of multiple ergodic averages". In: *ArXiv e-prints* (2006). arXiv: `0606362`.

[IK04] H. Iwaniec and E. Kowalski. *Analytic Number Theory*. AMS, 2004.

[Kuc21a] B. Kuca. "Further bounds in the polynomial Szemerédi theorem over finite fields". In: *Acta Arith.* **198** (2021), pp. 77–108.

[Kuc21b] B. Kuca. "Multidimensional polynomial Szemerédi theorem in finite fields for distinct-degree polynomials". In: *ArXiv e-prints* (2021). arXiv: `2103.12606`.

[Kuc21c] B. Kuca. "On several notions of complexity of polynomial progressions". In: *ArXiv e-prints* (2021). arXiv: `2104.07339`.

[Kuc21d] B. Kuca. "True complexity of polynomial progressions in finite fields". In: *Proc. Edinb. Math. Soc.* (2021), pp. 1–53.

[Lei02] A. Leibman. "Polynomial mappings of groups". In: *Israel J. Math.* **129** (2002), pp. 29–60.

[Lei05a] A. Leibman. "Convergence of multiple ergodic averages along polynomials of several variables". In: *Israel J. Math.* **146** (2005), pp. 303–315.

[Lei05b] A. Leibman. "Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold". In: *Ergodic Theory Dynam. Systems* **25**.1 (2005), pp. 201–213.

[Lei07] A. Leibman. "Orbits on a nilmanifold under the action of a polynomial sequence of translations". In: *Ergodic Theory Dynam. Systems* **27**.4 (2007), pp. 1239–1252.

[Lei09] A. Leibman. "Orbit of the diagonal in the power of a nilmanifold". In: *Trans. Amer. Math. Soc.* **362**.03 (2009), pp. 1619–1658.

[Luc06] J. Lucier. “Intersective sets given by a polynomial”. In: *Acta Arith.* **123** (2006), pp. 57–95.

[Man18] F. Manners. “Good bounds in certain systems of true complexity 1”. In: *Discrete Anal.* **21** (2018), 40 pp.

[Man21] F. Manners. “True complexity and iterated Cauchy-Schwarz”. In: *ArXiv e-prints* (2021). arXiv: `2109.05731`.

[NRS06] B. Nagle, V. Rödl, and M. Schacht. “The counting lemma for regular k-uniform hypergraphs”. In: *Random Structures Algorithms* **28**.2 (2006), pp. 113–179.

[Pel18] S. Peluse. “Three-term polynomial progressions in subsets of finite fields”. In: *Israel J. Math.* **228** (1 2018), pp. 379–405.

[Pel19] S. Peluse. “On the polynomial Szemerédi theorem in finite fields”. In: *Duke Math. J.* **168**.5 (2019), pp. 749–774.

[Pel20] S. Peluse. “Bounds for sets with no polynomial progressions”. In: *Forum Math. Pi* **8** (e16 2020).

[PP19] S. Peluse and S. Prendiville. “Quantitative bounds in the non-linear Roth theorem”. In: *ArXiv e-prints* (2019). arXiv: `1903.02592`.

[PP20] S. Peluse and S. Prendiville. “A polylogarithmic bound in the non-linear Roth theorem”. In: *Int. Math. Res. Nov. IMRN* (2020). rnaa261.

[Pre17] S. Prendiville. “Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case”. In: *Discrete Anal.* **5** (2017), 34 pp.

[Ric19] A. Rice. “A maximal extension of the best-known bounds for the Furstenberg-Sárközy theorem”. In: *Acta Arith.* **187** (2019), pp. 1–41.

[Rot53] K. F. Roth. “On certain sets of integers”. In: *J. Lond. Math. Soc.* **28**.1 (1953), pp. 104–109.

[RS04] V. Rödl and J. Skokan. “Regularity lemma for k-uniform hypergraphs”. In: *Random Structures Algorithms* **25**.1 (2004), pp. 1–42.

[Ruz84] I. Ruzsa. “Difference sets without squares”. In: *Period. Math. Hungar.* **15**.3 (1984), pp. 205–209.

[San11] T. Sanders. “On Roth’s theorem on progressions”. In: *Ann. of Math.* **56** (174 2011), pp. 619–636.

[San12] T. Sanders. “On certain other sets of integers”. In: *J. Anal. Math.* **56** (116 2012), pp. 53–82.

[Sár78a] A. Sárközy. “On difference sets of sequences of integers. I”. In: *Acta Math. Hungar.* **31**.1-2 (1978), pp. 125–149.

[Sár78b] A. Sárközy. “On difference sets of sequences of integers. III”. In: *Acta Math. Hungar.* **31** (1978), pp. 355–386.

[Sch20] T. Schoen. “Improved bound in Roth’s theorem on arithmetic progressions”. In: *ArXiv e-prints* (2020). arXiv: `2005.01145`.

[Shk06a] I. D. Shkredov. “On a generalization of Szemerédi’s theorem”. In: *Proc. Lond. Math. Soc.* **93**.3 (2006), pp. 723–760.

[Shk06b] I. D. Shkredov. “On a problem of Gowers”. In: *Izv. Math.* **70**.2 (2006), pp. 385–425.

[Sli03] S. Slijepcević. “A polynomial Sárközy-Furstenberg theorem with upper bounds”. In: *Acta Math. Hungar.* **98**.1-2 (2003), pp. 111–128.

[SSZ20] A. Sah, M. Sawhney, and Y. Zhao. “Patterns without a popular difference”. In: *ArXiv e-prints* (2020). arXiv: `2004.07722`.

[Sze75] E. Szemerédi. “On sets of integers containing k elements in arithmetic progression”. In: *Acta Arith.* **27**.1 (1975), pp. 199–245.

[Sze90] E. Szemerédi. “Integer sets containing no arithmetic progressions”. In: *Acta Math. Hungar.* **56** (1990), pp. 155–158.

[Tao06a] T. Tao. “A quantitative ergodic theory proof of Szemerédi’s theorem”. In: *Electron. J. Combin.* **13**.99 (2006), pp. 1–49.

[Tao06b] T. Tao. “A variant of the hypergraph removal lemma”. In: *J. Combin. Thy.* **113**.99 (2006), pp. 1257–1280.

[Tao06c] T. Tao. “The dichotomy between structure and randomness, arithmetic progressions, and the primes”. In: *Proceedings of the International Congress of Mathematicians.* Europ. Math. Soc., 2006, pp. 581–608.

[Tao07] T. Tao. “What is good mathematics?” In: *ArXiv e-prints* (2007). arXiv: `0702396`.

[Tow10] H. Towsner. “A model theoretic proof of Szemerédi’s theorem”. In: *ArXiv e-prints* (2010). arXiv: `1002.4456`.

[TV06] T. Tao and V. Vu. *Additive Combinatorics*. Cambridge Studies in Advanced Mathematics. Cambridge U. P., 2006.

[Van27] B. L. Van der Waerden. “Beweis einer Baudetschen Vermutung”. In: *Nieuw Arch. Wisk* **15** (1927), pp. 212–216.

[Wal12] M. Walsh. “Norm convergence of nilpotent ergodic averages”. In: *Ann. of Math.* **175** (3 2012), pp. 1667–1688.

[You19] K. Younis. “Lower bounds in the polynomial Szemerédi theorem”. In: *ArXiv e-prints* (2019). arXiv: `1908.06058`.

[Zie07] T. Ziegler. “Universal characteristic factors and Furstenberg averages”. In: *J. Amer. Math. Soc.* **20** (2007), pp. 53–97.

2 Further bounds in the polynomial Szemerédi theorem over finite fields[1]

Abstract

We provide upper bounds for the size of subsets of finite fields lacking the polynomial progression

$$(x,\ x+y,\ ...,\ x+(t-1)y,\ x+y^t,\ ...,\ x+y^{t+k-1}).$$

These are the first known upper bounds in the polynomial Szemerédi theorem for the case when polynomials are neither linearly independent nor homogeneous of the same degree. We moreover improve known bounds for subsets of finite fields lacking arithmetic progressions with a difference coming from the set of k-th power residues, i.e. configurations of the form

$$(x,\ x+y^k,\ ...,\ x+(t-1)y^k).$$

Both results follow from an estimate of the number of such progressions in an arbitrary subset of a finite field.

2.1 Introduction

Generalizing Szemerédi's theorem on arithmetic progressions in subsets of integers [Sze75], Bergelson and Leibman proved that each dense subset of $\mathbb{Z}$ contains a configuration of the form $(x,\ x+P_1(y),\ ...,\ x+P_t(y))$, where $y \in \mathbb{Z} \setminus \{0\}$ and $P_1, ..., P_t$ are polynomials with integer coefficients and zero constant term [BL96]. Their proof, based on ergodic theory, does not give

[1] This chapter is an adapted version of B. Kuca. "Further bounds in the polynomial Szemerédi theorem over finite fields". In: *Acta Arith.* **198** (2021), pp. 77–108. The differences between this chapter and the Acta Arithmetica paper include minor adaptations of the notation, a simplified proof of Lemma 2.4.6, and the omission of the concluding section from the Acta Arithmetica paper.

explicit quantitative bounds. Although no general bounds are known so far, they exist in certain special cases, for instance for $(x,\ x+y^k,\ \ldots,\ x+(t-1)y^k)$ with $t \geqslant 2$ and $k > 1$ [Pre17] or for $(x,\ x+y,\ x+y^2)$ [PP19]. In the finite field analogue of the question, when we are looking for bounds on the size of $A \subseteq \mathbb{F}_q$ lacking $(x,\ x+P_1(y),\ \ldots,\ x+P_t(y))$, bounds are known in the case of $P_1, \ldots, P_t$ being linearly independent over $\mathbb{Q}$ [Pel19].

In this paper, we give the first explicit upper bounds for the sizes of subsets of finite fields lacking certain polynomial progressions. Our main result is the following.

Theorem 2.1.1. *Let $t, k \in \mathbb{N}_+$, and p be a prime. Suppose that $A \subseteq \mathbb{F}_p$ lacks the progression*

$$(x,\ x+y,\ \ldots,\ x+(t-1)y,\ x+y^t,\ \ldots,\ x+y^{t+k-1}) \tag{2.1}$$

with $y \neq 0$. Then

$$|A| \ll \begin{cases} p^{1-c}, & t = 1, 2, \\ p\dfrac{(\log\log p)^4}{\log p}, & t = 3, \\ p(\log p)^{-c}, & t = 4, \\ p(\log\log p)^{-c}, & t > 4 \end{cases}$$

where all constants are positive, and the implied constant depends on k and t while c depends only on t. For $t > 4$, one can take the exponent c to equal $c = 2^{-2^{t+9}}$.

It is worth noting that the exponent c appearing in Theorem 2.1.1 for $t > 4$ is the same as the exponent that appeared in Gowers' bounds in Szemerédi theorem [Gow01].

One can think of (2.1) as the union of an arithmetic progression and a shifted geometric progression. The cases $t = 1$ and $t = 2$ are in fact identical, and the bound in this case comes from the work of Peluse [Pel19]. Our contribution is the $t > 2$ case, for which there are no previous bounds in the literature. This is the first polynomial progression for which quantitative bounds are known where polynomials in y are neither linearly independent nor homogeneous of the same degree. Theorem 2.1.1 is a special case of a more general result, which generalizes [Pel19] and uses it as a base case for induction.

Theorem 2.1.2. *Let $t, k \in \mathbb{N}_+$, $t \geqslant 3$, and $P_t,\ \ldots,\ P_{t+k-1}$ be polynomials in $\mathbb{Z}[y]$ such that*

$$a_t P_t + \ldots + a_{t+k-1} P_{t+k-1}$$

has degree at least t unless $a_t = \ldots = a_{t+k-1} = 0$ (in particular, P_t, …, P_{t+k-1} are $\mathbb{Q}$-linearly independent and each of them has degree at least t). Let $r_t(p)$ be the size of the largest subset of $\mathbb{F}_p$ lacking t-term arithmetic progressions and $s_t : [p_0, \infty) \to (0, 1]$ be a decreasing function satisfying $r_t(p) \leqslant p \cdot s_t(p)$ for all primes $p \geqslant p_0 > 0$, with $s_t(n) \to 0$ as $n \to \infty$. If $A \subseteq \mathbb{F}_p$ lacks

$$(x,\ x+y,\ \ldots,\ x+(t-1)y,\ x+P_t(y),\ \ldots,\ x+P_{t+k-1}(y)) \tag{2.2}$$

with $y \neq 0$, then

$$|A| \ll p \cdot s_t(Cp^c)$$

where the constants C, c, and the implied constant depend on t, k, and P_t, …, P_{t+k-1} but not on the choice of s_t.

The best bounds for r_t currently in the literature are of the form

$$r_t(p) \ll \begin{cases} p\dfrac{(\log\log p)^4}{\log p}, & t = 3 \text{ [Blo16]}, \\ p(\log p)^{-c}, & t = 4 \text{ [GT17]}, \\ p(\log\log p)^{-c}, & t > 4 \text{ [Gow01]} \end{cases}$$

yielding the bounds given in Theorem 2.1.1. The content of Theorem 2.1.2 is that up to the values of constants, our bounds are of the same shape as the bounds in Szemerédi theorem. One cannot hope to do better, as each set containing (2.2) necessarily contains an t-term arithmetic progression. The function s_t plays only an auxiliary role, allowing us to conveniently express known bounds in Szemerédi's theorem as functions defined over positive real numbers. For example, the aforementioned bound for r_3 allows us to take $s_3(p) = C\frac{(\log\log p)^4}{\log p}$ for some $C > 0$. Combined with Theorem 2.1.2, it yields the bound

$$|A| \ll p\frac{(\log\log(Cp^c))^4}{\log Cp^c} \ll p\frac{(\log\log p)^4}{\log p}$$

in the case $t = 3$ of Theorem 2.1.1.

We prove Theorem 2.1.2 by first proving an estimate for how many polynomial progressions a set $A \subseteq \mathbb{F}_p$ has. This counting result is the heart of this paper; once it is proved, deducing Theorem 2.1.2 is straightforward.

Theorem 2.1.3 (Counting theorem). *Let $t \in \mathbb{N}_+$ and P_t, …, P_{t+k-1} be poly-*

nomials in $\mathbb{Z}[y]$ *such that*

$$a_t P_t + \dots + a_{t+k-1} P_{t+k-1}$$

has degree at least t *unless* $a_t = \dots = a_{t+k-1} = 0$ *(in particular,* $P_t, \dots, P_{t+k-1}$ *are linearly independent and each of them has degree at least* t*). Suppose that* $f_0, \dots, f_{t+k-1} : \mathbb{F}_p \to \mathbb{C}$ *satisfy* $|f_j(x)| \leqslant 1$ *for each* $0 \leqslant j \leqslant t+k-1$ *and* $x \in \mathbb{F}_p$*. Then*

$$\begin{aligned} &\mathop{\mathbb{E}}_{x,y \in \mathbb{F}_p} \prod_{j=0}^{t-1} f_j(x+jy) \prod_{j=t}^{t+k-1} f_j(x+P_j(y)) \\ &= \mathop{\mathbb{E}}_{x,y \in \mathbb{F}_p} \prod_{j=0}^{t-1} f_j(x+jy) \left(\prod_{j=t}^{t+k-1} \mathbb{E} f_j \right) + O(p^{-c}) \end{aligned} \tag{2.3}$$

where all the constants are positive and depend on t, k *and polynomials* $P_t, \dots, P_{t+k-1}$ *but not on* $f_0, \dots, f_{t+k-1}$*.*

Using the language of probability theory, we can interpret this result as “discorrelation”: up to an error $O(p^{-c})$, the polynomials $P_t, \dots, P_{t+k-1}$ occur independently from t-term arithmetic progressions.

The condition imposed on the polynomials $P_t, \dots, P_{t+k-1}$ may seem artificial, but Theorem 2.1.3 fails if this condition is not satisfied. As an example of failure, consider the configuration $(x,\ x+y,\ x+2y,\ x+y^2)$. Because y^2 has degree 2, which is less than the length of the arithmetic progression, y^2 is contained in the span of $x^2, (x+y)^2, (x+2y)^2$. Thus, there exist quadratic polynomials Q_0, Q_1, Q_2 satisfying

$$Q_0(x) + Q_1(x+y) + Q_2(x+2y) + (x+y^2) = 0.$$

As a consequence, if we take $Q_3(t) = t$ and $f_j(t) = e_p(aQ_j(t))$ for $a \neq 0$, then

$$\mathop{\mathbb{E}}_{x,y} f_0(x) f_1(x+y) f_2(x+2y) f_3(x+y^2) = 1$$

while the right-hand side of (2.3) in this case is $O(p^{-c})$, as $\mathbb{E} f_3 = 0$. More generally, if a linear combination of $P_t, P_{t+1}, \dots, P_{t+k-1}$ has degree $d < t$, then there is a nontrivial algebraic relation connecting $(x,\ x+y,\ \dots,\ x+(t-1)y)$ with some of $P_t, \dots, P_{t+k-1}$, and this relation prevents discorrelation from happening.

To complement these results, we prove an upper bound for the size of subsets of $\mathbb{F}_p$ lacking progressions of the form

$$(x,\ x+y^k,\ ..,\ x+(t-1)y^k) \tag{2.4}$$

i.e. arithmetic progressions with $k-$th power common difference. An upper bound on subsets of $\mathbb{Z}$ lacking this configuration of the form $C\frac{N}{(\log\log N)^c}$, with constants depending on t and k, was proved by Prendiville [Pre17] using the density increment method, and it naturally carries over to subsets of finite fields. Our bound works only for finite fields, where it is of the same shape as Prendiville's for $t > 4$, albeit with a better exponent, and strictly improves on it for $t = 3, 4$.

Theorem 2.1.4 (Sets lacking arithmetic progressions with k-th power differences)**.** *Suppose $A \subseteq \mathbb{F}_p$ contains no arithmetic progression of length t and common difference coming from the set of k-th power residues. Then*

$$|A| \ll \begin{cases} p\dfrac{(\log\log p)^4}{\log p}, & t = 3, \\ p(\log p)^{-c}, & t = 4, \\ p(\log\log p)^{-c}, & t > 4. \end{cases}$$

The constant c depends only on t, and in fact for $t > 4$, we can take $c = 2^{-2^{t+9}}$. More generally,

$$|A| \ll p \cdot s_t(c' \cdot p^{c'})$$

where s_t is defined as in Theorem 2.1.2. The constants C, c' and the implied constants are positive and depend on k and t.

Again, up to the values of constants involved, our bounds are optimal in the sense that they are of the same shape as the bounds in Szemerédi theorem.

We derive the bounds in Theorem 2.1.4 using a simple argument that heavily exploits the density and equidistribution of k-th power residues in the finite fields. With this argument, we prove the following more general counting theorem which implies Theorem 2.1.4.

Theorem 2.1.5 (Counting theorem for linear forms with restricted variables)**.** *Let $L_1, ..., L_t$ be pairwise linearly independent linear forms in $x_1, ..., x_d$. Let $k_1, ..., k_d$ be positive integers. Moreover, if $k_j > 1$, assume that no linear form L_i is of the form $L_i(x_1, ..., x_d) = ax_j$. If $f_1, ..., f_t$ satisfy $|f_i(x)| \leqslant 1$ for each*

$1 \leqslant i \leqslant t$ and each $x \in \mathbb{F}_p$, then

$$\mathop{\mathbb{E}}_{x_1,\dots,x_d \in \mathbb{F}_p} \prod_{j=1}^{t} f_j(L_j(x_1^{k_1}, \dots, x_d^{k_d})) = \mathop{\mathbb{E}}_{x_1,\dots,x_d \in \mathbb{F}_p} \prod_{j=1}^{t} f_j(L_j(x_1, \dots, x_d)) + O(p^{-c}). \tag{2.5}$$

In particular,

$$\begin{aligned} &|\{(x_1, \dots, x_d) \in \mathbb{F}_p^d : L_i(x_1^{k_1}, \dots, x_d^{k_d}) \in A \text{ for } 1 \leqslant i \leqslant t\}| \\ &= |\{(x_1, \dots, x_d) \in \mathbb{F}_p^d : L_i(x_1, \dots, x_d) \in A \text{ for } 1 \leqslant i \leqslant t\}| + O(p^{-c}). \end{aligned}$$

2.1.1 Known results

In this section we enumerate known bounds for subsets of $\mathbb{F}_q$ or $[N]$ lacking polynomial progressions

$$(x,\ x + P_1(y),\ \dots,\ x + P_t(y))$$

where not all of $P_1, \dots, P_t$ are linear. There are some differences between the integral and finite field settings. Most importantly, finite fields contain significantly more polynomial progressions of a given form if at least one polynomial is nonlinear. That is because a nonlinear polynomial P of degree $d > 1$ has only $\Theta(N^{\frac{1}{d}})$ images in $[N]$, but it is a dense subset of $\mathbb{F}_q$, in the sense that there are at least $\frac{q}{d}$ images of P in $\mathbb{F}_q$.

The case $t = 1$ in natural numbers is often referred to as the Furstenberg-Sárḱozy theorem, and it is equivalent to finding the largest subset A of natural numbers whose difference set does not intersect the values of P evaluated at integers. This problem has been studied, among others, by Sárközy [Sár78a; Sár78b], Balog, Pelikán, Pintz, and Szemerédi [Bal+94], Slijepčević [Sli03], Lucier [Luc06], and Rice [Ric19]. They showed that A is sparse, i.e. has asymptotic density equal to 0, if and only if for each natural number n there exists $m \in \mathbb{N}$ for which n divides $P(m)$, getting explicit bounds on the way; such polynomials have been called *intersective*. When $P(y) = y^k$ for $k > 1$, a lower bound of the form $\Omega(N^c)$ for $0 < c < 1$ depending on k can be obtained by trivial greedy algorithm, and the value of c has been improved nontrivially by Ruzsa [Ruz84]. For finite fields $\mathbb{F}_p$, an elementary Fourier analytic argument gives upper bounds of the form $O(p^{\frac{1}{2}})$ with the implied constant depending on k, while the best known lower bounds are of the form $\Omega(\log p \log\log\log p)$ for infinitely many primes p [GR90].

In the case $t > 1$, bounds have only been known in two extremes. If P_1, ..., P_t are all homogeneous of the same degree, i.e. we have a configuration of the form

$$(x,\ x + c_1 y^k,\ \ldots,\ x + c_t y^k)$$

then Prendiville [Pre17] proved that all subsets of $[N]$ lacking this configuration have size $O\left(\frac{N}{(\log\log N)^c}\right)$ for some $c > 0$ depending on t and k. Theorem 2.1.4 improves this result over finite fields for configurations of length 3 and 4.

The other extreme is when P_1, ..., P_t are all linearly independent. This case has recently been tackled over finite fields by Peluse [Pel19] who has showed that subsets of $\mathbb{F}_q$ lacking such progressions have size $O(q^{1-c})$ for $c > 0$ depending on $P_1, \ldots, P_t$. In the case $t = 2$, a specific exponent is known due to works of Bourgain and Chang [BC17], Peluse [Pel18], and Dong, Li and Sawin [DLS20]. Recently, these results have been extended to the integers: Peluse and Prendiville showed that subsets of $[N]$ lacking $(x,\ x+y,\ x+y^2)$ have size $O(N(\log\log N)^{-c})$ [PP19], and Peluse then proved a bound of this form for subsets of $[N]$ lacking $(x,\ x + P_1(y),\ \ldots,\ x + P_t(y))$ with $P_1, \ldots, P_t$ all having distinct degrees and zero constant terms [Pel20].

2.1.2 Notation, terminology, and assumptions

Throughout the paper, p always denotes the characteristic and cardinality of the finite field $\mathbb{F}_p$ in which we are currently working.

A function f is *1-bounded* if $||f||_\infty \leqslant 1$. We always assume that f is a 1-bounded function from $\mathbb{F}_p$ to $\mathbb{C}$ unless explicitly stated otherwise. Sometimes, we use an expression $\mathbf{b}(t_1, \ldots, t_n)$ to denote a 1-bounded function depending only on the variables $t_1, \ldots, t_n$ whose exact form is irrelevant and may differ from line to line.

We denote constants by $0 < c < 1 < C$. The exact values of these constants are generally unimportant, only their relative size, therefore we shall often use the same symbol c, C to denote constants whose value changes from line to line or even in the same expression. If there are good reasons to distinguish between two constants in the same expression, we shall denote them as c, c' or C, C' respectively. If we need to fix a constant for the duration of an argument, we give it a numerical subscript, e.g. c_0. We also use asymptotic notation $f = O(g), g = \Omega(f), f \ll g$, or $g \gg f$ to denote that $|f(p)| \leqslant C|g(p)|$ for sufficiently large p. The constant may depend on parameters such as the length of the polynomial progression or the degrees and leading coefficients of polynomials

$P_1, ..., P_t$ involved. However, if the asymptotic notation is used in an expression involving arbitrary functions $f_0, ..., f_t$, the constant never depends on the choice of $f_0, ..., f_t$. While it is quite common in additive combinatorics to denote the dependence of the constant on these parameters by e.g. writing C_t when it depends on t, we refrain from doing so in order not to clutter the notation. Therefore the reader should always assume that constants depend on the shape and length of the polynomial progression, but never on the functions $f_0, ..., f_t$ weighting the progression. We shall reiterate this in the statements of our lemmas and theorems.

We often use expected values, which we denote by

$$\mathop{\mathbb{E}}_{x \in X} f(x) = \frac{1}{|X|} \sum_{x \in X} f(x).$$

If the set X is omitted from the notation, it is assumed that x is taken from $\mathbb{F}_p$ or from another specified set.

We denote the indicator function of the set A by 1_A. The map $\mathcal{C} : x \mapsto \overline{x}$ denotes the conjugation operator. Finally, we set $e_p(x) := e(x/p) = e^{2\pi i x/p}$.

2.2 Basic concepts from additive combinatorics

The purpose of this section is to describe a few basic and standard concepts that are used extensively throughout this paper. We only introduce here ideas that are essential for all the arguments. There are tools which shall only be applied in specific proofs, and these will be discussed in relevant sections.

2.2.1 Fourier transform

Given a function $f : \mathbb{F}_p \to \mathbb{C}$ and $\alpha \in \mathbb{F}_p$, we define its Fourier transform by the formula

$$\hat{f}(\alpha) := \mathop{\mathbb{E}}_{x} f(x) e_p(\alpha x).$$

We also call $\hat{f}(\alpha)$ the *Fourier coefficient of* f *at* α. We define the inner product on $\mathbb{F}_p$ as well as L^s and ℓ^s norms for functions from $\mathbb{F}_p$ to $\mathbb{C}$ to be

$$\langle f, q \rangle := \mathop{\mathbb{E}}_{x} f(x)\overline{g(x)},\ ||f||_{L^s} = \left(\mathop{\mathbb{E}}_{x} |f(x)|^s\right)^{\frac{1}{s}} \text{ and } ||f||_{\ell^s} = \left(\sum_{x} |f(x)|^s\right)^{\frac{1}{s}}.$$

for $1 \leqslant s < \infty$, and we set

$$||f||_\infty := ||f||_{L^\infty} = ||f||_{\ell^\infty} = \max\{|f(x)| : x \in \mathbb{F}_p\}.$$

2.2.2 Gowers norms

Let $\Delta_h f(x) = f(x+h)\overline{f(x)}$ denote the *multiplicative derivative* of f. The U^s norm of f is defined as

$$||f||_{U^s} := \left(\mathop{\mathbb{E}}_{x,h_1,\dots,h_s} \prod_{\underline{w}\in\{0,1\}^s} \mathcal{C}^{|w|} f(x+\underline{w}\cdot\underline{h}) \right)^{\frac{1}{2^s}} \tag{2.6}$$

where $|w| = w_1 + \dots + w_s$. If $f = 1_A$, then $||1_A||_{U^s}^{2^s}$ is the normalized count of s-dimensional parallelepipeds in A, i.e. configurations of the form

$$(x + w_1 h_1 + \dots + w_s h_s)_{\underline{w}\in\{0,1\}^s}.$$

It turns out that $||f||_{U^s}$ is a well-defined norm for $s > 1$ and a seminorm for $s = 1$ (for the proofs of these and other facts on Gowers norms described in this section, including Lemma 2.2.1, consult [Gre07] or [Tao12]). In fact, $||f||_{U^1} = |\mathbb{E}_x f(x)| = |\hat{f}(0)|$. Gowers norms enjoy several important properties that are used extensively in this paper. First, they are monotone:

$$||f||_{U^1} \leqslant ||f||_{U^2} \leqslant ||f||_{U^3} \leqslant \dots$$

Second, one can express a U^s norm of f in terms of a lower-degree Gowers norm of its multiplicative derivatives:

$$||f||_{U^s}^{2^s} = \mathop{\mathbb{E}}_{h_1,\dots,h_{s-k}} ||\Delta_{h_1,\dots,h_{s-k}} f||_{U^k}^{2^k}.$$

In particular, taking $k = 2$ gives:

$$||f||_{U^s}^{2^s} = \mathop{\mathbb{E}}_{h_1,\dots,h_{s-2}} ||\Delta_{h_1,\dots,h_{s-2}} f||_{U^2}^{4}.$$

The utility of this formula for us is that U^2 norm is much easier to understand than the U^s norms for $s > 2$. In particular, $||f||_{U^2} = ||\hat{f}||_{\ell^4}$, and from the fact that $\max_{\phi\in\mathbb{F}_p} |\hat{f}(\phi)| \leqslant ||\hat{f}||_{\ell^4} \leqslant \max_{\phi\in\mathbb{F}_p} |\hat{f}(\phi)|^{\frac{1}{2}}$ it follows that having a large U^2 norm is equivalent to having a large Fourier coefficient, which is the statement of U^2 *inverse theorem*. For $s > 2$, corresponding inverse theorems exist as well, but they are significantly more involved and we fortunately do not need them.

Gowers norms, introduced by Gowers in his celebrated proof of Szemerédi theorem, occur frequently in additive combinatorics because $||1_A||_{U^s}$ controls the number of $(s+1)$-term arithmetic progressions in A in the following way.

Lemma 2.2.1 (Generalized von Neumann theorem)**.** *Let $f_0, ..., f_s$ be 1-bounded. Then*

$$|\mathop{\mathbb{E}}_{x,y} f_0(x)f_1(x+y)\cdots f_s(x+sy)| \leqslant \min_{0\leqslant i\leqslant s} ||f_j||_{U^s}.$$

2.2.3 Counting arithmetic progressions in subsets of finite fields

In Theorems 2.1.3 and 2.1.5, we show that a certain counting operator can be expressed in terms of

$$\Lambda_t(f_0, ..., f_{t-1}) := \mathop{\mathbb{E}}_{x,y} f_0(x)f_1(x+y)\cdots f_{t-1}(x+(t-1)y)$$

which counts t-term arithmetic progressions weighted by f_0, ..., f_{t-1}. In particular, $\Lambda_t(1_A) = \Lambda_t(1_A, ..., 1_A)$ is a normalized count of t-term arithmetic progressions in A. Instead of giving the exact estimates for what this counting operator is, we want to bound it from below by an expression involving $N_t(\alpha)$, which is the smallest natural number such that $p > N_t(\alpha)$ implies that each subset of $\mathbb{F}_p$ of size at least αp contains an $t-$term arithmetic progression. The reason why we want to have the estimate for Λ_t in terms of N_t is because the functions N_t and $r_t'(n) := r_t(n)/n$ are essentially inverses, where $r_t(p)$ is the size of the largest subset of $\mathbb{F}_p$ not containing an t-term arithmetic progression. What we mean by this is that if r_t' is bounded from above by a decreasing function s_t, then - subject to certain conditions - N_t is bounded from above by s_t^{-1}. The following lemma makes this precise.

Lemma 2.2.2. *Let $r_t(p)$ be the size of the largest subset of $\mathbb{F}_p$ lacking t-term arithmetic progressions. Let $N_t(\alpha)$ be the smallest natural number such that $p > N_t(\alpha)$ implies that each subset of $\mathbb{F}_p$ of size at least αp has an t-term arithmetic progression. Suppose that $s_t : [p_0, \infty) \to (0, 1]$ is a decreasing function with $\lim_{n\to\infty} s_t(n) = 0$. Let M_t be its inverse defined on $(0, \alpha_0]$, where $\alpha_0 := s_t(p_0)$. Then $r_t(p) \leqslant ps_t(p)$ for $p \geqslant p_0$ if and only if $N_t(\alpha) \leqslant M_t(\alpha)$ for $0 < \alpha \leqslant \alpha_0$.*

Combining Lemma 2.2.2 with an averaging argument of Varnavides [Var59], we obtain the following lemma, the precise version of which has been borrowed from [RW19].

Lemma 2.2.3 (Averaging over progressions)**.** *Suppose $0 < \alpha_0 \leqslant 1$, and let $M_t : (0, \alpha_0] \to \mathbb{R}_+$ be a decreasing function satisfying $N_t \leqslant M_t$. Suppose that*

$A \subseteq \mathbb{F}_p$ has size $|A| = \alpha p$ for some $0 < \alpha \leqslant \alpha_0$. Then $|\Lambda_t(1_A)| \gg 1/M_t(\alpha/2)^2$, where the implied constant depends on t.

We conclude this section with the proof of Lemma 2.2.2.

Proof of Lemma 2.2.2. Assume that s_t is defined as in the statement of the lemma and that $r_t(p) \leqslant s_t(p)p$ for all prime $p \geqslant p_0$. Fix a prime number $p \geqslant p_0$ and $\alpha \in (0, \alpha_0]$. Suppose that $A \subseteq \mathbb{F}_p$ of size $|A| = \alpha p$ lacks a t-term arithmetic progression. The assumption of $p \geqslant p_0$ implies that $|A| \leqslant r_t(p) \leqslant s_t(p)p$, or $\alpha \leqslant s_t(p)$. From the monotonicity of s_t it follows that $p \leqslant M_t(\alpha)$.

Thus, if a subset $A \subseteq \mathbb{F}_p$ of size $|A| = \alpha p$ for $0 < \alpha \leqslant \alpha_0$ lacks a t-term arithmetic progression, it must be that either $p \leqslant p_0$ or $p \leqslant M_t(\alpha)$, implying $N_t(\alpha) \leqslant \max\{p_0, M_t(\alpha)\}$. The definition of p_0 and monotonicity of M_t imply that $p_0 = M_t(\alpha_0) \leqslant M_t(\alpha)$, and so $N_t(\alpha) \leqslant M_t(\alpha)$.

Conversely, suppose $N_t(\alpha) \leqslant M_t(\alpha)$ for $0 < \alpha \leqslant \alpha_0$. Suppose that a set $A \subseteq \mathbb{F}_p$ of size $|A| = \alpha p$ lacks an t-term arithmetic progression, and assume $0 < \alpha \leqslant \alpha_0$, $p \geqslant p_0$. Then $p \leqslant N_t(\alpha) \leqslant M_t(\alpha)$, and so $\alpha \leqslant s_t(p)$.

It thus follows that if a subset $A \subseteq \mathbb{F}_p$ of size $|A| = \alpha p$ for $p \geqslant p_0$ lacks an t-term arithmetic progression, then either $\alpha \leqslant s_t(p)$ or $\alpha > \alpha_0$. If the latter holds, then $\alpha > \alpha_0$ implies $M_t(\alpha) < M_t(\alpha_0) = p_0$, and so this case is impossible whenever $p \geqslant p_0$. Thus we must have that $\alpha \leqslant s_t(p)$ whenever $p \geqslant p_0$. □

2.3 Deriving upper bounds in Theorem 2.1.2

This section is devoted to the proof of Theorem 2.1.2 using Theorem 2.1.3 coupled with the notation from Section 2.2.3.

Proof of Theorem 2.1.2. Throughout this proof, all the constants are allowed to depend on t, k and P_t, ..., P_{t+k-1}. From Theorem 2.1.3 it follows that

$$\mathop{\mathbb{E}}_{x,y} \prod_{j=0}^{t-1} 1_A(x + jy) \prod_{j=t}^{t+k-1} 1_A(x + P_j(y)) = \left(\mathop{\mathbb{E}}_{x,y} \prod_{j=0}^{t-1} 1_A(x + jy) \right) \alpha^k + O(p^{-c})$$

If $A \subseteq \mathbb{F}_p$ for $p \geqslant p_0$ has size $|A| = \alpha p$ and lacks progressions (2.2), then the expression on the left-hand side is $O(p^{-1})$, and so

$$\left(\mathop{\mathbb{E}}_{x,y} \prod_{j=0}^{t-1} 1_A(x + jy) \right) \alpha^k \ll p^{-c}. \tag{2.7}$$

Let M_t be the inverse function for s_t on $(0, \alpha_0]$, where $\alpha_0 = s_t(p_0)$, and set $M = M_t(\alpha/2)$. The assumption $p \geqslant p_0$ and the fact that s_t is decreasing imply that $0 < \alpha \leqslant \alpha_0$. Applying Lemma 2.2.3 to (2.7) gives $\alpha^k M^{-2} \ll p^{-c}$. Behrend's construction implies that M grows faster than polynomially in α: that is, for each $C > 1$ there exists $c > 0$ such that $M \geqslant c\alpha^{-C}$ [Beh46]. Consequently, we have $M^{-3} \ll p^{-c}$ which implies that $M \gg p^c$ for a different constant $0 < c < 1$. From monotonicity of s_t it follows that $\alpha \leqslant 2s_t(cp^c)$. □

To illustrate the last bit of the above proof, we take Gowers's [Gow01] estimate

$$N_t(\alpha) \leqslant 2^{2^{\alpha^{-C}}}$$

for $t > 4$. Combined with $N_t(\alpha/2) \gg p^c$, it gives the inequality $2^{2^{C\alpha^{-C}}} \gg p^c$. After rearranging, it yields

$$\alpha \ll \frac{1}{(\log\log p)^c}.$$

Note that the function $s_t(p) = (\log_2 \log_2 p)^{-c}$ is precisely the inverse function of $M_t(\alpha) = 2^{2^{\alpha^{-C}}}$ for an appropriate choice of constants.

2.4 Proof of Theorem 2.1.3

Finally, we come to the main part of the paper, which is the proof of the counting theorem for the progression (2.2). Like before, all the constants here are allowed to depend on t, k and $P_t, ..., P_{t+k-1}$. First, we lexicographically order the set $\mathbb{N}_+^2$, i.e.

$$(t,k) < (t',k') \iff t < t' \text{ or } (t = t' \text{ and } k < k').$$

We induct on (t,k) by following the lexicographic order on $\mathbb{N}_+^2$. Let $\mathcal{S}(t,k)$ denote the statement of Theorem 2.1.3 for (t,k); that is, $\mathcal{S}(t,k)$ holds iff for all linearly independent polynomials $P_t, ..., P_{t+k-1}$ of degree at least t that do not span a polynomial of degree less than t there exists a constant $c > 0$ such that for all 1-bounded functions $f_0, ..., f_{t+k-1}$, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y} \prod_{j=0}^{t-1} f_j(x+jy) \prod_{j=t}^{t+k-1} f_j(x+P_j(y)) \\
&= \left(\mathop{\mathbb{E}}_{x,y} \prod_{j=0}^{t-1} f_j(x+jy)\right) \prod_{j=t}^{t+k-1} \mathbb{E} f_j + O(p^{-c}).
\end{aligned}$$

$\mathcal{S}(1,k)$ and $\mathcal{S}(2,k)$ follow from the work of Peluse [Pel19], and they shall serve as our base cases. It turns out that in the inductive step, we have to distinguish between two cases:

1. $\mathcal{S}(t,1)$, assuming the statement holds for all $(t',k') < (t,1)$ (although we shall only need to invoke $\mathcal{S}(t-1,2)$).

2. $\mathcal{S}(t,k)$ for $k>1$, assuming it holds for $\mathcal{S}(t',k')$ with $(t',k') < (t,k)$.

To handle the first case, we use the Cauchy-Schwarz inequality and a change of variables to reduce $\mathcal{S}(t,1)$ to $\mathcal{S}(t-1,2)$. However, this argument fails in the second case, for which we therefore employ a longer approach that resembles more closely the arguments from [Pel19; PP19].

Throughout this section, we denote the counting operator appearing in the statement of the Theorem 2.1.3 by Λ with appropriate subscripts. Thus,

$$\Lambda_{t,P_t,\ldots,P_{t+k-1}}(f_0,\ldots,f_{t+k-1}) := \mathop{\mathbb{E}}_{x,y} \prod_{j=0}^{t-1} f_j(x+jy) \prod_{j=t}^{t+k-1} f_j(x+P_j(y)).$$

In particular, Λ_t denotes the counting operator for t-term arithmetic progressions:

$$\Lambda_t(f_0,\ldots,f_{t-1}) := \mathop{\mathbb{E}}_{x,y} \prod_{j=0}^{t-1} f_j(x+jy).$$

When t,k, and $P_t,\ldots,P_{t+k-1}$ are clear out of the context, we shall suppress the subscripts and denote the operator just by Λ.

2.4.1 Proof of $\mathcal{S}(t,1)$ assuming $\mathcal{S}(t-1,2)$

As advertised earlier, we first prove the inductive step for $\mathcal{S}(t,1)$. Let P be a polynomial of degree at least t. Our goal is to show that the counting operator

$$\Lambda_{t,P}(f_0,\ldots,f_t) = \mathop{\mathbb{E}}_{x,y} \left(\prod_{j=0}^{t-1} f_j(x+jy)\right) f_t(x+P(y)) \tag{2.8}$$

is in fact controlled by an operator involving an arithmetic progression of length $t-1$ of difference functions of $f_1,\ldots,f_{t-1}$. To accomplish this, we first rewrite (2.8) as

$$\mathop{\mathbb{E}}_{x} f_0(x) \mathop{\mathbb{E}}_{y} \left(\prod_{j=1}^{t-1} f_j(x+jy)\right) f_t(x+P(y)).$$

Applying the Cauchy-Schwarz inequality in x together with 1-boundedness of f_0 and changing variables, we obtain that

$$|\Lambda_{t,P}(f_0,...,f_t)|^2 \leqslant \mathop{\mathbb{E}}_{x}\left|\mathop{\mathbb{E}}_{y}\left(\prod_{j=1}^{t-1} f_j(x+jy)\right) f_t(x+P(y))\right|^2$$
$$\leqslant \mathop{\mathbb{E}}_{x,y,h}\left(\prod_{j=1}^{t-1}\Delta_{jh}f_j(x+jy)\right)\overline{f_t(x+P(y))}f_t(x+P(y+h))$$

Translating $x\mapsto x-y$ and using the triangle inequality gives

$$|\Lambda_{t,P}(f_0,...,f_t)|^2 \leqslant \mathop{\mathbb{E}}_{h}\left|\mathop{\mathbb{E}}_{x,y}\left(\prod_{j=1}^{t-1}\Delta_{jh}f_j(x+(j-1)y)\right)\right.$$
$$\left.\overline{f_t(x+P(y)-y)}f_t(x+P(y+h)-y)\right|.$$

By the pigeonhole principle, there exists $h\neq 0$ such that

$$|\Lambda_{t,P}(f_0,...,f_t)|^2 \leqslant \left|\mathop{\mathbb{E}}_{x,y}\left(\prod_{j=1}^{t-1}\Delta_{jh}f_j(x+(j-1)y)\right)\right.$$
$$\left.\overline{f_t(x+P(y)-y)}f_t(x+P(y+h)-y)\right| + O(p^{-1})$$
$$= |\Lambda_{t-1,P_t,P_{t+1}}(g_0,...,g_{t-2},\overline{f_t},f_t)| + O(p^{-1})$$

where we set

$$P_t(y)=P(y)-y,\quad P_{t+1}(y)=P(y+h)-y \text{ and }\quad g_j(t)=\Delta_{(j+1)h}f_{j+1}(t).$$

From $h\neq 0$ it follows that P_t, P_{t+1} are linearly independent. Moreover, for any $(a,b)\neq(0,0)$, the polynomial aP_t+bP_{t+1} has degree at least $t-1$, attaining this degree precisely when $a+b=0$. We have thus reduced the study of $\Lambda_{t,P}$ to the analysis of $\Lambda_{t-1,P_t,P_{t+1}}$, and so we are in the $\mathcal{S}(t-1,2)$ case. Applying Theorem 2.1.3 for this case, we see that

$$|\Lambda_{t,P}(f_0,...,f_t)|^2 \leqslant |\Lambda_{t-1,P_t,P_{t+1}}(g_0,...,g_{t-2},\overline{f_t},f_t)| + O(p^{-1})$$
$$= |\Lambda_{t-1}(g_0,...,g_{t-2})|\cdot|\mathbb{E}\, f_t|^2 + O(p^{-c})$$
$$\leqslant |\mathbb{E}\, f_t|^2 + O(p^{-c})$$

and hence

$$|\Lambda_{t,P}(f_0, ..., f_t)| \leqslant |\mathbb{E} f_t| + O(p^{-c}).$$

We have established so far that the U^1 norm of f_t controls $\Lambda_{t,P}(f_0, ..., f_t)$ up to a power-saving error, i.e. $||f_t||_{U^1} = 0$ implies $|\Lambda_{t,P}(f_0, ..., f_t)| = O(p^{-c})$. To utilise this fact, we decompose $f_t = \mathbb{E} f_t + (f_t - \mathbb{E} f_t)$ and split $\Lambda_{t,P}$ accordingly. The term involving $f_t - \mathbb{E} f_t$ has size at most $O(p^{-c})$ because $\mathbb{E}(f_t - \mathbb{E} f_t) = 0$, and so

$$\Lambda_{t,P}(f_0, ..., f_t) = \Lambda_t(f_0, ..., f_{t-1}) \mathbb{E} f_t + O(p^{-c}),$$

as required.

2.4.2 Proof of $\mathcal{S}(t, k)$, $k > 1$

Our next goal is to prove $\mathcal{S}(t, k)$ whenever $k > 1$. The natural thing to try would be to prove this case in a similar manner we proved $\mathcal{S}(t, 1)$; that is, to apply the Cauchy-Schwarz inequality to the counting operator

$$\Lambda_{t,P_t,...,P_{t+k-1}}(f_0, ..., f_{t+k-1}) \tag{2.9}$$

and bound it by the counting operator of

$$\Lambda_{t-1,Q_t,R_t,...,Q_{t+k-1},R_{t+k-1}}(g_0, ..., g_{t-2}, \overline{f_t}, f_t, ..., \overline{f_{t+k-1}}, f_{t+k-1})$$

where

$$Q_j(y) = P_j(y) - y, \quad R_j(y) = P_j(y + h) - y \quad \text{and} \quad g_j(t) = \Delta_{(j+1)h} f_{j+1}(t).$$

However, this simple extension of the method used to prove $\mathcal{S}(t, 1)$ does not work because there is no guarantee that $Q_t, R_t, ..., Q_{t+k-1}, R_{t+k-1}$ are linearly independent (and in general, they may not be), nor that any nonzero linear combination of them has degree at least $t - 1$. To illustrate this problem, we look at

$$(x,\ x + y,\ x + 2y,\ x + y^3,\ x + y^4).$$

Applying the Cauchy-Schwarz inequality and translating by $x \mapsto x - y$, we control this configuration by the counting operator of the configuration

$$(x,\ x+y,\ x+y^3-y,\ x+(y+h)^3-y,\ x+y^4-y,\ x+(y+h)^4-y).$$

Note that the polynomials $y,\ y^3-y,\ (y+h)^3-y,\ y^4-y,\ (y+h)^4-y$ have degree at most 4, and there are 5 of them, hence there exist $a_1, ..., a_5, b$ not all zero such that

$$a_1 y + a_2(y^3-y) + a_3((y+h)^3-y) + a_4(y^4-y) + a_5((y+h)^4-y) = b.$$

Consequently, one cannot apply induction hypothesis to this configuration. One therefore needs to come up with a different method.

Throughout this section, we let

$$\Lambda := \Lambda_{t,P_t,...,P_{t+k-1}}.$$

Our general strategy for $\mathcal{S}(t,k)$, $k>1$ is to gradually replace each of $f_t, ..., f_{t+k-1}$ by additive characters. Our method follows very closely the techniques in [Pel20; Pel19; PP19], and we shall point the reader to the relevant statements in these papers for comparison. To replace arbitrary functions by characters, we introduce an inner induction loop, much like in the proof of Theorem 2.1 of [Pel19]. For $0 \leqslant r \leqslant k$, let $\mathcal{S}(t,k,r)$ denote the statement that for all polynomials P_t, ..., P_{t+k-1} satisfying the conditions of Theorem 2.1.3, there exists $c>0$ such that

$$\begin{aligned}&\Lambda_{t,P_t,...,P_{t+k-1}}(f_0, ..., f_{t+r-1}, e_p(a_{t+r}\cdot), ..., e_p(a_{t+k-1}\cdot))\\&= \Lambda_t(f_0, ..., f_{t-1}) \prod_{j=t}^{t+r-1} \mathbb{E} f_j \prod_{j=t+r}^{t+k-1} 1_{a_j=0} + O(p^{-c}).\end{aligned}$$

for all 1-bounded functions $f_0, ..., f_{t+r-1} : \mathbb{F}_p \to \mathbb{C}$. We note that $\mathcal{S}(t,k,r)$ is the special case of $\mathcal{S}(t,k)$ restricted to the situation when $f_j = e_p(a_j\cdot)$ for $t+r \leqslant j \leqslant t+k-1$, and $\mathcal{S}(t,k,k)$ is equivalent to $\mathcal{S}(t,k)$. We shall therefore deduce $\mathcal{S}(t,k)$ by inducting on $0 \leqslant r \leqslant k$ for fixed (t,k). We start by proving the base case $\mathcal{S}(t,k,0)$, which by the homomorphism property of additive characters and assumptions on the polynomials $P_t, ..., P_{t+k-1}$ reduces to the statement in the following lemma.

Lemma 2.4.1. *Let $a \in \mathbb{F}_p^\times$ and $t \in \mathbb{N}_+$. Suppose that $P \in \mathbb{Z}[y]$ has degree at* *least t and that the functions $f_0, \ldots, f_{t-1} : \mathbb{F}_p \to \mathbb{C}$ are 1-bounded. Then*

$$|\Lambda_{t,P}(f_0, \ldots, f_{t-1}, e_p(a\cdot))| \leqslant O(p^{-c}) \tag{2.10}$$

for a constant $c > 0$ depending on t and P but not on a or $f_0, \ldots, f_{t-1}$.

Proof. We prove the statement by induction on t. For $t = 1$, we have

$$|\mathop{\mathbb{E}}_{x,y} f_0(x)e_p(a(x + P(y))| = |\mathop{\mathbb{E}}_{x} f_0(x)e_p(ax)| \cdot |\mathop{\mathbb{E}}_{y} e_p(aP(y))| \ll p^{-c}$$

by the 1-boundedness of f_0 and Weyl differencing.

Suppose $t > 1$. Then an application of the Cauchy-Schwarz inequality to remove f_0 followed by a change of variables gives

$$|\Lambda_{t,P}(f_0, \ldots, f_{t-1}, e_p(a\cdot))| \leqslant |\mathop{\mathbb{E}}_{h} \Lambda_{t-1,Q_h}(e_p(-a\cdot)\Delta_h f_1, \ldots, \Delta_{(t-1)h} f_{t-1}, e_p(a\cdot))|$$

where $Q_h(y) := P(y + h) - P(y)$. For $h \neq 0$, the degree of Q_h satisfies

$$\deg Q_h = \deg P - 1 \geqslant t - 1.$$

By inductive hypothesis and triangle inequality, (2.10) holds for t. The lemma follows by induction. □

The heart of the proof of $\mathcal{S}(t, k)$ for $k > 1$ is thus to show that $\mathcal{S}(t, k, r+1)$ can be deduced from $\mathcal{S}(t, k, r)$. The next lemma states this more formally.

Lemma 2.4.2. *Let $t \geqslant 3$, $k \geqslant 2$ and $0 \leqslant r < k$ be natural numbers. Assume $\mathcal{S}(t, k, r)$ holds. Then $\mathcal{S}(t, k, r + 1)$ holds as well.*

The case $\mathcal{S}(t, k)$, $k > 1$ thus follows by inducting on r and the observation that $\mathcal{S}(t, k, k) = \mathcal{S}(t, k)$.

From now on, assume (t, k, r) is fixed. In the remainder of this section, we outline the proof of Lemma 2.4.2. We formulate consecutive steps of the proof as lemmas to be proved separately in the next section. Our first task in proving Lemma 2.4.2 is to show that Λ is controlled by some Gowers norm of f_{t+r}. This follows from the so-called PET induction scheme, which originally appeared in Bergelson and Leibman's ergodic-theoretic proof of the polynomial Szemerédi theorem [BL96] and was subsequently applied in the works of Prendiville and Peluse [Pre17; Pel19; PP19] and Tao and Ziegler [TZ08; TZ16; TZ18].

Lemma 2.4.3 (PET induction, Proposition 2.2 of [Pel19])**.** *Let $P_1, \ldots, P_l$ be nonconstant polynomials in $\mathbb{Z}[y]$ such that $P_i - P_j$ is nonconstant whenever*

$i \neq j$. Then for any $1 \leqslant j \leqslant l$ there exist $s \in \mathbb{N}$ and $0 < \beta \leqslant 1$, depending only on the degrees and leading coefficients of $P_1, ..., P_l$, such that

$$|\Lambda_{x,x+P_1(y),...,x+P_l(y)}(f_0, ..., f_l)| \leqslant ||f_j||_{U^s}^{\beta} + O(p^{-\beta}).$$

for all 1-bounded functions $f_0, ..., f_l : \mathbb{F}_p \to \mathbb{C}$.

Our statement differs slightly from the statement of Proposition 2.2 in [Pel19] in that Peluse did not mention explicitly our condition that the difference between any two polynomials P_i, P_j cannot be constant. However, she assumed throughout her paper that $P_1, ..., P_l$ were distinct polynomials with zero constant terms, which implies our condition. In our paper, the polynomials may have nonzero constant terms, in which case we replace $P_i(y)$ by $P_i'(y) := P_i(y) - P_i(0)$ and $f_i(t)$ by $f_i'(t) := f_i(t + P_i(0))$, so that $f_i(x + P_i(y)) = f_i'(x + P_i'(y))$. The facts that f_i and f_i' have the same Gowers norms and that P_1', ..., P_l' are all distinct polynomials with zero constant terms allows us to reduce to the case covered in Proposition 2.2 of [Pel19].

The notation has already become rather formidable, and it will become even more so in the further part of the proof. To make it more palatable, we let $P_j(y) := jy$ for $0 \leqslant j \leqslant t-1$ and $f_j(t) := e_p(a_j t)$ for $t+r+1 \leqslant j \leqslant t+k-1$ for the rest of Section 2.4.

Our next step is to replace the function f_{t+r} in the counting operator by

$$F(x) := \mathop{\mathbb{E}}_{y} \prod_{j=0}^{t+r-1} f_j(x + P_j(y) - P_{t+r}(y)) \prod_{j=t+r+1}^{t+k-1} f_j(x + P_j(y) - P_{t+r}(y)) \tag{2.11}$$

We do it slightly differently here than we did it in the original version of [Kuc21] - instead of using a certain decomposition result on functions, we employ a trick from Lemma 5.12 of [Pre]. We call F a “dual function” because

$$\Lambda_{t,P_t,...,P_{t+k-1}}(f_0, ..., f_{t+k-1}) = \langle F, \overline{f_{t+r}} \rangle.$$

Applying the Cauchy-Schwarz inequality and the definition of F, we see that

$$\left| \langle F, \overline{f_{t+r}} \rangle \right|^2 \leqslant \|F\|_{L^2}^2 = \Lambda_{t,P_t,...,P_{t+k-1}}(f_0, ..., f_{t+r-1}, \overline{F}, f_{t+r+1}, ..., f_{t+k-1}).$$

This gives us control over $\Lambda_{t,P_t,...,P_{t+k-1}}(f_0, ..., f_{t+k-1})$ by the U^s norm of F using Lemma 2.4.3.

In general, higher degree Gowers norms control lower degree norms but the converse is not true. For the special case of the dual function F, we however show that $||F||_{U^s}$ is indeed controlled by $||F||_{U^2}$ for any $s \in \mathbb{N}$. We achieve this in the lemma below which is an adaptation to our setting of the degree-lowering technique of Peluse, first utilised in Lemma 4.1 of [Pel19].

Lemma 2.4.4 (Degree lowering)**.** *Let F be defined as in (2.11). For each $s > 2$,*

$$||F||_{U^{s-1}} = \Omega(||F||_{U^s}^{2^{2s-1}}) - O(p^{-c})$$

for $c > 0$ depending on $t, k,$ and $P_t, ..., P_{t+k-1}$ but not on $f_0, ..., f_{t+k-1}$. As a consequence,

$$||F||_{U^2} = \Omega(||F||_{U^s}^{2^{(s-2)(s+2)}}) - O(p^{-c}).$$

Having a control by the U^2 norm of the dual function F is important because this norm is in turn controlled by the U^1 norms of the component functions $f_t, ..., f_{t+r-1}, f_{t+r+1}, ..., f_{t+k-1}$, which follows from Lemma 2.4.1 coupled with $\mathcal{S}(t, k-1)$. Recalling that $f_j(t) := e_p(a_j t)$ for $t+r+1 \leqslant j \leqslant t+k-1$ and so $||f_j||_{U^1} = 1_{a_j=0}$ for these values of j, we obtain the following lemma.

Lemma 2.4.5 (U^1 control of the dual)**.** *Let F be defined as in (2.11). Then*

$$||F||_{U^2} \leqslant \min_{t\leqslant j\leqslant t+r-1} ||f_j||_{U^1}^{\frac{1}{2}} \cdot \prod_{j=t+r+1}^{t+k-1} 1_{a_j=0} + O(p^{-c})$$

for some $c > 0$ depending on $t, k,$ and $P_t, ..., P_{t+k-1}$ but not on the functions $f_0, ..., f_{t+k-1}$.

Combining the estimates of two previous lemmas with argument above Lemma 2.4.4 gives the following bound, which differs slightly from the analogous bound in the original version of the paper [Kuc21].

Lemma 2.4.6 (U^1 control of Λ, cf. Lemma 4.2 of [Pel19] and Theorem 7.1 of [PP19])**.** *There exists a constant $c > 0$ and $s \in \mathbb{N}$ depending only on $t, k, P_t, ..., P_{t+k-1}$ but not on $f_0, ..., f_{t+k-1}$ such that*

$$|\Lambda(f_0, ..., f_{t+k-1})| \leqslant \min_{t\leqslant j\leqslant t+r-1} ||f_j||_{U^1}^c \cdot \prod_{j=t+r+1}^{t+k-1} 1_{a_j=0} + O(p^{-c}).$$

Having established Lemma 2.4.6, it is straightforward to prove $\mathcal{S}(t, k, r+1)$; however, the argument is slightly different for $r = 0$ and $r > 0$. If $r = 0$, then by Lemma 2.4.6 we have

$$|\Lambda(f_0, \dots, f_{t+k-1})| \leqslant \prod_{j=t+1}^{t+k-1} 1_{a_j=0} + O(p^{-c}).$$

If not all of a_{t+1}, ..., a_{t+k-1} are zero, then

$$|\Lambda(f_0, \dots, f_{t+k-1})| \ll p^{-c}.$$

Otherwise we are in the case $\mathcal{S}(t, 1)$. Combining these two alternatives gives $\mathcal{S}(t, k, 1)$.

If $r > 0$, we split each of f_t, ..., f_{t+r-1} into $f_j = \mathbb{E} f_j + (f_j - \mathbb{E} f_j)$, and decompose Λ accordingly. Then $\Lambda(f_0, \dots, f_{t+k-1})$ splits into the main term

$$\Lambda(f_0, \dots, f_{t-1}, \mathbb{E} f_t, \dots, \mathbb{E} f_{t+r-1}, f_{t+r}, \dots, f_{t+k-1})$$

and $2^r - 1$ error terms, each of which involves at least one $f_j - \mathbb{E} f_j$ for $t \leqslant j \leqslant t + r - 1$. Using Lemma 2.4.6, each of the error terms has size $O(p^{-c})$; hence

$$\begin{aligned}\Lambda(f_0, \dots, f_{t+k-1}) &= \Lambda(f_0, \dots, f_{t-1}, \mathbb{E} f_t, \dots, \mathbb{E} f_{t+r-1}, f_{t+r}, \dots, f_{t+k-1}) + O(p^{-c}) \\ &= \Lambda_{t,P_{t+r},\dots P_{t+k-1}}(f_0, \dots, f_{t-1}, f_{t+r}, \dots, f_{t+k-1}) \prod_{j=t}^{t+r-1} \mathbb{E} f_j \\ &\quad + O(p^{-c}).\end{aligned}$$

Applying the $\mathcal{S}(t, k, r)$ case, we can split $\Lambda_{t,P_{t+r},\dots P_{t+k-1}}$

$$\begin{aligned}&\Lambda_{t,P_{t+r},\dots P_{t+k-1}}(f_0, \dots, f_{t-1}, f_{t+r}, \dots, f_{t+k-1}) \\ &= \Lambda_t(f_0, \dots, f_{t-1}) \;\mathbb{E} f_{t+r} \prod_{j=t+r+1}^{t+k-1} 1_{a_j=0} + O(p^{-c})\end{aligned}$$

and hence

$$\Lambda(f_0, \dots, f_{t+k-1}) = \Lambda_t(f_0, \dots, f_{t-1}) \prod_{j=t}^{t+r} \mathbb{E} f_j \prod_{j=t+r+1}^{t+k-1} 1_{a_j=0} + O(p^{-c}).$$

This proves $\mathcal{S}(t, k, r + 1)$ for $r > 0$.

2.4.3 Proofs of Lemmas 2.4.4, 2.4.5 and 2.4.6

While in the previous section we outlined the proof of $\mathcal{S}(t,k)$ for $k>1$, here we derive the technical lemmas which are used in this proof.

Proof of Lemma 2.4.4. This proof follows the path of Proposition 6.6 in [PP19]. The main idea is to write the U^s norm of the dual function F as an average of the U^2 norms of derivatives of F, extract the maximum Fourier coefficients of $\Delta_{h_1,\dots,h_{s-2}}F$, and show that for a dense proportion of $(h_1,\dots,h_{s-2})$ these coefficients satisfy certain linear relations provided $||F||_{U^s}\gg p^{-c}$. If $s=3$ and $\phi(h)$ is the phase of the maximum Fourier coefficient of $\Delta_h F$, then we show that ϕ is constant on a dense proportion of h. For $s>3$, analogous relations are somewhat more complicated. These linear relations turn out to be sufficient to get a control of the U^s norm of F by its U^{s-1} norm with polynomial bounds.

Using the definition of Gowers norms, we have

$$\eta := ||F||_{U^s}^{2^s} = \mathop{\mathbb{E}}_{h_1,\dots,h_{s-2}} ||\Delta_{h_1,\dots,h_{s-2}}F||_{U^2}^4.$$

Let $H_1=\{(h_1,\dots,h_{s-2})\in\mathbb{F}_p^{s-2} : ||\Delta_{h_1,\dots,h_{s-2}}F||_{U^2}^4 \geqslant \frac{1}{2}\eta\}$. To simplify the notation, let $\underline{h}=(h_1,\dots,h_{s-2})$ and $\mathbb{E}_{\underline{h}} := \mathbb{E}_{\underline{h}\in\mathbb{F}_p^{s-2}}$. From the popularity principle (Lemma 1.2.10) it follows that $|H_1|\geqslant \frac{1}{2}\eta p^{s-2}$, and so

$$\frac{1}{4}\eta^2 \leqslant \mathop{\mathbb{E}}_{\underline{h}} ||\Delta_{\underline{h}}F||_{U^2}^4\cdot 1_{H_1}(\underline{h}). \tag{2.12}$$

The U^2 inverse theorem, stated in Section 2.2.2, implies that the square of the U^2 norm of a function is bounded by its maximum Fourier coefficient. Given $\Delta_{\underline{h}}F$, let $\widehat{\Delta_{\underline{h}}F}(\phi(\underline{h}))$ denote its maximum Fourier coefficient. Then the right hand side of (2.12) is bounded by

$$\begin{aligned}\mathop{\mathbb{E}}_{\underline{h}}\left|\widehat{\Delta_{\underline{h}}F}(\phi(\underline{h}))\right|^2 1_{H_1}(\underline{h}) &= \mathop{\mathbb{E}}_{\underline{h}}\left|\mathop{\mathbb{E}}_{x}\Delta_{\underline{h}}F(x)e_p(\phi(\underline{h})x)\right|^2 1_{H_1}(\underline{h})\\ &= \mathop{\mathbb{E}}_{x,x',\underline{h}}\Delta_{\underline{h}}F(x)\overline{\Delta_{\underline{h}}F(x')}e_p(\phi(\underline{h})(x-x'))1_{H_1}(\underline{h}).\end{aligned} \tag{2.13}$$

To simplify the already cumbersome notation, we denote $Q_j=P_j-P_{t+r}$ for $0\leqslant j\leqslant t+k-1$. Unpacking the definition of the dual function F, the expression (2.13) equals

$$\mathop{\mathbb{E}}_{x,x',\underline{h}} \Delta_{\underline{h}} \left(\mathop{\mathbb{E}}_{y} \prod_{\substack{0 \leqslant j \leqslant t+k-1,\\ j \neq t+r}} f_j(x + Q_j(y)) \right) \overline{\Delta_{\underline{h}} \left(\mathop{\mathbb{E}}_{y} \prod_{\substack{0 \leqslant j \leqslant t+k-1,\\ j \neq t+r}} f_j(x' + Q_j(y)) \right)} \tag{2.14}$$

$$e_p(\phi(\underline{h})(x - x'))1_{H_1}(\underline{h}).$$

After writing out the multiplicative derivatives, (2.14) is equal to

$$\mathop{\mathbb{E}}_{x,x',\underline{h}} \mathop{\mathbb{E}}_{\underline{y},\underline{y}' \in \mathbb{F}_p^{\{0,1\}^{s-2}}} \prod_{\substack{0 \leqslant j \leqslant t+k-1,\\ j \neq t+r}} \prod_{\underline{w} \in \{0,1\}^{s-2}} \mathcal{C}^{|w|} f_j(x + \underline{w} \cdot \underline{h} + Q_j(\underline{y}_{\underline{w}})) \tag{2.15}$$
$$\mathcal{C}^{|w|} \overline{f_j(x' + \underline{w} \cdot \underline{h} + Q_j(\underline{y}'_{\underline{w}}))} e_p(\phi(\underline{h})(x - x'))1_{H_1}(\underline{h}).$$

The product in (2.15) contains 2^{s-2} copies of f_j for each j and each of x and x'. In each of these copies the y-variable is different. We would like all the copies of f_j to be expressed in terms of the same y-variable. To achieve this, we modify (2.15) by applying the Cauchy-Schwarz inequality $s - 2$ times. First, (2.15) can be rewritten as

$$\mathop{\mathbb{E}}_{x,x',h_1,\dots,h_{s-3}} \mathop{\mathbb{E}}_{\underline{y},\underline{y}' \in \mathbb{F}_p^{\{0,1\}^{s-2}}} \mathbf{b}(x, x', h_1, \dots, h_{s-3}, \underline{y}, \underline{y}') \mathop{\mathbb{E}}_{h_{s-2}} \prod_{\substack{0 \leqslant j \leqslant t+k-1,\\ j \neq t+r}} \prod_{\substack{\underline{w} \in \{0,1\}^{s-2}\\ w_s = 1}} \tag{2.16}$$

$$\mathcal{C}^{|w|} f_j(x + \underline{w} \cdot \underline{h} + Q_j(\underline{y}_{\underline{w}})) \mathcal{C}^{|w|} \overline{f_j(x' + \underline{w} \cdot \underline{h} + Q_j(\underline{y}'_{\underline{w}}))} e_p(\phi(\underline{h})(x - x'))1_{H_1}(\underline{h}).$$

By the Cauchy-Schwarz inequality and change of variables, (2.16) is bounded by

$$\Bigg(\mathop{\mathbb{E}}_{x,x',h_1,\dots,h_{s-3},h_{s-2},h'_{s-2}} \mathop{\mathbb{E}}_{\underline{y},\underline{y}' \in \mathbb{F}_p^{\{0,1\}^{s-2}}} \prod_{\substack{0 \leqslant j \leqslant t+k-1,\\ j \neq t+r}} \prod_{\substack{\underline{w} \in \{0,1\}^{s-2}\\ w_s = 1}} \mathcal{C}^{|w|} \tag{2.17}$$

$$\left(f_j(x + \sum_{i=1}^{s-3} w_i h_i + w_{s-2} h_{s-2} + Q_j(\underline{y}_{\underline{w}}) \right) \overline{f_j \left(x + \sum_{i=1}^{s-3} w_i h_i + w_{s-2} h'_{s-2} + Q_j(\underline{y}_{\underline{w}}) \right)}$$

$$\overline{f_j \left(x' + \sum_{i=1}^{s-3} w_i h_i + w_{s-2} h_{s-2} + Q_j(\underline{y}'_{\underline{w}}) \right)} f_j \left(x' + \sum_{i=1}^{s-3} w_i h_i + w_{s-2} h'_{s-2} + Q_j(\underline{y}'_{\underline{w}})) \right)$$

$$e_p((\phi(h_1, \dots, h_{s-3}, h_{s-2}) - \phi(h_1, \dots, h_{s-3}, h'_{s-2}))(x - x'))$$

$$1_{H_1}(h_1, \dots, h_{s-3}, h_{s-2}) 1_{H_1}(h_1, \dots, h_{s-3}, h'_{s-2}))^{\frac{1}{2}}.$$

The presence of so many terms in (2.17) comes from the fact that in the process of applying the Cauchy-Schwarz inequality and changing variables, each expression $E(h_{s-2})$ (depending possibly on other variables as well) is replaced by $E(h_{s-2})\overline{E(h'_{s-2})}$. Therefore the number of expressions in the product doubles, making (2.17) rather lengthy. Applying Cauchy-Schwarz another $s-3$ times to $h_{s-3}, ..., h_1$ respectively, we bound (2.17) by

$$\left(\mathop{\mathbb{E}}_{x,x',y,y',\underline{h},\underline{h}'} \prod_{\substack{0\leqslant j\leqslant t+k-1,\\ j\neq t+r}} \prod_{\underline{w}\in\{0,1\}^{s-2}} (1_{H^1}(\underline{h}^{(\underline{w})})\mathcal{C}^{|w|} f_j(x+\underline{w}\cdot\underline{h}^{(\underline{w})}+Q_j(y)) \right. \tag{2.18}$$

$$\left. \overline{\mathcal{C}^{|w|} f_j(x'+\underline{w}\cdot\underline{h}^{(\underline{w})}+Q_j(y'))) } e_p\left(\sum_{\underline{w}\in\{0,1\}^{s-2}} (-1)^{|w|}\phi(\underline{h}^{(\underline{w})})(x-x')\right)\right)^{\frac{1}{2^{s-2}}}.$$

where

$$\underline{h}_i^{(\underline{w})} = \begin{cases} h_i, w_i = 0 \\ h'_i, w_i = 1 \end{cases}$$

The expression (2.18) can be simplified to

$$\left(\mathop{\mathbb{E}}_{\underline{h},\underline{h}'} \left| \mathop{\mathbb{E}}_{x,y} \left(\prod_{\substack{j=0,\\ j\neq t+r}}^{t+k-1} g_j(x+P_j(y)) \right) e_p\left(\psi(\underline{h},\underline{h}')(x+P_{t+r}(y))\right) \right|^2 1_{\square(H_1)}(\underline{h},\underline{h}') \right)^{\frac{1}{2^{s-2}}}$$

where

$$g_j(t) := \prod_{\underline{w}\in\{0,1\}^{s-2}} \mathcal{C}^{|w|} f_j(t+\underline{w}\cdot\underline{h}^{(\underline{w})}),$$

$$\square(A) := \{(\underline{h},\underline{h}') \in \mathbb{F}_p^{2(s-2)} : \forall \underline{w} \in \{0,1\}^{s-2} \underline{h}^{(\underline{w})} \in A\}$$

for $A \subseteq \mathbb{F}_p^{s-2}$ and

$$\psi(\underline{h},\underline{h}') := \sum_{\underline{w}\in\{0,1\}^{s-2}} (-1)^{|w|}\phi(\underline{h}^{(\underline{w})}).$$

Recall that for $t+r+1 \leqslant j \leqslant t+k-1$, we have defined f_j to be $f_j(x) = e_p(a_j x)$. Combining this with the assumption that $s > 2$, we have that

$$g_j(x+P_j(y)) = e_p\left(a_j \sum_{\underline{w}\in\{0,1\}^{s-2}} (-1)^{|w|}\underline{w}\cdot\underline{h}^{(\underline{w})}\right)$$

for these values of j. This expression depends only on $\underline{h}$ but not on x or P_j, and so we incorporate g_{t+r+1}, ..., g_{t+k-1} into the absolute value. We thus obtain the estimate

$$\mathop{\mathbb{E}}_{\underline{h},\underline{h}'} \left| \mathop{\mathbb{E}}_{x,y} \left(\prod_{0 \leqslant j \leqslant t+r-1} g_j(x+P_j(y)) \right) e_p\left(\psi(\underline{h},\underline{h}')(x+P_{t+r}(y))\right) \right|^2 \tag{2.19}$$
$$1_{\square(H_1)}(\underline{h},\underline{h}') \geqslant \left(\frac{\eta}{2}\right)^{2^{s-1}}.$$

We are now able to apply the induction hypothesis. By $\mathcal{S}(t,k,r)$, the expression inside the absolute values equals $O(p^{-c})$ unless $\psi(\underline{h},\underline{h}') = 0$. Therefore, the set

$$H_2 := \{(\underline{h},\underline{h}') \in \square(H_1) : \psi(\underline{h},\underline{h}') = 0\}$$

has size at least

$$\left(\left(\frac{\eta}{2}\right)^{2^{s-1}} - O(p^{-c}) \right) p^{2(s-2)}.$$

In particular, there exists $\underline{h} \in H_1$ such that the fiber

$$H_3 := \{\underline{h}' : (\underline{h},\underline{h}') \in H_2\}$$

has size at least

$$\left(\left(\frac{\eta}{2}\right)^{2^{s-1}} - O(p^{-c}) \right) p^{s-2}.$$

Fix this $\underline{h}$. We now show that the phases ϕ possess some amount of low-rank structure which we subsequently use to complete the proof of the lemma. By the definitions of H_2 and H_3, for each $\underline{h}' \in H_3$ we have $\psi(\underline{h},\underline{h}') = 0$. Define

$$\psi_i(\underline{h},\underline{h}') := (-1)^s \sum_{\substack{\underline{w} \in \{0,1\}^{s-2},\\ w_1 = \ldots = w_{i-1} = 1,\\ w_i = 0}} (-1)^{|w|} \phi(\underline{h}^{(\underline{w})}).$$

Note that, $\psi(\underline{h},\underline{h}') = \phi(h_1',...,h_{s-2}') - \psi_1(\underline{h},\underline{h}') - ... - \psi_{s-2}(\underline{h},\underline{h}')$. Crucially, ψ_i does not depend on $h_1',...,h_i'$. Thus, $\psi(\underline{h},\underline{h}') = 0$ implies that

$$\phi(h_1',...,h_{s-2}') = \sum_{i=1}^{s-2} \psi_i(\underline{h},\underline{h}').$$

That is to say, $\phi(h'_1, ..., h'_{s-2})$ can be decomposed into a sum of $s-2$ functions, each of which does not depend on h'_i for a different i.

To alleviate the pain that the reader may experience while struggling with the notation, we illustrate the aforementioned for $s = 3$ and 4. For $s = 3$,

$$\psi(h, h') = \phi(h) - \phi(h') = \psi_1(h) - \phi(h').$$

Hence $\psi(h, h') = 0$ implies that $\phi(h') = \phi(h)$. For $s = 4$,

$$\begin{aligned}\psi(\underline{h}, \underline{h}') &= \phi(h_1, h_2) - \phi(h'_1, h_2) - \phi(h_1, h'_2) + \phi(h'_1, h'_2) \\ &= \psi_1(\underline{h}, \underline{h}') - \psi_2(\underline{h}, \underline{h}') + \phi(h'_1, h'_2)\end{aligned}$$

and so $\psi(\underline{h}, \underline{h}') = 0$ implies that

$$\phi(h'_1, h'_2) = \phi(h_1, h'_2) + \phi(h'_1, h_2) - \phi(h_1, h_2) = \psi_2(\underline{h}, \underline{h}') - \psi_1(\underline{h}, \underline{h}').$$

We now estimate the expression

$$\mathop{\mathbb{E}}_{\underline{h}'} ||\Delta_{\underline{h}'} F||_{U^2}^4 1_{H_3}(\underline{h}') \tag{2.20}$$

from above and below. From below, it is bounded by

$$\frac{\eta}{2} \cdot \left(\left(\frac{\eta}{2}\right)^{2^{s-1}} - O(p^{-c}) \right) \geqslant \left(\frac{\eta}{2}\right)^{2^s} - O(p^{-c}).$$

The upper bound is more complicated, and it relies on the fact that we can decompose $\phi(\underline{h}')$ into a sum of ψ_i's such that ψ_i does not depend on h'_i. Using U^2-inverse theorem, (2.20) is bounded from above by:

$$\mathop{\mathbb{E}}_{\underline{h}'} \left| \widehat{\Delta_{\underline{h}'} F}(\phi(\underline{h}')) \right|^2 1_{H_3}(\underline{h}') = \mathop{\mathbb{E}}_{\underline{h}'} \left| \widehat{\Delta_{\underline{h}'} F} \left(\sum_{i=1}^{s-2} \psi_i(\underline{h}') \right) \right|^2 1_{H_3}(\underline{h}'). \tag{2.21}$$

By positivity, we can extend (2.21) to the entire $\mathbb{F}_p^{s-2}$; that is, we have

$$\mathop{\mathbb{E}}_{\underline{h}'} \left| \widehat{\Delta_{\underline{h}'} F} \left(\sum_{i=1}^{s-2} \psi_i(\underline{h}') \right) \right|^2 1_{H_3}(\underline{h}') \leqslant \mathop{\mathbb{E}}_{\underline{h}'} \left| \widehat{\Delta_{\underline{h}'} F} \left(\sum_{i=1}^{s-2} \psi_i(\underline{h}') \right) \right|^2. \tag{2.22}$$

Rewritting, we obtain that

$$\mathop{\mathbb{E}}_{\underline{h}'}\left|\widehat{\Delta_{\underline{h}'}F}\left(\sum_{i=1}^{s-2}\psi_i(\underline{h}')\right)\right|^2=\mathop{\mathbb{E}}_{\underline{h}'}\left|\mathop{\mathbb{E}}_{x}\Delta_{\underline{h}'}F(x)e_p\left(\sum_{i=1}^{s-2}\psi_i(\underline{h}')x\right)\right|^2$$
$$=\mathop{\mathbb{E}}_{x,\underline{h}',h_{s-1}}\Delta_{\underline{h}',h_{s-1}}F(x)e_p\left(\sum_{i=1}^{s-2}\psi_i(\underline{h}')h_{s-1}\right). \tag{2.23}$$

We apply Cauchy-Schwarz $s-2$ times to (2.23) to get rid of the phases $\psi_i(\underline{h}')$. In the first application, we start by rewriting (2.23) as

$$\mathop{\mathbb{E}}_{\substack{x,h_2',\dots,\\h_{s-2}',h_{s-1}}}\mathbf{b}(x,h_2',\dots,h_{s-2}',h_{s-1})\mathop{\mathbb{E}}_{h_1'}\Delta_{h_2',\dots,h_{s-2}',h_{s-1}}F(x+h_1')e_p\left(\sum_{i=2}^{s-2}\psi_i(\underline{h}')h_{s-1}\right) \tag{2.24}$$

and then we bound it by

$$\Big(\mathop{\mathbb{E}}_{x,h_1',h_1'',h_2',\dots,h_{s-2}',h_{s-1}}\Delta_{h_2',\dots,h_{s-2}',h_{s-1}}\left(F(x+h_1')\overline{F(x+h_1'')}\right) \tag{2.25}$$
$$e_p\left(\sum_{i=2}^{s-2}(\psi_i(h_1',h_2',\dots,h_{s-2}')-\psi_i(h_1'',h_2'\dots,h_{s-2}'))h_{s-1}\right)\Big)^{\frac{1}{2}}.$$

After repeatedly applying Cauchy-Schwarz in this manner, we get rid of all the phases and bound (2.25) by $||F||_{U^{s-1}}^2$. This proves the lemma. □

The second proof is simpler.

Proof of Lemma 2.4.5. By U^2-inverse theorem, $||F||_{U^2}^2\leqslant\max_{\alpha\in\mathbb{F}_p}|\hat{F}(\alpha)|$. By Lemma 2.4.1, this is $O(p^{-c})$ unless $\alpha=0$, in which case

$$\hat{F}(\alpha)=\Lambda_{t,P_t,\dots,P_{t+r-1},P_{t+r+1},\dots,P_{t+k-1}}(f_0,f_1,\dots,f_{t+r-1},f_{t+r+1},\dots,f_{t+k-1}).$$

Thus,

$$||F||_{U^2}^2\leqslant 1_{\alpha=0}|\Lambda_{t,P_t,\dots,P_{t+r-1},P_{t+r+1},\dots,P_{t+k-1}} \tag{2.26}$$
$$(f_0,f_1,\dots,f_{t+r-1},f_{t+r+1},\dots,f_{t+k-1})|+O(p^{-c})$$
$$\leqslant|\Lambda_t(f_0,\dots,f_{t-1})|\prod_{\substack{t\leqslant j\leqslant t+k-1,\\ j\neq t+r}}|\mathbb{E}\,f_j|+O(p^{-c})$$

where the second inequality follows from applying $\mathcal{S}(t,k-1)$. Recalling that $f_j(t)=e_p(a_jt)$ for $t+r+1\leqslant j\leqslant t+k-1$ and combining it with (2.26), we get that

$$||F||_{U^2}^2 \leqslant \min_{t\leqslant j\leqslant t+r-1} ||f_j||_{U^1} \cdot \prod_{j=t+r+1}^{t+k-1} 1_{a_j=0} + O(p^{-c}).$$

Taking square roots on both sides and applying Hölder's inequality proves the lemma. □

Next we prove Lemma 2.4.6 using the previous lemmas.

Proof of Lemma 2.4.6. We recall from the argument above Lemma 2.4.4 that

$$\begin{aligned}&\left|\Lambda_{t,P_t,\ldots,P_{t+k-1}}(f_0, \ldots, f_{t+k-1})\right|^2 \leqslant \|F\|_{L^2}^2 \\ &= \Lambda_{t,P_t,\ldots,P_{t+k-1}}(f_0, \ldots, f_{t+r-1}, \overline{F}, f_{t+r+1}, \ldots, f_{t+k-1}).\end{aligned}$$

By Lemma 2.4.3, there exist $s \in \mathbb{N}_+$ and $\beta > 0$ for which

$$\left|\Lambda_{t,P_t,\ldots,P_{t+k-1}}(f_0, \ldots, f_{t+r-1}, \overline{F}, f_{t+r+1}, \ldots, f_{t+k-1})\right| \leqslant \|F\|_{U^s}^{\beta} + O(p^{-\beta}).$$

Combining the above two inequalities with Lemmas 2.4.4 and 2.4.5, we deduce that

$$\begin{aligned}&\left|\Lambda_{t,P_t,\ldots,P_{t+k-1}}(f_0, \ldots, f_{t+r-1}, \overline{F}, f_{t+r+1}, \ldots, f_{t+k-1})\right| \\ &\leqslant \min_{t\leqslant i\leqslant t+r-1} \|f_i\|_{U^1}^c \prod_{j=t+r+1}^{t+k-1} 1_{a_j=0} + O(p^{-c}),\end{aligned}$$

for some $c > 0$ independent of $f_0, \ldots, f_{t+k-1}$, as claimed. □

2.5 Upper bounds for subsets of $\mathbb{F}_p$ lacking arithmetic progressions with k-th power common differences

We now switch gears, moving away from the progression (2.2) towards arithmetic progressions with common difference coming from the set of k-th powers. In this section, we prove Theorem 2.1.4 assuming Theorem 2.1.5. The argument goes much the same way as deriving Theorem 2.1.2 from Theorem 2.1.3.

First, we prove the following simple lemma which allows us to reduce to the case $k|p-1$.

Lemma 2.5.1. *Let $k \in \mathbb{N}_+$ and Q_k be the set of k-th power residues in $\mathbb{F}_p$. Then $Q_k = Q_{\gcd(k,p-1)}$.*

Proof. Since $\mathbb{F}_p^\times$ is a cyclic group under multiplication, we can write it as $\mathbb{F}_p^\times = \langle a | a^{p-1} = 1 \rangle$. Note that for each $k \in \mathbb{N}$, Q_k and $Q_{\gcd(k,p-1)}$ are subgroups of $\mathbb{F}_p^\times$ of cardinality $\frac{p-1}{\gcd(k,p-1)}$, generated respectively by a^k and $a^{\gcd(k,p-1)}$. The property $\gcd(k, p-1)|k$ moreover implies that Q_k is a subgroup of $Q_{\gcd(k,p-1)}$, and so they must be equal. $\square$

Proof of Theorem 2.1.4. The set of k-th powers in $\mathbb{F}_p$ is precisely Q_k, and by Lemma 2.5.1 it is the same as the set $Q_{\gcd(k,p-1)}$. Therefore we can assume that k divides $p-1$, otherwise we replace k with $\gcd(k, p-1)$. Suppose $A \subseteq \mathbb{F}_p$ for $p \geqslant p_0$ of size $|A| = \alpha p$ lacks t-term arithmetic progressions with difference coming from the set of k-th powers. From Theorem 2.1.5 it follows that

$$\begin{aligned} &\mathop{\mathbb{E}}_{x,y} 1_A(x)1_A(x+y)\cdots 1_A(x+(t-1)y)1_{Q_k}(y) \\ &= \frac{1}{k}\mathop{\mathbb{E}}_{x,y} 1_A(x)1_A(x+y)\cdots 1_A(x+(t-1)y) + O\left(p^{-c}\right). \end{aligned} \tag{2.27}$$

Since A lacks progressions with k-th power differences, the left-hand side of (2.27) is 0, and so we have

$$\mathop{\mathbb{E}}_{x,y} 1_A(x)1_A(x+y)\cdots 1_A(x+(t-1)y) = O\left(p^{-c}\right). \tag{2.28}$$

Applying Lemma 2.2.3 to (2.27) gives $M^{-2} \ll p^{-c}$ where $M = M_t(\frac{1}{2}\alpha)$ and M_t is the inverse function to s_t on $(0, \alpha_0]$, $\alpha_0 = s_t(p_0)$. Since M grows faster than polynomially in α^{-1} by Behrend's construction [Beh46], this gives $M_t \gg p^c$. Applying s_t to both sides and noting that s_t is decreasing, we obtain that $\alpha \leqslant 2s_t(Cp^c)$. $\square$

2.6 Counting theorem for the number of linear configurations in subsets of $\mathbb{F}_p$ with variables restricted to the set of k-th powers

This section is devoted to the proof of Theorem 2.1.5. We will first show that without loss of generality, we can assume that k_i divides $p-1$ for each $1 \leqslant i \leqslant d$. This will simplify the notation in the rest of the argument.

Lemma 2.6.1. *We have*

$$\begin{aligned}\mathop{\mathbb{E}}_{x_1,\dots,x_d}\prod_{i=1}^{t} f_j(L_i(x_1^{k_1},\dots,x_d^{k_d})) &= \mathop{\mathbb{E}}_{x_1,\dots,x_d}\prod_{i=1}^{t} f_j(L_i(x_1^{k_1'},\dots,x_d^{k_d'}))\\ &= k_1'\dots k_d' \mathop{\mathbb{E}}_{x_1,\dots,x_d}\prod_{i=1}^{t} f_j(L_i(x_1,\dots,x_d))\\ &\prod_{i=1}^{d} 1_{Q_{k_i'}}(x_i) + O\left(p^{-1}\right)\end{aligned}$$

where $k_i' := \gcd(k_i, p-1)$ *for each* $1 \leqslant i \leqslant d$.

Proof. By Lemma 2.5.1, $Q_k = Q_{\gcd(k,p-1)}$ for each $k \in \mathbb{N}_+$. Therefore the set of k_i-th power residues agrees with the set of k_i'-th power residues for each $1 \leqslant i \leqslant d$. Consequently, the set of tuples

$$\{(x_1^{k_1},\dots,x_d^{k_d}) : (x_1,\dots,x_d) \in \mathbb{F}_p^d\}$$

equals the set of tuples

$$\{(x_1^{k_1'},\dots,x_d^{k_d'}) : (x_1,\dots,x_d) \in \mathbb{F}_p^d\},$$

and moreover each tuple $(x_1^{k_1},\dots,x_d^{k_d})$ appears in $\mathbb{F}_p^d$ the same number of times as the tuple $(x_1^{k_1'},\dots,x_d^{k_d'})$. This implies the first equality, as the summations in both expressions are carried over the same sets of tuples the same number of times.

The second equality follows from the fact that each value of $y \in \mathbb{F}_p^\times$ equals $x_i^{k_i'}$ for precisely k_i' different values of $x_i \in \mathbb{F}_p$. The error term $O\left(p^{-1}\right)$ corresponds to the cases when at least one of the variables $x_1,\dots,x_d$ is 0. Using union bound, there are at most dp^{d-1} such cases, which together contribute at most $\frac{d}{p}$ to the expectation. □

We thus assume for the rest of this section that $k_1,\dots,k_d$ are coprime to $p-1$. With this assumption, we now describe a useful expression for $1_{Q_{k_i}}$ which is crucial in proving the error term in Theorem 2.1.5. Let a be a generator for the multiplicative group $\mathbb{F}_p^\times$. Define the map

$$\begin{aligned}\chi_{k_i} : \mathbb{F}_p^\times &\to \mathbb{C}\\ a^l &\mapsto e_{k_i}(l).\end{aligned}$$

The function χ_{k_i} is thus a *multiplicative character of order* k_i, i.e. a group homomorphism from $\mathbb{F}_p^\times$ to $\mathbb{C}^\times$ satisfying $\chi_{k_i}^{k_i} = 1$. We extend χ_{k_i} to $\mathbb{F}_p$ by setting $\chi_{k_i}(0) = 0$. Then χ_{k_i} picks out Q_{k_i}, in the sense that $\chi_{k_i}(x) = 1 \iff x \in Q_{k_i}$. Using the orthogonality of roots of unity, we can write

$$1_{Q_{k_i}}(x) = \frac{1 + \chi_{k_i}(x) + \chi_{k_i}(x)^2 + ... + \chi_{k_i}(x)^{k_i-1}}{k_i} - \frac{1}{k_i} 1_{\{0\}}(x). \tag{2.29}$$

We now use (2.29) to replace each $1_{Q_{k_i}}$ by a sum of characters in (2.5). Using the multilinearity of the operator, we obtain a main term of the same form as in (2.5), which corresponds to the terms in (2.29) having $1_{Q_{k_i}}$ replaced by $\frac{1}{k_i}$. Terms where $1_{Q_{k_i}}$ is replaced by $\frac{1}{k_i}1_{\{0\}}(x)$ are of size $O\left(p^{-1}\right)$, and there is a bounded number of them. It remains to deal with the terms that contain some $\frac{\chi_k^j(x)}{k}$ with $j > 0$ but have no $\frac{1}{k_i}1_{\{0\}}(x)$. Each such term is of the form

$$\mathop{\mathbb{E}}_{x_1,...,x_d} \prod_{i=1}^{t} f_j(L_i(x_1, ..., x_d)) \prod_{i \in S} \frac{\chi_{k_i}^{j_i}(x_i)}{k_i} \tag{2.30}$$

for a nonempty $S \subseteq \{1 \leqslant i \leqslant d : k_i > 1\}$ and $1 \leqslant j_i \leqslant k_i - 1$. From the fact that k_i divides d it follows that $\chi_{k_i}^{j_i}$ is also a character of order k_i, so without loss of generality we can take $j_i = 1$ for each $1 \leqslant i \leqslant d$.

Green and Tao proved that linear forms $L'_1(x_1, ..., x_d)$, ..., $L'_m(x_1, ..., x_d)$ are controlled by a Gowers norm [GT10; Tao12]: specifically, they showed that

$$\left| \mathop{\mathbb{E}}_{x_1,...,x_d} \prod_{j=1}^{t} f_j(L'_i(x_1, ..., x_d)) \right| \leqslant \min_{1 \leqslant j \leqslant t} ||f_j||_{U^s} \tag{2.31}$$

whenever for each $1 \leqslant i \leqslant t$ one can partition $\{L'_j : j \neq i\}$ into $s + 1$ classes such that L'_i does not lie in the span of each of them. The only case when such s may not exist is if two linear forms L'_i and L'_j are the same up to scaling. Otherwise we can partition linear forms into such classes: in the worst case, each of $\{L'_j : j \neq i\}$ forms a separate class, in which case $s = t - 2$. This extreme case occurs in arithmetic progressions, for instance: the operator

$$\mathop{\mathbb{E}}_{x,y} f_0(x) f_1(x + y) \cdots f_{t-1}(x + (t - 1)y)$$

is bounded by $||f_i||_{U^{t-1}}$ for each $0 \leqslant i \leqslant t - 1$.

We assumed specifically that no two linear forms L_i, L_j are scalar multiples, and that L_i is never a scalar multiple of e_j. From these assumptions we obtain the following lemma, which is essentially a restatement of Green and Tao's result tailored to our context.

Lemma 2.6.2. *For an arbitrary collection of characters χ_{k_i} of order k_i, we* *have the bound*

$$\left| \mathop{\mathbb{E}}_{x_1,\dots,x_d} \prod_{j=1}^{t} f_j(L_i(x_1,...,x_d)) \prod_{i\in S} \frac{\chi_{k_i}(x_i)}{k_i} \right| \leqslant \left(\prod_{i\in S} \frac{1}{k_i} \right) \min_{i\in S} ||\chi_{k_i}||_{U^s} \tag{2.32}$$

where $s-1$ is the CS-complexity of the system

$$\{L_1, ..., L_t\} \cup \{x_j : j \in S\} \tag{2.33}$$

In particular, one can take $s = t + |S| - 1 \leqslant t + d - 1$.

Proof. By assumption, all forms in the system

$$\{L_1, ..., L_t\} \cup \{x_j : 1 \leqslant j \leqslant d, k_j > 1\} \tag{2.34}$$

are pairwise linearly independent. Since (2.33) is a subset of (2.34), all forms in (2.33) are also pairwise linearly independent. Therefore the CS-complexity of this system is finite, and is at most $t + |S| - 2$ because the system (2.33) consists of $t + |S| - 1$ linear forms. □

It thus follows that the error term in (2.5) is controlled by Gowers norms of characters. The multiplicative property of characters makes it easy to bound their Gowers norms using tools such as Weil's bound, the following version of which we use.

Lemma 2.6.3 (Weil's bound, Corollary 11.24 of [IK04])**.** *Let χ be a nonprincipal multiplicative character of $\mathbb{F}_p$, and let $b_1, ..., b_{2r} \in \mathbb{F}_p$. If one of them is different from the others, then*

$$\left| \mathop{\mathbb{E}}_{x} \chi((x-b_1)\cdots(x-b_r))\overline{\chi}((x-b_{r+1})\cdots(x-b_{2r})) \right| \leqslant 2rp^{-\frac{1}{2}}.$$

With this corollary, we can easily estimate the Gowers norms of nonprincipal multiplicative characters.

Lemma 2.6.4 (Gowers norms of characters)**.** *If χ is a nonprincipal multiplicative character of $\mathbb{F}_p$ of order k and s is a natural number, then*

$$||\chi||_{U^s} \leqslant 2p^{-2^{-(s+1)}}.$$

The reader may also consult [FKM13] for a more general discussion of Gowers norms of functions on $\mathbb{F}_p$ of a strongly algebraic nature.

Proof. By definition, the U^s norm of χ is given by the following expression

$$\begin{aligned}
||\chi||_{U^s}^{2^s} &= \mathop{\mathbb{E}}_{h_1,\dots,h_s} \mathop{\mathbb{E}}_{x} \prod_{\underline{w}\in\{0,1\}^s} C^{|w|}\chi(x+\underline{w}\cdot\underline{h}) \\
&= \mathop{\mathbb{E}}_{h_1,\dots,h_s} \mathop{\mathbb{E}}_{x} \chi\left(\prod_{\substack{\underline{w}\in\{0,1\}^s,\\ |w| \text{ even}}} (x+\underline{w}\cdot\underline{h})\right) \overline{\chi}\left(\prod_{\substack{\underline{w}\in\{0,1\}^s,\\ |w| \text{ odd}}} (x+\underline{w}\cdot\underline{h})\right) \\
&\leqslant \mathop{\mathbb{E}}_{h_1,\dots,h_s} \left| \mathop{\mathbb{E}}_{x} \chi\left(\prod_{\substack{\underline{w}\in\{0,1\}^s,\\ |w| \text{ even}}} (x+\underline{w}\cdot\underline{h})\right) \overline{\chi}\left(\prod_{\substack{\underline{w}\in\{0,1\}^s,\\ |w| \text{ odd}}} (x+\underline{w}\cdot\underline{h})\right)\right|.
\end{aligned}$$

If $\underline{w}\cdot\underline{h}$ are not all equal, then by Lemma 2.6.3 we have

$$\left| \mathop{\mathbb{E}}_{x} \chi\left(\prod_{\underline{w}\in\{0,1\}^s, |w| \text{ even}} (x+\underline{w}\cdot\underline{h})\right) \overline{\chi}\left(\prod_{\underline{w}\in\{0,1\}^s, |w| \text{ odd}} (x+\underline{w}\cdot\underline{h})\right)\right| \leqslant 2^s p^{-\frac{1}{2}}.$$

The only possibility for $\underline{w}\cdot\underline{h}$ being equal for all $\underline{w}\in\{0,1\}^s$ is when $h_1 = \dots = h_s = 0$, which happens with probability p^{-s}. Thus

$$||\chi||_{U^s}^{2^s} \leqslant 2^s p^{-\frac{1}{2}} + p^{-s}$$

and so

$$||\chi||_{U^s} \ll p^{-2^{-(s+1)}}.$$

□

Applying the results of Lemma 2.6.4 to Lemma 2.6.2, we see that the error term in (2.5) is of the size $O\left(p^{-c}\right)$, which proves Theorem 2.1.5.

Acknowledgements

The author is indebted to Sean Prendiville for his unrelenting support, useful suggestions, inspiring discussions, and help with editing the paper, and to the anonymous referees for their suggestions on how to make the arguments simpler and clearer.

References

[Bal+94] A. Balog et al. "Difference sets without kth powers". In: *Acta Math. Hungar.* **65**.2 (1994), pp. 165–187.

[BC17] J. Bourgain and M.-C. Chang. “Nonlinear Roth type theorems in finite fields”. In: *Israel J. Math.* **221** (2017), pp. 853–867.

[Beh46] F. Behrend. “On sets of integers which contain no three terms in arithmetical progression”. In: *Proc. Natl. Acad. Sci. USA* **32** (12 1946), pp. 331–2.

[BL96] V. Bergelson and A. Leibman. “Polynomial extensions of van der Waerden’s and Szemerédi’s theorems”. In: *J. Amer. Math. Soc.* **9** (1996), pp. 725–753.

[Blo16] T. Bloom. “A quantitative improvement for Roth’s theorem on arithmetic progressions”. In: *J. Lond. Math. Soc.* **93** (2016), pp. 643–663.

[DLS20] D. Dong, X. Li, and W. Sawin. “Improved estimates for polynomial Roth type theorems in finite fields”. In: *J. Anal. Math.* **141** (2020), pp. 689–705.

[FKM13] É. Fouvry, E. Kowalski, and P. Michel. “An inverse theorem for Gowers norms of trace functions over $\mathbb{F}_p$”. In: *Math. Proc. Cambridge Philos. Soc.* **155**.2 (2013), pp. 277–295.

[Gow01] W. T. Gowers. “A new proof of Szemerédi’s theorem”. In: *Geom. Funct. Anal.* **11**.3 (2001), pp. 465–588.

[GR90] S. W. Graham and C. J. Ringrose. “Lower bounds for least quadratic non-residues”. In: *Analytic Number Theory. Progress in Mathematics* **85** (1990), pp. 269–309.

[Gre07] B. Green. “Montreal lecture notes on quadratic Fourier analysis”. In: *ArXiv e-prints* (2007). arXiv: 0604089.

[GT10] B. Green and T. Tao. “Linear equations in primes”. In: *Ann. of Math.* **171** (3 2010), pp. 1753–1850.

[GT17] B. Green and T. Tao. “New bounds for Szemerédi’s theorem, III: a polylogarithmic bound for $r_4(N)$”. In: *Mathematika* **63**.3 (2017), pp. 944–1040.

[IK04] H. Iwaniec and E. Kowalski. *Analytic Number Theory.* AMS, 2004.

[Kuc21] B. Kuca. “Further bounds in the polynomial Szemerédi theorem over finite fields”. In: *Acta Arith.* **198** (2021), pp. 77–108.

[Luc06] J. Lucier. “Intersective sets given by a polynomial”. In: *Acta Arith.* **123** (2006), pp. 57–95.

[Pel18] S. Peluse. "Three-term polynomial progressions in subsets of finite fields". In: *Israel J. Math.* **228** (1 2018), pp. 379–405.

[Pel19] S. Peluse. "On the polynomial Szemerédi theorem in finite fields". In: *Duke Math. J.* **168**.5 (2019), pp. 749–774.

[Pel20] S. Peluse. "Bounds for sets with no polynomial progressions". In: *Forum Math. Pi* **8** (e16 2020).

[PP19] S. Peluse and S. Prendiville. "Quantitative bounds in the non-linear Roth theorem". In: *ArXiv e-prints* (2019). arXiv: `1903.02592`.

[Pre] S. Prendiville. *Fourier methods in combinatorial number theory.* URL: `https://sites.google.com/view/web-add-comb/webinar-in-additive-combinatorics/lecture-series-fourier-methods-in-combinatorial-number-theory?authuser=0`.

[Pre17] S. Prendiville. "Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case". In: *Discrete Anal.* **5** (2017), 34 pp.

[Ric19] A. Rice. "A maximal extension of the best-known bounds for the Furstenberg-Sárközy theorem". In: *Acta Arith.* **187** (2019), pp. 1–41.

[Ruz84] I. Ruzsa. "Difference sets without squares". In: *Period. Math. Hungar.* **15**.3 (1984), pp. 205–209.

[RW19] L. Rimanić and J. Wolf. "Szemerédi's theorem in the primes". In: *Proc. Edinb. Math. Soc.* **62** (2 2019), pp. 443–457.

[Sár78a] A. Sárközy. "On difference sets of sequences of integers. I". In: *Acta Math. Hungar.* **31**.1-2 (1978), pp. 125–149.

[Sár78b] A. Sárközy. "On difference sets of sequences of integers. III". In: *Acta Math. Hungar.* **31** (1978), pp. 355–386.

[Sli03] S. Slijepcević. "A polynomial Sárközy-Furstenberg theorem with upper bounds". In: *Acta Math. Hungar.* **98**.1-2 (2003), pp. 111–128.

[Sze75] E. Szemerédi. "On sets of integers containing k elements in arithmetic progression". In: *Acta Arith.* **27**.1 (1975), pp. 199–245.

[Tao12] T. Tao. *Higher order Fourier analysis.* AMS, 2012.

[TZ08] T. Tao and T. Ziegler. "The primes contain arbitrarily long polynomial progressions". In: *Acta Math.* **201**.2 (2008), pp. 213–305.

[TZ16] T. Tao and T. Ziegler. “Concatenation theorems for anti-Gowers uniform functions and Host- Kra characteristic factors”. In: *Discrete Anal.* **13** (2016).

[TZ18] T. Tao and T. Ziegler. “Polynomial patterns in the primes”. In: *Forum Math. Pi* **6** (2018).

[Var59] P. Varnavides. “On certain sets of positive density”. In: *J. Lond. Math. Soc.* **s1-34**.3 (1959), pp. 358–360.

Blank Page

3 Multidimensional polynomial Szemerédi theorem in finite fields for polynomials of distinct degrees[1]

Abstract

We obtain a polynomial upper bound in the finite-field version of the multidimensional polynomial Szemerédi theorem for distinct-degree polynomials. That is, if $P_1, ..., P_t$ are nonconstant integer polynomials of distinct degrees and $\mathbf{v}_1, ..., \mathbf{v}_t$ are nonzero vectors in $\mathbb{F}_p^D$, we show that each subset of $\mathbb{F}_p^D$ lacking a nontrivial configuration of the form

$$(\mathbf{x},\ \mathbf{x}+\mathbf{v}_1 P_1(y),\ ...,\ \mathbf{x}+\mathbf{v}_t P_t(y))$$

has at most $O(p^{D-c})$ elements. In doing so, we apply the notion of Gowers norms along a vector adapted from ergodic theory, which extends the classical concept of Gowers norms on finite abelian groups.

3.1 Introduction

We prove the following bound in the finite field version of the multidimensional polynomial Szemerédi theorem of Bergelson and Leibman [BL96].

Theorem 3.1.1. *Let $D, t \in \mathbb{N}_+$, $\boldsymbol{v}_1, ..., \boldsymbol{v}_t \in \mathbb{Z}^D$ be nonzero vectors and $P_1, ..., P_t \in \mathbb{Z}[y]$ be polynomials satisfying $\deg P_1 < ... < \deg P_t$. There exist constants $c, C > 0$ and a threshold $p_0 \in \mathbb{N}$ such that for all primes $p > p_0$, each subset $A \subseteq \mathbb{F}_p^D$ of size at least Cp^{D-c} contains*

$$(\boldsymbol{x},\ \boldsymbol{x}+\boldsymbol{v}_1 P_1(y),\ ...,\ \boldsymbol{x}+\boldsymbol{v}_t P_t(y)) \tag{3.1}$$

[1] This chapter is an adapted version of B. Kuca. “Multidimensional polynomial Szemerédi theorem in finite fields for distinct-degree polynomials”. In: *ArXiv e-prints* (2021). arXiv: 2103. 12606.

for some $\boldsymbol{x} \in \mathbb{F}_p^D$ *and nonzero* $y \in \mathbb{F}_p$.

A special case of this statement is that each subset of $\mathbb{F}_p^2$ of size $\Omega(p^{2-c})$ contains a nontrivial configuration of the form

$$((x_1,\ x_2),\ (x_1+y,\ x_2),\ (x_1,\ x_2+y^2)), \tag{3.2}$$

previously proved in [HLY21], or a novel result that each subset of $\mathbb{F}_p^3$ of size $\Omega(p^{3-c})$ contains a nontrivial configuration of the form

$$((x_1,x_2,x_3),\ (x_1+y,x_2,x_3),\ (x_1,x_2+y^2,x_3),\ (x_1,x_2,x_3+y^3)). \tag{3.3}$$

A configuration (3.2) or (3.3) is nontrivial if $y \neq 0$.

Theorem 3.1.1 follows from the following result and its corollary.

Theorem 3.1.2. *Let* $D, t \in \mathbb{N}_+$, $\boldsymbol{v}_1, ..., \boldsymbol{v}_t \in \mathbb{Z}^D$ *be nonzero vectors and* $P_1, ..., P_t \in \mathbb{Z}[y]$ *be polynomials satisfying* $0 < \deg P_1 < ... < \deg P_t$. *There exists* $c > 0$ *and a threshold* $p_0 \in \mathbb{N}$ *such that for all primes* $p > p_0$ *and all 1-bounded functions* $f_0, ..., f_t : \mathbb{F}_p^D \to \mathbb{C}$, *we have*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{F}_p^D,\, y\in\mathbb{F}_p} f_0(\boldsymbol{x}) \prod_{i=1}^t f_i(\boldsymbol{x}+\boldsymbol{v}_i P_i(y)) = \mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{F}_p^D} f_0(\boldsymbol{x}) \prod_{i=1}^t \mathop{\mathbb{E}}_{n_i\in\mathbb{F}_p} f_i(\boldsymbol{x}+\boldsymbol{v}_i n_i) + O(p^{-c}).$$

Corollary 3.1.3. *Let* $D, t \in \mathbb{N}_+$, $\boldsymbol{v}_1, ..., \boldsymbol{v}_t \in \mathbb{Z}^D$ *be nonzero vectors and* $P_1, ..., P_t \in \mathbb{Z}[y]$ *be polynomials satisfying* $0 < \deg P_1 < ... < \deg P_t$. *There exists* $c > 0$ *and a threshold* $p_0 \in \mathbb{N}$ *such that for all primes* $p > p_0$ *and all nonnegative 1-bounded functions* $f : \mathbb{F}_p^D \to \mathbb{C}$, *we have*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{F}_p^D,\, y\in\mathbb{F}_p} f(\boldsymbol{x}) \prod_{i=1}^t f(\boldsymbol{x}+\boldsymbol{v}_i P_i(y)) \geqslant \left(\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{F}_p^D} f(\boldsymbol{x})\right)^{t+1} + O(p^{-c}).$$

A careful analysis of the proofs of Theorems 3.1.1 and 3.1.2 reveals that the bound in Theorem 3.1.1 and the error terms in Theorem 3.1.2 and Corollary 3.1.3 can be chosen uniformly for all the nonzero vectors $\mathbf{v}_1, ..., \mathbf{v}_t$. Similarly, these quantities do not depend on the specific form of the polynomials $P_1, ..., P_t$, only on their degrees. The threshold p_0 in both theorems does however depend on the vectors and the polynomials. We also remark that both results hold for $\mathbb{F}_q$ with q being a prime power, provided that the characteristic of $\mathbb{F}_q$ is sufficiently large in terms of the vectors $\mathbf{v}_1, ..., \mathbf{v}_t$ and the polynomials $P_1, ..., P_t$.

As an example, Theorem 3.1.2 implies that

$$\begin{aligned}&\mathop{\mathbb{E}}_{x_1,x_2,y\in\mathbb{F}_p} f_0(x_1,x_2)f_1(x_1+y,x_2)f_2(x_1,x_2+y^2)\\ &= \mathop{\mathbb{E}}_{\substack{x_1,x_2,\\ x_1',x_2'\in\mathbb{F}_p}} f_0(x_1,x_2)f_1(x_1',x_2)f_2(x_1,x_2') + O(p^{-c})\end{aligned}$$

for some $c>0$ uniformly in all 1-bounded functions $f_0,f_1,f_2:\mathbb{F}_p^2\to\mathbb{C}$. This particular statement has been proved in [HLY21] with an explicit constant $c=\frac{1}{8}$, but its natural analogue for (3.3) is novel:

$$\begin{aligned}&\mathop{\mathbb{E}}_{\substack{x_1,x_2,x_3,\\ y\in\mathbb{F}_p}} f_0(x_1,x_2,x_3)f_1(x_1+y,x_2,x_3)f_2(x_1,x_2+y^2,x_3)f_3(x_1,x_2,x_3+y^3)\\ &= \mathop{\mathbb{E}}_{\substack{x_1,x_2,x_3,\\ x_1',x_2',x_3'\in\mathbb{F}_p}} f_0(x_1,x_2,x_3)f_1(x_1',x_2,x_3)f_2(x_1,x_2',x_3)f_3(x_1,x_2,x_3') + O(p^{-c}).\end{aligned}$$

It then follows from Corollary 3.1.3 that if $A\subseteq\mathbb{F}_p^3$ has size $|A|=\alpha p^3$ for $\alpha\gg p^{-c/4}$, then

$$\left|\left\{\Big((x_1,x_2,x_3),(x_1+y,x_2,x_3),(x_1,x_2+y^2,x_3),(x_1,x_2,x_3+y^3)\Big)\in A^3\right\}\right|\gg\alpha^4p^{12}.$$

Theorems 3.1.1 and 3.1.2 extend results from [HLY21], which proves the same statements in the special case $t=2$, $D=2$, $\mathbf{v}_1=(1,0)$ and $\mathbf{v}_2=(0,1)$, i.e. for configurations of the form

$$((x_1,\ x_2),\ (x_1+P_1(y),\ x_2),\ (x_1,\ x_2+P_2(y))). \tag{3.4}$$

Theorems 3.1.1 and 3.1.2 also generalise results from the one-dimensional ($D=1$) case [BC17; DLS20; Pel18; Pel19; Kuc21]. Some of the results in the abovementioned papers also have integer analogues [Shk06a; Shk06b; Sár78a; Sár78b; Bal+94; Sli03; Luc06; Ric19; BM20; Pre17; PP19; PP20; Pel20]. In our paper, we develop multidimensional analogues of techniques pioneered in [Pel19] and later used in [Pel20; PP19; PP20; Kuc21], and we use the version of the PET induction scheme from [CFH11].

Notation

Throughout the paper, we fix $D\in\mathbb{N}_+$. We write elements of $\mathbb{F}_p^D$ as $\mathbf{x}=(x_1,...,x_D)$ and elements of $\mathbb{F}_p$ as x.

For a set X, we let $\mathbb{E}_{x\in X}=\frac{1}{|X|}\sum_{x\in X}$ denote the average over X. If $X=\mathbb{F}_p^D$ or $\mathbb{F}_p$, then we supress the mentioning of the set and let $\mathbb{E}_{\mathbf{x}}=\mathbb{E}_{\mathbf{x}\in\mathbb{F}_p^D}$ and $\mathbb{E}_x=\mathbb{E}_{x\in\mathbb{F}_p}$. Given a vector $\mathbf{v}_i\in\mathbb{F}_p^D$, we denote $V_i=\mathrm{Span}_{\mathbb{F}_p}\{\mathbf{v}_i\}$, and we define $\mathbb{E}(f|V_i)(\mathbf{x}) = \mathbb{E}_{\mathbf{x}+V_i} f = \mathbb{E}_y f(\mathbf{x}+\mathbf{v}_i y)$ to be the average of f along the coset $\mathbf{x}+V_i$. We also set $\mathbf{v}_0 = \mathbf{0}$ and $P_0 = 0$, and we call a function $f : \mathbb{F}_p^D \to \mathbb{C}$ *1-bounded* if $\|f\|_\infty = \max_{\mathbf{x}} |f(\mathbf{x})| \leqslant 1$.

We begin with specifying further pieces of notation used in this paper. For $k \in \mathbb{Z}$ and $P \in \mathbb{Z}[y]$, we set $\partial_k P(y) = P(y+k) - P(y)$, and for $f : \mathbb{F}_p^D \to \mathbb{C}$, we let $\Delta_{\mathbf{v}} f(\mathbf{x}) = f(\mathbf{x})\overline{f(\mathbf{x}+\mathbf{v})}$. We also let $\mathcal{C}z = \overline{z}$ be the conjugation operator. For $\underline{w} \in \{0,1\}^s$, we set $|w| = w_1 + ... + w_s$. Finally, the letter p denotes a sufficiently large prime, and we let $e_p(x) = e^{2\pi i x/p}$.

We use the asymptotic notation in the standard way. If $I \subseteq \mathbb{N}$ and $f, g : \mathbb{N} \to \mathbb{C}$, with g taking positive real values, we denote $f = O(g)$, $f \ll g$, $g \gg f$ or $g = \Omega(f)$ if there exists $C > 0$ such that $|f(n)| \leqslant Cg(n)$ for all $n \in I$. If the constant C depends on a paramter, we record this dependence with a subscript. All constants are allowed to depend on D, t, the polynomials $P_1, ..., P_t$ or the vectors $\mathbf{v}_1, ..., \mathbf{v}_t$, and this dependence is not recorded. An exception to this rule are results in Section 3.4, where we specify all the parameters that constants depend on.

Acknowledgments

We would like to thank Sean Prendiville for useful conversations and comments on an earlier version of this paper, and Donald Robertson for consultations on this project.

3.2 Gowers norms along a vector

To prove Conjecture 3.1.2, we need a notion of Gowers norms along a vector. Let $f : \mathbb{F}_p^D \to \mathbb{C}$, $s \in \mathbb{N}_+$ and $\mathbf{v} \in \mathbb{Z}^D$. We define the Gowers norm of f of degree s along $\mathbf{v}$ to be

$$\|f\|_{U^s(\mathbf{v})} = \left(\mathop{\mathbb{E}}_{\mathbf{x}, h_1, ..., h_s} \prod_{w \in \{0,1\}^s} \mathcal{C}^{|w|} f(\mathbf{x} + \mathbf{v}(w_1 h_1 + ... + w_s h_s)) \right)^{\frac{1}{2^s}}.$$

These norms are finitary analogues of Host-Kra seminorms from ergodic theory, corresponding to the transformation $T\mathbf{x} = \mathbf{x} + \mathbf{v}$ on $\mathbb{F}_p^D$. If $D = 1$, then the norm $U^s(\mathbf{v})$ equals the 1-dimensional Gowers norm U^s for any nonzero vector $\mathbf{v}$. If $D = 2$ and $\mathbf{v} = (1, 0)$, we have

$$\|f\|_{U^2(\mathbf{v})} = \left(\mathop{\mathbb{E}}_{x_1, x_2, h_1, h_2} f(x_1, x_2) \overline{f(x_1 + h_1, x_2) f(x_1 + h_2, x_2)} f(x_1 + h_1 + h_2, x_2) \right)^{\frac{1}{4}}.$$

Gowers norms along a vector satisfy a lot of the usual properties of Gowers norms. Letting $f_{\mathbf{x}}(n) = (\mathbf{x} + \mathbf{v}n)$, we can relate Gowers norms along a vector to 1-dimensional Gowers norms via the formula

$$\|f\|_{U^s(\mathbf{v})} = \left(\mathop{\mathbb{E}}_{\mathbf{x}} \|f_{\mathbf{x}}\|_{U^s}^{2^s}\right)^{\frac{1}{2^s}}. \tag{3.5}$$

The $U^1(\mathbf{v})$ norm is in fact a seminorm, and it is given by

$$\begin{aligned}\|f\|_{U^1(\mathbf{v})}^2 &= \mathop{\mathbb{E}}_{\mathbf{x},h} f(\mathbf{x})\overline{f(\mathbf{x}+\mathbf{v}h)} = \mathop{\mathbb{E}}_{\mathbf{x},h,k} f(\mathbf{x}+\mathbf{v}k)\overline{f(\mathbf{x}+\mathbf{v}h)} \\ &= \mathop{\mathbb{E}}_{\mathbf{x}}\left|\mathop{\mathbb{E}}_{h} f(\mathbf{x}+\mathbf{v}h)\right|^2 = \mathop{\mathbb{E}}_{\mathbf{x}}\left|\mathop{\mathbb{E}}_{\mathbf{x}+V} f\right|^2 = \|\mathbb{E}(f|V)\|_{L^2}^2,\end{aligned}$$

where $V = \mathrm{Span}\{\mathbf{v}\}$. Having large $U^1(\mathbf{v})$ norm thus tells us that f has large average on many cosets of $V = \mathrm{Span}_{\mathbb{F}_p}\{\mathbf{v}\}$. In particular, for $D = 2$ and $\mathbf{v} = (1, 0)$, we have

$$\|f\|_{U^1(\mathbf{v})}^2 = \mathop{\mathbb{E}}_{x_2}\left|\mathop{\mathbb{E}}_{x_1} f(x_1, x_2)\right|^2.$$

The identity $\|f\|_{U^1(\mathbf{v})} = \|\mathbb{E}(f|V)\|_{L^2}$ can be extended to higher values of s as follows: if f is V-measurable in the sense of being constant on cosets of V, then $\|f\|_{U^s(\mathbf{v})} = \|\mathbb{E}(f|V)\|_{L^{2^s}}$.

For $s \geqslant 2$, the seminorm $U^s(\mathbf{v})$ is a norm and satisfies the usual monotonicity property

$$\|f\|_{U^1(\mathbf{v})} \leqslant \|f\|_{U^2(\mathbf{v})} \leqslant \|f\|_{U^3(\mathbf{v})} \leqslant \dots$$

Both of these properties can be derived from the formula (3.5) and the corresponding properties for 1-dimensional Gowers norms. It is also straightforward to deduce from the definition that $U^s(\mathbf{v})$ satisfies the induction property

$$\|f\|_{U^s(\mathbf{v})} = \left(\mathop{\mathbb{E}}_{h_{k+1},\dots,h_s} \|\Delta_{\mathbf{v}h_{k+1},\dots,\mathbf{v}h_s} f\|_{U^k(\mathbf{v})}^{2^k}\right)^{\frac{1}{2^s}}.$$

We need a better understanding of the $U^2(\mathbf{v})$ norm. This norm can be related to Fourier analysis as follows. For $\mathbf{x}, \mathbf{v} \in \mathbb{F}_p^D$ and $k \in \mathbb{F}_p$, we define the Fourier transform of f along $\mathbf{v}$ as

$$\widehat{f}(\mathbf{x}; \mathbf{v}; k) = \mathop{\mathbb{E}}_{n} f(\mathbf{x} + \mathbf{v}n) e_p(-kn),$$

so that

$$f(\mathbf{x}+\mathbf{v}n) = \sum_{k\in\mathbb{F}_p} \widehat{f}(\mathbf{x};\mathbf{v};k)e_p(kn) \tag{3.6}$$

for any $n \in \mathbb{F}_p$.

In particular, letting $D = 2$ and $\mathbf{v} = (1,0)$, we get

$$\widehat{f}(\mathbf{x};\mathbf{v};k) = \mathop{\mathbb{E}}_{n} f(x_1+n, x_2)e_p(-kn),$$

We observe that $|\widehat{f}(\mathbf{x};\mathbf{v};k)| = |\widehat{f}(\mathbf{x}';\mathbf{v};k)|$ whenever $\mathbf{x}-\mathbf{x}' \in V$. With these definitions, we have

$$\|f\|_{U^2(\mathbf{v})}^4 = \mathop{\mathbb{E}}_{\mathbf{x},k} |\widehat{f}(\mathbf{x};\mathbf{v};k)|^4 \leqslant \mathop{\mathbb{E}}_{\mathbf{x}} |\widehat{f}(\mathbf{x};\mathbf{v},\phi(\mathbf{x}))|^2 \tag{3.7}$$

whenever f is 1-bounded, where $\phi(\mathbf{x})$ is an element of $\mathbb{F}_p$ for which

$$\max_k |\widehat{f}(\mathbf{x};\mathbf{v};k)| = |\widehat{f}(\mathbf{x};\mathbf{v};\phi(\mathbf{x}))|.$$

The Fourier transform can also be used to give an alternative description of the $U^1(\mathbf{v})$ norm; specifically,

$$\|f\|_{U^1(\mathbf{v})}^2 = \mathop{\mathbb{E}}_{\mathbf{x}} |\widehat{f}(\mathbf{x};\mathbf{v};0)|^2$$

since $\widehat{f}(\mathbf{x};\mathbf{v};0) = \mathbb{E}(f|V)(\mathbf{x})$.

We also need the following variant of the classical exponential sums estimates, which can be found e.g. in [Kow] as Theorem 3.2.

Lemma 3.2.1. *Let $P \in \mathbb{Z}[y]$ be a polynomial with $\deg P = d$ satisfying $1 \leqslant d < p$. Then*

$$\left|\mathop{\mathbb{E}}_{y} e_p(P(y))\right| \leqslant (d-1)p^{-1/2}.$$

3.3 The outline of the argument

We fix integers $0 \leqslant m \leqslant t$, nonzero vectors $\mathbf{v}_1, ..., \mathbf{v}_t \in \mathbb{Z}^D$, and polynomials $P_1, ..., P_t \in \mathbb{Z}[y]$ satisfying $0 < \deg P_1 < \deg P_2 < ... < \deg P_t$.

Roughly speaking, our proof of Theorem 3.1.1 goes by induction on t, and it follows the three-step strategy of [Pel19]. Like in [Pel19], we start by obtaining a global Gowers norm control on the operator

$$\mathop{\mathbb{E}}_{\mathbf{x},y}\prod_{i=0}^{t} f_i(\mathbf{x}+\mathbf{v}_i P_i(y)). \tag{3.8}$$

We then perform a degree-lowering argument to show that we can in fact control this operator by a U^1-type norm. Finally we use the properties of this norm to show that

$$\left|\mathop{\mathbb{E}}_{\mathbf{x},y}\prod_{i=0}^{t} f_i(\mathbf{x}+\mathbf{v}_i P_i(y)) - \mathop{\mathbb{E}}_{\mathbf{x}}\prod_{i=0}^{t}\mathbb{E}(f_i|V_i)(\mathbf{x})\right| \ll p^{-c}$$

for some $c>0$.

One difference between our argument and that of [Pel19] is that in contrast to the $D=1$ case, where we would control (3.8) by Gowers norms of each of the function $f_0,...,f_t$, in the $D>1$ case we can only bound (3.8) in terms of some Gowers norm $U^s(\mathbf{v}_t)$ of the function f_t. Moreover, obtaining such a bound is only possible under the extra assumption that $\deg P_t > \max(\deg P_1,...,\deg P_{t-1},0)$, whereas the $D=1$ case only requires linear independence of $P_1,...,P_t$. The PET induction procedure that produces such a bound has been developed in [CFH11], and we adapt the results of this paper in Section 3.4.

In the $D=1$ case, the U^1 Gowers norm is of "L^1 type", in the sense that $\|f\|_{U^1} = |\mathbb{E}_x f(x)| \leqslant \|f\|_{L^1}$ for any $f:\mathbb{F}_p\to\mathbb{C}$. However, in the $D>1$ case the $U^1(\mathbf{v})$ norm is of "L^2 type" for any nonzero vector $\mathbf{v}\in\mathbb{F}_p^D$, in the sense that $\|f\|_{U^1(\mathbf{v})} = \|\mathbb{E}(f|V)\|_{L^2}$. As a consequence, it turns out that we need to obtain a Gowers norm control of the L^2 norm of the function

$$G_t(\mathbf{x}) = \mathop{\mathbb{E}}_{y}\prod_{i=1}^{t} f_i(\mathbf{x}+\mathbf{v}_i P_i(y))$$

rather than the operator (3.8). Together with an application of the Cauchy-Schwarz inequality, this implies the Gowers norm control of (3.8).

To be able to perform induction on t, we need to consider more general operators

$$G_{m,t}(\mathbf{x}) = \mathop{\mathbb{E}}_{y}\prod_{i=1}^{m} f_i(\mathbf{x}+\mathbf{v}_i P_i(y)) \prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x})P_i(y))1_{\mathcal{U}}(\mathbf{x})$$

for some phase functions $\phi_{m+1},...,\phi_t:\mathbb{F}_p^D\to\mathbb{F}_p$ and $\mathcal{U}\subseteq\mathbb{F}_p^D$. By applying a trick from Lemma 5.12 of [Pre] and a variant of the PET induction procedure outlined in Section 3.4, we show that this operator is controlled by the $U^s(\mathbf{v}_m)$ of the dual function

$$F_{m,t}(\mathbf{x}) = \mathop{\mathbb{E}}_{y,k}\left(\prod_{i=1}^{m-1} f_i(\mathbf{x}+\mathbf{v}_iP_i(y)-\mathbf{v}_mP_m(y+k))\right.$$
$$\overline{f_i(\mathbf{x}+\mathbf{v}_iP_i(y+k)-\mathbf{v}_mP_m(y+k))})f_m(\mathbf{x}+\mathbf{v}_m(P_m(y)-P_m(y+k))$$
$$\left(\prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x}-\mathbf{v}_mP_m(y+k))\partial_k P_i(y))\right)1_{\mathcal{U}}(\mathbf{x}-\mathbf{v}_mP_m(y+k)).$$

A degree-lowering argument then shows us that the $U^s(\mathbf{v}_m)$ norm of $F_{m,t}$ can be bounded from above by a small power of $\|F_{m,t}\|_{U^1(\mathbf{v}_m)}$ and an error term $O(p^{-c})$. This norm can in turn be bounded from above by the norms $\|f_i\|_{U^1(\mathbf{v}_i)}$ for $1 \leqslant i \leqslant m$, from which we deduce that

$$\|G_{m,t}\|_{L^2}^2 = \mathop{\mathbb{E}}_{\mathbf{x}}\left|\prod_{i=1}^{m}\mathbb{E}(f_i|V_i)(\mathbf{x})\right|^2 \prod_{i=m+1}^{t} 1_{\phi_i(\mathbf{x})=0}1_{\mathcal{U}}(\mathbf{x}) + O(p^{-c})$$

Theorem 3.1.2 follows by taking $m=t$ and $\mathcal{U}=\mathbb{F}_p^D$.

The proof that the L^2 norm of $G_{m,t}$ is controlled by the norms $\|f_i\|_{U^1(\mathbf{v}_i)}$ for $1 \leqslant i \leqslant m$, with all the degree-lowering arguments, occupies the entirety of Section 3.5. In Section 3.6, we conclude the proof of Theorem 3.1.2 and use it to derive Theorem 3.1.1.

3.4 Controlling counting operators by Gowers norms

The material in this section follows closely Sections 4 and 5 of [CFH11]. We say that two nonconstant polynomials $P, Q \in \mathbb{Z}[y]$ are equivalent, denoted $P \sim Q$, if they have the same degree and the same highest-degree coefficient; equivalently, $P \sim Q$ iff $\deg(P-Q) < \min\{\deg P, \deg Q\}$.

Let $t, m \in \mathbb{Z}$ and $\mathcal{P}_j = (P_{j1}, ..., P_{jm}) \in \mathbb{Z}[y]^m$ for $1 \leqslant j \leqslant t$ and assume that at least one of $P_{j1}, ..., P_{jm}$ is nonconstant for each j. We want to determine when the operator

$$\mathop{\mathbb{E}}_{\mathbf{x},y}\prod_{i=1}^{m} f_i(\mathbf{x}+\mathbf{v}_1P_{1i}(y)+...+\mathbf{v}_tP_{ti}(y)) \tag{3.9}$$

is controlled by a Gowers norm for some nonzero vectors $\mathbf{v}_1, ..., \mathbf{v}_t \in \mathbb{Z}^D$, and the tuple $\mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_t)$ is a compact way of encoding information about the polynomials appearing in (3.9).

Let $d = \max_{j,i} \deg P_{ji}$. We define

$$\mathcal{P}'_j = \{P_{ji} : \deg P_{j'i} = 0 \text{ for } j < j' \leqslant t\},$$

and we let w_{jk} be the number of distinct equivalence classes of polynomials of degree k in $\mathcal{P}'_j$. The *type* of the family $(\mathcal{P}_1, ..., \mathcal{P}_t)$ is the matrix

$$\begin{pmatrix} w_{11} & \dots & w_{1d} \\ w_{21} & \dots & w_{2d} \\ \vdots & \dots & \vdots \\ w_{t1} & \dots & w_{td}. \end{pmatrix}$$

Given two $t \times d$ matrices $W = (w_{jk})$ and $W' = (w'_{jk})$, we order them in the reversed lexicographic way; that is, $W < W'$ if $w_{td} < w'_{td}$, or $w_{td} = w'_{td}$ and $w_{t(d-1)} < w'_{t(d-1)}$, ..., or $w_{tk} = w_{tk}$ for all $1 \leqslant k \leqslant d$ and $w_{(t-1)d} < w'_{(t-1)d}$, and so on.

The family $(\mathcal{P}_1, ..., \mathcal{P}_t)$ is *nice* if

1. $\deg P_{tm} \geqslant \deg P_{ti}$ for $1 \leqslant i \leqslant m$;
2. $\deg P_{tm} > \deg P_{ji}$ for $1 \leqslant j < t$ and $1 \leqslant i \leqslant m$;
3. $\deg(P_{tm} - P_{ti}) > \deg(P_{jm} - P_{ji})$ for $1 \leqslant j < t$ and $1 \leqslant i < m$.

The arguments from Sections 4 and 5 of [CFH11], after appropriate adaptations to the finite field setting, can be used to show the following.

Proposition 3.4.1. *Let $m, t, d \in \mathbb{N}_+$ and $\mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_t)$ be a nice family of polynomials of degree d. There exist $s \in \mathbb{N}_+$ and $C, c > 0$ depending only on m, t, d such that for any nonzero vectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_t \in \mathbb{Z}^D$ and any 1-bounded functions $f_1, ..., f_m : \mathbb{F}_p \to \mathbb{C}$, we have the bound*

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{x},y} \prod_{i=1}^m f_i(\boldsymbol{x} + \boldsymbol{v}_1 P_{1i}(y) + ... + \boldsymbol{v}_t P_{ti}(y)) \right| \leqslant \|f_m\|^c_{U^s(\boldsymbol{v}_t)} + Cp^{-c}.$$

Proof. Suppose first that $\deg P_{tm} = 1$. By the definition of nice families of polynomials, this means that $P_{ti}(y) = a_i y$ for distinct nonzero integers $a_{t1}, ..., a_{tm}$, and $P_{ji} = 0$ for all $1 \leqslant j < t$. We therefore end up with the expression

$$\mathop{\mathbb{E}}_{\mathbf{x},y} f_1(\mathbf{x} + \mathbf{v}_t a_1 y) \cdots f_m(\mathbf{x} + \mathbf{v}_t a_m y).$$

We let $f_{i,\mathbf{x}}(y) = f_i(\mathbf{x} + \mathbf{v}_t y)$, so that

$$\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\mathbf{x},y} f_1(\mathbf{x} + \mathbf{v}_t a_1 y) \cdots f_m(\mathbf{x} + \mathbf{v}_t a_m y)\right| \\
=&\left|\mathop{\mathbb{E}}_{\mathbf{x},n,y} f_1(\mathbf{x} + \mathbf{v}_t(n + a_1 y)) \cdots f_m(\mathbf{x} + \mathbf{v}_t(n + a_m y))\right| \\
\leqslant&\mathop{\mathbb{E}}_{\mathbf{x}}\left|\mathop{\mathbb{E}}_{n,y} f_{1,\mathbf{x}}(n + a_1 y) \cdots f_{m,\mathbf{x}}(n + a_m y)\right|.
\end{aligned}$$

Using the standard 1-dimensional estimates together with the distinctness of $a_1, ..., a_m$, the relation (3.5), and the Hölder inequality, we deduce that

$$\left|\mathop{\mathbb{E}}_{\mathbf{x},y} f_1(\mathbf{x} + \mathbf{v}_t a_1 y) \cdots f_m(\mathbf{x} + \mathbf{v}_t a_m y)\right| \leqslant \mathop{\mathbb{E}}_{\mathbf{x}} \|f_{m,\mathbf{x}}\|_{U^{m-2}} \leqslant \|f_m\|_{U^{m-2}(\mathbf{v}_t)}.$$

Suppose now that $\deg P_{tm} = d > 1$. We can assume that for each $1 \leqslant i \leqslant m$, the polynomial map $y \mapsto \mathbf{v}_1 P_{1i}(y) + ... + \mathbf{v}_t P_{ti}(y)$ is nonconstant, otherwise we incorporate f_i into f_0. We proceed in three steps. First, we apply the Cauchy-Schwarz inequality to bound

$$\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\mathbf{x},y} \prod_{i=1}^m f_i(\mathbf{x} + \mathbf{v}_1 P_{1i}(y) + ... + \mathbf{v}_t P_{ti}(y))\right|^2 \\
\leqslant&\left|\mathop{\mathbb{E}}_{\mathbf{x},y,h} \prod_{i=1}^m f_i(\mathbf{x} + \mathbf{v}_1 P_{1i}(y) + ... + \mathbf{v}_t P_{ti}(y))\overline{f_i(\mathbf{x} + \mathbf{v}_1 P_{1i}(y+h) + ... + \mathbf{v}_t P_{ti}(y+h))}\right|.
\end{aligned}$$

Second, we translate $\mathbf{x} \mapsto \mathbf{x} - \mathbf{v}_1 Q_1(y) - ... - \mathbf{v}_m Q_m(y)$ for appropriately chosen polynomials $Q_1, ..., Q_m \in \mathbb{Z}[y]$, set $\tilde{P}_{ji;h}(y) = P_{ji}(y+h) - Q_i(y)$ and use the Cauchy-Schwarz inequality to bound

$$\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\mathbf{x},y} \prod_{i=1}^m f_i(\mathbf{x} + \mathbf{v}_1 P_{1i}(y) + ... + \mathbf{v}_t P_{ti}(y))\right|^2 \\
\leqslant&\mathop{\mathbb{E}}_{h}\left|\mathop{\mathbb{E}}_{\mathbf{x},y} \prod_{i=1}^m f_i(\mathbf{x} + \mathbf{v}_1 \tilde{P}_{1i;0}(y) + ... + \mathbf{v}_t \tilde{P}_{ti;0}(y))\overline{f_i(\mathbf{x} + \mathbf{v}_1 \tilde{P}_{1i;h}(y) + ... + \mathbf{v}_t \tilde{P}_{ti;h}(y))}\right|.
\end{aligned} \tag{3.10}$$

We choose the polynomial $Q_1, ..., Q_m$ in such a way that for all except at most $m-1$ differences $h \in \mathbb{Z}$, the family $S_h\mathcal{P} = (S_h\mathcal{P}_1, ..., S_h\mathcal{P}_t)$, where $S_h\mathcal{P}_j = (\tilde{P}_{j1;0}, \tilde{P}_{j1;h}, ..., \tilde{P}_{jm;0}, \tilde{P}_{jm;h})$, has a type strictly smaller than $\mathcal{P}$. We do this as follows. Let $l = \min\{j : \mathcal{P}'_j \text{ is nonempty}\}$. If $l < t$, then we take Q_l to be a polynomial of the smallest degree in $\mathcal{P}'_l$ and set $Q_j = 0$ for $j \neq l$. If $l = t$, i.e. $\mathcal{P}'_1, ..., \mathcal{P}'_{t-1}$ are all empty, then we split into two cases. If $P_{ti} \sim P_{tm}$ for all $1 \leqslant i \leqslant m$, then we set $Q_j = P_{jm}$ for all $1 \leqslant j \leqslant t$. Otherwise we choose $1 \leqslant i \leqslant m$ such that P_{ti} has the smallest degree of all $P_{t1}, ..., P_{tm}$ and let $Q_j = P_{ji}$ for all $1 \leqslant j \leqslant t$. The assumption that $\mathcal{P}'_1, ..., \mathcal{P}'_{t-1}$ are empty implies that $P_{t1}, ..., P_{tm}$ are all nonconstant, and hence P_{ti} has positive degree.

By Lemmas 4.4 and 5.4 of [CFH11], for all except at most $m-1$ values of $h \in \mathbb{Z}$, the family $S_h\mathcal{P}$ is nice and has a strictly smaller type than $\mathcal{P}$. Finally, by Lemma 4.5 and 5.5 of [CFH11], there exists a constant $M = M(d, m, t)$ independent of the choice of the polynomial family $\mathcal{P}$ or vectors $\mathbf{v}_1, ..., \mathbf{v}_t$, such that after repeating the abovementioned procedure at most M times, we end up with the polynomial family of degree 1. Inducting on the type of $\mathcal{P}$ and applying the induction hypothesis to the right-hand side of (3.10), we deduce that there exists $c_0, C_0, s > 0$ depending only on m, d, t for which

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{\mathbf{x},y}\prod_{i=1}^m f_i(\mathbf{x}+\mathbf{v}_1 P_{1i}(y)+...+\mathbf{v}_t P_{ti}(y))\right| &\leqslant \left(\mathop{\mathbb{E}}_{h}\|f_m\|_{U^s}^{c_0}+C_0p^{-c_0}+(m-1)p^{-1}\right)^{\frac{1}{2}}\\ &\leqslant \|f_m\|_{U^s}^{c}+Cp^{-c}\end{aligned}$$

for some $c, C > 0$ that only depend on m, d, t. □

We will amply use the following corollary of Proposition 3.4.1, which plays the same role as Proposition 2.2 of [Pel19] in that paper.

Corollary 3.4.2. *Let $1 \leqslant m \leqslant t$ be integers, $P_1, ..., P_t \in \mathbb{Z}[y]$ be polynomials satisfying $0 < \deg P_1 < \deg P_2 < ... < \deg P_m$, and $\phi_{m+1}, ..., \phi_t : \mathbb{F}_p^D \to \mathbb{F}_p$. There exist $s \in \mathbb{N}_+$ and $c > 0$ such that for any nonzero vectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_m \in \mathbb{Z}^D$ and any 1-bounded functions $f_1, ..., f_m, g_1, ..., g_m : \mathbb{F}_p^D \to \mathbb{C}$, we have the bound*

$$\begin{aligned}&\left|\mathop{\mathbb{E}}_{\boldsymbol{x},y,k}\prod_{i=1}^m f_i(\boldsymbol{x}+\boldsymbol{v}_i P_i(y))\overline{g_i(\boldsymbol{x}+\boldsymbol{v}_i P_i(y+k))}\prod_{i=m+1}^t e_p(\phi_i(\boldsymbol{x})\partial_k P_i(y))\right|\\ &\leqslant \|g_m\|_{U^s(\boldsymbol{v_t})}^c+O(p^{-c}).\end{aligned}$$

Proof. By applying the Cauchy-Schwarz inequality in $\mathbf{x}, k$ to

$$\mathop{\mathbb{E}}_{\mathbf{x},y,k}\prod_{i=1}^m f_i(\mathbf{x}+\mathbf{v}_i P_i(y))\overline{g_i(\mathbf{x}+\mathbf{v}_i P_i(y+k))}\prod_{i=m+1}^t e_p(\phi_i(\mathbf{x})\partial_k P_i(y)) \tag{3.11}$$

in $\mathbf{x}$ and k and setting $k_1 = k$, we observe that

$$\begin{aligned}&\left|\mathop{\mathbb{E}}_{\mathbf{x},y,k}\prod_{i=1}^m f_i(\mathbf{x}+\mathbf{v}_i P_i(y))\overline{g_i(\mathbf{x}+\mathbf{v}_i P_i(y+k))}\prod_{i=m+1}^t e_p(\phi_i(\mathbf{x})\partial_k P_i(y))\right|^2\\ &\leqslant \mathop{\mathbb{E}}_{\mathbf{x},y,k_1,k_2}\prod_{i=1}^m\prod_{\underline{w}\in\{0,1\}^2}\mathcal{C}^{|w|}g_i(\mathbf{x}+\mathbf{v}_i P_i(y+w_1k_1+w_2k_2))\prod_{i=m+1}^t e_p(\phi_i(\mathbf{x})\partial_{k_1,k_2}P_i(y)).\end{aligned}$$

Importantly, the degree of the polynomial $\partial_{k_1,k_2} P_i$ in y is 1 less than the degree of $\partial_{k_1} P_i$. If $d = \max\{\deg P_{m+1}, ..., \deg P_t\}$, then we get rid of the phases $e_p(\phi_i(\mathbf{x})\partial_k P_i(y))$ by applying the Cauchy Schwarz inequality $d+1$ times to (3.11). Thus,

$$\left| \mathop{\mathbb{E}}_{\mathbf{x},y,k} \prod_{i=1}^m f_i(\mathbf{x}+\mathbf{v}_i P_i(y))\overline{g_i(\mathbf{x}+\mathbf{v}_i P_i(y+k))} \prod_{i=m+1}^t e_p(\phi_i(\mathbf{x})\partial_k P_i(y)) \right|^{2^{d+2}}$$
$$\leqslant \mathop{\mathbb{E}}_{\mathbf{x},y,k_1,...,k_{d+2}} \prod_{i=1}^m \prod_{\underline{w}\in\{0,1\}^{d+2}} \mathcal{C}^{|w|} g_i(\mathbf{x}+\mathbf{v}_i P_i(y+w_1k_1+...+w_{d+2}k_{d+2})).$$

For $(1-O(p^{-1}))p^{d+2}$ values of $(k_1, ..., k_{d+2})$, the expressions $(w_1k_1+...+w_{d+2}k_{d+2})_{w\in\{0,1\}^{d+2}}$ are all distinct. By the pigeonhole principle, there exists a tuple $(k_1, ..., k_{d+2}) \in \mathbb{F}_p^{d+2}$ satisfying this property for which moreover

$$\left| \mathop{\mathbb{E}}_{\mathbf{x},y,k} \prod_{i=1}^m f_i(\mathbf{x}+\mathbf{v}_i P_i(y))\overline{g_i(\mathbf{x}+\mathbf{v}_i P_i(y+k))} \prod_{i=m+1}^t e_p(\phi_i(\mathbf{x})\partial_k P_i(y)) \right|^{2^{d+2}}$$
$$\leqslant \mathop{\mathbb{E}}_{\mathbf{x},y} \prod_{i=1}^m \prod_{\underline{w}\in\{0,1\}^{d+2}} \mathcal{C}^{|w|} g_i(\mathbf{x}+\mathbf{v}_i P_i(y+w_1k_1+...+w_{d+2}k_{d+2})) + O(p^{-1}).$$

Let $P_{i;\underline{w}}(y) = P_i(y+w_1k_1+...+w_{d+2}k_{d+2})$. For every $\underline{w}, \underline{w}' \in \{0,1\}^{d+2}$, the polynomials $P_{i;\underline{w}}$ and $P_{i;\underline{w}'}$ are distinct and equivalent. Therefore the family $(\mathcal{P}_1, ..., \mathcal{P}_m)$ corresponding to the operator

$$\mathop{\mathbb{E}}_{\mathbf{x},y} \prod_{i=1}^m \prod_{\underline{w}\in\{0,1\}^{d+2}} \mathcal{C}^{|w|} g_{i;\underline{w}}(\mathbf{x}+\mathbf{v}_i P_{i;\underline{w}}(y))$$

is nice. The result then follows from Proposition 3.4.1. □

3.5 Degree lowering

In this section, we fix integers $0 \leqslant m \leqslant t$, nonzero vectors $\mathbf{v}_1, ..., \mathbf{v}_t \in \mathbb{Z}^D$, and polynomials $P_1, ..., P_t \in \mathbb{Z}[y]$ satisfying $0 < \deg P_1 < \deg P_2 < ... < \deg P_t$. The main result of this section is the proposition below, from which we deduce Theorem 3.1.2 in the next section. This result plays in our argument a similar part as Lemma 4.1 of [Pel19] in that paper.

Proposition 3.5.1. *There exists a constant $c > 0$ with the following property: for all 1-bounded functions $f_1, ..., f_m : \mathbb{F}_p^D \to \mathbb{F}_p$, phase functions $\phi_{m+1}, ..., \phi_t$:*

$\mathbb{F}_p^D \to \mathbb{F}_p$ and subsets $\mathcal{U} \subseteq \mathbb{F}_p^D$, the function

$$G_{m,t}(\boldsymbol{x}) = \mathop{\mathbb{E}}_{y} \prod_{i=1}^{m} f_i(\boldsymbol{x} + \boldsymbol{v}_i P_i(y)) \prod_{i=m+1}^{t} e_p(\phi_i(\boldsymbol{x}) P_i(y)) 1_{\mathcal{U}}(\boldsymbol{x}),$$

satisfies

$$\|G_{m,t}\|_{L^2}^2 = \mathop{\mathbb{E}}_{\boldsymbol{x}} \left| \prod_{i=1}^{m} \mathbb{E}(f_i|V_i)(\boldsymbol{x}) \right|^2 1_{\mathcal{U}'}(\boldsymbol{x}) + O(p^{-2c}),$$

where

$$\mathcal{U}' = \{\boldsymbol{x} \in \mathcal{U} : \phi_{m+1}(\boldsymbol{x}) = ... = \phi_t(\boldsymbol{x}) = 0\}.$$

In particular,

$$\|G_{m,t}\|_{L^2} \leqslant \min_{1 \leqslant i \leqslant m} \|f_i\|_{U^1(\boldsymbol{v}_i)} + O(p^{-c}),$$

if $t \geqslant 1$, and if $\|G_{m,t}\|_{L^2} \geqslant \delta \gg p^{-c}$, then $|\mathcal{U}'| = \Omega(\delta^2 p^D)$.

We prove Proposition 3.5.1 by induction on m. We start with the base case $m = 0$. If $t = 0$, then the statement is trivially true, otherwise it follows from Lemma 3.2.1. The proof for $m \in \mathbb{N}_+$ requires several technical lemmas which concern the properties of the dual function $F_{m,t}$.

Lemma 3.5.2. *Let $m \geqslant 1$, $f_1, ..., f_m : \mathbb{F}_p^D \to \mathbb{C}$ be 1-bounded, $\phi_{m+1}, ..., \phi_t : \mathbb{F}_p^D \to \mathbb{F}_p$ and $\mathcal{U} \subseteq \mathbb{F}_p^D$. Let*

$$\begin{aligned} F_{m,t}(\boldsymbol{x}) = \mathop{\mathbb{E}}_{y,k} \Bigg(&\prod_{i=1}^{m-1} f_i(\boldsymbol{x} + \boldsymbol{v}_i P_i(y) - \boldsymbol{v}_m P_m(y+k)) \\ &\overline{f_i(\boldsymbol{x} + \boldsymbol{v}_i P_i(y+k) - \boldsymbol{v}_m P_m(y+k))} f_m(\boldsymbol{x} + \boldsymbol{v}_m(P_m(y) - P_m(y+k)) \Bigg) \\ &\left(\prod_{i=m+1}^{t} e_p(\phi_i(\boldsymbol{x} - \boldsymbol{v}_m P_m(y+k)) \partial_k P_i(y)) \right) 1_{\mathcal{U}}(\boldsymbol{x} - \boldsymbol{v}_m P_m(y+k)). \end{aligned}$$

For each integer $s > 1$, there exists $c > 0$ independent of the choice of functions f_i, ϕ_i and the set $\mathcal{U}$, for which

$$\|F_{m,t}\|_{U^s(\boldsymbol{v}_m)} \ll \|F_{m,t}\|_{U^{s-1}(\boldsymbol{v}_m)}^c + p^{-c}.$$

Lemma 3.5.2 plays an analogous role in our argument to Proposition 6.6 of [PP19] in that paper, or Lemma 2.4.4 in Chapter 2. Multiple applications of Lemma 3.5.2 and the Hölder inequality give the following corollary.

Lemma 3.5.3. *Let $F_{m,t}$ be as in Lemma 3.5.2. For every $s \in \mathbb{N}_+$, there exists a constant $c > 0$ independent of the choice of functions f_i, ϕ_i and the set $\mathcal{U}$, for which*

$$\|F_{m,t}\|_{U^s(\boldsymbol{v}_m)} \ll \|F_{m,t}\|_{U^1(\boldsymbol{v}_m)}^c + p^{-c}.$$

Proof. The statement is trivially true for $s = 1$, so suppose that $s > 1$. Applying Lemma 3.5.2, we obtain that

$$\|F_{m,t}\|_{U^s(\mathbf{v}_m)} \ll \|F_{m,t}\|_{U^{s-1}(\mathbf{v}_m)}^{c_0} + p^{-c_0}. \tag{3.12}$$

By induction hypothesis, there exists $c_1 > 0$ for which

$$\|F_{m,t}\|_{U^s(\mathbf{v}_m)} \ll \|F_{m,t}\|_{U^1(\mathbf{v}_m)}^{c_1} + p^{-c_1}. \tag{3.13}$$

Combining (3.12) and (3.13) with the Hölder inequality, we get the result with $c = c_0 c_1$.

□

Finally, we show that the $\|F_{m,t}\|_{U^1(\mathbf{v}_m)}$ norm is bounded by the norms $\|f_1\|_{U^1(\mathbf{v}_1)}$, ..., $\|f_m\|_{U^1(\mathbf{v}_m)}$, a result analogous to Lemma 9 of [Kuc21].

Lemma 3.5.4. *Let $F_{m,t}$ be as in Lemma 3.5.2. There exists a constant $c > 0$ independent of the choice of functions f_i, ϕ_i and the set $\mathcal{U}$, for which $\|F_{m,t}\|_{U^1(\boldsymbol{v}_m)} \leqslant \min_{1 \leqslant i \leqslant m} \|f_i\|_{U^1(\boldsymbol{v}_i)} + O(p^{-c})$.*

Our induction scheme works as follows. For $m \in \mathbb{N}$, the (m,t) case of Proposition 3.5.1 is used to prove the $(m+1,t)$ cases of Lemmas 3.5.2 and 3.5.3 as well as the $(m+1,t+1)$ case of Lemma 3.5.4. It follows that once the (m,t) cases of Proposition 3.5.1 are proved for all $t \geqslant m$, the $(m+1,t)$ cases of Lemmas 3.5.2, 3.5.3 and 3.5.4 are proved for all $t \geqslant m+1$. The $(m+1,t)$ case of Proposition 3.5.1 is then derived with the help of the $(m+1,t)$ cases of Lemmas 3.5.3 and 3.5.4.

Proof of Proposition 3.5.1 in the case $m \geqslant 1$. We recall that

$$\begin{aligned}\|G_{m,t}\|_{L^2}^2 = \mathop{\mathbb{E}}_{\mathbf{x},y,k} \prod_{i=1}^m f_i(\mathbf{x} + \mathbf{v}_i P_i(y)) \overline{f_i(\mathbf{x} + \mathbf{v}_i P_i(y+k))} \\ \prod_{i=m+1}^t e_p(\phi_i(\mathbf{x}) \partial_k P_i(y)) 1_{\mathcal{U}}(\mathbf{x}).\end{aligned}$$

Translating $\mathbf{x} \mapsto \mathbf{x} - \mathbf{v}_m P_m(y+k)$, we observe that

$$\|G_{m,t}\|_{L^2}^2 = \langle F_{m,t}, f_m \rangle, \tag{3.14}$$

where

$$F_{m,t}(\mathbf{x}) = \mathop{\mathbb{E}}_{y,k} \Bigg(\prod_{i=1}^{m-1} f_i(\mathbf{x} + \mathbf{v}_i P_i(y) - \mathbf{v}_m P_m(y+k)) \\ \overline{f_i(\mathbf{x} + \mathbf{v}_i P_i(y+k) - \mathbf{v}_m P_m(y+k))}) f_m(\mathbf{x} + \mathbf{v}_m(P_m(y) - P_m(y+k)) \Bigg) \\ \prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x} - \mathbf{v}_m P_m(y+k)) \partial_k P_i(y)) 1_{\mathcal{U}}(\mathbf{x} - \mathbf{v}_m P_m(y+k)).$$

is as in the statement of Lemma 3.5.2. Applying the Cauchy-Schwarz inequality to (3.14), we obtain

$$\|G_{m,t}\|_{L^2}^4 \leqslant \|F_{m,t}\|_{L^2}^2 = \mathop{\mathbb{E}}_{\mathbf{x},y,k} \prod_{i=1}^{m-1} f_i(\mathbf{x} + \mathbf{v}_i P_i(y)) \overline{f_i(\mathbf{x} + \mathbf{v}_i P_i(y+k))} \\ f_m(\mathbf{x} + \mathbf{v}_m P_m(y)) \overline{F_{m,t}(\mathbf{x} + \mathbf{v}_m P_m(y+k))} \prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x}) \partial_k P_i(y)) 1_{\mathcal{U}}(\mathbf{x}).$$

We thank Sean Prendiville for showing us the trick that we have just used to bound $\|G_{m,t}\|_{L^2}$ in terms of $\|F_{m,t}\|_{L^2}$.

By Corollary 3.4.2 applied to the sum above, there exists $s \in \mathbb{N}_+$ and $0 < c_0 < 1$, independent from the choice of $f_1, ..., f_m, \phi_{m+1}, ..., \phi_t, \mathcal{U}$, such that

$$\|G_{m,t}\|_{L^2}^4 \leqslant \|F_{m,t}\|_{U^s(\mathbf{v}_m)}^{c_0} + O(p^{-c_0}). \tag{3.15}$$

We then apply Lemma 3.5.3 and Lemma 3.5.4, to bound

$$\|F_{m,t}\|_{U^s(\mathbf{v}_m)} \ll \min_{1 \leqslant i \leqslant m} \|f_i\|_{U^1(\mathbf{v}_i)}^{c_1} + p^{-c_1} \tag{3.16}$$

for some $0 < c_1 < 1$. Combining (3.15) and (3.16), letting $c = c_0 c_1 / 4$ and using the Hölder inequality, we get the bound

$$\|G_{m,t}\|_{L^2}^2 \ll \min_{1 \leqslant i \leqslant m} \|f_i\|_{U^1(\mathbf{v}_i)}^{2c} + p^{-2c}. \tag{3.17}$$

Splitting each $f_1, ..., f_m$ into $f_i = \mathbb{E}(f_i|V_i) + (f_i - \mathbb{E}(f_i|V_i))$, observing that

$\mathbb{E}(f_i|V_i)(\mathbf{x}+\mathbf{v}_iP_i(y)) = \mathbb{E}(f_i|V_i)(\mathbf{x})$ and using the bound (3.17), we deduce that

$$\|G_{m,t}\|_{L^2}^2 = \mathop{\mathbb{E}}_{\mathbf{x}} \left| \mathop{\mathbb{E}}_{y} \prod_{i=1}^{m} \mathbb{E}(f_i|V_i)(\mathbf{x}+\mathbf{v}_iP_i(y)) \prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x})P_i(y)) \right|^2 1_{\mathcal{U}}(\mathbf{x}) + O(p^{-2c})$$
$$= \mathop{\mathbb{E}}_{\mathbf{x}} \left| \prod_{i=1}^{m} \mathbb{E}(f_i|V_i)(\mathbf{x}) \right|^2 \cdot \left| \mathop{\mathbb{E}}_{y} \prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x})P_i(y)) \right|^2 1_{\mathcal{U}}(\mathbf{x}) + O(p^{-2c})$$

As a consequence of Lemma 3.2.1 applied to the inner average over y, we obtain

$$\|G_{m,t}\|_{L^2}^2 = \mathop{\mathbb{E}}_{\mathbf{x}} \prod_{i=1}^{m} |\,\mathbb{E}(f_i|V_i)(\mathbf{x})|^2 1_{\mathcal{U}'}(\mathbf{x}) + O(p^{-2c})$$

for

$$\mathcal{U}' = \{\mathbf{x} \in \mathcal{U} : \phi_{m+1}(\mathbf{x}) = ... = \phi_t(\mathbf{x}) = 0\}.$$

It follows from the 1-boundedness of $f_1, ..., f_m$ that $|\mathcal{U}'| \geqslant \delta^2 p^D$ whenever $\|G_{m,t}\|_{L^2} \geqslant \delta \gg p^{-c}$. The 1-boundedness of $f_1, ..., f_m$ and the Hölder inequality further imply that

$$\|G_{m,t}\|_{L^2} \leqslant \min_{1\leqslant i\leqslant m} \|f_i\|_{U^1(\mathbf{v}_i)} + O(p^{-c}).$$

$\square$

We now proceed to prove Lemma 3.5.2, which contains the bulk of the technicalities in this paper.

Proof of Lemma 3.5.2. We recall that

$$F_{m,t}(\mathbf{x}) = \mathop{\mathbb{E}}_{y,k} \Bigg(\prod_{i=0}^{m-1} f_i(\mathbf{x}+\mathbf{v}_iP_i(y) - \mathbf{v}_mP_m(y+k))$$
$$\overline{f_i(\mathbf{x}+\mathbf{v}_iP_i(y+k) - \mathbf{v}_mP_m(y+k))}) f_m(\mathbf{x}+\mathbf{v}_m(P_m(y) - P_m(y+k)) \Bigg)$$
$$\prod_{i=m+1}^{t} e_p\left(\phi_i(\mathbf{x} - \mathbf{v}_mP_m(y+k))\partial_k P_i(y)\right) 1_{\mathcal{U}}(\mathbf{x} - \mathbf{v}_mP_m(y+k)),$$

For simplicity, we set $F = F_{m,t}$ and $f_0 = 1_{\mathcal{U}}$ as well as recall that $\mathbf{v}_0 = \mathbf{0}$ and

$P_0 = 0$, so that

$$F(\mathbf{x}) = \mathop{\mathbb{E}}_{y,k}\left(\prod_{i=0}^{m-1} f_i(\mathbf{x}+\mathbf{v}_iP_i(y)-\mathbf{v}_mP_m(y+k))\right.$$
$$\left.\overline{f_i(\mathbf{x}+\mathbf{v}_iP_i(y+k)-\mathbf{v}_mP_m(y+k)))}f_m(\mathbf{x}+\mathbf{v}_m(P_m(y)-P_m(y+k))\right)$$
$$\prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x}-\mathbf{v}_mP_m(y+k))\partial_kP_i(y)).$$

We let $\delta = \|F\|_{U^s(\mathbf{v}_m)}$. We also denote $\underline{h} = (h_1, ..., h_{s-2})$ and $\mathbb{E}_{\underline{h}} = \mathbb{E}_{\underline{h}\in\mathbb{F}_p^{s-2}}$. From the induction formula for Gowers norms and the inequality (3.7), we deduce that

$$\delta^{2^s} = \|F\|_{U^s(\mathbf{v}_m)}^{2^s} = \mathop{\mathbb{E}}_{\underline{h}}\|\Delta_{\underline{h}\mathbf{v}_m}F\|_{U^2(\mathbf{v}_m)}^4 \leqslant \mathop{\mathbb{E}}_{\underline{h},\mathbf{x}}|\widehat{\Delta_{\underline{h}\mathbf{v}_m}F}(\mathbf{x};\mathbf{v}_m;\phi_m(\mathbf{x};\underline{h}))|^2$$

for some $\phi_m(\mathbf{x};\underline{h}) \in \mathbb{F}_p$. We can assume that $\phi_m(\mathbf{x};\underline{h})$ is the same for all $\mathbf{x}$ lying in the same coset of V_m since $|\widehat{\Delta_{\underline{h}\mathbf{v}_m}F}(\mathbf{x};\mathbf{v}_m;k)| = |\widehat{\Delta_{\underline{h}\mathbf{v}_m}F}(\mathbf{x};\mathbf{v}_m;k+n\mathbf{v}_m)|$ for all $k, n \in \mathbb{F}_p$. Thus, $\phi_m(\cdot;\underline{h})$ is V_m-measurable for each fixed $\underline{h}$. We let

$$H_1 = \{\underline{h}\in\mathbb{F}_p^{s-2} : \|\Delta_{\underline{h}\mathbf{v}_m}F\|_{U^2(\mathbf{v}_m)}^4 \geqslant \delta^{2^s}/2\}$$

and

$$\mathcal{U}_{\underline{h}} = \{\mathbf{x}\in\mathbb{F}_p^D : |\widehat{\Delta_{\underline{h}\mathbf{v}_m}F}(\mathbf{x};\mathbf{v}_m;\phi_m(\mathbf{x};\underline{h}))|^2 \geqslant \delta^{2^s}/4\}.$$

It follows from the popularity principle (Lemma 1.2.10) that

$$\begin{aligned}\delta^{2^s} &\ll \mathop{\mathbb{E}}_{\mathbf{x},\underline{h}}\left|\widehat{\Delta_{\underline{h}\mathbf{v}_m}F}(\mathbf{x};\mathbf{v}_m;\phi_m(\mathbf{x};\underline{h}))\right|^2 1_{\mathcal{U}}(\mathbf{x})1_{H_{\mathbf{x}}}(\underline{h})\\
&\leqslant \mathop{\mathbb{E}}_{\mathbf{x},\underline{h},n,n'}\Delta_{\underline{h}\mathbf{v}_m}F(\mathbf{x}+n\mathbf{v}_m)\overline{\Delta_{\underline{h}\mathbf{v}_m}F(\mathbf{x}+n'\mathbf{v}_m)}\\
&\quad e_p(\phi_m(\mathbf{x};\underline{h})(n'-n))1_{\mathcal{U}_{\underline{h}}}(\mathbf{x})1_{H_1}(\underline{h}).\end{aligned}$$

After expanding $\Delta_{\underline{h}\mathbf{v}_m}F(\mathbf{x}+n\mathbf{v}_m)$ and $\overline{\Delta_{\underline{h}\mathbf{v}_m}F(\mathbf{x}+n'\mathbf{v}_m)}$, the right-hand side of the above equals

$$\mathop{\mathbb{E}}_{\mathbf{x},n,n',\underline{h}} \mathop{\mathbb{E}}_{\underline{y},\underline{y}',\underline{k},\underline{k}'\in\mathbb{F}_p^{\{0,1\}^{s-2}}} \prod_{\underline{w}\in\{0,1\}^{s-2}} \mathcal{C}^{|w|}\left[\left(\prod_{i=0}^{m-1} f_i(\mathbf{x}+\mathbf{v}_iP_i(y_{\underline{w}})+\mathbf{v}_m(n+\underline{w}\cdot\underline{h}-P_m(y_{\underline{w}}+k_{\underline{w}}))\right)\right.$$

$$\overline{f_i(\mathbf{x}+\mathbf{v}_iP_i(y_{\underline{w}}+k_{\underline{w}})+\mathbf{v}_m(n+\underline{w}\cdot\underline{h}-P_m(y_{\underline{w}}+k_{\underline{w}})))}$$

$$\overline{f_i(\mathbf{x}+\mathbf{v}_iP_i(y'_{\underline{w}})+\mathbf{v}_m(n'+\underline{w}\cdot\underline{h}-P_m(y'_{\underline{w}}+k'_{\underline{w}})))}$$

$$f_i(\mathbf{x}+\mathbf{v}_iP_i(y'_{\underline{w}}+k'_{\underline{w}})+\mathbf{v}_m(n'+\underline{w}\cdot\underline{h}-P_m(y'_{\underline{w}}+k'_{\underline{w}}))))$$

$$f_m(\mathbf{x}+\mathbf{v}_m(n+\underline{w}\cdot\underline{h}+P_m(y_{\underline{w}})-P_m(y_{\underline{w}}+k_{\underline{w}})))$$

$$\left.\overline{f_m(\mathbf{x}+\mathbf{v}_m(n'+\underline{w}\cdot\underline{h}+P_m(y'_{\underline{w}})-P_m(y'_{\underline{w}}+k'_{\underline{w}})))}\right]$$

$$e_p\left(\phi_m(\mathbf{x};\underline{h})(n'-n)+\sum_{i=m+1}^{t}\sum_{\underline{w}\in\{0,1\}^{s-2}}(-1)^{|w|}(\phi_i(\mathbf{x}+\mathbf{v}_m(n+\underline{w}\cdot\underline{h}-P_m(y_{\underline{w}}+k_{\underline{w}})))\partial_{k_{\underline{w}}}P_i(y_{\underline{w}})\right.$$

$$\left.-\phi_i(\mathbf{x}+\mathbf{v}_m(n'+\underline{w}\cdot\underline{h}-P_m(y'_{\underline{w}}+k'_{\underline{w}})))\partial_{k'_{\underline{w}}}P_i(y'_{\underline{w}})\right)1_{\mathcal{U}_{\underline{h}}}(\mathbf{x})1_{H_1}(\underline{h}).$$

It does not suit us that the expression above contains a product of many copies of $f_1,...,f_m$ whose arguments include different y,y',k,k' variables. We want all the copies of $f_0,...,f_m$ to be expressed in the same y,y',k,k' variables. We shall achieve this by applying the Cauchy-Schwarz inequality $s-2$ times to the expression above. Letting $\underline{\tilde{h}}=(h_1,...,h_{s-3},h'_{s-2})$ and applying the Cauchy-Schwarz inequality in all variables except h_{s-2}, we bound the expression above by the square root of

$$\mathop{\mathbb{E}}_{\substack{\mathbf{x},n,n',h_1,\\ ...,h_{s-3},h_{s-2},h'_{s-2}}} \mathop{\mathbb{E}}_{\substack{\underline{y},\underline{y}',\underline{k},\\ \underline{k}'\in\mathbb{F}_p^{\{0,1\}^{s-2}}}} \prod_{\substack{\underline{w}\in\{0,1\}^{s-2},\\ w_s=1}} \mathcal{C}^{|w|}\left(\prod_{i=0}^{m}\tilde{f}_i\right) e_p\left((\phi_m(\mathbf{x};\underline{h})-\phi_m(\mathbf{x};\underline{\tilde{h}}))(n'-n)\right.$$

$$+\sum_{i=m+1}^{t}\sum_{\substack{\underline{w}\in\{0,1\}^{s-2}\\ w_s=1}}(-1)^{|w|}((\phi_i(\mathbf{x}+\mathbf{v}_m(n+\underline{w}\cdot\underline{h}-P_m(y_{\underline{w}}+k_{\underline{w}})))$$

$$-\phi_i(\mathbf{x}+\mathbf{v}_m(n+\underline{w}\cdot\underline{\tilde{h}}-P_m(y_{\underline{w}}+k_{\underline{w}}))))\partial_{k_{\underline{w}}}P_i(y_{\underline{w}})$$

$$-(\phi_i(\mathbf{x}+\mathbf{v}_m(n'+\underline{w}\cdot\underline{h}-P_m(y_{\underline{w}}+k_{\underline{w}})))$$

$$\left.-\phi_i(\mathbf{x}+\mathbf{v}_m(n'+\underline{w}\cdot\underline{\tilde{h}}-P_m(y_{\underline{w}}+k_{\underline{w}}))))\partial_{k'_{\underline{w}}}P_i(y'_{\underline{w}}))\right)1_{\mathcal{U}_{\underline{h}}}(\mathbf{x})1_{H_1}(\underline{h})1_{H_1}(\underline{\tilde{h}}),$$

where

$$\begin{aligned}
&\tilde{f}_i(\mathbf{x},\mathbf{y},\underline{k},n,n',\underline{w},h_1,...,h_{s-3},h_{s-2},h'_{s-2})\\
&= f_i(\mathbf{x}+\mathbf{v}_iP_i(y_{\underline{w}})+\mathbf{v}_m(n+\underline{w}\cdot\underline{h}-P_m(y_{\underline{w}}+k_{\underline{w}})))\\
&\overline{f_i(\mathbf{x}+\mathbf{v}_iP_i(y_{\underline{w}})+\mathbf{v}_m(n+\underline{w}\cdot\underline{\tilde{h}}-P_m(y_{\underline{w}}+k_{\underline{w}})))}\\
&\overline{f_i(\mathbf{x}+\mathbf{v}_iP_i(y_{\underline{w}}+k_{\underline{w}})+\mathbf{v}_m(n+\underline{w}\cdot\underline{h}-P_m(y_{\underline{w}}+k_{\underline{w}})))}\\
&f_i(\mathbf{x}+\mathbf{v}_iP_i(y_{\underline{w}}+k_{\underline{w}})+\mathbf{v}_{,}(n+\underline{w}\cdot\underline{\tilde{h}}-P_m(y_{\underline{w}}+k_{\underline{w}})))\\
&\overline{f_i(\mathbf{x}+\mathbf{v}_iP_i(y'_{\underline{w}})+\mathbf{v}_m(n'+\underline{w}\cdot\underline{h}-P_m(y'_{\underline{w}}+k'_{\underline{w}})))}\\
&f_i(\mathbf{x}+\mathbf{v}_iP_i(y'_{\underline{w}})+\mathbf{v}_m(n'+\underline{w}\cdot\underline{\tilde{h}}-P_m(y'_{\underline{w}}+k'_{\underline{w}})))\\
&f_i(\mathbf{x}+\mathbf{v}_iP_i(y'_{\underline{w}}+k'_{\underline{w}})+\mathbf{v}_m(n'+\underline{w}\cdot\underline{h}-P_m(y'_{\underline{w}}+k'_{\underline{w}})))\\
&\overline{f_i(\mathbf{x}+\mathbf{v}_iP_i(y'_{\underline{w}}+k'_{\underline{w}})+\mathbf{v}_m(n'+\underline{w}\cdot\underline{\tilde{h}}-P_m(y'_{\underline{w}}+k'_{\underline{w}})))}
\end{aligned}$$

for $0\leqslant i\leqslant m-1$ and

$$\begin{aligned}
&\tilde{f}_m(\mathbf{x},\mathbf{y},\underline{k},n,n',\underline{w},h_1,...,h_{s-3},h_{s-2},h'_{s-2})\\
&= f_m(\mathbf{x}+\mathbf{v}_m(n+\underline{w}\cdot\underline{h}+P_m(y_{\underline{w}})-P_m(y_{\underline{w}}+k_{\underline{w}})))\\
&\overline{f_m(\mathbf{x}+\mathbf{v}_m(n+\underline{w}\cdot\underline{\tilde{h}}+P_m(y_{\underline{w}})-P_m(y_{\underline{w}}+k_{\underline{w}})))}\\
&\overline{f_m(\mathbf{x}+\mathbf{v}_m(n'+\underline{w}\cdot\underline{h}+P_m(y'_{\underline{w}})-P_m(y'_{\underline{w}}+k'_{\underline{w}})))}\\
&f_m(\mathbf{x}+\mathbf{v}_m(n'+\underline{w}\cdot\underline{\tilde{h}}+P_m(y'_{\underline{w}})-P_m(y'_{\underline{w}}+k'_{\underline{w}}))).
\end{aligned}$$

Applying the Cauchy-Schwarz inequality another $s-3$ times, each time in all variables except h_{s-3} during the first application, h_{s-4} during the second application, etc., we obtain the bound

$$\begin{aligned}
\delta^{2^{2s-2}} \ll \mathop{\mathbb{E}}_{\mathbf{x},\underline{h},\underline{h}'}\Bigg|\mathop{\mathbb{E}}_{n,y,k}\left(\prod_{i=0}^{m-1} g_i(\mathbf{x}+\mathbf{v}_iP_i(y)+\mathbf{v}_mn)\overline{g_i(\mathbf{x}+\mathbf{v}_iP_i(y+k)+\mathbf{v}_mn)}\right)&\\
g_m(\mathbf{x}+\mathbf{v}_m(n+P_m(y)))e_p\Bigg(\psi_m(\mathbf{x};\underline{h},\underline{h}')(n+P_m(y+k))&\\
+\sum_{i=m+1}^{t}\psi_i(\mathbf{x}+\mathbf{v}_mn;\underline{h},\underline{h}')\partial_kP_i(y)\Bigg)\Bigg|^2 1_{\mathcal{U}_{(\underline{h},\underline{h}')}}(\mathbf{x})1_{H_2}(\underline{h},\underline{h}')&
\end{aligned}$$

where

$$\underline{h}_i^{(\underline{w})} = \begin{cases} h_i, & \text{if } w_i=0\\ h'_i, & \text{if } w_i=1\end{cases}$$

$$g_i(\mathbf{x}) := \prod_{\underline{w}\in\{0,1\}^{s-2}} \mathcal{C}^{|w|} f_i\left(\mathbf{x}+\mathbf{v}_m \underline{1}\cdot \underline{h}^{(\underline{w})}\right) \quad \text{with} \quad \underline{1}=(1,...,1)\in\mathbb{F}_p^{s-2},$$

$$\begin{aligned} H_2 &= \{(\underline{h},\underline{h}')\in\mathbb{F}_p^{2(s-2)} : \forall \underline{w}\in\{0,1\}^{s-2} \underline{h}^{(\underline{w})}\in H_1\}\\ \mathcal{U}_{(\underline{h},\underline{h}')} &= \bigcap_{\underline{w}\in\{0,1\}^{s-2}} \mathcal{U}_{\underline{h}^{(\underline{w})}}, \end{aligned}$$

and

$$\psi_i(\mathbf{x};\underline{h},\underline{h}') = \begin{cases} \sum\limits_{\underline{w}\in\{0,1\}^{s-2}} (-1)^{|w|}\phi_m(\mathbf{x};\underline{h}^{(\underline{w})}) & \text{for } i=m\\ \sum\limits_{\underline{w}\in\{0,1\}^{s-2}} (-1)^{|w|}\phi_i(\mathbf{x}+\mathbf{v}_m\underline{1}\cdot\underline{h}^{(\underline{w})}) & \text{for } m+1\leqslant i\leqslant t. \end{cases}$$

Applying the Cauchy-Schwarz inequality in y, n to the expectation inside the absolute value, performing minor changes of variables and recalling that $\phi_m(\cdot,\underline{h})$ and $\mathcal{U}_{(\underline{h},\underline{h}')}$ are V_m-measurable, we get the bound

$$\begin{aligned} \delta^{2^{2s-2}} \ll & \mathop{\mathbb{E}}_{\mathbf{x},\underline{h},\underline{h}',n,y,k} \left(\prod_{i=0}^{m-1} g_i(\mathbf{x}+\mathbf{v}_iP_i(y)+\mathbf{v}_m n)\overline{g_i(\mathbf{x}+\mathbf{v}_iP_i(y+k)+\mathbf{v}_m n)}\right)\\ & e_p\left(\sum_{i=m}^{t}\psi_i(\mathbf{x}+\mathbf{v}_m n;\underline{h},\underline{h}')\partial_k P_i(y)\right) 1_{\mathcal{U}_{(\underline{h},\underline{h}')}}(\mathbf{x})1_{H_2}(\underline{h},\underline{h}')\\ & = \mathop{\mathbb{E}}_{\mathbf{x},\underline{h},\underline{h}'}\left|\mathop{\mathbb{E}}_{y}\prod_{i=0}^{m-1} g_i(\mathbf{x}+\mathbf{v}_iP_i(y))e_p\left(-\sum_{i=m}^{t}\psi_i(\mathbf{x};\underline{h},\underline{h}')P_i(y)\right)\right|^2\\ & 1_{\mathcal{U}_{(\underline{h},\underline{h}')}}(\mathbf{x})1_{H_2}(\underline{h},\underline{h}'). \end{aligned}$$

We then use the 1-boundedness of g_0 and the fact that $g_0(\mathbf{x}+\mathbf{v}_0P_0(y)) = g_0(\mathbf{x})$ is independent of y to conclude that

$$\delta^{2^{2s-2}} \ll \mathop{\mathbb{E}}_{\mathbf{x},\underline{h},\underline{h}'}\left|\mathop{\mathbb{E}}_{y}\prod_{i=1}^{m-1} g_i(\mathbf{x}+\mathbf{v}_iP_i(y))e_p\left(-\sum_{i=m}^{t}\psi_i(\mathbf{x};\underline{h},\underline{h}')P_i(y)\right)\right|^2 1_{\mathcal{U}_{(\underline{h},\underline{h}')}}(\mathbf{x})1_{H_2}(\underline{h},\underline{h}').$$

By the popularity principle, the set

$$\begin{aligned} H_3 = \{&(\underline{h},\underline{h}')\in H_2 :\\ &\mathop{\mathbb{E}}_{\mathbf{x}}\left|\mathop{\mathbb{E}}_{y}\prod_{i=1}^{m-1} g_i(\mathbf{x}+\mathbf{v}_iP_i(y))e_p\left(-\sum_{i=m}^{t}\psi_i(\mathbf{x};\underline{h},\underline{h}')P_i(y)\right)\right|^2 1_{\mathcal{U}_{(\underline{h},\underline{h}')}}(\mathbf{x}) \gg \delta^{2^{2s-2}}\} \end{aligned}$$

has $\Omega(\delta^{2^{2s-2}}p^{2s-4})$ elements. In particular, there exists $\underline{h}\in H_2$ for which the fiber

$$H_4 := \{\underline{h}' : (\underline{h}, \underline{h}') \in H_3\}$$

has $\Omega(\delta^{2^{2s-2}} p^{s-2})$ elements. We fix such $\underline{h}$.

Applying Proposition 3.5.1 in the case $(m-1, t)$, we conclude that for each $\underline{h}' \in H_4$, the set

$$\mathcal{U}'_{\underline{h}'} = \{\mathbf{x} \in \mathcal{U}_{(\underline{h}, \underline{h}')} : \psi_m(\mathbf{x}; \underline{h}, \underline{h}') = 0\}$$

has $\Omega(\delta^{2^{2s-2}} p^D)$ elements as long as $\delta \gg p^{-c_1}$ for a constant $c_1 > 0$ given by the case $(m-1, t)$ of Proposition 3.5.1.

We now show that the phases ψ_m possess some linear structure that we subsequently use to complete the proof. We define

$$\eta_i(\mathbf{x}; \underline{h}, \underline{h}') := (-1)^{s-1} \sum_{\substack{\underline{w} \in \{0,1\}^{s-2}, \\ w_1 = \ldots = w_{i-1} = 1, \\ w_i = 0}} (-1)^{|w|} \phi(\mathbf{x}; \underline{h}^{(\underline{w})}),$$

so that

$$\psi_m(\mathbf{x}; \underline{h}, \underline{h}') = (-1)^s \left(\phi_m(\mathbf{x}; \underline{h}') - \eta_1(\mathbf{x}; \underline{h}, \underline{h}') - \ldots - \eta_{s-2}(\mathbf{x}; \underline{h}, \underline{h}')\right).$$

Crucially, η_i does not depend on $h'_1, \ldots, h'_i$. Thus, $\psi(\mathbf{x}; \underline{h}, \underline{h}') = 0$ implies that

$$\phi_m(\mathbf{x}; \underline{h}') = \sum_{i=1}^{s-2} \eta_i(\mathbf{x}; \underline{h}, \underline{h}').$$

That is to say, $\phi_m(\mathbf{x}; \underline{h}')$ can be decomposed into a sum of $s-2$ functions, each of which does not depend on h'_i for a different i.

We illustrate the aforementioned definitions for $s = 3$ and 4. For $s = 3$,

$$\psi_m(\mathbf{x}; h, h') = \phi_m(\mathbf{x}; h) - \phi_m(\mathbf{x}; h') = \eta_1(\mathbf{x}; h) - \phi_m(\mathbf{x}; h').$$

Hence $\psi_m(\mathbf{x}; h, h') = 0$ implies that $\phi_m(\mathbf{x}; h') = \phi_m(\mathbf{x}; h)$. For $s = 4$,

$$\begin{aligned}\psi_m(\mathbf{x}; \underline{h}, \underline{h}') &= \phi_m(\mathbf{x}; h_1, h_2) - \phi_m(\mathbf{x}; h_1, h'_2) - \phi_m(\mathbf{x}; h'_1, h_2) + \phi_m(\mathbf{x}; h'_1, h'_2) \\ &= -\eta_1(\mathbf{x}; \underline{h}, \underline{h}') - \eta_2(\mathbf{x}; \underline{h}, \underline{h}') + \phi_m(\mathbf{x}; h'_1, h'_2)\end{aligned}$$

and so $\psi_m(\mathbf{x};\underline{h},\underline{h}')=0$ implies that

$$\begin{aligned}\phi_m(\mathbf{x};h_1',h_2') &= -\phi_m(\mathbf{x};h_1,h_2)+\phi_m(\mathbf{x};h_1,h_2')+\phi_m(\mathbf{x};h_1',h_2)\\ &= \eta_1(\mathbf{x};\underline{h},\underline{h}')+\eta_2(\mathbf{x};\underline{h},\underline{h}').\end{aligned}$$

To bound the $U^s(\mathbf{v}_m)$ norm of F by its $U^{s-1}(\mathbf{v}_m)$ norm, we estimate the expression

$$\mathop{\mathbb{E}}_{\underline{h}',\mathbf{x}}\left|\widehat{\Delta_{\underline{h}'\mathbf{v}_m}F}(\phi_m(\mathbf{x};\underline{h}'))\right|^2 1_{\mathcal{U}'_{\underline{h}'}}(\mathbf{x})1_{H_4}(\underline{h}') \tag{3.18}$$

from above and below. For each $\underline{h}'\in H_4$ and $\mathbf{x}\in\mathcal{U}'_{\underline{h}'}$, we have

$$\left|\widehat{\Delta_{\underline{h}'\mathbf{v}_m}F}(\phi_m(\mathbf{x};\underline{h}'))\right|^2\gg\delta^{2^s}.$$

Together with the lower bounds on the size of H_4 and $\mathcal{U}'_{\underline{h}'}$ whenever $\underline{h}'\in H_4$, we deduce that (3.18) is bounded from below by $\Omega(\delta^{4^s})$.

The upper bound is more complicated, and it relies on the fact that we can decompose $\phi_m(\underline{h}')$ into a sum of η_i's such that η_i does not depend on h_i'. From the definitions of H_4 and $\mathcal{U}'_{\underline{h}'}$ it follows that

$$\begin{aligned}&\mathop{\mathbb{E}}_{\underline{h}',\mathbf{x}}\left|\widehat{\Delta_{\underline{h}'\mathbf{v}_m}F}(\phi_m(\mathbf{x};\underline{h}'))\right|^2 1_{\mathcal{U}'_{\underline{h}'}}(\mathbf{x})1_{H_4}(\underline{h}')\\ &=\mathop{\mathbb{E}}_{\underline{h}',\mathbf{x}}\left|\widehat{\Delta_{\underline{h}'\mathbf{v}_m}F}\left(\sum_{i=1}^{s-2}\eta_i(\mathbf{x};\underline{h},\underline{h}')\right)\right|^2 1_{\mathcal{U}'_{\underline{h}'}}(\mathbf{x})1_{H_4}(\underline{h}').\end{aligned} \tag{3.19}$$

By positivity, we can extend (3.19) to the entire $\mathbb{F}_p^{s-2}$; that is, we have

$$\mathop{\mathbb{E}}_{\underline{h}',\mathbf{x}}\left|\widehat{\Delta_{\underline{h}'\mathbf{v}_m}F}\left(\sum_{i=1}^{s-2}\eta_i(\mathbf{x};\underline{h},\underline{h}')\right)\right|^2 1_{\mathcal{U}'_{\underline{h}'}}(\mathbf{x})1_{H_4}(\underline{h}')\leqslant\mathop{\mathbb{E}}_{\underline{h}',\mathbf{x}}\left|\widehat{\Delta_{\underline{h}'\mathbf{v}_m}F}\left(\sum_{i=1}^{s-2}\eta_i(\mathbf{x};\underline{h},\underline{h}')\right)\right|^2.$$

Rewritting, we obtain that

$$\begin{aligned}\mathop{\mathbb{E}}_{\underline{h}',\mathbf{x}}\left|\widehat{\Delta_{\underline{h}'\mathbf{v}_m}F}\left(\sum_{i=1}^{s-2}\eta_i(\mathbf{x};\underline{h},\underline{h}')\right)\right|^2 &= \mathop{\mathbb{E}}_{\underline{h}',\mathbf{x}}\left|\mathop{\mathbb{E}}_{n}\Delta_{\underline{h}'\mathbf{v}_m}F(\mathbf{x}+n\mathbf{v}_m)e_p\left(\sum_{i=1}^{s-2}\eta_i(\mathbf{x};\underline{h},\underline{h}')n\right)\right|^2\\ &=\mathop{\mathbb{E}}_{\underline{h}',\mathbf{x},n,k}\Delta_{\underline{h}'\mathbf{v}_m,k\mathbf{v}_m}F(\mathbf{x}+n\mathbf{v}_m)e_p\left(\sum_{i=1}^{s-2}\eta_i(\mathbf{x};\underline{h},\underline{h}')k\right).\end{aligned} \tag{3.20}$$

We apply the Cauchy-Schwarz inequality $s-2$ times to (3.20) to get rid of the phases $\eta_i(\mathbf{x};\underline{h},\underline{h}')$. In the first application, we apply the inequality in all variables but h'_1, thus bounding (3.20) by

$$\mathop{\mathbb{E}}_{\substack{h'_2,\dots,h'_{s-2},\\ \mathbf{x},n,k}} \left| \mathop{\mathbb{E}}_{h'_1} \Delta_{h'_2\mathbf{v}_m,\dots,h'_{s-2}\mathbf{v}_m,k\mathbf{v}_m} F(\mathbf{x}+\mathbf{v}_m(n+h'_1)) e_p\left(\sum_{i=2}^{s-2} \eta_i(\mathbf{x};\underline{h},\underline{h}')k\right)\right|^2 \tag{3.21}$$

$$= \mathop{\mathbb{E}}_{\substack{h'_1,h''_1,h'_2,\dots,h'_{s-2},\\ \mathbf{x},n,k}} \Delta_{h'_2\mathbf{v}_m,\dots,h'_{s-2}\mathbf{v}_m,k\mathbf{v}_m}\left(F(\mathbf{x}+\mathbf{v}_m(n+h'_1))\overline{F(\mathbf{x}+\mathbf{v}_m(n+h''_1))}\right)$$
$$e_p\left(\sum_{i=2}^{s-2}(\psi_i(\mathbf{x};\underline{h},\underline{h}') - \psi_i(\mathbf{x};\underline{h},\tilde{\underline{h}}'))k\right),$$

where $\tilde{\underline{h}}' = (h''_1, h'_2, \dots, h'_{s-2})$. After repeatedly applying the Cauchy-Schwarz inequality in this manner, we get rid of all the phases and bound (3.21) by $||F||^2_{U^{s-1}(\mathbf{v}_m)}$. Thus, $||F||_{U^{s-1}(\mathbf{v}_m)} \gg \delta^{2^{2s-1}}$ as long as $\delta \gg p^{-c_1}$. Taking $c = \min(c_1, 1/2^{2s-1})$, it follows that

$$||F||_{U^s(\mathbf{v}_m)} \ll ||F||^c_{U^{s-1}(\mathbf{v}_m)} + p^{-c}.$$

□

Proof of Lemma 3.5.4. We set $F = F_{m,t}$. By definition,

$$\|F\|^2_{U^1(\mathbf{v}_m)} = \mathop{\mathbb{E}}_{\mathbf{x}}\left|\mathop{\mathbb{E}}_{\mathbf{x}+V_m} F\right|^2,$$

where

$$\mathop{\mathbb{E}}_{\mathbf{x}+V_m} F = \mathop{\mathbb{E}}_{n} F(\mathbf{x}+\mathbf{v}_m n) = \mathop{\mathbb{E}}_{n,y,k}\left(\prod_{i=1}^{m-1} f_i(\mathbf{x}+\mathbf{v}_i P_i(y)+\mathbf{v}_m n)\right.$$
$$\left.\overline{f_i(\mathbf{x}+\mathbf{v}_i P_i(y+k)+\mathbf{v}_m n)}\right) f_m(\mathbf{x}+\mathbf{v}_m(n+P_m(y))) \tag{3.22}$$
$$\prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x}+\mathbf{v}_m n)\partial_k P_i(y)) 1_{\mathcal{U}}(\mathbf{x}+\mathbf{v}_m n). \tag{3.23}$$

We first prove the statement when $m = t = 1$. In that case,

$$\|F\|^2_{U^1(\mathbf{v}_m)} \leqslant \mathop{\mathbb{E}}_{\mathbf{x},n,y,k} \overline{f_m(\mathbf{x}+\mathbf{v}_m(n+P_m(y)))} f_m(\mathbf{x}+\mathbf{v}_m(n+P_m(y+k))).$$

Replacing both instances of f_m by their Fourier transform along $\mathbf{v}_m$, we observe that

$$\|F\|^2_{U^1(\mathbf{v}_m)} \leqslant \mathop{\mathbb{E}}_{\mathbf{x}} \sum_{l_1,l_2} \overline{\widehat{f_m}(\mathbf{x};\mathbf{v}_m;l_1)}\widehat{f_m}(\mathbf{x};\mathbf{v}_m;l_2)$$
$$\mathop{\mathbb{E}}_{n,y,k} e_p(n(l_2-l_1)+l_2P(y+k)-l_1P(y)).$$

Using Lemma 3.2.1 and Parseval's identity $\sum_l \left|\widehat{f_m}(\mathbf{x};\mathbf{v}_m;l)\right|^2 = \mathbb{E}_{\mathbf{x}+V_m}|f_m|^2$, we deduce that

$$\|F\|^2_{U^1(\mathbf{v}_m)} \leqslant \mathop{\mathbb{E}}_{\mathbf{x}} \left|\widehat{f_m}(\mathbf{x};\mathbf{v}_m;0)\right|^2 + O(p^{-1/2}) = \|f_m\|^2_{U^1(\mathbf{v}_m)} + O(p^{-1/2}).$$

We assume now that $t>1$. Applying the Cauchy-Schwarz inequality in k to (3.22) and performing several changes of variables, we bound

$$\left|\mathop{\mathbb{E}}_{\mathbf{x}+V_m} F\right|^2 \leqslant \mathop{\mathbb{E}}_{n}\left|\mathop{\mathbb{E}}_{y}\prod_{i=1}^{m-1} f_i(\mathbf{x}+\mathbf{v}_iP_i(y)+\mathbf{v}_mn)\right)$$
$$\left.\prod_{i=m+1}^{t} e_p(-\phi_i(\mathbf{x}+\mathbf{v}_mn)P_i(y))1_{\mathcal{U}}(\mathbf{x}+\mathbf{v}_mn)\right|^2,$$

and so

$$\|F\|^2_{U^1(\mathbf{v}_m)} \leqslant \mathop{\mathbb{E}}_{\mathbf{x}}\left|\mathop{\mathbb{E}}_{y}\prod_{i=1}^{m-1} f_i(\mathbf{x}+\mathbf{v}_iP_i(y))\prod_{i=m+1}^{t} e_p(-\phi_i(\mathbf{x})P_i(y))1_{\mathcal{U}}(\mathbf{x})\right|^2.$$

Applying the $(m-1,t-1)$ case of Proposition 3.5.1 to

$$G_{m-1,t-1}(\mathbf{x}) = \mathop{\mathbb{E}}_{y}\prod_{i=1}^{m-1} f_i(\mathbf{x}+\mathbf{v}_iP_i(y))\prod_{i=m+1}^{t} e_p(-\phi_i(\mathbf{x})P_i(y))1_{\mathcal{U}}(\mathbf{x}),$$

which is where we use $t>1$, we deduce that

$$\|F\|_{U^1(\mathbf{v}_m)} \leqslant \min_{1\leqslant i\leqslant m-1}\|f_i\|_{U^1}(\mathbf{v}_i) + O(p^{-c_1}) \tag{3.24}$$

for some $c_1>0$.

It remains to show that $\|F\|_{U^1(\mathbf{v}_m)} \leqslant \|f_m\|_{U^1}(\mathbf{v}_m)+O(p^{-c_1})$. Once again, we look at $\|F\|_{U^1(\mathbf{v}_m)}$, splitting each $f_1,\ldots,f_{m-1}$ into $f_i = \mathbb{E}(f_i|V_i)+(f_i-\mathbb{E}(f_i|V_i))$.

Using (3.24) and the fact $\mathbb{E}(f_i|V_i)(\mathbf{x}+\mathbf{v}_i P_i(y)) = \mathbb{E}(f_i|V_i)(\mathbf{x})$, we get that

$$\|F\|^2_{U^1(\mathbf{v}_m)} = \mathop{\mathbb{E}}_{\mathbf{x}}\Bigg|\mathop{\mathbb{E}}_{n,y,k}\prod_{i=1}^{m-1}|\,\mathbb{E}(f_i|V_i)(\mathbf{x}+\mathbf{v}_m n)|^2 f_m(\mathbf{x}+\mathbf{v}_m(n+P_m(y)))$$
$$\prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x}+\mathbf{v}_m n)\partial_k P_i(y))1_{\mathcal{U}}(\mathbf{x}+\mathbf{v}_m n)\Bigg|^2 + O(p^{-c_1})$$

We let $g(\mathbf{x}) = \prod_{i=1}^{m-1}|\,\mathbb{E}(f_i|V_i)(\mathbf{x})|^2 1_{\mathcal{U}}(\mathbf{x})$, so that

$$\|F\|^2_{U^1(\mathbf{v}_m)} = \mathop{\mathbb{E}}_{\mathbf{x}}\Bigg|\mathop{\mathbb{E}}_{n,y} g(\mathbf{x}+\mathbf{v}_m n) f_m(\mathbf{x}+\mathbf{v}_m(n+P_m(y)))$$
$$\prod_{i=m+1}^{t} e_p(-\phi_i(\mathbf{x}+\mathbf{v}_m n)P_i(y))\mathop{\mathbb{E}}_{k}\prod_{i=m+1}^{t} e_p(\phi_i(\mathbf{x}+\mathbf{v}_m n)P_i(y+k))\Bigg|^2.$$

Using Lemma 3.2.1 and the linear independence of $P_{m+1}, ..., P_t$, the expectation in k is of size $O(p^{-1/2})$ unless $\phi_{m+1}(\mathbf{x}+\mathbf{v}_m n) = ... = \phi_t(\mathbf{x}+\mathbf{v}_m n) = 0$, and so

$$\|F\|^2_{U^1(\mathbf{v}_m)} = \mathop{\mathbb{E}}_{\mathbf{x}}\Big|\mathop{\mathbb{E}}_{n,y} g(\mathbf{x}+\mathbf{v}_m n) f_m(\mathbf{x}+\mathbf{v}_m(n+P_m(y)))\Big|^2 + O(p^{-c})$$

for $c = \min(c_1, 1/2)$. To get rid of g, we apply the Cauchy-Schwarz inequality in n to the inner expectation and obtain

$$\|F\|^2_{U^1(\mathbf{v}_m)} \leqslant \mathop{\mathbb{E}}_{\mathbf{x},y,n,k} \overline{f_m(\mathbf{x}+\mathbf{v}_m(n+P_m(y)))} f_m(\mathbf{x}+\mathbf{v}_m(n+P_m(y+k))) + O(p^{-c})$$

We conclude the proof exactly the same way as in the $m=t=1$ case.

$\square$

3.6 Estimating the number of progressions from below

With all the results from Section 3.5, we are finally able to prove Theorem 3.1.2.

Proof of Theorem 3.1.2. Let $f_0, ..., f_t : \mathbb{F}_p^D \to \mathbb{C}$ be 1-bounded, $P_1, ..., P_t \in \mathbb{Z}[y]$ be polynomials satisfying $0 < \deg P_1 < ... < P_t$, and $\mathbf{v}_1, ..., \mathbf{v}_t \in \mathbb{Z}^D$ be nonzero vectors. By Proposition 3.5.1, we have

$$\left|\mathop{\mathbb{E}}_{\mathbf{x},y}\prod_{i=0}^{t} f_i(\mathbf{x}+\mathbf{v}_iP_i(y))\right| \leqslant \left(\mathop{\mathbb{E}}_{\mathbf{x}}\left|\mathop{\mathbb{E}}_{y}\prod_{i=0}^{t} f_i(\mathbf{x}+\mathbf{v}_iP_i(y))\right|^2\right)^{\frac{1}{2}}$$
$$\leqslant \min_{1\leqslant i\leqslant t}\|f_i\|_{U^1(\mathbf{v}_i)} + O(p^{-c})$$

for a constant $c>0$ independent of $f_0,...,f_t$. The statement follows by splitting $f_1,...,f_t$ as $f_i=\mathbb{E}(f_i|V_i)+(f_i-\mathbb{E}(f_i|V_i))$, recalling that $\|f_i-\mathbb{E}(f_i|V_i)\|_{U^1(\mathbf{v}_i)}=0$ and noting that $f_0=\mathbb{E}(f_0|V_0)$. □

Corollary 3.1.3 follows from the following lemma, which is a special case of Lemma [Chu11].

Lemma 3.6.1. *Let $\boldsymbol{v}_1,...,\boldsymbol{v}_t\in\mathbb{Z}^D$ be nonzero vectors and $f:\mathbb{F}_p^D\to\mathbb{C}$ be nonnegative. Then*

$$\mathop{\mathbb{E}}_{\boldsymbol{x}}\prod_{i=0}^{t}\mathbb{E}(f|V_i)(\boldsymbol{x}) \geqslant \left(\mathop{\mathbb{E}}_{\boldsymbol{x}\in\mathbb{F}_p^D} f(\boldsymbol{x})\right)^{t+1}.$$

Proof of Theorem 3.1.1. Suppose that $A\subseteq\mathbb{F}_p^D$ has size $|A|=\alpha p^D$. Theorem 3.1.1 and Lemma 3.6.1 imply that there exists a constant $c>0$ for which A contains $\Omega(\alpha^{t+1}p^D)-O(p^{D-c})$ nontrivial configurations (3.1). It follows that if $\alpha\gg p^{-c/(t+1)}$, then A contains a nontrivial configuration (3.1).

□

References

[Bal+94] A. Balog et al. “Difference sets without kth powers”. In: *Acta Math. Hungar.* **65**.2 (1994), pp. 165–187.

[BC17] J. Bourgain and M.-C. Chang. “Nonlinear Roth type theorems in finite fields”. In: *Israel J. Math.* **221** (2017), pp. 853–867.

[BL96] V. Bergelson and A. Leibman. “Polynomial extensions of van der Waerden’s and Szemerédi’s theorems”. In: *J. Amer. Math. Soc.* **9** (1996), pp. 725–753.

[BM20] T. Bloom and J. Maynard. “A new upper bound for sets with no square differences”. In: *ArXiv e-prints* (2020). arXiv: `2011.13266`.

[CFH11] Q. Chu, N. Frantzikinakis, and B. Host. “Ergodic averages of commuting transformations with distinct degree polynomial iterates”. In: *Proc. Lond. Math. Soc.* **102** (2011), pp. 801–842.

[Chu11] Q. Chu. “Multiple recurrence for two commuting transformations”. In: *Ergodic Theory Dynam. Systems* **31** (2011), pp. 771–792.

[DLS20] D. Dong, X. Li, and W. Sawin. “Improved estimates for polynomial Roth type theorems in finite fields”. In: *J. Anal. Math.* **141** (2020), pp. 689–705.

[HLY21] R. Han, M. T. Lacey, and F. Yang. “A polynomial Roth theorem for corners in finite fields”. In: *ArXiv e-prints* (2021). arXiv: `2012.11686`.

[Kow] E. Kowalski. *Exponential sums over finite fields: elementary methods.* URL: `https://people.math.ethz.ch/~kowalski/exp-sums.pdf`.

[Kuc21] B. Kuca. “Further bounds in the polynomial Szemerédi theorem over finite fields”. In: *Acta Arith.* **198** (2021), pp. 77–108.

[Luc06] J. Lucier. “Intersective sets given by a polynomial”. In: *Acta Arith.* **123** (2006), pp. 57–95.

[Pel18] S. Peluse. “Three-term polynomial progressions in subsets of finite fields”. In: *Israel J. Math.* **228** (1 2018), pp. 379–405.

[Pel19] S. Peluse. “On the polynomial Szemerédi theorem in finite fields”. In: *Duke Math. J.* **168**.5 (2019), pp. 749–774.

[Pel20] S. Peluse. “Bounds for sets with no polynomial progressions”. In: *Forum Math. Pi* **8** (e16 2020).

[PP19] S. Peluse and S. Prendiville. “Quantitative bounds in the non-linear Roth theorem”. In: *ArXiv e-prints* (2019). arXiv: `1903.02592`.

[PP20] S. Peluse and S. Prendiville. “A polylogarithmic bound in the non-linear Roth theorem”. In: *Int. Math. Res. Nov. IMRN* (2020). rnaa261.

[Pre] S. Prendiville. *Fourier methods in combinatorial number theory.* URL: `https://sites.google.com/view/web-add-comb/webinar-in-additive-combinatorics/lecture-series-fourier-methods-in-combinatorial-number-theory?authuser=0`.

[Pre17] S. Prendiville. “Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case”. In: *Discrete Anal.* **5** (2017), 34 pp.

[Ric19] A. Rice. "A maximal extension of the best-known bounds for the Furstenberg-Sárközy theorem". In: *Acta Arith.* **187** (2019), pp. 1–41.

[Sár78a] A. Sárközy. "On difference sets of sequences of integers. I". In: *Acta Math. Hungar.* **31**.1-2 (1978), pp. 125–149.

[Sár78b] A. Sárközy. "On difference sets of sequences of integers. III". In: *Acta Math. Hungar.* **31** (1978), pp. 355–386.

[Shk06a] I. D. Shkredov. "On a generalization of Szemerédi's theorem". In: *Proc. Lond. Math. Soc.* **93**.3 (2006), pp. 723–760.

[Shk06b] I. D. Shkredov. "On a problem of Gowers". In: *Izv. Math.* **70**.2 (2006), pp. 385–425.

[Sli03] S. Slijepcević. "A polynomial Sárközy-Furstenberg theorem with upper bounds". In: *Acta Math. Hungar.* **98**.1-2 (2003), pp. 111–128.

4 On several notions of complexity of polynomial progressions[1]

Abstract

For a polynomial progression

$$(x,\ x+P_1(y),\ ...,\ x+P_t(y)),$$

we define four notions of complexity: Host-Kra complexity, Weyl complexity, true complexity and algebraic complexity. The first two describe the smallest characteristic factor of the progression, the third one refers to the smallest-degree Gowers norm controlling the progression, and the fourth one concerns algebraic relations between terms of the progressions. We conjecture that these four notions are equivalent, which would give a purely algebraic criterion for determining the smallest Host-Kra factor or the smallest Gowers norm controlling a given progression. We prove this conjecture for all progressions whose terms only satisfy homogeneous algebraic relations and linear combinations thereof, as well a family of progressions of the form

$$(x,\ x+y,\ ...,\ x+(t-1)y,\ x+y^d).$$

The former family includes, but is not limited to, arithmetic progressions, progressions with linearly independent polynomials $P_1, ..., P_t$ and progressions whose terms satisfy no quadratic or higher-order relations. For progressions that satisfy only linear relations, such

[1]This chapter is a merger of the papers B. Kuca. "On several notions of complexity of polynomial progressions". In: *ArXiv e-prints* (2021). arXiv: 2104.07339 and B. Kuca. "True complexity of polynomial progressions in finite fields". In: *Proc. Edinb. Math. Soc.* (2021), pp. 1–53. Most of the material together with the structure of the chapter comes from the former paper, however, several results have previously appeared in the latter paper. The material from the latter paper has been clearly designated as such.

as

$$(x, x+y^2, x+2y^2, x+y^3, x+2y^3),$$

we derive several combinatorial and dynamical corollaries: (1) an estimate for the count of such progressions in subsets of $\mathbb{Z}/N\mathbb{Z}$ or totally ergodic dynamical systems; (2) a lower bound for multiple recurrence; (3) and a popular common difference result in $\mathbb{Z}/N\mathbb{Z}$. Lastly, we show that Weyl complexity and algebraic complexity always agree, which gives a straightforward algebraic description of Weyl complexity.

4.1 Introduction

A polynomial $P \in \mathbb{R}[y]$ is *integral* if $P(\mathbb{Z}) \subseteq \mathbb{Z}$ and $P(0) = 0$. For $t \in \mathbb{N}_+$, an *integral polynomial progression* of length $t+1$ is a tuple $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ given by

$$\vec{P}(x,y) = (x,\ x+P_1(y),\ ...,\ x+P_t(y))$$

for distinct integral polynomials $P_1, ..., P_t$. We moreover say that a set $A \subseteq \mathbb{N}$ contains $\vec{P}(x,y)$ for some $x, y \in \mathbb{N}$ if $\vec{P}(x,y) \in A^{t+1}$. The condition $P_i(0) = 0$ in the definition of integral polynomials can be easily disposed of if the polynomials $P_i - P_i(0)$ are all distinct, but we assume it for the sake of clarity of the exposition.

A major result on integral polynomial progressions is the polynomial Szemerédi theorem by Bergelson and Leibman, which extends the famous theorem of Szemerédi on arithmetic progressions.

Theorem 4.1.1 (Polynomial Szemerédi theorem)**.** *[BL96] Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression, and suppose that $A \subseteq \mathbb{N}$ is dense*[2]*. Then A contains $\vec{P}(x,y)$ for some $x, y \in \mathbb{N}_+$.*

Theorem 4.1.1 can be deduced from the following ergodic theoretic statement using the Furstenberg correspondence principle.

Theorem 4.1.2. *[BL96; HK05a] Let $(X, \mathcal{X}, \mu, T)$ be an invertible measure-preserving dynamical system, $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polyno-*

[2] Meaning that $\limsup\limits_{N\to\infty} \frac{|A\cap[N]|}{N} > 0$, where $[N] = \{1, ..., N\}$.

mial progression. If $\mu(A) > 0$ *for* $A \in \mathcal{X}$*, then*

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \mu(A \cap T^{P_1(n)}A \cap \cdots \cap T^{P_t(n)}A) > 0,$$

where $[N] = \{1, ..., N\}$ *and* $\mathbb{E}_{x\in X} = \frac{1}{|X|}\sum_{x\in X}$ *for any set* X*.*

To prove Theorem 4.1.1, one thus needs to understand limits of multiple ergodic averages of the form

$$\mathop{\mathbb{E}}_{n\in[N]} T^{P_1(n)}f_1 \cdots T^{P_t(n)}f_t \tag{4.1}$$

for $f_1, ..., f_t \in L^\infty(\mu)$. By a remarkable result of Host and Kra [HK05a; HK05b], there exists a family of factors[3] $(\mathcal{Z}_s)_{s\in\mathbb{N}}$, called henceforth *Host-Kra factors*, with the property that weak or L^2 limits of expressions of the form (4.1) remain unchanged if we project any of the functions f_i onto one of the factors $\mathcal{Z}_s$ for some s dependent on $\vec{P}$ and i.

Definition 4.1.3 (Characteristic factors)**.** *Let* $(X, \mathcal{X}, \mu, T)$ *be an invertible measure-preserving dynamical system,* $t \in \mathbb{N}_+$ *and* $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ *be an integral polynomial progression.*

Suppose that $1 \leqslant i \leqslant t$*. A factor* $\mathcal{Y}$ *of* $\mathcal{X}$ *is* characteristic for the L^2-convergence of $\vec{P}$ at i *if for all choices of* $f_1, ..., f_t \in L^\infty(\mu)$*, the* L^2*-limit of (4.1) is 0 whenever* $\mathbb{E}(f_i|\mathcal{Y}) = 0$*.*

Similarly, suppose that $0 \leqslant i \leqslant t$*. A factor* $\mathcal{Y}$ *of* $\mathcal{X}$ *is* characteristic for the weak convergence of $\vec{P}$ at i *if for all choices of* $f_0, ..., f_t \in L^\infty(\mu)$*, the weak limit of (4.1), i.e. the expression*

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_X f_0 \cdot T^{P_1(n)}f_1 \cdots T^{P_t(n)}f_t d\mu, \tag{4.2}$$

is 0 whenever $\mathbb{E}(f_i|\mathcal{Y}) = 0$*.*

Theorem 4.1.4 ([HK05a; Lei05a])**.** *Let* $t \in \mathbb{N}_+$*. For each integral polynomial progression* $\vec{P} \in \mathbb{R}[x, y]^{t+1}$*, there is* $s \in \mathbb{N}$ *such that for all invertible ergodic systems* $(X, \mathcal{X}, \mu, T)$*, the factor* $\mathcal{Z}_s$ *is characteristic for the* L^2 *convergence of* $\vec{P}$ *at* i *for all* $0 \leqslant i \leqslant t$*.*

The utility of Host-Kra factors, as laid out in [HK05b], comes from the fact that they are inverse limits of *nilsystems*, and so understanding (4.1) for

[3] The definitions of factors, Weyl systems, nilsystems, and other concepts from ergodic theory and higher order Fourier analysis used in the introduction will be provided in subsequent sections.

arbitrary systems comes down to proving certain equidistribution results on spaces called nilmanifolds that possess rich algebraic structure.Importantly, $\mathcal{Z}_s$ is a factor of $\mathcal{Z}_{s+1}$ for each $s \in \mathbb{N}$, hence it is natural to inquire about the smallest value of s for which the factor $\mathcal{Z}_s$ is characteristic for $\vec{P}$ at i. This leads to the following definition, versions of which have previously been examined in [BLL07; Fra08; Lei09; Fra16].

Definition 4.1.5 (Host-Kra complexity)**.** *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. Fix $0 \leqslant i \leqslant t$. The progression $\vec{P}$ has* Host-Kra complexity *s at i, denoted $\mathcal{HK}_i(\vec{P})$, if s is the smallest natural number such that the factor $\mathcal{Z}_s$ is characteristic for the weak convergence of $\vec{P}$ at i for all invertible totally ergodic dynamical systems $(X, \mathcal{X}, \mu, T)$. We say $\vec{P}$ has* Host-Kra complexity *s if $\max_i \mathcal{HK}_i(\vec{P}) = s$.*

Investigating complexity has been of particular interest for a class of dynamical systems called Weyl systems since early Host-Kra complexity results relied on reducing the case of a general totally ergodic system to the case of a Weyl system [FK05; Fra08; Lei09].

Definition 4.1.6 (Weyl system)**.** *A* Weyl system *is an ergodic system $(X, \mathcal{X}, \mu, T)$, where X is a compact abelian Lie group and T is a unipotent affine transformation on X, i.e. $Tx = \phi(x) + a$ for $a \in X$ and an automorphism ϕ of X satisfying $(\phi - \mathrm{Id_X})^s = 0$ for some $s \in \mathbb{N}_+$.*

This leads to another notion of complexity, a variant of which has previously appeared in [BLL07; Fra08; Lei09; Fra16].

Definition 4.1.7 (Weyl complexity)**.** *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. Fix $0 \leqslant i \leqslant t$. The progression $\vec{P}$ has* Weyl complexity *s at i, denoted $\mathcal{W}_i(\vec{P})$, if s is the smallest natural number such that the factor $\mathcal{Z}_s$ is characteristic for the weak convergence of $\vec{P}$ at i for all Weyl systems $(X, \mathcal{X}, \mu, T)$. We say $\vec{P}$ has* Weyl complexity *s if $\max_i \mathcal{W}_i(\vec{P}) = s$.*

In previous works [BLL07; Lei09; Fra08; Fra16], the aforementioned notions of complexity have been defined for a polynomial family $\mathcal{P} = \{P_1, ..., P_t\}$ rather than for a progression $\vec{P}$. However, we want to extend the definitions of complexity to "index 0", i.e. the x term in $\vec{P}$, which is why we prefer to define it for $\vec{P}$ rather than $\mathcal{P}$. Similarly, complexity has previously been defined for L^2 convergence rather than weak convergence. However, the existence of L^2 limit (Theorem 4.1.4) and the Cauchy-Schwarz inequality imply that weak limit exists and equals the strong limit.

Host-Kra factors are deeply related to a family of seminorms called *Gowers-Host-Kra seminorms*. For $s \in \mathbb{N}_+$ and $f \in L^\infty(\mu)$, the Gowers-Host-Kra seminorm of f of degree s is denoted by $|||f|||_s$ and satisfies the property

$$|||f|||_{s+1} = 0 \iff \mathbb{E}(f|\mathcal{Z}_s) = 0$$

as well as the monotonicity property

$$|||f|||_1 \leqslant |||f|||_2 \leqslant |||f|||_3 \leqslant \dots \tag{4.3}$$

Gowers-Host-Kra seminorms have natural finitary analogues. For the transformation $Tx = x+1$ on $X = \mathbb{Z}/N\mathbb{Z}$ with N prime and the uniform probability measure μ, the weak limit (4.2) becomes

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+P_1(y))\cdots f_t(x+P_t(y)). \tag{4.4}$$

The Gowers-Host-Kra seminorm of any $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is a norm (for $s > 1$) called the *Gowers norm* and denoted by U^s, and it takes the form

$$||f||_{U^s} = \left(\mathop{\mathbb{E}}_{x,h_1,\dots,h_s\in\mathbb{Z}/N\mathbb{Z}} \prod_{w\in\{0,1\}^s} \mathcal{C}^{|w|} f(x+w_1h_1+\dots+w_sh_s)\right)^{\frac{1}{2^s}}, \tag{4.5}$$

where $\mathcal{C} : z \mapsto \overline{z}$ is the conjugation operator and $|w| = w_1 + \cdots + w_s$. As a result, $||f||_{U^s} = 0$ for some $s > 1$ if and only if $||f||_{U^2} = 0$ if and only if $f = 0$, and so inquiring about the smallest characteristic factor of this system in the sense of Definition 4.1.3 makes little sense. We can however ask which Gowers norm "controls" $\vec{P}$ in a more finitary way, and this leads to another notion of complexity, originally introduced in the works of Gowers and Wolf on systems of linear forms [GW10; GW11a; GW11b; GW11c] and subsequently investigated in [GT10; Alt21; Man18; Man21].

Definition 4.1.8 (True complexity). *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. Fix $0 \leqslant i \leqslant t$. The progression $\vec{P}$ has* true complexity *s at i, denoted $\mathcal{T}_i(\vec{P})$, if s is the smallest natural number with the following property: for every $\varepsilon > 0$, there exist $\delta > 0$ and $N_0 \in \mathbb{N}$ such that for all primes $N > N_0$ and all functions $f_0, \dots, f_t : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ satisfying* $\max_i ||f_i||_\infty \leqslant 1$, *we have*

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+P_1(y))\cdots f_t(x+P_t(y))\right| < \varepsilon$$

whenever $||f_i||_{U^{s+1}} < \delta$. *We say* $\vec{P}$ *has* true complexity s *if* $\max_i \mathcal{T}_i(\vec{P}) = s$.

We have so far defined three notions of complexity, that of Host-Kra, Weyl and true complexity. They are all defined in terms of ergodic theory or higher order Fourier analysis and have to do with "controlling" expressions like (4.1) and (4.4) by characteristic factors, Gowers-Host-Kra seminorms and Gowers norms. We shall now introduce one more notion, defined purely in terms of algebraic properties of polynomial progressions, and conjecture that all four concepts of complexity are in fact the same.

Definition 4.1.9 (Algebraic relations and algebraic complexity)**.** *Let* $t \in \mathbb{N}_+$ *and* $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ *be an integral polynomial progression. An* algebraic relation *of degree* $(j_0, ..., j_t)$ *satisfied by* $\vec{P}$ *is a tuple* $(Q_0, ..., Q_t) \in \mathbb{R}[u]^{t+1}$ *such that*

$$Q_0(x) + Q_1(x + P_1(y)) + ... + Q_t(P_t(y)) = 0, \tag{4.6}$$

where $\deg Q_i = j_i$ *for each* $0 \leqslant i \leqslant t$. *The progression* $\vec{P}$ *has* algebraic complexity s *at* i *for some* $0 \leqslant i \leqslant t$*, denoted* $\mathcal{A}_i(\vec{P})$*, if* s *is the smallest natural number such that for any algebraic relation* $(Q_0, ..., Q_t)$ *satisfied by* $\vec{P}$*, the degree of* Q_i *is at most* s. *It has* algebraic complexity s *if* $\max_i \mathcal{A}_i(\vec{P}) = s$.

Conjecture 4.1.10 (Four notions of complexity are the same)**.** *Let* $t \in \mathbb{N}_+$ *and* $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ *be an integral polynomial progression. Fix* $0 \leqslant i \leqslant t$. *Then*

$$\mathcal{HK}_i(\vec{P}) = \mathcal{W}_i(\vec{P}) = \mathcal{T}_i(\vec{P}) = \mathcal{A}_i(\vec{P}) \leqslant t - 1.$$

The heuristic for Conjecture 4.1.10 is as follows: evaluating expressions like (4.2) and (4.4) comes down to understanding the distribution of certain polynomial sequences on nilmanifolds, and the only obstructions to equidistribution come from algebraic relations of the form (4.6).

Several substatements of Conjecture 4.1.10, such as the equivalence of Weyl and Host-Kra complexity and the upper bound on complexities, have previously been conjectured in [BLL07; Lei09; Fra08; Fra16]. Similarly, the equivalence of true and algebraic complexity has been studied and proved for linear configurations [GW10; GW11a; GW11b; GW11c; Alt21; Man18; Man21] as well as certain subclasses of polynomial progressions [Pel19; Kuc21a]. However, we have not seen the full statement of Conjecture 4.1.10 anywhere in the literature. In particular, we have not found a conjecture relating Host-Kra and Weyl complexity to algebraic complexity, even though the aforementioned papers researching the topic mention that algebraic relations form a source of obstructions preventing a progression from having a characteristic small-degree Host-Kra factor.

Before we state our main result, we have to distinguish between two large families of progressions.

Definition 4.1.11 (Homogeneous and inhomogeneous relations and progressions)**.** *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. An* algebraic relation $(Q_0, ..., Q_t) \in \mathbb{R}[u]^{t+1}$ *is* homogeneous *of degree d if it is of the form*

$$(Q_0(u), ..., Q_t(u)) = (a_0 u^d, ..., a_t u^d)$$

for some $a_0, ..., a_t \in \mathbb{R}$ (some but not all of which may be zero), and inhomogeneous *otherwise. The progression $\vec{P}$ is* homogeneous *if all the algebraic relations that it satisfies are linear combinations of its homogeneous algebraic relations, and it is called* inhomogeneous *otherwise.*

An example of a homogeneous progression is $(x,\ x+y,\ x+2y,\ x+y^3)$, which only satisfies a homogeneous relation

$$x - 2(x+y) + (x+2y) = 0. \tag{4.7}$$

Other examples include arithmetic progressions, progressions with $P_1, ..., P_t$ being linearly independent such as $(x,\ x+y,\ x+y^2)$, or progressions whose terms satisfy no quadratic relations, such as $(x,\ x+y^2,\ x+2y^2,\ x+y^3,\ x+2y^3)$. By contrast, the progression $(x,\ x+y,\ x+2y,\ x+y^2)$ is inhomogeneous because it satisfies both (4.7) and the inhomogeneous relation

$$x^2 + 2x - 2(x+y)^2 + (x+2y)^2 - 2(x+y^2) = 0 \tag{4.8}$$

that cannot be broken down into a sum of homogeneous relations. These two progressions will accompany us as running examples throughout the paper.

Our main result is the following.

Theorem 4.1.12 (Conjecture 4.1.10 holds for homogeneous progressions)**.** *Let $t \in \mathbb{N}+$. If $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ is a homogeneous polynomial progression, then it satisfies Conjecture 4.1.10.*

Having defined Host-Kra complexity using totally ergodic systems, we would like to extend our results to ergodic systems. We have however encountered an algebraic obstacle in doing so that prevents us from performing this generalisation for all homogeneous progressions. We introduce a subfamily of homogeneous polynomial progressions for which this extension is possible, borrowing the terminology of Frantzikinakis from [Fra08].

Definition 4.1.13 (Eligible progressions). *A homogeneous polynomial progression $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ is* eligible *if for every $r \in \mathbb{N}_+$ and every $0 \leqslant j \leqslant r-1$, the family*

$$\tilde{\vec{P}}(x,y) = (x,\ x + \tilde{P}_{1,j}(y),\ \dots,\ x + \tilde{P}_{t,j}(y)),$$

where $\tilde{P}_{i,j}(y) = \frac{P_i(r(y-1)+j)-P_i(j)}{r}$, is homogeneous, and $\mathcal{A}_i(\vec{P}) = \mathcal{A}_i(\tilde{\vec{P}})$ for every $0 \leqslant i \leqslant t$.

The condition in Definition 4.1.13 may seem artificial at first glance, but this turns out to be the condition that we need to pass from totally ergodic to ergodic systems. While we believe that all homogeneous progressions satisfy this condition, we have not been able to prove this.

We now state the corollary that gives us the smallest characteristic Host-Kra factor for eligible progressions on ergodic systems The main difference is that if a system has complexity 0, then the $\mathcal{Z}_0$ factor has to be replaced by the rational Kronecker factor $\mathcal{K}_{rat}$.

Corollary 4.1.14. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an eligible homogeneous polynomial progression, and suppose that $\mathcal{A}_i(\vec{P}) = s$ for some $0 \leqslant i \leqslant t$ and $s \in \mathbb{N}$. For all invertible ergodic dynamical systems $(X, \mathcal{X}, \mu, T)$, the factor $\mathcal{Z}_s$ is characteristic for the weak or L^2 convergence of $\vec{P}$ at i if $s > 0$, and $\mathcal{K}_{rat}$ is characteristic for the weak or L^2 convergence of $\vec{P}$ at i if $s = 0$.*

Since all polynomial progressions of algebraic complexity at most 1 are homogeneous and eligible, the following corollary follows.

Corollary 4.1.15. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be polynomial progression of algebraic complexity at most 1. For all invertible ergodic dynamical systems $(X, \mathcal{X}, \mu, T)$, the factor $\mathcal{Z}_1$ is characteristic for the weak or L^2 convergence of $\vec{P}$ at i if $\mathcal{A}_i(\vec{P}) = 1$, and $\mathcal{K}_{rat}$ is characteristic for the weak or L^2 convergence of $\vec{P}$ at i if $\mathcal{A}_i(\vec{P}) = 0$.*

We also include here the following result on a certain family of inhomogeneous progressions. This result originally appeared as Theorem 1.10 in [Kuc21b].

Theorem 4.1.16. *Let $t, d \in \mathbb{N}_+$ satisfy $t \geqslant 3$ and $2 \leqslant d \leqslant t-1$, and*

$$\vec{P}(x,y) = (x,\ x+y,\ ...,\ x+(t-1)y,\ x+y^d). \tag{4.9}$$

Then

$$\mathcal{A}_i(\vec{P}) = \mathcal{T}_i(\vec{P}) = \begin{cases} \max(t-2, d\left\lfloor \frac{t-1}{d} \right\rfloor),\ 0 \leqslant i \leqslant t-1 \\ \left\lfloor \frac{t-1}{d} \right\rfloor,\ i = t. \end{cases}$$

Theorem 4.1.12, Corollary 4.1.14 and Theorem 4.1.16 can be viewed as extensions of [HK05a; HK05b; FK05; FK06; Fra08; BLL07; Lei09], which find characteristic factors for linear configurations, linearly independent polynomials, progressions of length 4, examine Weyl complexity for arbitrary integral polynomial progression, and give an upper bound for Host-Kra complexity for general integral progressions. Theorems 4.1.12 and 4.1.16 also partly extend [GW10; GW11a; GW11b; GW11c; GT10; Alt21; Man18; Man21; Pel19; Kuc21a], which among other things determine true complexity for certain families of linear forms and integral polynomial progressions.

From the fact that all progressions of algebraic complexity 1 are homogeneous and eligible, we deduce the following counting result.

Corollary 4.1.17. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression of algebraic complexity at most 1. Suppose that $Q_1, ..., Q_d \in \mathbb{R}[y]$ are integral polynomials such that $P_i(y) = \sum_{j=1}^d a_{ij} Q_j(y)$ for $a_{ij} \in \mathbb{Z}$ for each $0 \leqslant i \leqslant t$ and $1 \leqslant j \leqslant d$. Let $L_i(y_1, ...y_d) = \sum_{j=1}^d a_{ij} y_j$. Then the following is true.*

1. *For any $f_0, ..., f_t : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ with $\max_i ||f_i||_\infty \leqslant 1$, we have*

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} \prod_{i=0}^t f_i(x+P_i(y)) = \mathop{\mathbb{E}}_{x,y_1,...,y_d\in\mathbb{Z}/N\mathbb{Z}} \prod_{i=0}^t f_i(x+L_i(y_1,...,y_d)) + o(1),$$

   where the error term $o(1)$ is taken as $N \to \infty$ over primes and does not depend on the choice of $f_0, ..., f_t$.

2. *For any invertible totally ergodic dynamical system $(X, \mathcal{X}, \mu, T)$ and $f_0, ..., f_t \in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_X \prod_{i=0}^{t} T^{P_i(n)} f_i d\mu = \lim_{N\to\infty} \mathop{\mathbb{E}}_{n_1,\dots,n_d\in[N]} \int_X \prod_{i=0}^{t} T^{L_i(n_1,\dots,n_d)} f_i d\mu.$$

We shall illustrate Corollary 4.1.17 for the specific example of

$$\vec{P}(x,y) = (x,\ x+y^2,\ x+2y^2,\ x+y^3,\ x+2y^3).$$

Taking $Q_1(y) = y^2$ and $Q_2(y) = y^3$ as in the statement of Corollary 4.1.17, we let $\vec{L}(x,y_1,y_2) = (x,\ x+y_1,\ x+2y_1,\ x+y_2,\ x+2y_2)$. For any $A \subseteq \mathbb{Z}/N\mathbb{Z}$, we then have

$$\begin{aligned} &|\{(x,y)\in(\mathbb{Z}/N\mathbb{Z})^2 : (x,\ x+y^2,\ x+2y^2,\ x+y^3,\ x+2y^3)\in A^5\}| \\ =&|\{(x,y_1,y_2)\in(\mathbb{Z}/N\mathbb{Z})^3 : (x,\ x+y_1,\ x+2y_1,\ x+y_2,\ x+2y_2)\in A^5\}|/N + o(N^2) \end{aligned}$$

upon setting $f_0 = \dots = f_t = 1_A$. If $(X,\mathcal{X},\mu,T)$ is a totally ergodic system and $A\in\mathcal{X}$, then we similarly obtain that

$$\begin{aligned} &\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \mu(A\cap T^{n^2}A\cap T^{2n^2}A\cap T^{n^3}A\cap T^{2n^3}A) \\ &= \lim_{N\to\infty} \mathop{\mathbb{E}}_{n,m\in[N]} \mu(A\cap T^{n}A\cap T^{2n}A\cap T^{m}A\cap T^{2m}A). \end{aligned}$$

For progressions of algebraic complexity 1, we also prove the following result, which generalises Theorem C of [Fra08], Theorem 1.12 of [GT10], and results from [BHK05]. In additive combinatorics, problems of this type are known as finding popular common differences; in ergodic theory, one speaks of establishing lower bounds for multiple recurrence.

Theorem 4.1.18. *Let $t\in\mathbb{N}_+$ and $\vec{P}\in\mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression of algebraic complexity at most 1, with the following property: there exist linearly independent integral polynomials $Q_1,\dots,Q_k$ such that*

$$\{a_1Q_1+\dots+a_kQ_k :\ a_1,\dots,a_k\in\mathbb{Z}\} = \{b_1P_1+\dots+b_tP_t :\ b_1,\dots,b_t\in\mathbb{Z}\}. \tag{4.10}$$

Then the following is true.

1. *Let $(X,\mathcal{X},\mu,T)$ be an ergodic invertible measure preserving system and*

$A \in \mathcal{X}$. Suppose that $\mu(A) > 0$. Then for every $\varepsilon > 0$, the set

$$\{n \in \mathbb{N} : \mu(A \cap T^{P_1(n)}A \cap \cdots \cap T^{P_t(n)}A) > \mu(A)^{t+1} - \varepsilon\}$$

is syndetic.

2. *Suppose that $A \subseteq \mathbb{N}$ has upper density $\alpha > 0$. Then for every $\varepsilon > 0$, the set*

$$\{n \in \mathbb{N} : \mu(A \cap (A + P_1(n)) \cap \cdots \cap (A + P_t(n))) > \alpha^{t+1} - \varepsilon\}$$

is syndetic.

3. *For any $\alpha, \varepsilon > 0$ and prime N, and any subset $A \subseteq \mathbb{Z}/N\mathbb{Z}$ of size $|A| \geqslant \alpha N$, we have*

$$|\{n \in \mathbb{Z}/N\mathbb{Z} : |A \cap (A + P_1(n)) \cap \cdots \cap (A + P_t(n))| > (\alpha^{t+1} - \varepsilon)N\}| \gg_{\alpha,\varepsilon} N.$$

The definition of homogeneity (Definition 4.1.11) is equivalent to a certain linear algebraic property that will be described in details in Section 4.4; this property makes it possible to explicitly describe closures of orbits of nilsequences evaluated at terms of homogeneous polynomial progressions, from which we deduce Theorem 4.1.12. Homogeneous polynomial progressions are moreover the largest family of integral polynomial progressions for which such an explicit description is possible, and even the simplest examples of inhomogeneous progressions lead to complications absent in the homogeneous case. The following result makes this precise. As with all other results in this section, all the concepts in Theorem 4.1.19 are explained in subsequent sections.

Theorem 4.1.19 (Dichotomy between homogeneous and inhomogeneous progressions)**.** *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. Suppose that G is a connected, simply-connected, nilpotent Lie group with a rational filtration $G_\bullet$ and Γ is a cocompact lattice. There exists a subnilmanifold G^P/Γ^P of G^{t+1}/Γ^{t+1} with the following property.*

1. *If $\vec{P}$ is homogeneous, then for every irrational polynomial sequence $g : \mathbb{Z} \to G$ adapted to $G_\bullet$, the sequence*

$$g^P(x,y) = (g(x),\ g(x + P_1(y)),\ \ldots,\ g(x + P_t(y)))$$

is equidistributed on G^P/Γ^P.

2. *If $\vec{P}$ is inhomogeneous, then for every irrational polynomial sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$, the closure of g^P is a union of finitely many translates of a subnilmanifold of G^P/Γ^P. For every $\vec{P}$, we can moreover find a filtered nilmanifold G/Γ and an irrational polynomial sequence $g : \mathbb{Z} \to G$ such that g^P is equidistributed on a proper subnilmanifold of G^P/Γ^P.*

While we have not been able to prove full Conjecture 4.1.10 for inhomogeneous progressions, we are able to say a bit more about the relationship between various notions of complexity in the general case.

Theorem 4.1.20. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. Fix $0 \leqslant i \leqslant t$. Then*

$$\mathcal{W}_i(\vec{P}) = \mathcal{A}_i(\vec{P}) \leqslant \min(\mathcal{T}_i(\vec{P}), \mathcal{HK}_i(\vec{P})).$$

Of the various statements made in Theorem 4.1.20, it is the equivalence of Weyl and algebraic complexities that is a new statement here. The fact that Host-Kra complexity bounds Weyl complexity is a simple consequence of definitions and shall be explained in Section 4.12. We now show that the fact that algebraic complexity is bounded from above by true complexity. This result has originally been stated as Theorem 1.13 in [Kuc21b].

Proof. Suppose that $\mathcal{A}_i(\vec{P}) = s$ for some $0 \leqslant i \leqslant t$. We claim that $\mathcal{T}_i(\vec{P}) \geqslant s$. By assumption, there exists an algebraic relation

$$Q_0(x) + Q_1(x + P_1(y)) + ... + Q_t(x + P_t(y)) = 0,$$

for some $Q_0, ..., Q_t \in \mathbb{Z}[u]$, where Q_i has degree $s+1$. Let $f_j(u) = e_N(Q_j(u))$ for each $0 \leqslant j \leqslant t$, where $e_N(u) = e^{2\pi i u/N}$. The functions f_j are clearly 1-bounded. It follows from the properties of additive characters that

$$\begin{aligned}&\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x) f_1(x+P_1(y)) \cdots f_t(x+P_t(y)) \\ =& \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} e_p(Q_0(x) + Q_1(x+P_1(y)) + ... + Q_t(x+P_t(y))) = 1.\end{aligned}$$

To prove the result, we want to show that $\|f_i\|_{U^s} \to 0$ as $N \to \infty$, which will imply that the U^s norm cannot control the i-th term of the configuration. The definition (4.5) of Gowers norms can be restated as

$$\|f_i\|_{U^s}^{2^s} = \mathop{\mathbb{E}}_{x,h_1,...,h_s\in\mathbb{Z}/N\mathbb{Z}} \Delta_{h_1,...,h_s} f_i(x), \tag{4.11}$$

where $\Delta_h f(x) := f(x+h)\overline{f(x)}$ and $\Delta_{h_1,\ldots,h_s} = \Delta_{h_1}\cdots\Delta_{h_s}$. Since Q_i has degree s and e_N is an additive character, the function $\Delta_{h_1,\ldots,h_s} f_i(x)$ is of the form $e_N(Q(h_1,...,h_s))$ for a nonconstant polynomial Q. By properties of exponential sums, the sum in (4.11) is of size $O(N^{-c_s})$, and so

$$\|f_i\|_{U^s} \ll N^{-c_s}.$$

Thus U^s norm does not control the i-th term of the configuration, implying that $\mathcal{T}_i(\vec{P}) \geqslant s$. □

Outline of the chapter

We start the chapter by introducing basic ergodic theoretic definitions and results concerning nilsystems in Section 4.2, and we explain why analyzing expressions like (4.2) comes down to answering equidistribution questions on nilmanifolds. We then show in Section 4.3 that in studying equidistribution on nilmanifolds, we can restrict ourselves to nilmanifolds that are quotients of connected groups at the expense of replacing a linear sequence by a polynomial one.

Section 4.4 explains key differences between homogeneous and inhomogeneous progressions, and in particular it shows the upper bound on algebraic complexity for homogeneous progressions in Theorem 4.1.12. Definitions introduced in this section allow us to state in the infinitary version of an equidistribution result for homogeneous polynomial progressions on nilmanifolds (Theorem 4.5.3) in Section 4.5, from which we deduce that for homogeneous progressions, Host-Kra complexity is bounded from above by algebraic complexity (Corollary 4.5.4). We further use Theorem 4.5.3 to deduce Corollaries 4.1.14 and 4.1.17(ii).

In Section 4.6, we introduce finitary analogues of tools from Section 4.2. These are needed in Section 4.7, in which we show that proving the equivalence of true and algebraic complexity for homogeneous progression comes down to proving Theorem 4.6.9, a finitary version of Theorem 4.5.3. We also explain in Section 4.7 how to prove Corollary 4.1.17(i). Theorem 4.6.9, the main technical part of this paper, is derived in Section 4.8. Unfortunately, Theorem 4.6.9 fails for inhomogeneous progressions, as explained in Section 4.9. In Section 4.10, we propose a method to handle inhomogeneous progressions. While we succeed in proving an analogue of Theorem 4.5.3 for the inhomogeneous progression $(x,\ x+y,\ x+2y,\ x+y^2)$ in Proposition 4.10.1, we have been unable to extend this construction to all inhomogeneous progressions. In Section 4.11, however, we give an alternative argument that allows us to prove Theorem 4.1.16. Subsequently, we show in 4.12 that Weyl and algebraic complexity are always equal, which is the main statement of Theorem 4.1.20. We conclude the paper by proving Theorem 4.1.18 in Section 4.13.

Acknowledgments

We are indebted to Donald Robertson for his comments on earlier versions of the paper and fruitful conversations on the project while it was carried out. We would also like to thank Sean Prendiville for introducing us to the topic of complexity, Tuomas Sahlsten for hosting a reading group on the dynamical proof of Szemerédi theorem, and Jonathan Chapman for useful discussions on algebraic relations between terms of polynomial progressions.

4.2 Infinitary nilmanifold theory

4.2.1 Basic definitions from ergodic theory

Let $(X, \mathcal{X}, \mu, T)$ be an invertible measure-preserving dynamical system (henceforth, we shall simply call it a *system*). The background in ergodic theory that we need can be found in [HK05b; HK18], among others; here, we only reiterate the most important definitions.

Definition 4.2.1. *A* factor *of a system* $(X, \mathcal{X}, \mu, T)$ *can be defined in three equivalent ways:*

1. *it is a* T*-invariant sub-*σ*-algebra of* $\mathcal{X}$*;*
2. *it is a system* $(Y, \mathcal{Y}, \nu, S)$ *together with a* factor map $\pi : X' \to Y'$*, i.e. a measurable map defined for a measurable* T*-invariant set* X' *of full measure, satisfying* $S \circ \pi = \pi \circ T$ *on* X' *and* $\mu \circ \pi^{-1} = \nu$*;*
3. *it is a* T*-invariant subalgebra of* $L^\infty(\mu)$*.*

For $r \in \mathbb{N}$, we let $\mathcal{K}_r$ be the factor spanned by all T^r-invariant functions in $L^\infty(\mu)$. In particular, $\mathcal{K}_1 = \mathcal{I}$ is the factor spanned by T-invariant functions, and the *rational Kronecker factor* $\mathcal{K}_{rat} = \bigvee_{r\in\mathbb{N}} \mathcal{K}_r$ is the factor spanned by all the functions in $L^\infty(\mu)$ that are T^r-invariant for some $r \in \mathbb{N}$. A system is *ergodic* if $\mathcal{K}_1 = \mathcal{I}$ is the trivial factor spanned by constant functions, and it is *totally ergodic* if $\mathcal{K}_{rat}$ is the trivial factor.

Of particular interest to us is a sequence of factors $(\mathcal{Z}_s)_{s\in\mathbb{N}}$ defined in [HK05b], which we refer to as *Host-Kra factors*. In accordance with Definition

4.2.1, we shall sometimes think of $\mathcal{Z}_s$ as a sub-σ-algebra of $\mathcal{X}$, and at other times we will consider a factor map $\pi_s : X \to Z_s$ and a factor $(Z_s, \mathcal{Z}_s, \lambda, S)$ of $(X, \mathcal{X}, \mu, T)$. If we concurrently talk about Host-Kra factors of two distinct spaces X and Y, we may write $Z_s(X)$ and $Z_s(Y)$ to mean Host-Kra factors of X and Y respectively. We do not explicitly use the definition of Host-Kra factors anywhere in the paper, and so we leave the interested reader to look it up in [HK05b; HK18]. Instead, we rely on two properties of this family of factors that concern their utility and structure respectively. First, these factors are characteristic for the convergence of polynomial progressions, as proved in Theorem 4.1.4. Rephrasing Theorem 4.1.4 in terms of Definition 4.1.5, we can say that each integral polynomial progression has a finite Host-Kra complexity. Second, each factor $\mathcal{Z}_s$ is an inverse limit[4] of s-step nilsystems [HK05b], which are objects of primary importance to us.

4.2.2 Nilsystems

Let G be a Lie group with connected component G^o and identity 1. A *filtration* on G of degree s is a chain of subgroups

$$G = G_0 = G_1 \geqslant G_2 \geqslant \ldots \geqslant G_s \geqslant G_{s+1} = G_{s+2} = \ldots = 1$$

satisfying $[G_i, G_j] \leqslant G_{i+j}$ for each $i, j \in \mathbb{N}$. We denote it as $G_\bullet = (G_i)_{i=0}^\infty$. A natural example of a filtration is the lower central series, given by $G_{k+1} = [G, G_k]$ for each $k > 1$, where the *commutator* of two elements $a, b \in G$ is defined as $[a, b] = a^{-1}b^{-1}ab$, and $[A, B]$ is the subgroup of G generated by all the commutators $[a, b]$ with $a \in A, b \in B$. The group G is *s-step nilpotent* if $G_{s+1} = 1$, where G_{s+1} is the s-th element of the lower central series of G. The only 0-step nilpotent group is the trivial group, and 1-step nilpotent groups are precisely abelian groups.

For the rest of the paper, we let G be a nilpotent Lie group and $\Gamma \leqslant G$ be a cocompact lattice. We call the quotient $X = G/\Gamma$ a *nilmanifold*. The group G acts on X by left translation, and for each $a \in G$, we call the map $T_a(g\Gamma) = (ag)\Gamma$ a *nilrotation*. Setting $\mathcal{G}/\Gamma$ to be the Borel σ-algebra of X and ν to be the Haar measure with respect to left translation, we call the system $(G/\Gamma, \mathcal{G}/\Gamma, \nu, T_a)$ a *nilsystem*.

A subgroup $H \leqslant G$ is *rational* if $H/(H \cap \Gamma)$ is closed in G/Γ. A filtration $G_\bullet$ is rational if G_i is a rational subgroup for each $i \in \mathbb{N}$. We shall assume

[4]The system $(X, \mathcal{X}, \mu, T)$ is an *inverse limit* of a sequence of factors $(X, \mathcal{X}_i, \mu, T)$ if $\mathcal{X}_i$ form an increasing sequence of factors of $\mathcal{X}$ such that $\mathcal{X} = \bigvee_{i\in\mathbb{N}} \mathcal{X}_i$ up to sets of measure zero.

throughout the paper that each filtration that we discuss is rational. The reason for making this assumption is twofold. In the study of true complexity, we use nilmanifolds coming from the inverse theorem for Gowers norms, which are endowed with rational filtrations [GTZ12]. While studying Host-Kra complexity, we work with a filtration whose rationality is equivalent to the rationality of the lower central series filtration, the latter fact being justified e.g. in Section 1 of [GT12].

In the case when $(G/\Gamma, \mathcal{G}/\Gamma, \nu, T_a)$ is an ergodic nilsystem, which will always be our case anyway, we can make two simplifying assumptions about the group G. By passing to a universal cover, we assume that G is simply connected. Replacing the nilsystem with several simpler nilsystems, we further assume that G is spanned by G^o and a. These assumptions, justified in Chapter 11 of [HK18], hold for the rest of the paper.

We also denote $\Gamma_i = G_i \cap \Gamma$ and $\Gamma^o = G^o \cap \Gamma$. The rationality of G_i in G means that Γ_i is cocompact in G_i.

Since our argument in the proof of Host-Kra complexity relies on reducing to the case of the system being a totally ergodic nilsystem, we now state several equivalent conditions of total ergodicity for nilsystems.

Proposition 4.2.2 (Conditions for total ergodicity of nilsystems, Corollary 7 and 8 of [HK18]). *Let $(G/\Gamma, \mathcal{G}/\Gamma, \nu, T_a)$ be an ergodic nilsystem. There exists $r \in \mathbb{N}_+$ such that $T_a^j(G^o/\Gamma^o)$ is totally ergodic with respect to T_a^r for all $0 \leqslant j < r$.*

Moreover, the following are equivalent:

1. *T_a is totally ergodic;*
2. *G/Γ is connected;*
3. *$G = G^o\Gamma$.*

Nilsystems allow a particularly simple description of factors. If $G_\bullet$ is the lower central series filtration, then

$$Z_s = \frac{G}{G_{s+1}\Gamma} \tag{4.12}$$

for all $s \in \mathbb{N}_+$ (see Chapter 11 of [HK18]). For $s = 0$, we have $Z_0 = G/(G^o\Gamma) \cong (\mathbb{Z}/r\mathbb{Z})$, where r is the smallest positive integer for which $a^r \in G^o$. It follows from Proposition 4.2.2 that Z_0 is trivial if and only if the nilsystem is totally ergodic.

Let $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. By Theorem 4.1.4, there exists $s \in \mathbb{N}$ such that for every ergodic system $(X, \mathcal{X}, \mu, T)$ and all choices of $f_0, ..., f_t \in L^\infty(\mu)$, we have

$$\begin{aligned} &\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_X f_0 \cdot T^{P_1(n)} f_1 \cdot ... \cdot T^{P_t(n)} f_t d\mu \\ &= \lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_{Z_s} \mathbb{E}(f_0|\mathcal{Z}_s) \cdot S^{P_1(n)}\, \mathbb{E}(f_1|\mathcal{Z}_s) \cdot ... \cdot S^{P_t(n)}\, \mathbb{E}(f_t|\mathcal{Z}_s) d\lambda, \end{aligned} \tag{4.13}$$

where $(Z_s, \mathcal{Z}_s, S, \lambda)$ is the appropriate Host-Kra factor of $(X, \mathcal{X}, \mu, T)$. Using the fact that Z_s is an inverse limit of ergodic s-step nilsystems, we can approximate the average (4.13) arbitrarily well by projections onto ergodic nilsystems. Hence we are left with understanding averages of the form

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_{G/\Gamma} \tilde{f}_0(b\Gamma) \cdot \tilde{f}_1(a^{P_1(n)} b\Gamma) \cdot ... \cdot \tilde{f}_t(a^{P_t(n)} b\Gamma) d\nu(b\Gamma) \tag{4.14}$$

where $\tilde{f}_i$ is the projection of f_i onto an ergodic s-step nilsystem $(G/\Gamma, \mathcal{G}/\Gamma, \nu, T_a)$ for all $0 \leqslant i \leqslant t$. If T is totally ergodic, then so is the nilrotation T_a.

4.2.3 Polynomial sequences

Let $G_\bullet$ be a filtration on G of degree s. A *polynomial sequence* $g : \mathbb{Z} \to G$ adapted to $G_\bullet$ is a sequence

$$g(n) = \prod_{i=0}^{s} g_i^{\binom{n}{i}}. \tag{4.15}$$

with the property that $g_i \in G_i$ for each i. Such sequences form a group denoted as $\text{poly}(\mathbb{Z}, G_\bullet)$ by Proposition 6.2 of [GT12]. One may ask why we define polynomial sequence as (4.15) rather than in the seemingly more natural form

$$g(n) = \prod_{i=0}^{s} g_i^{n^i}. \tag{4.16}$$

The reason is that if g is written in the form (4.15), then we have the following nice statement.

Lemma 4.2.3 (Lemma 2.8 of [CS12]). *Suppose that $g \in \text{poly}(\mathbb{Z}, G_\bullet)$. The sequence $g(n) = \prod_{i=0}^{s} g_i^{\binom{n}{i}}$ takes values in $H \leqslant G$ if and only if $g_0, ..., g_s \in H$.*

Proof. The converse direction is straightforward, and we prove the forward direction by induction on $0 \leqslant k \leqslant s$. For $k = 0$, we observe that $g_0 = g(0) \in H$. Suppose that the statement holds for k, i.e. $g_0, ..., g_k \in H$. Then

$g(k+1) = \left(\prod_{i=0}^{k} g_i^{\binom{k}{i}}\right) g_{k+1}$. Since $g(k+1), g_0, ..., g_k$ are all in H, it follows that $g_{k+1} \in H$. □

Lemma 4.2.3 is not true if g is written in the form (4.16); for instance, $g(n) = \binom{n}{2} = \frac{1}{2}n^2 - \frac{1}{2}n$ takes values in $\mathbb{Z}$ even though $\frac{1}{2}, -\frac{1}{2} \notin \mathbb{Z}$.

In a similar manner, we define for any $D \in \mathbb{N}_+$ the group $\mathrm{poly}(\mathbb{Z}^D, G_\bullet)$ of D-parameter polynomial sequence $g : \mathbb{Z}^D \to G$ adapted to $G_\bullet$, i.e. sequences of the form

$$g(n_1, ..., n_D) = \prod_{i=0}^{s} \prod_{i_1+...+i_D=i} g_{i_1,...,i_D}{}^{\binom{n_1}{i_1} \cdots \binom{n_D}{i_D}}$$

for $g_{i_1,...,i_D} \in G_{i_1+...i_D}$.

4.2.4 Infinitary equidistribution theory on nilmanifolds

For the rest of Section 4.2, we assume that G is connected. For $D \in \mathbb{N}_+$, a polynomial sequence $g \in \mathrm{poly}(\mathbb{Z}^D, G_\bullet)$ is *equidistributed* on G/Γ if

$$\mathop{\mathbb{E}}_{n \in [N]^D} F(g(n)\Gamma) \to \int_{G/\Gamma} F d\nu$$

for any continuous function $F : G/\Gamma \to \mathbb{C}$. The following notion from [GT12] is useful when discussing equidistribution because it allows us to formulate Theorem 4.2.5 below in a way that makes it easy to state its quantitative version (Theorem 4.6.7) later on.

Definition 4.2.4 (Horizontal characters). *A* horizontal character *on G is a continuous group homomorphism $\eta : G \to \mathbb{R}$ for which $\eta(\Gamma) \leqslant \mathbb{Z}$.*

In particular, each horizontal character vanishes on $[G, G]$.

Equidistribution on nilmanifolds was studied by Leibman, who provided a useful criterion for when a polynomial sequence is equidistributed on a nilmanifold.

Theorem 4.2.5 (Leibman's equidistribution theorem, [Lei05b]). *Let $D \in \mathbb{N}_+$ and $g \in \mathrm{poly}(\mathbb{Z}^D, G_\bullet)$. The following are equivalent:*

1. *g is equidistributed in G/Γ;*
2. *the projection of g onto $G/[G,G]$ is equidistributed in $G/[G,G]\Gamma$;*
3. *if $\eta : G \to \mathbb{R}$ is a horizontal character for which $\eta \circ g$ is constant, then η is trivial.*

We shall also need a stronger notion of equidistribution, that of irrational sequences, coming from [GT10].

Definition 4.2.6. *Suppose that $G_\bullet$ is a filtration on G and $i \in \mathbb{N}_+$, and let*

$$G_i^\nabla = \langle G_{i+1}, [G_j, G_{i-j}], 1 \leqslant j < i \rangle$$

An i-th level character *is a continuous group homomorphism $\eta_i : G_i \to \mathbb{R}$ that vanishes on G_i^∇ and satisfies $\eta_i(\Gamma_i) \subseteq \mathbb{Z}$. An element g_i of G_i is* irrational *if $\eta_i(g_i) \notin \mathbb{Z}$ for any nontrivial i-th level character η_i. A sequence $g(n) = \prod_{i=0}^{s} g_i^{\binom{n}{i}}$ is* irrational *if g_i is irrational for all $i \in \mathbb{N}_+$.*

All irrational sequences are equidistributed (Lemma 3.7 of [GT10]), but not vice versa. For instance, let $g(n) = a_1 n + ... + a_s n^s$ be a real-valued polynomial. It is a polynomial sequence in $\mathbb{R}$ adapted to the filtration $G_1 = ... = G_s = \mathbb{R}$, $G_{s+1} = 0$. With respect to this filtration, the i-th level characters of degree $i \neq s$ are all trivial because $G_i = G_{i+1} = G_i^\nabla$ while s-th level characters are precisely the maps $\eta_s(x) = kx$ for some $k \in \mathbb{Z}$. Thus, g is irrational iff $a_s \notin \mathbb{Q}$, and g is equidistributed iff there exists $1 \leqslant i \leqslant s$ with $a_i \notin \mathbb{Q}$. It is clear in this case that irrational implies equidistributed, but not vice versa.

We want to emphasise that whether a sequence is irrational or not depends on what filtration we are using, whereas the notion of equidistribution does not depend on the filtration.

4.3 Reducing to the case of connected groups

The expression (4.14) indicates that to understand Host-Kra complexity of a polynomial progression $\vec{P}$, we have to understand the distribution of orbits

$$(b\Gamma,\ a^{P_1(n)}b\Gamma,\ ...,\ a^{P_t(n)}b\Gamma) \tag{4.17}$$

inside a connected nilmanifold G^{t+1}/Γ^{t+1}. The point of this section is to show that we can replace linear orbits $(a^n b\Gamma)_{n\in\mathbb{N}}$ on G/Γ by polynomial orbits $(g_b(n)\Gamma^o)_{n\in\mathbb{N}}$ on G^o/Γ^o for some irrational polynomial sequence $g_b : \mathbb{Z} \to G^o$ with respect to a certain naturally defined filtration $G^o_\bullet$ on G^o. This way, we want to reduce the question of finding the closure for (4.17) inside $(G/\Gamma)^{t+1}$ to finding the closure for

$$(g_b(m)\Gamma^o,\ g_b(m + P_1(n))\Gamma^o,\ ...,\ g_b(m + P_t(n))\Gamma^o) \tag{4.18}$$

inside $(G^o/\Gamma^o)^{t+1}$. The connectedness of G^o then allows us to use tools from Section 4.2.4. We start with the following simple lemma that allows us to introduce another variable.

Lemma 4.3.1. *Let $(G/\Gamma, \mathcal{G}/\Gamma, \nu, T_a)$ be a nilsystem and $F : (G/\Gamma)^{t+1} \to \mathbb{R}$. Then*

$$\begin{aligned}&\mathop{\mathbb{E}}_{n\in[N]} \int_{G/\Gamma} F(b\Gamma, a^{P_1(n)}b\Gamma, ..., a^{P_t(n)}b\Gamma) d\nu(b\Gamma)\\ &= \mathop{\mathbb{E}}_{m,n\in[N]} \int_{G/\Gamma} F(a^m b\Gamma, a^{m+P_1(n)}b\Gamma, ..., a^{m+P_t(n)}b\Gamma) d\nu(b\Gamma).\end{aligned}$$

Proof. Since T_a is measure preserving, we have

$$\begin{aligned}&\int_{G/\Gamma} F(b\Gamma, a^{P_1(n)}b\Gamma, ..., a^{P_t(n)}b\Gamma) d\nu(b\Gamma)\\ &= \int_{G/\Gamma} F(a^m b\Gamma, a^{m+P_1(n)}b\Gamma, ..., a^{m+P_t(n)}b\Gamma) d\nu(b\Gamma)\end{aligned}$$

for any $m, n \in \mathbb{N}$. Consequently,

$$\begin{aligned}&\int_{G/\Gamma} F(b\Gamma, a^{P_1(n)}b\Gamma, ..., a^{P_t(n)}b\Gamma) d\nu(b\Gamma)\\ &= \mathop{\mathbb{E}}_{m\in[N]} \int_{G/\Gamma} F(a^m b\Gamma, a^{m+P_1(n)}b\Gamma, ..., a^{m+P_t(n)}b\Gamma) d\nu(b\Gamma),\end{aligned}$$

from which the lemma follows. □

The main result of this section is the following.

Proposition 4.3.2. *Let $(G/\Gamma, \mathcal{G}/\Gamma, \nu, T_a)$ be a totally ergodic nilsystem and $b \in G^o$. Suppose that $G_\bullet$ is the lower central series filtration on G and $G^o_\bullet = G_\bullet \cap G^o$. Then there exists an irrational sequence $g_b \in \mathrm{poly}(\mathbb{Z}, G^o_\bullet)$ such that $g_b(n)\Gamma = a^n b\Gamma$.*

We observe that with this filtration on G^o, we have $G^o_k = G_k$ for $k \geqslant 2$. That follows from the fact that the groups G_k are connected for $k \geqslant 2$ (Lemma 5 of [HK18]), and hence are contained in G^o.

We lose no generality in assuming that $b \in G^o$; Proposition 4.2.2 and the connectedness of G/Γ imply that for all $b \in G$ there exists $b' \in G^o$ such that $b\Gamma = b'\Gamma$.

Proof. The connectedness of G/Γ implies that $G = G^o\Gamma$, and so there exist $\alpha \in G^o$ and $\gamma \in \Gamma$ such that $a = \alpha\gamma^{-1}$. Then

$$a^n b\Gamma = (\alpha\gamma^{-1})^n b\Gamma = (\alpha\gamma^{-1})^n b\gamma^n\Gamma.$$

It follows from the normality of G^o and the fact that α and b are elements of G^o that the sequence $g_b(n) = (\alpha\gamma^{-1})^n b\gamma^n$ takes values in G^o. Since the sequences $h_1(n) = a^n b$ and $h_2(n) = \gamma^n$ are adapted to $G_\bullet$, and the set $\mathrm{poly}(\mathbb{Z}, G_\bullet)$ is a group, we deduce that $g_b = h_1 h_2$ is adapted to $G^o_\bullet = G_\bullet \cap G^o$.

We want a more precise description of g_b, and for this we shall use some results from Sections 11-13 of [Lei09]. Let $g = g_b$ for the identity $b = 1$; that is, $g(n) = (\alpha\gamma^{-1})^n\gamma^n$. Leibman showed in Section 11.2 of [Lei09] that

$$g(n) = \prod_{1\leqslant k_1\leqslant s}(A^{k-1}\alpha)^{q_{k_1}(n)} \prod_{1\leqslant k_2<k_1<s}[A^{k_1-1}\alpha, A^{k_2-1}\alpha]^{q_{k_1,k_2}(n)}..., \tag{4.19}$$

where $Ax = [x, \gamma]$ and $q_{k_1,...,k_r}$ are integral polynomials with $\deg q_{k_1,...,k_r} \leqslant k_1 + ... + k_r$. More explicitly, we have

$$g(n) = \alpha^n(A\alpha)^{\binom{n}{2}}(A^2\alpha)^{\binom{n}{3}}\cdots[A\alpha, \alpha]^{\binom{n}{3}}[A^2\alpha, \alpha]^{\binom{n}{4}} \tag{4.20}$$

$$\cdots[A^2\alpha, A\alpha]^{4\binom{n+1}{5}}[A^3\alpha, A\alpha]^{5\binom{n+1}{6}}\cdots \tag{4.21}$$

The coefficients of g can be analysed using a family of subgroups of G^o introduced in Section 12 of [Lei09]. For $k_1, ..., k_l \in \mathbb{N}_+$, we let $G^o_{(k_1,...,k_l)}$ be the subgroup of G^o generated by all l-fold commutators[5] of elements of the form $A^{k_1-1}h_1$, ..., $A^{k_l-1}h_l$ for $h_1, ..., h_l \in G^o$. We then define

$$G^o_{k,l} = \langle G^o_{(k_1,...,k_i)} : i \geqslant l,\ k_1 + ... + k_l \geqslant k\rangle$$

for integers $1 \leqslant l \leqslant k$ and set $G^o_{k,l} = G^o_{l,l}$ whenever $l > k$.

The following lemma lists some basic properties of the groups $G^o_{k,l}$ that we shall use.

Lemma 4.3.3. *For any integers $1 \leqslant l \leqslant k$,*

1. *$G^o_{k,l}$ is normal in G;*
2. *$[G^o_{k,l}, G^o_{i,j}] \leqslant G^o_{k+i,l+j}$ for any integers $1 \leqslant i \leqslant j$;*
3. *$A^j G^o_{k,l} \leqslant G^o_{k+j,l}$ for any $j \in \mathbb{N}$;*
4. *$G^o_{k+1,l}$ and $G^o_{k,l+1}$ are subgroups of $G^{k,l}$, and the quotient groups $G^o_{k,l}/G^o_{k+1,l}$ and $G^o_{k,l}/G^o_{k,l+1}$ are abelian;*
5. *for $k \geqslant 2$, $G_k = G^0_k = G^o_{k,1} = \langle A^{k-1}G^o, G^o_{k,2}\rangle = \langle AG^o_{k-1}, G^o_{k,2}\rangle$;*

[5] A 1-fold commutator is any element $h \in G$. For $l > 1$, an l-fold commutator is an element of the form $[h_i, h_j]$, where h_i is an i-fold commutator, h_j is an j-fold commutator and $i + j = l$.

6. $(G^o)^\nabla_k = \langle G^o_{k,2}, G^o_{k+1}\rangle$

Proof. Properties (i)-(iv) are proved in Lemma 12.2 of [Lei09]. For $k \geqslant 2$, the statement $G_k = G^o_k$ in (v) is true by definition, and the statement $G_k = G^o_{k,1}$ is proved in Lemma 12.3 of [Lei09]. To finish the proof of (v), it remains to show that $G^o_{k,1} = \langle A^{k-1}G^o, G^o_{k,2}\rangle = \langle AG^o_{k-1}, G^o_{k,2}\rangle$ for $k \geqslant 2$. For $k = 2$, this is true by definition of $G^o_{k,1}$ and the fact that $G^o_{k,2} \geqslant G^o_{k,3} \geqslant \ldots$, which follows from part (iv). We assume that the statement is true for some $k \geqslant 2$. That G^0_{k+1} contains $\langle AG^o_k, G^o_{k+1,2}\rangle$ follows from the fact that both AG^o_k and $G^o_{k+1,2}$ are contained in the $(k+1)$-th element of the lower central series of G, which is precisely G^o_{k+1}. For the other direction, we observe that

$$\begin{aligned} G^o_{k+1} = [G_k, G] = [G^o_k, \langle G^o, \gamma\rangle] &\leqslant \langle [G^o_k, G^o], [G^o_k, \gamma]\rangle \\ &\leqslant \langle [A^{k-1}G^o, G^o], [G^o_{k,2}, G^o], AG^o_k\rangle \leqslant \langle G^o_{k+1,2}, AG^0_k\rangle. \end{aligned}$$

A similar argument shows that $G^o_{k+1} = \langle A^k G^o, G^o_{k+1,2}\rangle$.

Before we prove property (vi), we recall that $(G^o)^\nabla_k = \langle G_{k+1}, [G_j, G_{k-j}] : 1 \leqslant j < k\rangle$. That (vi) holds for $k = 1$ can be verified by inspection. For $k \geqslant 2$, we observe that $[A^{j-1}G^o, A^{k-j-1}G^o] \leqslant [G^0_j, G^0_{k-j}]$, and so

$$G^o_{k,2} \leqslant \langle [G^o_j, G^o_{k-j}] : 1 \leqslant j < k\rangle;$$

when coupled with property (v), this implies that $(G^o)^\nabla_k \geqslant \langle G^o_{k,2}, G^o_{k+1}\rangle$. For the converse, we have

$$[G^0_j, G^o_{k-j}] = [\langle A^{j-1}G^o, G^o_{j,2}\rangle, \langle A^{k-j-1}G^o, G^o_{k-j,2}\rangle] \leqslant \langle G^o_{k,2}, G^o_{k,3}, G^o_{k,4}\rangle \leqslant G^o_{k,2},$$

for each $1 \leqslant j < k$, from which it follows that $(G^o)^\nabla_k \leqslant \langle G^o_{k,2}, G^o_{k+1}\rangle$. □

Letting $g(n) = \prod_{i=1}^s g_i^{\binom{n}{i}}$, we observe from (4.19), (4.20) as well as parts (v) and (vi) of Lemma 4.3.3 that

$$g_i = A^{i-1}\alpha \mod (G^o)^\nabla_i. \tag{4.22}$$

For an arbitrary $b \in G^o$, we have $g_b(n) = a^n b\gamma^n = b(\alpha_b\gamma^{-1})^n\gamma^n$, where $\alpha_b = \alpha[\alpha, b]Ab$, as observed in Section 11.3 of [Lei09]. Letting $g_b(n) = \prod_{i=0}^s g_{b,i}^{\binom{n}{i}}$, it is therefore true that

$$g_{b,i} = A^{i-1}\alpha_b = A^{i-1}\alpha \mod (G^o)^\nabla_i \tag{4.23}$$

for all $i \in \mathbb{N}_+$.

For $i = 1$, we have $g_{b,1} = \alpha \mod G^o_2$, and we claim that $g_{b,i}$ is irrational. The ergodicity of a implies that for almost every b, the sequence $n \mapsto a^n b$ is equidistributed in G/Γ, and so the same is true for the sequence g_b in G^o/Γ^o. Consequently, the projection $\pi(g_b) : \mathbb{Z} \to G^o/(G^o_2\Gamma^o)$ is equidistributed as well. Since $\pi(g_b(n)) = \pi(b) + \pi(\alpha)n$, it follows that $\pi(\alpha)$ is an irrational element of G^o/G^o_2, and so $g_{b,1}$ is an irrational element of G^o.

Before proving that $g_{b,i}$ are irrational for $i > 1$, we discuss some properties of the map $A : G \to G$. From the definition of the filtration $G^o_\bullet$ we observe that $AG^o_i \leqslant G^o_{i+1}$ for all $i \geqslant 1$ (this is also a consequence of parts (iv) and (v) of Lemma 4.3.3). Therefore the map $A_i := A|_{G^o_i}$ takes values in G^o_{i+1}, and moreover $A_i(\Gamma_i) \leqslant \Gamma_{i+1}$. We also observe that the projection $\overline{A}_i : G^o_i \to G^o_{i+1}/(G^o)^\nabla_{i+1}$ is a (continuous) group homomorphism because

$$A(xy) = [xy, \gamma] = [x, \gamma][[x, \gamma], y][y, \gamma] = Ax[Ax, y]Ay = AxAy \mod G^o_{2i+1,2}$$

for any $x, y \in G^o_i$ and $G^o_{2i+1,2} \leqslant G^o_{i+1,2} \leqslant (G^o)^\nabla_{i+1}$ by parts (iv) and (vi) of Lemma 4.3.3. From part (v) of Lemma 4.3.3 it follows that $\overline{A}_i$ is surjective. Finally, we note using parts (iii) and (v) of Lemma 4.3.3 that $A_i((G^o)^\nabla_i) \leqslant (G^o)^\nabla_{i+1}$.

Suppose that $g_{b,i}$ is irrational but $g_{b,i+1}$ is not for some $1 \leqslant i < s$. Then there exists a nontrivial $(i+1)$-th level character $\eta_{i+1} : G^o_{i+1} \to \mathbb{R}$ such that $\eta_{i+1}(g_{b,i+1}) \in \mathbb{Z}$. From (4.23) and the fact that η_{i+1} vanishes on $(G^o)^\nabla_{i+1}$, we deduce that $\eta_{i+1}(g_{b,i+1}) = \eta_{i+1}(A^i\alpha)$. We also let $\overline{\eta}_{i+1} : G^o_{i+1}/(G^o)^\nabla_{i+1} \to \mathbb{R}$ be the induced map.

Let $\eta_i := \eta_{i+1} \circ A_i : G^o_i \to \mathbb{R}$. It is an i-th level character as a consequence of four facts: the vanishing of η_{i+1} on $(G^o)^\nabla_{i+1}$, the inclusion $(G^o_{i+1,2}) \leqslant (G^o)^\nabla_{i+1}$ (both of which imply that $\eta_i = \overline{\eta}_{i+1} \circ \overline{A}_i$ is a continuous group homomorphism), the inclusion $A_i((G^o)^\nabla_i) \leqslant (G^o)^\nabla_{i+1}$, and the fact that $\eta_i(\Gamma_i) \leqslant \mathbb{Z}$. It moreover satisfies

$$\eta_i(g_{b,i}) = \eta_i(A^{i-1}\alpha) = \eta_{i+1}(A^i\alpha) = \eta_{i+1}(g_{b,i+1}),$$

implying that $\eta_i(g_{b,i}) \in \mathbb{Z}$. The nontriviality of η_{i+1} implies that $\overline{\eta}_{i+1}$ and $\overline{A}_i$ are surjective maps onto nontrivial groups; hence η_i is nontrivial. This contradicts the irrationality of $g_{b,i}$. By induction, $g_{b,1}$, ..., $g_{b,s}$ are all irrational, implying that g_b is irrational. This finishes the proof of Proposition 4.3.2. □

Proposition 4.3.2 is vaguely reminiscent of Proposition 3.1 of [FK05] in that we replace a linear sequence by a polynomial object on a simpler space. These two results are not equivalent, however, in that in Proposition 4.3.2, we end up with a polynomial sequence on a nilmanifold of a connected group whereas in Proposition 3.1 of [FK05], one obtains a unipotent affine transformation on a torus.

Lemma 4.3.4. *Let $G_\bullet$ and $G_\bullet^o$ be as given in Proposition 4.3.2. Then $Z_i(G/\Gamma) = Z_i(G^o/\Gamma^o) = G^o/(G^o_{i+1}\Gamma^o)$ for each $i \in \mathbb{N}$.*

Proof. We do the cases $i = 0$ and $i > 0$ separately. For $i > 0$, we recall from (4.12) that $Z_i(G/\Gamma) = G/G_{i+1}\Gamma$. Since $G/\Gamma = G^o/\Gamma^o$ by connectedness of G/Γ, and $G_j = G_j^o$ for $j \geqslant 2$, it follows that

$$Z_i(G^o/\Gamma^o) = Z_i(G/\Gamma) = G/G_{i+1}\Gamma = G^o/G^o_{i+1}\Gamma^o.$$

For $i = 0$, we have $Z_i(G/\Gamma) = G/G^o\Gamma = 1 = G^o/G^o\Gamma^o = Z_i(G^o/\Gamma^o)$. □

4.4 Homogeneous and inhomogeneous polynomial progressions

The central message of this paper is that homogeneous polynomial progressions satisfy certain linear algebraic properties that make them pliable for our analysis. In this section, we explicitly describe these properties.

Let $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ be an integral polynomial progression. Let V_k be the subspace of $\mathbb{R}[x, y]$ given by

$$\begin{aligned} V_k &= \mathrm{Span}_{\mathbb{R}}\{(x + P_i(y))^j : \ 0 \leqslant i \leqslant t, \ 1 \leqslant j \leqslant k\} \\ &= \mathrm{Span}_{\mathbb{R}}\left\{\binom{x + P_i(y)}{j} : \ 0 \leqslant i \leqslant t, \ 1 \leqslant j \leqslant k\right\}, \end{aligned}$$

and similarly let

$$W_k = \mathrm{Span}_{\mathbb{R}}\left\{\binom{x + P_i(y)}{k} : \ 0 \leqslant i \leqslant t\right\}.$$

We also set

$$\begin{aligned} V^* = \mathrm{Span}_{\mathbb{R}}\{(Q_0, ..., Q_t) \in \mathbb{R}[u]^{t+1} : \ & Q_0(0) = ... = Q_t(0) = 0, \\ & Q_0(x) + Q_1(x + P_1(y)) + ... + Q_t(x + P_t(y)) = 0\} \end{aligned}$$

to be the space of all algebraic relations with zero constant terms satisfied by $\vec{P}$. We recall that an algebraic relation $(Q_0, ..., Q_t)$ is homogeneous if there exists $d \in \mathbb{N}_+$ and $a_0, ..., a_d \in \mathbb{R}$ not all zero such that $Q_i(u) = a_i u^d$ for each $0 \leqslant i \leqslant t$. We call $\vec{P}$ homogeneous if V^* is spanned by homogeneous algebraic relations, and inhomogeneous otherwise.

The concepts of integral polynomial progression and homogeneity, as well as our results in this paper, could be extended to multiparameter polynomial progressions of the form

$$(x,\ x + P_1(y_1, ..., y_r),\ ...,\ x + P_t(y_1, ..., y_r));$$

however, we do not pursue this generalisation so as not to obfuscate the notation.

Some important examples of homogeneous progressions include:

1. linear progressions $(x,\ x + a_1 y,\ ...,\ x + a_t y)$ for distinct nonzero integers $a_1, ..., a_t$, as well as their multiparameter generalizations;

2. progressions of algebraic complexity 0, i.e. progressions where the polynomials $P_1, ..., P_t$ are integral and linearly independent;

3. progressions of algebraic complexity 1, such as $(x,\ x+y,\ x+y^2,\ x+y+y^2)$, which satisfy no quadratic or higher-order algebraic relation.

Another, less obvious example of a homogeneous progression is $(x,\ x+y,\ x+2y,\ x+y^3)$, already mentioned in the introduction, which only satisfies the homogeneous relation

$$x - 2(x+y) + (x+2y) = 0. \tag{4.24}$$

This progression should be contrasted with $(x,\ x+y,\ x+2y,\ x+y^2)$, which is inhomogeneous because it satisfies both (4.24) and the inhomogeneous relation

$$x^2 + 2x - 2(x+y)^2 + (x+2y)^2 - 2(x+y^2) = 0 \tag{4.25}$$

that cannot be written down as a sum of homogeneous relations. More generally, progressions of the form

$$(x,\ x+y,\ ...,\ x+(t-1)y,\ x+P_t(y))$$

are all inhomogeneous whenever $1 < \deg P_t < t$.

For $k \in \mathbb{N}_+$, we define

$$W_k^c = W_k \cap \sum_{j \neq k} W_j \quad \text{and} \quad W^c = \sum_k W_k^c,$$

as well as the family of quotient spaces

$$W_k' = W_k / W_k^c = W_k / \left(W_k \cap \sum_{j \neq k} W_j \right).$$

The space W_k^c captures all the polynomials in W_k that "participat" in inhomogeneous algebraic relations, an intuition made more precise by the result below and the examples discussed below Proposition 4.4.2. The notation W_k^c is supposed to signify the fact that W_k^c is a complement of the subspace W_k' inside W_k.

Proposition 4.4.1 (Equivalent conditions for homogeneity)**.** *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ be an integral polynomial progression. The following are equivalent:*

1. *$\vec{P}$ is homogeneous;*
2. *W_k^c is trivial for each $k \in \mathbb{N}_+$;*
3. *$W_k' = W_k$ for each $k \in \mathbb{N}_+$.*

Proof. The equivalence of (ii) and (iii) follows trivially from the definition of W_k', and we focus on showing the equivalence of (i) and (ii) instead. The inhomogeneity of $\vec{P}$ implies the existence of a nontrivial algebraic relation $(Q_0(u), ..., Q_t(u)) = (\sum_k a_{0k} u^k, ..., \sum_k a_{tk} u^k)$ that is not a sum of homogeneous algebraic relations. What this means is that there exists $k \in \mathbb{N}_+$ for which

$$R(x, y) = a_{0k} x^k + a_{1k}(x + P_1(y))^k + ... + a_{tk}(x + P_t(y))^k \neq 0.$$

Since

$$Q_0(x) + Q_1(x + P_1(y)) + \cdots + Q_t(x + P_t(y)) = 0,$$

we have

$$R(x, y) = -\sum_{j \neq k} \sum_{i=0}^{t} a_{ij}(x + P_i(y))^j \in \sum_{j \neq k} W_j,$$

and so $W_k^c = W_k \cap \sum_{j \neq k} W_j$ is nonempty. Thus (ii) implies (i) by contrapositive. The argument can be reversed, and so (i) and (ii) are in fact equivalent. $\square$

For homogeneous progression, it is quite straightforward to obtain an upper bound on algebraic complexity.

Proposition 4.4.2. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a homogeneous polynomial progression. Then $\mathcal{A}_i(\vec{P}) \leqslant t-1$ for each $0 \leqslant i \leqslant t$.*

This bound is sharp, as evidenced by the example of arithmetic progressions.

Proof. By homogeneity of $\vec{P}$, the only algebraic relations of degree t that $\vec{P}$ could satisfy are of the form

$$a_0 \binom{x}{t} + a_1 \binom{x + P_i(y)}{t} + \ldots + a_t \binom{x + P_t(y)}{t} = 0. \tag{4.26}$$

A relation (4.26), together with the formula

$$\binom{x+P_i(y)}{t} = \binom{x}{t} + \binom{x}{t-1} P_i(y) + \binom{x}{t-2}\binom{P_i(y)}{2} + \ldots + \binom{P_i(y)}{t},$$

implies

$$a_1 \binom{P_i(y)}{k} + \ldots + a_t \binom{P_t(y)}{k} = 0$$

for $1 \leqslant k \leqslant t$. From the invertibility of Vandermonde matrix it follows that this is only possible when $a_1 = \ldots = a_t = 0$, in which case $a_0 = 0$ as well. Hence $\vec{P}$ satisfies no nontrivial relation of degree t.

$\square$

Proposition 4.4.1 implies that homogeneous progressions satisfy

$$V_k = \bigoplus_{i=1}^{k} W_i = \bigoplus_{i=1}^{k} W_i'. \tag{4.27}$$

In the inhomogeneous case, we instead have

$$V_k = \sum_{i=1}^{k} W_i = \left(\bigoplus_{i=1}^{k} W_i'\right) \oplus (W^c \cap V_k) \tag{4.28}$$

for some nontrivial subspace $W^c \cap V_k$. The nontriviality of this subspace is the main source of difficulty preventing us from generalising Theorem 4.1.12 to inhomogeneous progressions.

Given the rather abstract nature of the spaces W_k, W_k' and W_k^c, we illustrate their definitions with concrete examples. For the homogeneous progression $(x,\ x+y,\ x+2y,\ x+y^3)$, we have

$$\begin{aligned} W_1' &= W_1 = \mathrm{Span}_{\mathbb{R}}\{x, y, y^3\} \\ \text{and}\quad W_2' &= W_2 = \mathrm{Span}_{\mathbb{R}}\left\{\binom{x}{2}, xy + \binom{y}{2}, y^2, xy^3 + \binom{y^3}{2}\right\}, \end{aligned}$$

while for the inhomogeneous progression $(x,\ x+y,\ x+2y,\ x+y^2)$, we have

$$W_1 = \mathrm{Span}_{\mathbb{R}}\{x, y, y^2\} \quad \text{and} \quad W_2 = \mathrm{Span}_{\mathbb{R}}\left\{\binom{x}{2}, xy + \binom{y}{2}, y^2, xy^2 + \binom{y^2}{2}\right\}$$

but

$$\begin{aligned} W_1' = \mathrm{Span}_{\mathbb{R}}\{x, y\}, \quad W_2' &= \mathrm{Span}_{\mathbb{R}}\left\{\binom{x}{2}, xy + \binom{y}{2}, xy^2 + \binom{y^2}{2}\right\} \\ \text{and}\quad W^c &= \mathrm{Span}_{\mathbb{R}}\{y^2\}. \end{aligned}$$

The nontriviality of W^c for the latter progression is intrinsically related to the algebraic relation (4.25).

The spaces V_k and W_k are subspaces of $\mathbb{R}[x,y]$. We also need an analogous family of subspaces of $\mathbb{R}^{t+1}$. Let

$$\begin{aligned} \mathcal{P}_k &= \mathrm{Span}_{\mathbb{R}}\left\{\left(\binom{x}{j}, \binom{x+P_1(y)}{j}, \ldots, \binom{x+P_t(y))}{j}\right) : x, y \in \mathbb{R},\ 1 \leqslant j \leqslant k\right\} \\ &= \mathrm{Span}_{\mathbb{R}}\left\{\left(\binom{x}{k}, \binom{x+P_1(y)}{k}, \ldots, \binom{x+P_t(y))}{k}\right) : x, y \in \mathbb{R}\right\} \\ &= \mathrm{Span}_{\mathbb{R}}\left\{(x^k, (x+P_1(y))^k, \ldots, (x+P_t(y))^k) : x, y \in \mathbb{R}\right\}. \end{aligned}$$

We shall also use the notation

$$\begin{aligned} \vec{P}^k(x,y) &= (x^k, (x+P_1(y))^k, \ldots, (x+P_t(y))^k) \\ \text{and}\ \binom{\vec{P}(x,y)}{k} &= \left(\binom{x}{k}, \binom{x+P_1(y)}{k}, \ldots, \binom{x+P_t(y)}{k}\right). \end{aligned}$$

The equivalence of three formulations of $\mathcal{P}_k$ may not be obvious at first glance. The first two formulations are equivalent because if $(a_0, \ldots, a_t)$ is the coefficient of $\binom{x}{i}\binom{y}{l}$ in $\binom{\vec{P}(x,y)}{j}$, then it will be the coefficient of $\binom{x}{i+k-j}\binom{y}{l}$ in $\binom{\vec{P}(x,y)}{k}$ whenever $j \leqslant k$. The equivalence of the last two formulations follows by induction on k.

Let $t_k = \dim W_k$ and $t'_k = \dim W'_k$ for each $k \in \mathbb{N}$. The spaces W_k and $\mathcal{P}_k$ can be related as follows. Let $\{Q_{k,1}, ..., Q_{k,t_k}\}$ be a basis for W_k. Then

$$\left(\binom{x}{k}, \binom{x+P_1(y))}{k}, ..., \binom{x+P_t(y)}{k}\right) = \sum_{j=1}^{t_k} \vec{v}_{k,j} Q_{k,j}(x,y)$$

for some linearly independent vectors $\vec{v}_{k,1}, ..., \vec{v}_{k,t_k} \in \mathbb{R}^{t+1}$. We let $\tau_k(Q_{k,j}) = \vec{v}_{k,j}$, and extend this map to all of W_k by linearity. This map depends on the choice of the basis for W_k. It is surjective by the definition of $\mathcal{P}_k$ and injective by the linear independence of $\vec{v}_{k,1}, ..., \vec{v}_{k,t_k}$. Hence it is an isomorphism. In particular, Proposition 4.4.1 implies that $W'_k \cong \mathcal{P}_k$ whenever $\vec{P}$ is homogeneous.

For instance, for $(x,\ x+y,\ x+2y,\ x+y^3)$, the isomorphisms τ_1 and τ_2 are given by

$$\tau_1(x) = (1,1,1,1), \quad \tau_1(y) = (0,1,2,0), \quad \tau_1(y^3) = (0,0,0,1)$$

and

$$\begin{aligned} \tau_2\left(\binom{x}{2}\right) &= (1,1,1,1), \quad \tau_2\left(xy + \binom{y}{2}\right) = (0,1,2,0),\\ \tau_2(y^2) &= (0,0,1,0), \quad \tau_2\left(xy^3 + \binom{y^3}{2}\right) = (0,0,0,1). \end{aligned}$$

We treat $\mathbb{R}^{t+1}$ as an $\mathbb{R}$-algebra with coordinatewise multiplication $\vec{v} \cdot \vec{w} = (v(0)w(0), ..., v(t)w(t))$ for $\vec{v} = (v(0), ..., v(t))$ and $\vec{w} = (w(0), ..., w(t))$. We similarly let $A \cdot B = \{\vec{a} \cdot \vec{b} : \vec{a} \in A, \vec{b} \in B\}$ be the product set of A and B for any $A, B \subseteq \mathbb{R}^{t+1}$. With these definitions, we observe that $\mathcal{P}_{i+j} \leqslant \mathcal{P}_i \cdot \mathcal{P}_j$, but the converse is in general not true. We also set $\vec{e}_i$ to be the coordinate vector with 1 in the i-th place and 0 elsewhere.

4.5 Relating Host-Kra complexity to algebraic complexity

Having introduced the notation for the spaces $\mathcal{P}_i$, we are ready to show precisely how determining Host-Kra complexity for homogeneous progressions can be reduced to a certain equidistribution problem on nilmanifolds. We start by defining a group which contains the orbit (4.18). Groups of this form have previously been defined in [Lei09; GT10; CS12; Kuc21b], among others.

Definition 4.5.1 (Leibman group)**.** *Let $t \in \mathbb{N}_+$ and G be a connected group with a filtration $G_\bullet$ of degree s. For an integral polynomial progression $\vec{P} \in \mathbb{R}[x,y]^{t+1}$, we define the associated* Leibman group *to be*

$$G^P = \langle g_i^{\vec{v}_i} : g_i \in G_i, \vec{v}_i \in \mathcal{P}_i, 1 \leqslant i \leqslant s \rangle,$$

where $h^{\vec{v}} = (h^{v(0)}, ..., h^{v(t)})$ for any $h \in G$ and $\vec{v} = (v(0), ..., v(t)) \in \mathbb{R}^{t+1}$. We also set $\Gamma^P = G^P \cap G^{t+1}$. If $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$, then we denote

$$g^P(x,y) = (g(x), g(x + P_1(y)), ..., g(x + P_t(y)))$$

and observe that g^P takes values in G^P.

Lemma 4.5.2. *Let $t \in \mathbb{N}_+$ and G be a connected group with a filtration $G_\bullet$ of degree s. Suppose that $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ is an integral polynomial progression with $\mathcal{A}_i(\vec{P}) = s'$ for some $s' \in \mathbb{N}$ and $0 \leqslant i \leqslant t$. Then G^P contains $1^i \times G_{s'+1} \times 1^{t-i}$.*

Proof. The assumption $\mathcal{A}_i(\vec{P}) = s'$ implies that $(x + P_i(y))^{s'+1}$ is linearly independent from $(x + P_k(y))^{s'+1}$ for $k \neq i$, hence $\mathcal{P}_{s'+1}$ contains $\vec{e}_i$. The Lemma then follows by the definition of G^P. □

We are now ready to state an infinitary version of the main technical result in the paper. This result constitutes the first part of Theorem 4.1.19.

Theorem 4.5.3. *Let $t \in \mathbb{N}_+$ and G be a connected group with filtration $G_\bullet$. Suppose that $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ is irrational and that $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ is a homogeneous polynomial progression. Then g^P is equidistributed on the nilmanifold G^P/Γ^P.*

Importantly, Theorem 4.5.3 fails for inhomogeneous progressions in that for each inhomogeneous progression $\vec{P}$, we can find a nilmanifold G/Γ, a filtration $G_\bullet$, and an irrational sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ for which the orbit of g^P is contained in a proper subnilmanifold of G^P/Γ^P. An example of this is given in Section 4.9.

We have all the tools to prove Theorem 4.5.3 by now. However, we will later need a finitary version of Theorem 4.5.3, and so instead of proving twice what is essentially the same result, we shall only give the finitary proof later on and deduce Theorem 4.5.3 from it. For now, however, we can show how the $\mathcal{HK}_i(\vec{P}) \leqslant \mathcal{A}_i(\vec{P})$ part of Theorem 4.1.12 follows from Theorem 4.5.3.

Corollary 4.5.4. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a homogeneous polynomial progression. For any $0 \leqslant i \leqslant t$, we have*

$$\mathcal{HK}_i(\vec{P}) \leqslant \mathcal{A}_i(\vec{P}).$$

The converse inequality will follow from showing that algebraic complexity equals Weyl complexity, and that Weyl complexity is less than or equal to Host-Kra complexity, both of which are done in Section 4.12.

Proof of Corollary 4.5.4 using Theorem 4.5.3. Let $\mathcal{A}_i(\vec{P}) = s$. Let $(X, \mathcal{X}, \mu, T)$ be a totally ergodic system, $f_0, ..., f_t \in L^\infty(\mu)$, and suppose that $\mathbb{E}(f_i|\mathcal{Z}_s) = 0$. By Theorem 4.1.4, the expression

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_X f_0 \cdot T^{P_1(n)} f_1 \cdots T^{P_t(n)} f_t d\mu \tag{4.29}$$

remains unchanged if we project the functions $f_0, ..., f_t$ onto the factor $\mathcal{Z}_{s_0}$ for some $s_0 \in \mathbb{N}$. If $s_0 < s$, then $\mathbb{E}(f_i|\mathcal{Z}_{s_0}) = 0$ and the limit (4.29) is 0, so we can assume that $s_0 \geqslant s$. Since the factor $\mathcal{Z}_{s_0}$ is an inverse limit of s_0-step nilsystems, we can approximate X arbitrarily well by s_0-step nilsystems. More precisely, there exists an increasing sequence of factors $\mathcal{X}_k$ satisfying $\mathcal{Z}_s = \bigvee_{k\in\mathbb{N}} \mathcal{X}_k$, and such that $(X, \mathcal{X}_k, \mu, T)$ is isomorphic to an s_0-step nilsystem. Therefore, for every $\varepsilon' > 0$ there exists $k \in \mathbb{N}$ and $\mathcal{X}_k$-measurable functions $\tilde{f}_i \in L^\infty(\mu)$ satisfying $\|\tilde{f}_i - f_i\|_{L^1(\mu)} < \varepsilon'$. Letting $M = \sup_i \|f_i\|_{L^\infty(\mu)} < \infty$ and $\varepsilon = \varepsilon' M^t$, and using the triangle inequality together with the fact that T is measure preserving, we deduce that up to an error term $O(\varepsilon)$, the limit (4.29) is equal to the same limit with the functions f_i replaced by functions $\tilde{f}_i$ defined on a s_0-step nilsystem.

Let $(G/\Gamma, \mathcal{G}/\Gamma, \nu, T_a)$ be a totally ergodic nilsystem, and $G_\bullet$ be the lower central series filtration on G. Using (4.12), it suffices to show that if $f_0, ..., f_t \in L^\infty(\nu)$ and f_i vanishes on each coset of $G_{s+1}\Gamma$, then

$$\lim_{N\to\infty} \mathop{\mathbb{E}}_{n\in[N]} \int_{G/\Gamma} f_0(b\Gamma) \cdot f_1(a^{P_1(n)}b\Gamma) \cdots f_t(a^{P_t(n)}b\Gamma) d\nu(b\Gamma) = 0.$$

Let $G^o_\bullet$ be the filtration on G^o given by $G^o_\bullet = G_\bullet \cap G^o$, and let $g_b \in \text{poly}(\mathbb{Z}, G^o_\bullet)$ be the irrational sequence defined in Proposition 4.3.2 for which $a^n b\Gamma = g_b(n)\Gamma$. The irrationality of g_b, Lemma 4.3.1 and Theorem 4.5.3 imply that

$$\lim_{N\to\infty}\mathop{\mathbb{E}}_{n\in[N]}\int_{G/\Gamma} f_0(b\Gamma)\cdot f_1(a^{P_1(n)}b\Gamma)\cdots f_t(a^{P_t(n)}b\Gamma)d\nu(b\Gamma)$$
$$=\int_{G^o/\Gamma^o}\lim_{N\to\infty}\mathop{\mathbb{E}}_{m,n\in[N]} f_0(g_b(m)\Gamma^o)\cdot f_1(g_b(m+P_1(n))\Gamma^o)\cdots f_t(g_b(m+P_t(n))\Gamma^o)d\nu(b\Gamma^o)$$
$$=\int_{(G^o)^P/(\Gamma^o)^P} f_0\otimes\cdots\otimes f_t d\nu^P,$$

where $(G^o)^P$ is the Leibman group for $\vec{P}$ and ν^P is the Haar measure on $(G^o)^P/(\Gamma^o)^P$.

The assumption that f_i vanishes on each coset of $G_{s+1}\Gamma$ in G/Γ together with Lemma 4.3.4 imply that f_i vanishes on each coset of $G^o_{s+1}\Gamma^o$ inside G^o/Γ^o. By Lemma 4.5.2, the group $(G^o)^P$ contains $H = 1^i \times G^o_{s+1}\times 1^{t-i}$; therefore

$$\left|\int_{(G^o)^P/(\Gamma^o)^P} f_0\otimes\cdots\otimes f_t\right| \leqslant \int_{(G^o)^P/H(\Gamma^o)^P}\left|\int_{xH(\Gamma^o)^P} f_0\otimes\cdots\otimes f_t\right|$$
$$\leqslant \left(\prod_{j\neq i}||f_j||_\infty\right)\int_{(G^o)^P/H(\Gamma^o)^P}\left|\int_{x_iG^o_{s+1}\Gamma^o} f_i\right| = 0,$$

implying that $\mathcal{Z}_s$ is characteristic for the weak convergence of $\vec{P}$ at i. □

Corollary 4.5.4 implies that if a progression $\vec{P}$ satisfies $\mathcal{A}_i(\vec{P}) = s$, then $\mathcal{Z}_s$ is characteristic for the weak or L^2 convergence of $\vec{P}$ at i for any totally ergodic system. We now prove Corollary 4.1.14, which extends this result to ergodic systems for eligible progressions, with a slight modification in the $s=0$ case. The proof is almost identical to the proof of Proposition 4.1 in [Fra08].

Proof of Corollary 4.1.14. Let $\vec{P}\in\mathbb{R}[x,y]^{t+1}$ be an eligible homogeneous progression with $\mathcal{A}_i(\vec{P}) = s$ and $(X,\mathcal{X},\mu,T)$ be ergodic. By Theorem 4.1.4, there exists a Host-Kra factor that is characteristic for the weak and L^2 convergence of $\vec{P}$. Since each Host-Kra factor is an inverse limit of nilsequences, we can approximate X by an ergodic nilsystem $(G/\Gamma, \mathcal{G}/\Gamma, \nu, T_a)$. The compactness of G/Γ and the assumption that G is generated by the connected component G^o and a imply that $a^r\in G^o$ for some $r\in\mathbb{N}_+$; and hence

$$\mathop{\mathbb{E}}_{n\in[rN]}\prod_{i=1}^t T_a^{P_i(n)}f_i = \mathop{\mathbb{E}}_{j\in[r]}\mathop{\mathbb{E}}_{n\in[N]}\prod_{i=1}^t T_a^{P_i(r(n-1)+j)}f_i \tag{4.30}$$
$$= \mathop{\mathbb{E}}_{j\in[r]}\mathop{\mathbb{E}}_{n\in[N]}\prod_{i=1}^t (T_a^r)^{\tilde{P}_{i,j}(n)}(T_a^{P_i(j)}f_i),$$

where $\tilde{P}_{i,j}(n) = \frac{P_i(r(n-1)+j)-P_i(j)}{r}$. This is where we use the fact that $\vec{P}$ is eligible. The definition of eligibility implies that for any $0 \leqslant j < r$, the progression

$$\vec{\tilde{P}}_j(x,y) = (x,\ x+\tilde{P}_{1,j}(y),\ \ldots,\ x+\tilde{P}_{t,j}(y))$$

is homogeneous and that $\mathcal{A}_i(\vec{\tilde{P}}_j) = \mathcal{A}_i(\vec{P})$ for every $0 \leqslant i < r$.

If $s > 0$, suppose that $\mathbb{E}(f_i|\mathcal{Z}_s(T_a)) = 0$. Then the equality $\mathcal{Z}_s(T_a) = \mathcal{Z}_s(T_a^r)$ and the T_a-invariance of $\mathcal{Z}_s$ imply that $\mathbb{E}(T_a^{P_i(j)} f_i|\mathcal{Z}_s(T_a^r)) = 0$. We deduce from Corollary 4.5.4 and the total ergodicity of T_a^r on each connected components of G/Γ that the expression in (4.30) converges to 0 as $N \to \infty$.

If $s = 0$, suppose that $\mathbb{E}(f_i|\mathcal{K}_{rat}(T_a)) = 0$. The total ergodicity of T_a^r implies that $\mathcal{K}_{rat}(T_a) = \mathcal{Z}_0(T_a^r)$, and so $\mathbb{E}(T_a^{P_i(j)} f_i|\mathcal{Z}_0(T_a^r)) = 0$. Again, it follows from Corollary 4.5.4 and the total ergodicity of T_a^r on each connected components of G/Γ that the expression in (4.30) converges to 0 as $N \to \infty$.

□

We now show that progressions of algebraic complexity at most 1 are eligible, which together with Corollary 4.1.14 immediately implies Corollary 4.1.15.

Lemma 4.5.5. *Let* $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ *be an algebraic progression with* $\max_i \mathcal{A}_i(\vec{P}) \leqslant 1$. *Then* $\vec{P}$ *is homogeneous and eligible.*

Proof. From the definition of inhomogeneous relations it follows that each inhomogeneous relations must have degree at least 2. Thus, the fact that $\vec{P}$ has algebraic complexity at most 1 immediately implies that it is homogeneous.

To prove that $\vec{P}$ is eligible, we fix $r \in \mathbb{N}_+$ and $0 \leqslant j < r$. We show that the progression

$$\vec{\tilde{P}}(x,y) = (x,\ x+\tilde{P}_{1,j}(y),\ \ldots,\ x+\tilde{P}_{t,j}(y))$$

also has algebraic complexity at most 1, from which the eligibility of $\vec{P}$ will follow easily. Indeed, suppose that $\vec{\tilde{P}}$ satisfies an algebraic relation of degree 2:

$$\sum_{i=0}^{t} a_{i2}(x + P_i(r(y-1)-j) - P_i(j))^2 + a_{i1}(x + P_i(r(y-1)-j) - P_i(j)) = 0.$$

Setting $u = r(y-1) - j$ for brevity and rearranging we deduce that

$$\sum_{i=0}^{t} \Big(a_{i2}(x + P_i(u))^2 + (a_{i1} - 2P_i(j))(x + P_i(u)) + a_{i2}P_i(j)^2 - a_{i1}P_i(j) \Big) = 0.$$

The homogeneity of $\vec{P}$ implies that

$$\sum_{i=0}^{t} a_{i2}(x + P_i(u))^2 = 0,$$

and the fact that $\vec{P}$ has algebraic complexity at most 1 further implies that $a_{02} = ... = a_{t2} = 0$. Thus, $\vec{\tilde{P}}$ satisfies no algebraic relation of degree 2. It follows by induction that $\vec{\tilde{P}}$ satisfies no algebraic relation of degree $d > 2$ since each such relation

$$Q_0(x) + Q_1(x + P_1(y)) + ... + Q_t(x + P_t(y)) = 0 \tag{4.31}$$

would induce an algebraic relation of degree $d-1$ by partially differentiating (4.31) with respect to x. This establishes the claim that $\vec{\tilde{P}}$ has algebraic complexity at most 1. Thus, every algebraic relation satisfied by $\vec{\tilde{P}}$ is of the form

$$a_0 x + a_1(x + P_1(r(y-1)+j) - P_1(j)) + ... + a_t(x + P_t(r(y-1)+j) - P_t(j)) = 0$$

and corresponds to an algebraic relation

$$a_0 x + a_1(x + P_1(y)) + ... + a_t(x + P_t(y)) = 0$$

satisfied by $\vec{P}$. This one-to-one correspondence between the algebraic relations satisfied by $\vec{\tilde{P}}$ and $\vec{P}$ implies the eligibility of $\vec{P}$. $\square$

Theorem 4.5.3 allows us to prove the second part of Corollary 4.1.17.

Proof of Corollary 4.1.17(ii). Let $(X, \mathcal{X}, \mu, T)$ be a totally ergodic system, and suppose that $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ is an integral progression with algebraic complexity at most 1. This implies that $\vec{P}$ is homogeneous since each inhomogeneous algebraic relation must have degree at least 2. For each $0 \leqslant i \leqslant t$, let $P_i(y) = \sum_{j=1}^{d} a_{i,j} Q_j(y)$ and $L_i(y_1, ... y_d) = \sum_{j=1}^{d} a_{i,j} y_j$ for some $a_{i,j} \in \mathbb{Z}$ and integral polynomials $Q_1, ..., Q_d$. Letting

$$\vec{L}(x, y_1, ..., y_d) = (x,\ x + L_1(y_1, ..., y_d),\ ...,\ x + L_t(y_1, ..., y_d)),$$

we observe that $\vec{P}(x,y)=\vec{L}(x,Q_1(y),...,Q_d(y))$. It follows that $\vec{L}$ also has an algebraic complexity at most 1, since each algebraic relation of degree $(j_0,...,j_t)$ between terms of $\vec{L}$ would immediately imply an algebraic relation of the same degree between terms of $\vec{P}$ after substituting $y_i=Q_i(y)$.

Using the same argument as in the proof of Corollary 4.5.4, we reduce the question of understanding

$$\lim_{N\to\infty}\mathop{\mathbb{E}}_{n\in[N]}\int_X\prod_{i=0}^{t}T^{P_i(n)}f_i d\mu \tag{4.32}$$

to understanding

$$\lim_{N\to\infty}\mathop{\mathbb{E}}_{x,y\in[N]}F(g^P(x,y)) \tag{4.33}$$

for each Lipschitz function $F:(G/\Gamma)^{t+1}\to\mathbb{C}$ and an irrational sequence $g\in\mathrm{poly}(\mathbb{Z},G_\bullet)$ for some filtration $G_\bullet$ on G. Following the same method to analyse

$$\lim_{N\to\infty}\mathop{\mathbb{E}}_{y_1,...,y_d\in[N]}\int_X\prod_{i=0}^{t}T^{L_i(y_1,...,y_d)}f_i d\mu, \tag{4.34}$$

we deduce that understanding (4.34) comes down to estimating

$$\lim_{N\to\infty}\mathop{\mathbb{E}}_{x,y_1,...,y_d\in[N]}F(g^L(x,y_1,...,y_d)), \tag{4.35}$$

where

$$g^L(x,y_1,...,y_d)=(g(x),\ g(x+L_1(y_1,...,y_d)),\ ...,\ g(x+L_t(y_1,...,y_d))).$$

By Theorem 4.5.3, the limit in (4.33) equals $\int_{G^P/\Gamma^P}F$. Similarly, an ergodic version of Theorem 11 of [GT10] states that

$$\lim_{N\to\infty}\mathop{\mathbb{E}}_{x,y_1,...,y_d\in[N]}F(g^L(x,y_1,...,y_d))=\int_{G^L/\Gamma^L}F,$$

where

$$G^L=\left\langle g_i^{\vec{v}_i}:1\leqslant i\leqslant s, g_i\in G_i, \vec{v}_i\in\mathcal{L}_i\right\rangle$$

is the *Leibman group* originally defined in [Lei09], $\Gamma^L = G^L \cap \Gamma^{t+1}$, and

$$\mathcal{L}_i = \operatorname{Span}_{\mathbb{R}} \left\{ \left(\binom{x}{i}, \binom{x + L_1(y_1, ..., y_d)}{i}, \ ..., \ \binom{x + L_t(y_1, ..., y_d)}{i} \right) : \ x, y_1, ..., y_d \in \mathbb{R} \right\}.$$

From the fact that $\max_i \mathcal{A}_i(\vec{P}) \leqslant 1$ we deduce that $\mathcal{P}_2 = \mathcal{P}_3 = ... = \mathbb{R}^{t+1}$, and so $G^P = \langle h_1^{\vec{v}_1}, G_2^{t+1} : h_1 \in G_1, \vec{v}_1 \in \mathcal{P}_1 \rangle$. Similarly, the fact that $\vec{L}$ has algebraic complexity at most 1 reveals that $G^L = \langle h_1^{\vec{v}_1}, G_2^{t+1} : h_1 \in G_1, \vec{v}_1 \in \mathcal{L}_1 \rangle$. We moreover observe that $\mathcal{P}_1 = \mathcal{L}_1$; from this it follows that $G^P = G^L$, and so the limits in (4.33) and (4.35) are equal. This implies that (4.32) and (4.34) equal as well.

□

4.6 Finitary nilmanifold theory

Before we can prove a finitary version of Theorem 4.5.3, we need to introduce necessary finitary concepts required for this task. Most concepts and definitions in this and next section are taken from [GT10; GT12; CS12]. Throughout this section, we assume that G is connected, and that each nilmanifold G/Γ comes with a filtration $G_\bullet$ and a Mal'cev basis χ adapted to $G_\bullet$. We call a nilmanifold endowed with filtration and a Mal'cev basis *filtered*. A *Mal'cev basis* is a basis for the Lie algebra of G with some special properties; since we do not explicitly work with the notion of Mal'cev basis or its rationality in the paper, we refer the reader to [GT12] for definitions of these concepts. What matters for us is that each Mal'cev basis induces a diffemomorphism $\psi : G \to \mathbb{R}^m$, called *Mal'cev coordinate map*, which satisfies the following properties:

1. $\psi(\Gamma) = \mathbb{Z}^m$;
2. $\psi(G_i) = \{0\}^{m-m_i} \times \mathbb{R}^{m_i}$, where $m_i = \dim G_i$.

Thus, ψ provides a natural coordinate system on G that respects the filtration $G_\bullet$ and the lattice Γ. Similarly to ψ, we define maps $\psi_i : G_i \to \mathbb{R}^{m_i - m_{i+1}}$ by assigning to each element of G_i its Mal'cev coordinates indexed by $m - m_i + 1$, ..., $m - m_{i+1}$. With this definition, we have $\psi_i(x) = 0$ if and only if $x \in G_{i+1}$, and $\psi_i(x) \in \mathbb{Z}^{m_i - m_{i+1}}$ if and only if $x \in \Gamma_i$.

Definition 4.6.1 (Complexity of nilmanifolds)**.** *A filtered nilmanifold G/Γ has complexity M if the degree s of the filtration $G_\bullet$, the dimension m of the group G, and the rationality*[6] *of the Mal'cev basis χ are all bounded by M.*

[6] A Mal'cev basis $\chi = \{\mathcal{X}_1, ..., \mathcal{X}_m\}$ is M-rational if the structure constants c_{ijk} in $[\mathcal{X}_i, \mathcal{X}_j] = \sum_k c_{ijk} \mathcal{X}_k$ are rationals of height at most M.

We remark that complexity of nilmanifolds has nothing to do with the four notions of complexity of polynomial progressions that we examine. Neither does complexity of nilsequences defined below.

We endow nilmanifolds with the following metric.

Definition 4.6.2 (Metrics on G and G/Γ, Definition 2.2 of [GT12])**.** *Let G/Γ be a nilmanifold with Mal'cev basis χ and Mal'cev coordinate map ψ. Then $d(x, y)$ is the largest metric on G satisfying $d(x, y) \leqslant \|\psi(xy^{-1})\|_\infty$, and the metric on G/Γ is given by $d(x\Gamma, y\Gamma) = \inf\{d(x\gamma, d\gamma') : \gamma, \gamma' \in \Gamma\}$.*

We will control the size of functions on nilmanifolds with the following norm.

Definition 4.6.3 (Lipschitz norm, Definition 1.2 of [GT12])**.** *For a function $F : G/\Gamma \to \mathbb{C}$, we define its* Lipschitz norm *to be*

$$\|F\|_{\mathrm{Lip}} = \|F\|_\infty + \sup\left\{\frac{|F(y\Gamma) - F(z\Gamma)}{d(y\Gamma, z\Gamma)} : y, z \in G\right\},$$

*and we say that F is M-*Lipschitz *if $\|F\|_{\mathrm{Lip}} \leqslant M$.*

Definition 4.6.4 (Nilsequences)**.** *A function $f : \mathbb{Z} \to \mathbb{C}$ is a* nilsequence *of degree s and complexity M if $f(n) = F(g(n)\Gamma)$, where $F : G/\Gamma \to \mathbb{R}$ is an M-Lipschitz function on a filtered nilmanifold G/Γ of degree s and complexity M, and $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$.*

We note that although the complexity of a nilsequence bounds its degree from above, we usually work with nilsequences having fixed (small) degrees and bounded but large complexities, therefore it is important for us to keep account of both the degree and complexity of the nilsequence when working with it.

The notion of a M-Lipschitz function $F : G/\Gamma \to \mathbb{C}$ is taken with regards to the following metric on G/Γ.

Definition 4.6.5 (Quantitative equidistribution)**.** *Let $D \in \mathbb{N}_+$ and $\delta > 0$. A sequence $g \in \mathrm{poly}(\mathbb{Z}^D, G)$ is (δ, N)-*equidistributed *on G/Γ if*

$$\left|\mathop{\mathbb{E}}_{n \in [N]^D} F(g(n)\Gamma) - \int_{G/\Gamma} F\right| \leqslant \delta ||F||_{\mathrm{Lip}}$$

for all Lipschitz functions $F : G/\Gamma \to \mathbb{C}$, where $||F||_{\mathrm{Lip}}$ is the Lipschitz norm on F with respect to the metric from Definition 4.6.2.

It has been shown in Theorem 4.2.5 that equidistribution is related to horizontal characters. Given the Mal'cev coordinate map $\psi : G \to \mathbb{R}^m$, each horizontal character can be written in the form $\eta(x) = k\cdot\psi(x)$ for some $k \in \mathbb{Z}^m$. We call $|\eta| = |k| = |k_1| + ... + |k_m|$ the *modulus* of η. Similarly, each i-th level character $\eta_i : G_i \to \mathbb{R}$ is of the form $\eta_i(x) = k \cdot \psi_i(x)$ for some $k \in \mathbb{Z}^{m_i - m_{i+1}}$, and we define its modulus to be $|\eta_i| = |k| = |k_1| + ... + |k_{m_i - m_{i+1}}|$.

We shall also need to quantify the notion of polynomials that are "almost constant" mod $\mathbb{Z}$, using a definition from [GT12]. In what follows, $||x||_{\mathbb{R}/\mathbb{Z}} = \min\{|x-n| : n \in \mathbb{Z}\}$ is the *circle norm* of $x \in \mathbb{R}$.

Definition 4.6.6 (Smoothness norm). *Let*

$$Q(n_1, ..., n_D) = \sum_{i=0}^{d} \sum_{i_1+...+i_D=i} a_{i_1,...,i_D} \binom{n_1}{i_1} \cdots \binom{n_D}{i_D}$$

be a polynomial in $\mathbb{R}[n_1, ..., n_D]$. *For* $N \in \mathbb{N}_+$, *we define the* smoothness norm *of* Q *to be*

$$||Q||_{C^\infty[N]} = \max\{N^{i_1+...+i_D} ||a_{i_1,...,i_D}||_{\mathbb{R}/\mathbb{Z}} : i_1, ..., i_D \in \mathbb{N},\ 1 \leqslant i_1 + \cdots + i_D \leqslant d\}.$$

In particular, $||Q||_{C^\infty[N]}$ is bounded from above as $N \to \infty$ if and only if Q is constant mod $\mathbb{Z}$.

With these definitions, we are ready to state a quantitative version of Theorem 4.2.5

Theorem 4.6.7 (Quantitative Leibman's equidistribution theorem, Theorem 2.9 of [GT12]). *Let* $\delta > 0$, $M \geqslant 2$ *and* $D, N \in \mathbb{N}_+$ *with* $D \leqslant M$. *Let* G/Γ *be a filtered nilmanifold of complexity* M *and* $g \in \mathrm{poly}(\mathbb{Z}^D, G_\bullet)$. *Then there exists* $C_M > 0$ *such that at least one of the following is true:*

1. g *is* (δ, N)*-equidistributed in* G/Γ*;*
2. *there exists a nontrivial horizontal character* η *of modulus* $|\eta| \ll \delta^{-C_M}$ *for which* $||\eta \circ g||_{C^\infty[N]} \ll \delta^{-C_M}$.

We now need to quantify the notion of irrationality.

Definition 4.6.8 (Quantitative irrationality). *Let* G/Γ *be a filtered nilmanifold of degree* s*, and suppose* $A, N > 0$. *An element* $g_i \in G_i$ *is* (A, N)-irrational *if for every nontrivial* i*-th level character* $\eta : G_i \to \mathbb{R}$ *of modulus* $|\eta| \leqslant A$*, we have* $||\eta(g_i)||_{\mathbb{R}/\mathbb{Z}} \geqslant A/N^i$. *It is* A*-irrational if for every nontrivial* i*-th level character* $\eta : G_i \to \mathbb{R}$ *of modulus* $|\eta| \leqslant A$*, we have* $\eta \circ g_i \notin \mathbb{Z}$. *We say that a* *sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ is (A,N)-irrational (respectively A-irrational) if g_i is (A,N)-irrational (respectively A-irrational) for each $1 \leqslant i \leqslant s$. Similarly, we say that the nilsequence $n \mapsto F(g(n)\Gamma)$ is (A,N)- or A-irrational if the polynomial sequence g is.*

Clearly, (A,N)-irrationality is stronger than A-rationality, but for some of our applications the latter notion will be sufficient.

We are now ready to state the finitary version of Theorem 4.5.3, which is the main technical result of this paper, and derive Theorem 4.5.3 from it.

Theorem 4.6.9. *Let $t \in \mathbb{N}_+$ and $A, M, N \geqslant 2$. Let G/Γ be a filtered nilmanifold of complexity M. Suppose that $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ is (A,N)-irrational, $F : (G/\Gamma)^{t+1} \to \mathbb{C}$ is M-Lipschitz, and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ is a homogeneous polynomial progression. Then*

$$\mathop{\mathbb{E}}_{x,y\in[N]} F(g^P(x,y)\Gamma^{t+1}) = \int_{G^P/\Gamma^P} F + O_M(A^{-c_M})$$

for some $c_M > 0$.

Proof of Theorem 4.5.3 using Theorem 4.6.9. Let $F : (G/\Gamma)^{t+1} \to \mathbb{R}$ be a continuous function. By the Stone-Weierstrass theorem, Lipschitz functions on a compact set form a dense subset of the algebra of continuous functions. Approximating F by a sequence of Lipschitz functions if necessary, we can assume without loss of generality that F is Lipschitz. We let M be the maximum of the complexity of G/Γ and the Lipschitz norm of F.

Let $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ be an irrational sequence. For each $N \in \mathbb{N}_+$, we let A_N be the maximal real number A for which g is (A_N, N)-irrational. We claim that $A_N \to \infty$ as $N \to \infty$. If not, then there exists some number $A > 0$ and an index $i \in \mathbb{N}_+$ with the property that g_i is not (A,N)-irrational for all $N \in \mathbb{N}_+$. We fix this i. It follows that there exists a sequence of nontrivial i-th level characters $\eta_N : G_i \to \mathbb{R}$ of modulus at most A such that $||\eta_N(g_i)||_{\mathbb{R}/\mathbb{Z}} < A/N^i$. Since there are only finitely many i-th level characters of modulus bounded by A, we conclude that there exists a nontrivial i-th level character η of modulus at most A such that $||\eta(g_i)||_{\mathbb{R}/\mathbb{Z}} < A/N^i$ for all $N \in \mathbb{N}_+$. Taking $N \to \infty$, we see that $\eta(g_i) \in \mathbb{Z}$, contradicting the irrationality of g_i.

It therefore follows from Theorem 4.6.9 that

$$\mathop{\mathbb{E}}_{x,y\in[N]} F(g^P(x,y)\Gamma^{t+1}) = \int_{G^P/\Gamma^P} F + O_M(A_N^{-c_M})$$

Since M is constant, letting $N \to \infty$ sends the error term to 0, implying that g^P is equidistributed on G^P/Γ^P as claimed. □

4.7 Reducing true complexity to an equidistribution question

In Sections 4.3-4.6, we have shown how the question of determining Host-Kra complexity for homogeneous progressions can be reduced to showing that g^P is equidistributed on G^P/Γ^P. Determining true complexity for homogeneous progression comes down to the exact same equidistribution question. All the arguments in this section can be viewed as finitary analogues of arguments in previous sections.

Since we are now primarily concerned with functions from $\mathbb{Z}/N\mathbb{Z}$ to $\mathbb{C}$, we shall need an N-periodic version of certain previously defined concepts. In this section, N is always a prime, and the group G is connected. A function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is called *1-bounded* whenever $||f||_\infty \leqslant 1$.

Definition 4.7.1 (Periodic sequences)**.** *Let $G_\bullet$ be a filtration on G. A sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ is* N-periodic *if $g(n+N)g(n)^{-1} \in \Gamma$ for each $n \in \mathbb{Z}$, and it is periodic if it is N-periodic for some $N > 0$. A nilsequence $n \mapsto F(g(n)\Gamma)$ is N-periodic (resp. periodic) if g is.*

Given a homogeneous polynomial progression $\vec{P} \in \mathbb{R}[x, y]^{t+1}$, we want to show that $\mathcal{A}_i(\vec{P}) = \mathcal{T}_i(\vec{P})$ for each $0 \leqslant i \leqslant t$. The forward inequality has been derived in Section 4.1; it is the reverse inequality that poses a challenge. We thus want to prove the following.

Theorem 4.7.2. *Let $t \in \mathbb{N}_+$, $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ be a homogeneous polynomial progression, $0 \leqslant i \leqslant t$, and suppose that $\mathcal{A}_i(\vec{P}) = s$. For every $\varepsilon > 0$, there exist $\delta > 0$ and $N_0 \in \mathbb{N}$ such that for all primes $N > N_0$ and all 1-bounded functions $f_0, ..., f_t : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, we have*

$$\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x + P_1(y)) \cdots f_t(x + P_t(y)) \right| < \varepsilon$$

whenever $||f_i||_{U^{s+1}} < \delta$.

We know that each progression is controlled by *some* Gowers norm. The result below plays the same role in deriving Theorem 4.7.2 as Theorem 4.1.4 plays in the proof of Corollary 4.5.4.

Proposition 4.7.3 (Proposition 2.2 of [Pel19])**.** *Let $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ be an integral polynomial progression. There exists $s \in \mathbb{N}_+$ with the following property:*

for every $\varepsilon > 0$, there exist $\delta > 0$ and $N_0 \in \mathbb{N}$ such that for all primes $N > N_0$ and all 1-bounded functions $f_0, ..., f_t : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, we have

$$\left| \mathop{\mathbb{E}}_{x,y \in \mathbb{Z}/N\mathbb{Z}} f_0(x) f_1(x + P_1(y)) \cdots f_t(x + P_t(y)) \right| < \varepsilon$$

whenever $||f_i||_{U^{s+1}} < \delta$ for some $0 \leqslant i \leqslant t$.

Next, we want to perform a finitary analogue of the approximation-by-nilsystems argument. This can be achieved with the help of a periodic version of celebrated arithmetic regularity lemma from [GT10] in which the same polynomial sequence g is used in the decomposition of several functions. This lemma has originally appeared before as Lemma 2.13 of [Kuc21b].

Lemma 4.7.4. *Let $s, t \in \mathbb{N}+$, $\varepsilon > 0$, and $\mathcal{F} : \mathbb{R}_+ \to \mathbb{R}_+$ be a growth function. There exists $M = O_{\varepsilon,\mathcal{F}}(1)$, a filtered nilmanifold G/Γ of degree s and complexity at most M, and an N-periodic, $\mathcal{F}(M)$-irrational sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ satisfying $g(0) = 1$ such that for all 1-bounded functions $f_0, ..., f_t : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, there exist decompositions*

$$f_i = f_{i,nil} + f_{i,sml} + f_{i,unf}$$

where

1. *$f_{i,nil}(n) = F_i(g(n)\Gamma)$ for M-Lipschitz function $F_i : G/\Gamma \to \mathbb{C}$,*
2. *$||f_{i,sml}||_2 \leqslant \varepsilon$,*
3. *$||f_{i,unf}||_{U^{s+1}} \leqslant \frac{1}{\mathcal{F}(M)}$,*
4. *the functions $f_{i,nil}$, $f_{i,sml}$ and $f_{i,unf}$ are 4-bounded,*

In the proof we follow closely the argument leading up to Theorem 5.1 of [CS12], but we take extra care to ensure that for every i, the nilsequences $f_{i,nil}$ are defined in terms of the same polynomial sequence g.

Proof of Theorem 4.7.2. Fix $\varepsilon > 0$ and a growth function $\mathcal{F} : \mathbb{R}_+ \to \mathbb{R}_+$. We pick another growth function $\mathcal{F}_0$ that grows sufficiently slowly with respect to $\mathcal{F}$. By Theorem 3.4 of [CS12], there exists $0 < M_0 = O_{s,\varepsilon,\mathcal{F}_0}(1)$ such that for each i there is a filtered nilmanifold G_i/Γ_i of complexity M_0 and degree s, a p-periodic sequence $g_i \in \mathrm{poly}(\mathbb{Z}, (G_i)_\bullet)$, and an M_0-Lipschitz function $F_i' : G_i/\Gamma_i \to \mathbb{C}$ for which f_i decomposes into

$$f_i = f_{i,nil} + f_{i,sml} + f_{i,unf}$$

where the properties (ii), (iii), (iv) in Lemma 4.7.4 hold with M_0 in place of M and $\mathcal{F}_0$ in place of $\mathcal{F}$, and moreover $f_{i,nil}(n) = F_i'(g_i(n)\Gamma_i)$. By redefining F_i' and increasing its Lipschitz norm by a factor $O_{M_0}(1)$ if necessary, we can also assume that $g_i'(0) = 1$ for all $1 \leqslant i \leqslant t$.

We let

$$G = G_1 \times ... \times G_t, \quad \Gamma = \Gamma_1 \times ... \times \Gamma_t, \quad \text{and} \quad g(n) = (g_1(n), ..., g_t(n)),$$

and we define $F_i(x_1\Gamma_1, ..., x_t\Gamma_t) := F_i'(x_i\Gamma_i)$. With this definition, we can realize each $f_{i,nil}$ as a p-periodic nilsequence $f_{i,nil}(n) = F_i(g(n)\Gamma)$ of degree s and complexity $M_0 t$ on the same nilmanifold G/Γ using the same p-periodic sequence g for all $1 \leqslant i \leqslant t$.

The next step is to obtain irrationality on the nilsequences $f_{1,nil}$, ..., $f_{t,nil}$. In doing so, we apply the proof of Theorem 5.1 of [CS12], which we rerun here for completeness. Given a growth function $\mathcal{F}_1$ to be chosen later, we use Proposition 5.2 of [CS12] to obtain $M_1 \in [M_0, O_{M_0,t,\mathcal{F}_1}(1)]$ and a p-periodic polynomial $g' \in \operatorname{poly}(\mathbb{Z}, G'_\bullet)$ on some nilmanifold G'/Γ' of complexity $O_{M_1}(1)$ satisfying $g'(n)\Gamma = g(n)\Gamma$. By abuse of notation, we let F_i denote now its restriction to G'/Γ' for each $1 \leqslant i \leqslant t$. It is $O_{M_1}(1)$-Lipschitz on G'/Γ'. Therefore the nilsequence $f_{i,nil}$ has complexity $M \leqslant \mathcal{F}_2(M_1)$ for some function $\mathcal{F}_2$. Letting $\mathcal{F}_1(x) = \mathcal{F}(\mathcal{F}_2(x))$ thus guarantees that g' is $\mathcal{F}(M)$-irrational. To guarantee $\|f_{i,nil}\|_{U^s} \leqslant \frac{1}{\mathcal{F}(M)}$, we pick $\mathcal{F}_0$ so that $\mathcal{F}_0(M_0) \geqslant \mathcal{F}(M)$ using $M = O_{M_1}(1) = O_{M_0,t,\mathcal{F}}(1)$. Combining all the bounds, we have $M = O_{s,t,\varepsilon,\mathcal{F}}(1)$, as desired.

In their statement of Theorem 3.4 of [CS12], the authors only considered functions from $\mathbb{F}_p$ to $[0, 1]$. However, the statement works for arbitrary 1-bounded functions from $\mathbb{F}_p$ to $\mathbb{C}$ by splitting them into the real and imaginary part, and the positive and negative part. This way, we split a 1-bounded function from $\mathbb{F}_p$ to $\mathbb{C}$ into four 1-bounded functions from $\mathbb{F}_p$ to $[0, 1]$, implying the 4-boundedness of $f_{i,nil}$, $f_{i,sml}$ and $f_{i,unf}$.

□

The last piece that we need is a finitary, periodic version of Theorem 4.6.9.

Proposition 4.7.5. *Let $t \in \mathbb{N}_+$ and $A, M, N \geqslant 2$. Let G/Γ be a filtered nilmanifold and complexity M. Suppose that $g \in \operatorname{poly}(\mathbb{Z}, G_\bullet)$ is an A-irrational, N-periodic polynomial sequence, $F : (G/\Gamma)^{t+1} \to \mathbb{C}$ is M-Lipschitz and 1-*

bounded, and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ is a homogeneous polynomial progression. Then

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} F(g^P(x,y)\Gamma^{t+1}) = \int_{G^P/\Gamma^P} F + O_M(A^{-c_M})$$

for some $c_M > 0$.

Proof of Proposition 4.7.5 using Theorem 4.6.9. Let $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ be A-irrational and N-periodic. We claim that g is (A, Nk)-irrational for all sufficiently large $k \in \mathbb{N}_+$. If not, then there exists $1 \leqslant i \leqslant s$ such that for each $k \in \mathbb{N}_+$ there exists an i-th level character $\eta_{i,k} : G_i \to \mathbb{R}$ of complexity at most A satisfying $||\eta_{i,k}(g_i)||_{\mathbb{R}/\mathbb{Z}} < A/(Nk)^i$. The N-periodicity of g_i implies that $g_i^{N^i} \in \Gamma_i \bmod G_{i+1}^\nabla$ (Lemma 5.3 of [CS12]); hence $\eta_{i,k}(g_i) \in \frac{1}{N^i}\mathbb{Z}$. Thus, $\eta_{i,k}(g_i) \in \mathbb{Z}$ whenever $k^i > A$. In particular, since we can take k arbitrarily large, there exists a nontrivial i-th level character $\eta_{i,k}$ of complexity at most A for which $\eta_{i,k}(g_i) \in \mathbb{Z}$, contradicting the A-irrationality of g. Hence g is (A, Nk)-irrational for all sufficiently large $k \in \mathbb{N}_+$.

Applying Theorem 4.6.9, we deduce that

$$\begin{aligned}\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} F(g^P(x,y)\Gamma^{t+1}) &= \mathop{\mathbb{E}}_{x,y\in[Nk]} F(g^P(x,y)\Gamma^{t+1}) + O(1/k)\\ &= \int_{G^P/\Gamma^P} F + O_M(A^{-c_M}) + O(1/k)\end{aligned}$$

for all sufficiently large $k \in \mathbb{N}_+$. Taking $k \to \infty$ finishes the proof. □

We now proceed to prove Theorem 4.7.2. The proof that we give here is an adaptation of the proof Theorem 8.1 of [Kuc21b], and is analogous to the derivation of Corollary 4.5.4 from Theorem 4.5.3. Its logic is very similar to the proof of Theorem 7.1 in [GT10], with minor modifications that allow us to get a control of weights on different terms of the progression by different Gowers norms.

Proof. Fix $\varepsilon > 0$. By Proposition 4.7.3, there exists an integer $s_0 \geqslant 1$, a threshold $N_0 \in \mathbb{N}$, and a real number $\delta > 0$ such that for all primes $N > N_0$,

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+P_1(y))\cdots f_t(x+P_t(y))\right| \leqslant \varepsilon$$

for all 1-bounded functions $f_0, ..., f_t : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, at least one of which satisfies $\|f_j\|_{U^{s_0+1}} \leqslant \delta$. We let $\mathcal{F} : \mathbb{R}_+ \to \mathbb{R}_+$ be a growth function depending on ε to be fixed later. If $s \geqslant s_0$, then we are done, so suppose $s < s_0$.

Suppose that $f_0, ..., f_t : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ are 1-bounded functions, and suppose moreover that $\|f_i\|_{U^{s+1}} \leqslant \delta$. We use Lemma 4.7.4 to find $M = O_{\varepsilon,\mathcal{F}}(1)$, a filtered nilmanifold G/Γ of degree s_0 and complexity M, an N-periodic, $\mathcal{F}(M)$-irrational sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ with $g(0) = 1$, and decompositions

$$f_j = f_{j,nil} + f_{j,sml} + f_{j,unf} \tag{4.36}$$

satisfying the conditions of Lemma 4.7.4. Decomposing each f_j this way, we get 3^{t+1} terms. All the terms involving $f_{j,sml}$ can be bounded by $O(\varepsilon)$. By choosing $\mathcal{F}$ growing sufficiently fast depending on δ, we can assume that $\|f_{j,unf}\|_{s_0+1} \leqslant \frac{1}{\mathcal{F}(M)} \leqslant \frac{\delta}{4}$, which together with 4-boundedness of $f_{j,unf}$ implies that terms involving $f_{j,unf}$ contribute at most $O(\varepsilon)$.

This leaves us with

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+P_1(y))\cdots f_t(x+P_t(y))\\
&= \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_{0,nil}(x)f_{1,nil}(x+P_1(y))\cdots f_{t,nil}(x+P_t(y)) + O(\varepsilon)\\
&= \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} F(g^P(x,y)\Gamma^P) + O(\varepsilon),
\end{aligned}$$

where $F((u_0, ..., u_t)\Gamma^P) = F_0(u_0\Gamma)\cdots F_t(u_t\Gamma)$. By Proposition 4.7.5, we have

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} F(g^P(x,y)\Gamma^P) = \int_{G^P/\Gamma^P} F + o_{\mathcal{F}(M)\to\infty,M,\varepsilon}(1).$$

By the assumption of algebraic independence, the polynomial $(x+P_i(y))^{s+1}$ is not a linear combination of $x^{s+1}, ..., (x+P_{i-1}(y))^{s+1}, (x+P_{i+1}(y))^{s+1}, ..., (x+P_t(y))^{s+1}$. Consequently, the space $\mathcal{P}_{s+1}$ contains the vector $\vec{e}_i$ that has 1 in the i-th coordinate and 0 elsewhere. This implies that the group

$$G_{s+1}^{\vec{e}_i} = \langle h^{\vec{e}_i} : h \in G_{s+1}\rangle = \{1\}^{i-1} \times G_{s+1} \times \{1\}^{t-i}$$

is contained in G^P. In fact, $G_{s+1}^{\vec{e}_i}$ is a normal subgroup of G^P due to the normality of G_{s+1} in G. Therefore,

$$\int_{G^P/\Gamma^P} F = \int_{G^P/\Gamma^P} F_{\leqslant s},$$

where $F_{\leqslant s}((u_0, ..., u_t)\Gamma^P) = \left(\prod_{\substack{0\leqslant j\leqslant t,\\ j\neq i}} F_j(u_j\Gamma)\right) F_{i,\leqslant s}(u_i\Gamma)$ and $F_{i,\leqslant s}$ is the average of F_i over cosets of G_{s+1}:

$$F_{i,\leqslant s}(u\Gamma) = \int_{G_{s+1}/\Gamma_{s+1}} F_i(uw\Gamma)dw.$$

It is straightforward to see that $F_{i,\leqslant s}$ is $O(1)$-bounded and M-Lipschitz. We moreover have the bound

$$|F_{\leqslant s}((u_0, ..., u_t)\Gamma^P)| \ll |F_{i,\leqslant s}(u_i\Gamma)|$$

which implies that

$$\left| \int_{G^P/\Gamma^P} F \right| \leqslant \int_{G/\Gamma} |F_{i,\leqslant s}| \leqslant \left(\int_{G/\Gamma} |F_{i,\leqslant s}|^2 \right)^{\frac{1}{2}}$$

by an application of the triangle and Cauchy-Schwarz inequality.

The function $F_{i,\leqslant s}$ is a conditional expectation of F_i with respect to the σ-algebra of sets that are preimages of Borel subsets of $G/(G_{s+1}\Gamma)$ under the canonical projection, and is therefore invariant on G_{s+1}-cosets. It follows from the standard properties of conditional expectations that

$$\int_{G/\Gamma} |F_{i,\leqslant s}|^2 = \int_{G/\Gamma} F_{i,\leqslant s}\overline{F_{i,\leqslant s}} = \int_{G/\Gamma} F_i\overline{F_{i,\leqslant s}}.$$

By the $\mathcal{F}(M)$-irrationality of g, we have

$$\int_{G/\Gamma} F_i\overline{F_{i,\leqslant s}} = \mathop{\mathbb{E}}_{n\in\mathbb{Z}/N\mathbb{Z}} (F_i\overline{F_{i,\leqslant s}})(g(n)\Gamma) + o_{\mathcal{F}(M)\to\infty,M,\varepsilon}(1).$$

We let $\psi(n) = \overline{F_{i,\leqslant s}}(g(n)\Gamma)$. By the G_{s+1}-invariance of $F_{i,\leqslant s}$, this is a nilsequence of degree $\leqslant s$ and complexity M. By (4.36), we have

$$F_i(g(n)\Gamma) = f_i(n) - f_{i,sml}(n) - f_{i,unf}(n).$$

We then split $\mathbb{E}_{n\in\mathbb{Z}/N\mathbb{Z}} F_i(g(n)\Gamma)\psi(n)$ into three terms. Using the Cauchy-Schwarz inequality, the term involving $f_{i,sml}$ can be bounded as

$$\left| \mathop{\mathbb{E}}_{n\in\mathbb{Z}/N\mathbb{Z}} f_{i,sml}(n)\psi(n) \right| \ll \varepsilon.$$

To evaluate the contribution coming from f_i, we use $\|f_i\|_{U^{s+1}} \leqslant \delta$ and the converse to the inverse theorem for Gowers norms (Proposition 1.4 of Appendix

G of [GTZ11]) to conclude that

$$\left|\mathop{\mathbb{E}}_{n\in\mathbb{Z}/N\mathbb{Z}} f_i(n)\psi(n)\right| = o_{\delta\to 0,M,\varepsilon}(1).$$

Similarly, we use $\|f_{i,unf}\|_{U^{s_0+1}} \leqslant \delta$ and $s_0 \geqslant s$ to conclude that

$$\left|\mathop{\mathbb{E}}_{n\in\mathbb{Z}/N\mathbb{Z}} f_{i,unf}(n)\psi(n)\right| = o_{\mathcal{F}(M)\to\infty,M,\varepsilon}(1).$$

Combining all these estimates, we have

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+P_1(y))\cdots f_t(x+P_t(y))\right| = O(\varepsilon) + o_{\mathcal{F}(M)\to\infty,M,\varepsilon}(1) + o_{\delta\to 0,M,\varepsilon}(1).$$

By choosing $\mathcal{F}$ growing sufficiently fast and δ sufficiently small depending on ε, we obtain

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+P_1(y))\cdots f_t(x+P_t(y))\right| \ll \varepsilon,$$

which proves the theorem. □

Finally, we use Proposition 4.7.5 to derive part (i) of Corollary 4.1.17. This result has originally appeared as Theorem 9.1 in [Kuc21b].

Proof of Corollary 4.1.17(i). We set

$$\vec{L}(x,y_1,...,y_d) = (x,\ x+L_1(y_1,...,y_d),\ ...,\ x+L_t(y_1,...,y_d))$$

and

$$\mathcal{L}_i = \mathrm{Span}\{\vec{L}^i(x,y_1,...,y_d):\ x,y_1,...,y_d\in\mathbb{Z}\}$$

to be the analogues of $\vec{P}$ and $\mathcal{P}_i$ for the progression $\vec{\mathcal{L}}$. We observe from the definition of $L_1,...,L_t$ that

$$\vec{P}(y) = \vec{L}(Q_1(y),...,Q_d(y)) \tag{4.37}$$

for some linearly polynomials $Q_1,...,Q_d\in\mathbb{R}[y]$. We also let

$$G^L = \left\langle h_i^{\vec{v}_i}:\ h_i\in G_i, \vec{v}_i\in\mathcal{L}_i : 1\leqslant i\leqslant s\right\rangle$$

denote the Leibman group for $\vec{\mathcal{L}}$.

From the assumption of algebraic complexity 1 it follows that the squares x^2, $(x+P_1(y))^2$, ..., $(x+P_t(y))^2$ are linearly independent, implying that $\mathcal{P}_2 = \mathbb{R}^{t+1}$. From (4.37), it follows that $\mathcal{P}_1 = \mathcal{L}_1$ and $\mathcal{P}_i \subseteq \mathcal{L}_i$ for $i > 1$. Together with the fact that $\mathcal{P}_2 = \mathbb{R}^{t+1}$, this implies that $\mathcal{L}_i = \mathcal{P}_i$ for all $i \in \mathbb{N}_+$, and so the groups $G^P = G^L$ are in fact the same for any group G.

Given $\varepsilon > 0$, we take $\delta > 0$ and $N_0 \in \mathbb{N}$ as in Definition 4.1.8 for both $\vec{P}$ and $\vec{L}$, and we let $\mathcal{F} : \mathbb{R}_+ \to \mathbb{R}_+$ be a growth function to be fixed later. We moreover assume from now on that $N > N_0$.

By Lemma 4.7.4, there exist $M = O_{\varepsilon,\mathcal{F}}(1)$, a filtered nilmanifold G/Γ of degree 1 and complexity M, and an N-periodic, $\mathcal{F}(M)$-irrational sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ with $g(0) = 1$ such that there exist decompositions

$$f_i = f_{i,nil} + f_{i,sml} + f_{i,unf}$$

satisfying the conditions of Lemma 4.7.4. By taking $\mathcal{F}$ growing fast enough with respect to δ, we can assume that $\|f_{i,unf}\|_{U^2} \leqslant \frac{1}{\mathcal{F}(M)} \leqslant \delta/4$ for each i.

By applying the aforementioned decomposition to $f_0, ..., f_t$, each of the operators

$$\Lambda_P(f_0, ..., f_t) = \mathop{\mathbb{E}}_{x,y \in \mathbb{Z}/N\mathbb{Z}} f_0(x) f_1(x + P_1(y)) \cdots f_t(x + P_t(y))$$
$$\Lambda_L(f_0, ..., f_t) = \mathop{\mathbb{E}}_{x,y_1,...,y_d \in \mathbb{Z}/N\mathbb{Z}} f_0(x) f_1(x + L_1(y_1, ..., y_d)) \cdots f_t(x + L_t(y_1, ..., y_d))$$

splits into 3^{t+1} terms. The expressions involving at least one $f_{i,sml}$ can be bounded by $O(\varepsilon)$. All progressions of algebraic complexity 1 are homogeneous, and hence $\vec{P}$ has true complexity 1 by Theorem 1.3.11. Because of this and the way we picked δ, the expressions involving at least one $f_{i,unf}$ can be bounded by $O(\varepsilon)$ as well. We thus have

$$\Lambda_P(f_0, ..., f_t) = \Lambda_P(f_{0,nil}, ..., f_{t,nil}) + O(\varepsilon),$$

and similarly for $\Lambda_L(f_0, ..., f_t)$. By Proposition 4.7.5, we have

$$\Lambda_P(f_{0,nil}, ..., f_{t,nil}) = \int_{G^P/\Gamma^P} F + o_{\mathcal{F}(M)\to\infty, M, \varepsilon}(1),$$

where $F((u_0, ..., u_t)\Gamma^P) = F_0(u_0\Gamma) \cdots F_t(u_t\Gamma)$. Likewise, Theorem 4.1 of [CS12] implies that

$$\Lambda_L(f_{0,nil}, ..., f_{t,nil}) = \int_{G^L/\Gamma^L} F + o_{\mathcal{F}(M)\to\infty,M,\varepsilon}(1).$$

Using the fact that $G^P = G^L$ and combining all the estimates so far, we obtain the equality

$$\Lambda_P(f_0, ..., f_t) = \Lambda_L(f_0, ..., f_t) + O(\varepsilon) + o_{\mathcal{F}(M)\to\infty,M,\varepsilon}(1).$$

The result follows by letting $\mathcal{F}$ grow sufficiently fast with respect to ε, and by taking $\varepsilon \to 0$ as $N \to \infty$. □

Note that the only facts that we use in the proof above is that the progressions $\vec{P}$ and $\vec{L}$ are controlled by *some* Gowers norm (so that we can apply the regularity lemma), progressions of algebraic complexity 1 are homogeneous and the Leibman groups G^P and G^L are the same. It is the last two facts that follow from the algebraic independence of degree 2 of $\vec{P}$. We do not strictly require the information that $\vec{P}$ and $\vec{L}$ are controlled by the U^2 norm.

4.8 The proof of Theorem 4.6.9

To complete the proofs of Corollary 4.5.4 and Theorem 4.7.2, it remains to derive Theorem 4.6.9. Before we prove Theorem 4.6.9 for an arbitrary homogeneous progression, we want to deduce the theorem in the special case of $\vec{P} = (x,\ x+y,\ x+2y,\ x+y^3)$. This will help illustrate the method, and we will later compare this progression with $(x,\ x+y,\ x+2y,\ x+y^2)$ to see what is failing in the inhomogeneous case. The method is an adaptation of the proof of Theorem 1.11 from [GT10], however the linear algebraic component coming from the fact that we are dealing with polynomial progressions is much more involved.

Proposition 4.8.1. *Let $A, M, N \geqslant 2$. Let G/Γ be a filtered nilmanifold of degree 2 and complexity M. Suppose that $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ is an (A, N)-irrational sequence satisfying $g(0) = 1$, $F : (G/\Gamma)^{t+1} \to \mathbb{C}$ is M-Lipschitz, and $\vec{P} = (x,\ x+y,\ x+2y,\ x+y^3)$. Then*

$$\mathop{\mathbb{E}}_{x,y\in[N]} F(g^P(x,y)\Gamma^4) = \int_{G^P/\Gamma^P} F + O_M(A^{-c_M})$$

for some $c_M > 0$.

The assumption that G has a filtration of degree 2 is made to simplify the exposition, and because all the difficulties that emerge in higher-step cases are already present here.

We shall need the following lemma.

Lemma 4.8.2. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be a homogeneous polynomial progression, $\varepsilon > 0$, and $s, N \in \mathbb{N}_+$. Let $W_i \leqslant \mathbb{R}[x,y]$ be as defined in Section 4.4, and for each $1 \leqslant i \leqslant s$, let $Q_{i,1}, \dots, Q_{i,t_i}$ be a basis for W_i composed of integral polynomials. Suppose that a_{ij} are real numbers such that the polynomial*

$$Q(x,y) = \sum_{i=1}^{s}\sum_{j=1}^{t_i} a_{ij} Q_{i,j}(x,y)$$

satisfies $||Q||_{C^\infty[N]} \leqslant \varepsilon$. Then there exists a positive integer $q = O(1)$ with the property that $||qa_{sj}||_{\mathbb{R}/\mathbb{Z}} \ll \varepsilon N^{-s}$ for all $1 \leqslant j \leqslant t_s$.

Proof. For $s \in \mathbb{N}_+$, we let W_s, V_s be as in Section 4.4. We also define

$$\tilde{W}_s = \mathrm{Span}_{\mathbb{R}}\{(x+P_i(y))^s : 0 \leqslant i \leqslant t\} \quad \text{and} \quad U_s = \mathrm{Span}_{\mathbb{R}}\left\{\binom{x}{i}\binom{y}{j} : i+j < s\right\}.$$

We want to show first that $\dim W_s/U_s = \dim W_s = t_s$, i.e. that the polynomials $Q_{s,1}, \dots, Q_{s,t_s}$ remain linearly independent when we subtract from them the monomials in the Taylor basis of degree less than s. While this claim may plausibly hold for any polynomial progression, we prove it for homogeneous progressions since this is the only case in which we need this result. The homogeneity of $\vec{P}$ implies that $W_s \cong V_s/V_{s-1} \cong \tilde{W}_s$. Therefore $W_s/U_s \cong V_s/U_sV_{s-1} \cong \tilde{W}_s/U_s \cong \tilde{W}_s$, where the last isomorphism follows from the fact no polynomial in $\tilde{W}_s$ has a nonzero monomial of degree less than s. The claim $\dim W_s/U_s = t_s$ follows.

Let $Q(x,y) = \sum_{k,l} c_{kl}\binom{x}{k}\binom{y}{l}$ and $\tilde{Q}(x,y) = \sum\limits_{k+l\geqslant s} c_{kl}\binom{x}{k}\binom{y}{l}$. Thus, $\tilde{Q} = Q \mod U_s$, and it satisfies $||\tilde{Q}||_{C^\infty[N]} \leqslant \varepsilon$. Setting $Q_{i,j}(x,y) = \sum\limits_{k,l} b_{klij}\binom{x}{k}\binom{y}{l}$, we similarly let $\tilde{Q}_{i,j}(x,y) = \sum\limits_{k+l\geqslant s} b_{klij}\binom{x}{k}\binom{y}{l}$. We deduce from $\dim W_k/U_k = t_k = \dim W_k$ that $\tilde{Q}_{s,1}, \dots, \tilde{Q}_{s,t_s}$ are linearly independent.

From the definitions of Q and b_{klij} it follows that $c_{kl} = \sum\limits_{i,j} b_{klij}a_{ij}$, and that $||c_{kl}||_{\mathbb{R}/\mathbb{Z}} \leqslant \varepsilon N^{-(k+l)} \leqslant \varepsilon N^{-s}$ whenever $k+l \geqslant s$.

Let u be the number of pairs (k,l) with $k+l \geqslant s$ for which $c_{kl} \neq 0$. The fact that $\dim W_s/U_s = t_s$ implies that $u \geqslant t_s$. Indexing these pairs as $(k_1,l_1), \dots, (k_u,l_u)$ in some arbitrary fashion, we obtain an $u \times s$ matrix

$B = (b_{k_r l_r ij})_r$ as well as a t_s-dimensional column vector $a = (a_{sj})_j$ and a u-dimensional column vector $c = (c_{k_r l_r})_r$ such that $Ba = c$. The linear independence of $\tilde{Q}_{s,1}, ..., \tilde{Q}_{s,t_s}$ implies that there exists an invertible $t_s \times t_s$ submatrix $\tilde{B}$ of B and a t_s-dimensional column vector $\tilde{c}$ such that $\tilde{B}a = \tilde{c}$. Since the entries of $\tilde{B}$ are integers of size $O(1)$, the entries of $\tilde{B}^{-1}$ are rational numbers of height $O(1)$. Therefore, there exists a positive integer $q = O(1)$ for which the entries of the matrix $q\tilde{B}^{-1}$ are integers of size $O(1)$. The equality $a = \tilde{B}^{-1}\tilde{c}$ and the condition $||c_{kl}||_{\mathbb{R}/\mathbb{Z}} \leqslant \varepsilon N^{-s}$ whenever $k + l \geqslant s$ imply that $||qa_{sj}||_{\mathbb{R}/\mathbb{Z}} \ll \varepsilon N^{-s}$ for $1 \leqslant j \leqslant t_s$, as claimed. □

Proof of Proposition 4.8.1. Let $\vec{P} = (x,\ x + y,\ x + 2y,\ x + y^3)$. We set

$$\vec{v}_1 = (1, 1, 1, 1), \quad \vec{v}_2 = (0, 1, 2, 0), \quad \vec{v}_3 = (0, 0, 0, 1) \quad \text{and} \quad \vec{v}_4 = (0, 0, 1, 0)$$

and observe that

$$\vec{P}(x, y) = \vec{v}_1 x + \vec{v}_2 y + \vec{v}_3 y^3$$
$$\binom{\vec{P}(x, y)}{2} = \vec{v}_1 \binom{x}{2} + \vec{v}_2 \left(xy + \binom{y}{2}\right) + \vec{v}_3 \left(xy^3 + \binom{y^3}{2}\right) + \vec{v}_4 y^2.$$

Thus, we have

$$\mathcal{P}_1 = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \quad \text{and} \quad \mathcal{P}_2 = \mathcal{P}_3 = ... = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\} = \mathbb{R}^4$$

as well as

$$G^P = G^{\vec{v}_1} G^{\vec{v}_2} G^{\vec{v}_3} G_2^4,$$

where $H^{\vec{w}} = \langle h^{\vec{w}} : h \in H\rangle$.

We shall prove Proposition 4.8.1 by applying Theorem 4.6.7. Suppose that g^P is not $(c_M A^{-C_M}, N)$-equidistributed on G^P/Γ^P for some constants $0 < c_M < 1 < C_M$. By Theorem 4.6.7, there exists a nontrivial horizontal character $\eta : G^P \to \mathbb{R}$ of modulus at most cA, for which $||\eta \circ g^P||_{C^\infty[N]} \leqslant cA$ for some constant $c > 0$ that depends on c_M and C_M. The constant C_M is chosen in such a way as to match the exponents in the case (ii) of Theorem 4.6.7. We however have control over how we choose the constant c_M, and we shall pick it small enough to show that g^P not being $(c_M A^{-C_M}, N)$-equidistributed contradicts the (A, N)-irrationality of g.

Rewriting the expression for $\eta \circ g^P$, we see that

$$\eta \circ g^P(x,y) = \eta(g_1^{\vec{v}_1})x + \eta(g_1^{\vec{v}_2})y + \eta(g_1^{\vec{v}_3})y^3 + \eta(g_2^{\vec{v}_1})\binom{x}{2} + \eta(g_2^{\vec{v}_2})\left(xy + \binom{y}{2}\right) + \eta(g_2^{\vec{v}_3})\left(xy^3 + \binom{y^3}{2}\right) + \vec{v}_4 y^2.$$

Applying Lemma 4.8.2 and the assumption $||\eta \circ g^P||_{C^\infty[N]} \leqslant cA$, and choosing c_M in such a way that $c > 0$ is sufficiently small, we deduce that there exists a positive integer $q = O(1)$ such that $||q\eta(g_i^{\vec{v}_j})||_{\mathbb{R}/\mathbb{Z}} < AN^{-i}$ for all pairs

$$(i,j) \in \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(2,4)\}.$$

We aim to show that η is trivial by showing that it vanishes on all of G^P. First, we want to show that η vanishes on G_2^4. Suppose that $\eta|_{G_2^4} \neq 0$, and define $\xi_{2,1} : G_2 \to \mathbb{R}$ by $\xi_{2,1}(h_2) = q\eta(h_2^{(1,1,1,1)})$. We claim that $\xi_{2,1}$ is a 2^{nd} level character. To prove this, we need to show that $\xi_{2,1}$ is a continuous group homomorphism, it vanishes on G_3, it sends (Γ_2) to $\mathbb{Z}$, and it vanishes on $[G_1, G_1]$. The first statement follows from the fact that η is a continuous group homomorphism, the second is true since G_3 is trivial, and the third follows from the fact that $q \in \mathbb{Z}$, $\eta(\Gamma^P) \leqslant \mathbb{Z}$ and $(1,1,1,1) \in \mathbb{Z}^4$. To see the last statement, we note that for any $h_1, h_1' \in G_1$, we have

$$[h_1^{\vec{v}_1}, h_1'^{\vec{v}_1}] = [h_1, h_1']^{\vec{v}_1}.$$

Since $h_1^{\vec{v}_1}, h_1'^{\vec{v}_1}$ are both elements of G^P, we have

$$\xi_{2,1}([h_1, h_1']) = \eta([h_1, h_1']^{\vec{v}_1}) = \eta([h_1^{\vec{v}_1}, h_1'^{\vec{v}_1}]) = 0,$$

implying that $\xi_{2,1}$ vanishes on $[G_1, G_1]$. Thus, $\xi_{2,1}$ is a 2-rd level character.

Performing a similar analysis while looking at the coefficients of $\binom{x}{2}$, $xy + \binom{y}{2}$, $xy^3 + \binom{y^3}{2}$ and y^2 respectively, we conclude that for all $1 \leqslant j \leqslant 4$, the maps $\xi_{2,j}(h_2) = q\eta(h_2^{\vec{v}_j})$ from G_2 to $\mathbb{R}$ are 2^{nd} level characters. The nontriviality of η on G_2^4 and the fact that $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$ and $\vec{v}_4$ span $\mathcal{P}_2 = \mathbb{R}^4$ imply that for at least one value $1 \leqslant i \leqslant 4$, the character η does not vanish on $G_2^{\vec{v}_i}$. We fix this i. From $||\xi_{2,i}(g_i)||_{\mathbb{R}/\mathbb{Z}} = ||q\eta(g_i^{\vec{v}_j})||_{\mathbb{R}/\mathbb{Z}} < AN^{-i}$ and the (A,N)-irrationality of g_2 we deduce that $|\xi_{2,i}| > A$. Together with the bounds $q = O(1)$ and $|\vec{v}_1| = O(1)$, this implies that $|\eta| > c'A$ for some constant $c' > 0$. Choosing c_M in such a way that $c < c'$ gives the desired contradiction. Hence η vanishes on G_2^4.

This leaves us with

$$\eta \circ g^P(x,y) = \eta(g_1^{\vec{v}_1})x + \eta(g_1^{\vec{v}_2})y + \eta(g_1^{\vec{v}_3})y^3.$$

By analysing the coefficients of x, y and y^3 as above, we see that η vanishes on elements of the form $h_1^{\vec{v}_i}$ with $h_1 \in G_1$ and $1 \leqslant i \leqslant 3$. Thus, η vanishes on all of G^P. This contradicts the nontriviality of η, and so g^P is $(c_M A^{-C_M}, N)$-equidistributed on G^P/Γ^P. $\square$

We now prove Theorem 4.6.9 in full generality.

Proof of Theorem 4.6.9. Let $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression, $G_\bullet$ be a filtration of degree s and $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$. By (4.27), we can find a family $\{Q_{i,j} : \ 1 \leqslant i \leqslant s, \ 1 \leqslant j \leqslant t_i\}$ of linearly independent integral polynomials such that $Q_{i,1}$, ..., Q_{i,t_i} is a basis for $W_i = W_i'$ for $1 \leqslant i \leqslant s$. It is crucial that these polynomials are linearly independent, which follows from homogeneity of $\vec{P}$. For each i, let $\tau_i : W_i \to \mathcal{P}_i$ be the map associated with $Q_{i,1}$, ..., Q_{i,t_i} as defined in Section 4.4. We also let $\vec{v}_{i,j} \in \mathbb{Z}^{t+1}$ be the vectors such that $\tau_i(Q_{i,j}) = \vec{v}_{i,j}$.

As in the proof of Proposition 4.8.1, suppose that g^P is not $(c_M A^{-C_M}, N)$-equidistributed on G^P/Γ^P for some constants $0 < c_M < 1 < C_M$. We apply Theorem 4.6.7 again to conclude that there exists a nontrivial horizontal character $\eta : G^P \to \mathbb{R}$ of modulus at most cA, satisfying $||\eta \circ g^P||_{C^\infty[N]} \leqslant cA$ for some constant $c > 0$ that depends on c_M and C_M. The constant C_M is chosen in such a way as to match the exponents in the case (ii) of Theorem 4.6.7, but the choice of c_M is up to us again. We shall pick it small enough to show that the failure of g^P to be $(c_M A^{-C_M}, N)$-equidistributed contradicts the (A, N)-irrationality of g.

Thus,

$$\eta \circ g^P(x,y) = \sum_{i=1}^{s} \sum_{j=1}^{t_i} \eta(g_i^{\vec{v}_{i,j}}) Q_{i,j}(x,y).$$

Using Lemma 4.8.2 and the assumption $||\eta \circ g^P||_{C^\infty[N]} \leqslant cA$, and choosing c_M in such a way that $c > 0$ is sufficiently small, we deduce that there exists a positive integer $q = O(1)$ such that $||q\eta(g_i^{\vec{v}_{i,j}})||_{\mathbb{R}/\mathbb{Z}} < AN^{-i}$ for all $1 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant t_i$.

Our goal now is to show by downward induction on i that η vanishes on the group

$$H_i = \langle h_i^{\vec{v}_{i,j}} : h_i \in G_i, 1 \leqslant j \leqslant t_i \rangle$$

for all $i \in \mathbb{N}_+$. This is trivially true for $i \geqslant s+1$. Suppose that η vanishes on H_{i+1} for some $1 \leqslant i \leqslant s$ but that it does not vanish on H_i. We define the maps $\xi_{i,j} : G_i \to \mathbb{R}$ by $\xi_{i,j}(h_i) = \eta(qh_i^{\vec{v}_{i,j}})$ and claim that they are i-th level characters. They are continuous group homomorphisms because η is, and they vanish on G_{i+1} by induction hypothesis. Since $q \in \mathbb{Z}$ and $\vec{v}_{i,j}$ have integer entries, we also have $\xi_{i,j}(\Gamma_i) \subseteq \mathbb{Z}$. It remains to show that $\xi_{i,j}$ vanishes on $[G_l, G_{i-l}]$ for all $1 \leqslant l < i$. The fact that $\mathcal{P}_i \subseteq \mathcal{P}_l \cdot \mathcal{P}_{i-l}$ implies the existence of $\vec{u}_l \in \mathcal{P}_l$ and $\vec{u}_{i-l} \in \mathcal{P}_{i-l}$ for which $\vec{v}_{i,j} = \vec{u}_l \cdot \vec{u}_{i-l}$, and so we have

$$[G_l^{\vec{u}_l}, G_{i-l}^{\vec{u}_{i-l}}] = [G_l, G_{i-l}]^{\vec{u}_l \cdot \vec{u}_{i-l}} \bmod G_{i+1}^{t+1},$$

from which it follows that $\xi_{i,j}|_{[G_l, G_{i-l}]} = 0$. Therefore each $\xi_{i,j}$ is an i-th level character.

The nontriviality of η on H_i and the fact that $\mathcal{P}_i$ is spanned by the vectors $\vec{v}_{i,1}$, ..., $\vec{v}_{i,t_i}$ imply that for at least one value $1 \leqslant j \leqslant t_i$, the character η does not vanish on $G_i^{\vec{v}_{i,j}}$, and so $\xi_{i,j}$ is nontrivial. From $||\xi_{i,j}(g_i)||_{\mathbb{R}/\mathbb{Z}} = ||q\eta(g_i^{\vec{v}_{i,j}})||_{\mathbb{R}/\mathbb{Z}} < AN^{-i}$ and the (A, N)-irrationality of g_i we deduce that $|\xi_{i,j}| > A$. Together with the bounds $q = O(1)$ and $|\vec{v}_{i,j}| = O(1)$, this implies that $|\eta| > c'A$ for some constant $c' > 0$. We choose c_M in such a way that $c < c'$; this contradicts the nontriviality of η on H_i. This proves the inductive step; hence η vanishes on all of G^P, contradicting the nontriviality of η. It follows that g^P is $(c_M A^{-C_M}, N)$-equidistributed on G^P/Γ^P. $\square$

4.9 The failure of Theorem 4.6.9 in the inhomogeneous case

Having derived Theorem 4.6.9, we want to show why an analogous statement fails in the inhomogeneous case. We let

$$\vec{P}(x, y) = (x,\ x+y,\ x+2y,\ x+y^2), \tag{4.38}$$

with a square instead of a cube in the last position. It is an inhomogeneous progression because of the inhomogeneous relation (4.8). Suppose that $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ is an irrational polynomial sequence with $g(0) = 1$ on a connected group G with a filtration $G_\bullet$ of degree 2. We shall try to show that g^P is equidistributed on G^P/Γ^P the same way as we argued in Proposition 4.8.1, and we indicate where and why the argument fails.

Once again, we let

$$\vec{v}_1 = (1,1,1,1), \quad \vec{v}_2 = (0,1,2,0), \quad \vec{v}_3 = (0,0,0,1) \quad \text{and} \quad \vec{v}_4 = (0,0,1,0),$$

and we observe that $\mathcal{P}_1 = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ and $\mathcal{P}_2 = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$. Hence $G^P = G^{\vec{v}_1} G^{\vec{v}_2} G^{\vec{v}_3} G_2^4$. Suppose that g^P is not $(c_M A^{-C_M}, N)$-equidistributed on G^P/Γ^P for some constants $0 < c_M < 1 < C_M$. Theorem 4.6.7 once again implies the existence of a nontrivial horizontal character $\eta : G^P \to \mathbb{R}$ of modulus at most cA, for which $||\eta \circ g^P||_{C^\infty[N]} \leqslant cA$ for some constant $c > 0$ that depends on c_M and C_M.

Rewriting the expression for $\eta \circ g^P$, we see that

$$\begin{aligned}
\eta \circ g^P(x,y) &= \eta(g_1^{\vec{v}_1})x + \eta(g_1^{\vec{v}_2})y + \eta(g_1^{\vec{v}_3})y^2 \\
&\quad + \eta(g_2^{\vec{v}_1})\binom{x}{2} + \eta(g_2^{\vec{v}_2})\left(xy + \binom{y}{2}\right) + \eta(g_2^{\vec{v}_3})\left(xy^2 + \binom{y^2}{2}\right) + \vec{v}_4 y^2 \\
&= \eta(g_1^{\vec{v}_1})x + \eta(g_1^{\vec{v}_2})y + (\eta(g_1^{\vec{v}_3}) + \eta(g_2^{\vec{v}_4}))y^2 \\
&\quad + \eta(g_2^{\vec{v}_1})\binom{x}{2} + \eta(g_2^{\vec{v}_2})\left(xy + \binom{y}{2}\right) + \eta(g_2^{\vec{v}_3})\left(xy^2 + \binom{y^2}{2}\right).
\end{aligned}$$

Applying Lemma 4.8.2 and the assumption $||\eta \circ g^P||_{C^\infty[N]} \leqslant cA$, and choosing c_M in such a way that $c > 0$ is sufficiently small, we deduce that there exists a positive integer $q = O(1)$ such that

$$||q\eta(g_i^{\vec{v}_j})||_{\mathbb{R}/\mathbb{Z}} < AN^{-i} \tag{4.39}$$

for all pairs

$$(i,j) \in \{(1,1),(1,2),(2,1),(2,2),(2,3)\}.$$

By looking at the coefficient of $\binom{x}{2}$, $xy + \binom{y}{2}$ and $xy^2 + \binom{y^2}{2}$, we deduce that the maps

$$h_2 \mapsto q\eta(h_2^{\vec{v}_1}),\ q\eta(h_2^{\vec{v}_2}),\ q\eta(h_2^{\vec{v}_3})$$

are trivial 2^{nd} level characters; the argument goes the exact same way as in the proof of Proposition 4.8.1. Thus, η vanishes on all elements of the form $h_2^{\vec{w}_2}$ with $h_2 \in G_2$ and

$$\vec{w}_2 \in \mathcal{P}_2' = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}.$$

By looking at the coefficients of x and y, we similarly show that η vanishes on all elements of the form $h_1^{\vec{w}_1}$ with $h_1 \in G_1$ and

$$\vec{w}_1 \in \mathcal{P}_1' = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_1, \vec{v}_2\}.$$

We are left with

$$\eta \circ g^P(x,y) = \left(\eta(g_1^{\vec{v}_3}) + \eta(g_2^{\vec{v}_4})\right) y^2.$$

We would like to be able to say that η vanishes on all elements of the form $h_1^{\vec{w}_1}$ and $h_2^{\vec{w}_2}$ with $h_i \in G_i$ and $\vec{w}_i \in \mathcal{P}_i$; this would imply that η is trivial. For this to be case, it would suffice to show that both $\eta(g_1^{\vec{v}_3})$ and $\eta(g_2^{\vec{v}_4})$ satisfy an estimate (4.39), and then use (A, N)-irrationality of g_1 and g_2 to conclude that the characters $h_1 \mapsto q\eta(h_1^{\vec{v}_3})$ and $h_2 \mapsto q\eta(h_2^{\vec{v}_4})$ are trivial. Alas, this need not be true. In Proposition 4.8.1, the number $\eta(h_1^{\vec{v}_3})$ was the coefficient of y^3 while $\eta(h_2^{\vec{v}_4})$ was the coefficient of y^2, from which it followed that they both satisfied (4.39). Now, however, all we can show is that

$$||q(\eta(g_1^{\vec{v}_3}) + \eta(g_2^{\vec{v}_4}))||_{\mathbb{R}/\mathbb{Z}} < AN^{-1} \tag{4.40}$$

because $\eta(g_1^{\vec{v}_3}) + \eta(g_2^{\vec{v}_4})$ is the coefficient of y^2. But it need not follow that either of $\eta(g_1^{\vec{v}_3})$ and $\eta(g_2^{\vec{v}_4})$ satisfies (4.39); in particular, g^P may take values in a proper rational subgroup of G^P.

We illustrate this with a specific example. Suppose that $G = G_1 = \mathbb{R}^2$, $G_2 = 0 \times \mathbb{R}$, $G_3 = 0 \times 0$. The sequence $g(n) = (an, b\binom{n}{2})$ is adapted to the filtration $G_\bullet$, and it is irrational if and only if a and b are irrational. We identify G^4 with $\mathbb{R}^8$ via the map

$$G^4 \to \mathbb{R}^8$$
$$((x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)) \mapsto (x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4).$$

Setting

$$\vec{v}_{11} = \vec{e}_1 + \vec{e}_2 + \vec{e}_3 + \vec{e}_4, \quad \vec{v}_{12} = \vec{e}_2 + 2\vec{e}_3, \quad \vec{v}_{13} = \vec{e}_4,$$
$$\vec{v}_{21} = \vec{e}_5 + \vec{e}_6 + \vec{e}_7 + \vec{e}_8, \quad \vec{v}_{22} = \vec{e}_6 + 2\vec{e}_7, \quad \vec{v}_{23} = \vec{e}_8, \quad \vec{v}_{24} = \vec{e}_7,$$

we observe that $G^P = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_{11}, \vec{v}_{12}, \vec{v}_{13}, \vec{v}_{21}, \vec{v}_{22}, \vec{v}_{23}, \vec{v}_{24}\}$.

With these definitions, the coefficient of y^2 in g^P becomes $a\vec{v}_{13} + b\vec{v}_{24} = a\vec{e}_4 + b\vec{e}_7$. If $a, b, 1$ are rationally independent, then the closure of g^P is the image of the 7-dimensional subspace G^P in $(\mathbb{R}/\mathbb{Z})^8$. If a and b are rationally dependent, then the closure of g^P is the image in $(\mathbb{R}/\mathbb{Z})^8$ of the 6-dimensional subspace

$$\tilde{G} = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_{11}, \vec{v}_{12}, a\vec{v}_{13} + b\vec{v}_{24}, \vec{v}_{21}, \vec{v}_{22}, \vec{v}_{23}\}.$$

Finally, if some rational linear combination of a and b is a rational number q/r in its lower terms with $r > 1$, then the closure of g^P is a union of at most r translates of a 6-dimensional subtorus of G^P/Γ^P. For instance, if $a = \sqrt{2}$ and $b = \sqrt{2} + \frac{1}{3}$, then we define

$$\tilde{G} = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_{11}, \vec{v}_{12}, \vec{v}_{13} + \vec{v}_{24}, \vec{v}_{21}, \vec{v}_{22}, \vec{v}_{23}\}, \tag{4.41}$$

and observe that the sequences g_0^P, g_1^P, g_2^P defined by $g_i^P(x, y) = g^P(x, 3y+i)$ are equidistributed on $\tilde{G}/\tilde{\Gamma}$, $\frac{1}{3}\vec{v}_{24} + \tilde{G}/\tilde{\Gamma}$ and $\frac{1}{3}\vec{v}_{24} + \tilde{G}/\tilde{\Gamma}$ respectively. In particular, for inhomogeneous progressions it is not true that the group $\tilde{G}$ depends only on the filtration $G_\bullet$ and the progression $\vec{P}$.

While annihilating the coefficients of $\eta \circ g^P$, we were able to deal with the coefficients of x and y as well as $\binom{x}{2}$, $xy + \binom{y}{2}$ and $xy^2 + \binom{y^2}{2}$, which span the spaces W_1' and W_2' respectively. The problematic coefficient was that of y^2, belonging to the space W^c. We have remarked below (4.28) in Section 4.4 that the nontriviality of the subspace W^c prevents us from running the same argument as in Proposition 4.8.1 and Theorem 4.6.9 for inhomogeneous progressions; the problem with the coefficient of y^2 that we have encountered here illustrates this point. The reader should see from here how to generalise the aforementioned example to other inhomogeneous progression; this generalised construction proves part (ii) of Theorem 4.1.19.

4.10 Finding closure in the inhomogeneous case

Section 4.9 shows that we cannot always hope for the sequence g^P to equidistribute in G^P/Γ^P for an inhomogeneous progression $\vec{P}$. Here, we provide an inductive recipe for finding the closure of g^P in the case of $\vec{P}(x, y) = (x,\ x+y,\ x+2y,\ x+y^2)$. We believe that this argument could be generalised to an arbitrary inhomogeneous progressions; while trying to do so, however, we have encountered significant technical issues of linear algebraic nature that we have not been able to overcome.

Since the argument that we present here is already complicated enough, we prove it in an infinitary setting so as to avoid confusion coming from various quantitative parameters. In effect, we show the following.

Proposition 4.10.1. *Let G be a connected group with filtration $G_\bullet$ of degree s, and $\vec{P}(x,y) = (x,\ x+y,\ x+2y,\ x+y^2)$. Suppose that $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ is irrational. There exists a subgroup $\tilde{G} \leqslant G^P$ and a decomposition $g^P = \tilde{g}\gamma$, where $\tilde{g}$ takes values in $\tilde{G}$ and is equidistributed on $\tilde{G}/\tilde{\Gamma}$ whereas γ is periodic. Moreover, the group $\tilde{G}$ contains the subgroup*

$$K = \langle h_i^{\vec{w}_i} : h_i \in G_i, \vec{w}_i \in \mathcal{P}_i', 1 \leqslant i \leqslant s \rangle,$$

where

$$\begin{aligned} \mathcal{P}_1' &= \mathrm{Span}_{\mathbb{R}}\{(1,1,1,1),(0,1,2,0)\}, \\ \mathcal{P}_2' &= \mathrm{Span}_{\mathbb{R}}\{(1,1,1,1),(0,1,2,0),(0,0,0,1)\}, \\ \mathcal{P}_3' &= \mathcal{P}_4' = ... = \mathbb{R}^4. \end{aligned}$$

We will need the following lemma, which is similar in spirit to Lemma 4.8.2.

Lemma 4.10.2. *Let $a_1, ..., a_s$ be nonzero real numbers. Let $Q_1, ..., Q_s \in \mathbb{Q}[x,y]$ be linearly independent polynomials, and suppose that $Q = a_1 Q_1 + ... + a_r Q_s$ takes values in $\mathbb{Q}$. Then $a_i \in \mathbb{Q}$ for all $1 \leqslant i \leqslant s$.*

Proof. Let b_{kli} be the coefficient of $\binom{x}{k}\binom{y}{l}$ in Q_i. Then

$$c_{kl} = a_1 b_{kl1} + ... + a_s b_{kls}$$

is the coefficient of $\binom{x}{k}\binom{y}{l}$ in Q, and so it is rational. Indexing the pairs $(k_1, l_1), ..., (k_u, l_u)$ in some arbitrary fashion, we obtain an $u \times s$ matrix $B = (b_{k_r l_r i})_{ri}$ as well as an s-dimensional column vector $a = (a_i)_i$ and a u-dimensional column vector $c = (c_{j_l k_l})_l$ such that $Ba = c$. The linear independence of $Q_1, ..., Q_r$ implies that B has full rank, and so there exists an invertible $s \times s$ submatrix $\tilde{B}$ of B and an s-dimensional column vector $\tilde{c}$ such that $\tilde{B}a = \tilde{c}$. Since the entries of $\tilde{B}$ are rational, so are the entries of $\tilde{B}^{-1}$. The equality $a = \tilde{B}^{-1}\tilde{c}$ then implies that $a_i \in \mathbb{Q}$ for each $1 \leqslant i \leqslant s$. $\square$

Proof of Proposition 4.10.1. For each $i \geqslant 3$, we find a basis $\{Q_{i,1}, Q_{i,2}, Q_{i,3}, Q_{i,4}\}$ for W_i. The absence of an inhomogeneous algebraic relation of degree 3 or higher implies that

$$\sum_{i=3}^{s} W_i = \bigoplus_{i=3}^{s} W_i,$$

from which it follows that the polynomials in the set $\{Q_{i,j} : 3 \leqslant i \leqslant s,\ 1 \leqslant j \leqslant 4\}$ is linearly independent. For $3 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant 4$, we let $\vec{v}_{i,j} = \tau_i(Q_{i,j})$. We also set

$$\vec{v}_1 = (1,1,1,1), \quad \vec{v}_2 = (0,1,2,0), \quad \vec{v}_3 = (0,0,0,1) \quad \text{and} \quad \vec{v}_4 = (0,0,1,0).$$

We want to find a subgroup $\tilde{G}$ of G^P on which we can guarantee equidistribution. Starting with

$$H^{(1)} = \langle h_1^{\vec{v}_3},\ h_2^{\vec{v}_4} : h_1 \in G_1, h_2 \in G_2 \rangle,$$

we inductively define a chain of subgroups

$$H^{(1)} \geqslant H^{(2)} \geqslant H^{(3)} \geqslant \ldots$$

as well as groups $G^{(k)} = \langle K, H^{(k)} \rangle$ and $\Gamma^{(k)} = \Gamma^P \cap G^{(k)}$. We note that $G^{(1)} = G^P$.

We also inductively define sequences $g^{(k)}$ and $h^{(k)}$, starting with $h^{(1)}(y) = g_1^{\vec{v}_3 y^2} g_2^{\vec{v}_4 y^2}$ and $g^{(1)} = g^P$. If $g^{(k)}$ is equidistributed in $G^{(k)}/\Gamma^{(k)}$, then we terminate the procedure. Otherwise Theorem 4.2.5 implies the existence of a nontrivial horizontal character $\eta^{(k)} : G^{(k)} \to \mathbb{R}$ that vanishes on all of $G^{(k)}$ except $H^{(k)}$, and for which $\eta^{(k)} \circ g^{(k)} = \eta^{(k)} \circ h^{(k)}$ takes values in $\mathbb{Z}$. We then take $G^{(k+1)} = \ker \eta^{(k)}$ and $H^{(k+1)} = \ker \eta^{(k)}|_{H^{(k)}}$, and we factorize $h^{(k)} = h^{(k+1)}\gamma^{(k+1)}$ using an infinitary version of Proposition 9.2 of [GT12], where $\eta^{k+1} \circ h^{(k+1)} = 0$ and $\gamma^{(k+1)}$ is periodic. We define

$$g^{(k+1)}(x,y) = g^{(k)}(x,y)(\gamma^{(k+1)}(y))^{-1}$$

and observe that

$$g^{(k+1)}(x,y) = g_1^{\vec{v}_1 x + \vec{v}_2 y} h^{(k+1)}(y) g_2^{\vec{v}_1\binom{x}{2} + \vec{v}_2(xy + \binom{y}{2}) + \vec{v}_3(xy^2 + \binom{y^2}{2})} \prod_{i=3}^{s}\prod_{j=1}^{4} g_i^{\vec{v}_{i,j} Q_{i,j}} \bmod [G_1, G_2]^4.$$

The sequence $g^{(k+1)}$ takes values in $G^{(k+1)}$. We also write

$$h^{(k)}(y) = a^{(k)}(y)^{\vec{v}_4} b^{(k)}(y)^{\vec{v}_3},$$

with $a^{(k)}$ being G_2-valued and $b^{(k)}$ being G_1-valued. Letting $a^{(k)}(y) = \prod_{i=1}^{s} a_i^{(k)\binom{y}{i}}$

and similarly for $b^{(k)}$, we claim that $a_2^{(k)}$ and $b_2^{(k)}$ are irrational elements of G_2 and G_1 respectively with regard to the filtration $G_\bullet$ on G. Finally, we claim that

$$H^{(k)} = G_2^{\vec{v}_4} \mod G_1^{\vec{v}_3} \quad \text{and} \quad H^{(k)} = G_1^{\vec{v}_3} \mod G_2^{\vec{v}_4}$$

First, we observe that all these properties hold at $k = 1$. We assume that they hold for some $k \geqslant 1$, from which we aim to deduce that they also hold at $(k+1)$-th level.

If $g^{(k)}$ is equidistributed in $G^{(k)}/\Gamma^{(k)}$, then we are done. Otherwise there exists a nontrivial horizontal character $\eta^{(k)} : G^{(k)} \to \mathbb{R}$ for which $\eta^{(k)} \circ g^{(k)}$ is $\mathbb{Z}$-valued. We have

$$\begin{aligned}\eta^{(k)} \circ g^{(k)}(x,y) = {} & \eta^{(k)}(g_1^{\vec{v}_1})x + \eta^{(k)}(g_1^{\vec{v}_2})y + \eta^{(k)}(h^{(k)}(y)) \\ & + \eta^{(k)}(g_2^{\vec{v}_1})\binom{x}{2} + 2\eta^{(k)}(g_2^{\vec{v}_2})\left(xy + \binom{y}{2}\right) + \eta^{(k)}(g_2^{\vec{v}_3})\left(xy^2 + \binom{y^2}{2}\right) \\ & + \sum_{i=3}^{k}\sum_{j=1}^{4} \eta^{(k)}(g_i^{\vec{v}_{i,j}})Q_{i,j}(x,y).\end{aligned}$$

By looking at the coefficients of $Q_{i,j}$ for $3 \leqslant i \leqslant s$, applying Lemma 4.10.2, and following the same method as in the proof of Theorem 4.6.9, we see that $\eta^{(k)}$ vanishes on elements of the form $h_i^{\vec{v}_{i,j}}$ for $h_i \in G_i$, $3 \leqslant i \leqslant s$ and $1 \leqslant j \leqslant 4$, and so $\eta^{(k)}$ vanishes on all of $G_3 \times G_3 \times G_3 \times G_3$. This leaves us with

$$\begin{aligned}\eta^{(k)} \circ g^{(k)}(x,y) = {} & \eta^{(k)}(g_1^{\vec{v}_1})x + \eta^{(k)}(g_1^{\vec{v}_2})y + \eta^{(k)}(h^{(k)}(y)) \\ & + \eta^{(k)}(g_2^{\vec{v}_1})\binom{x}{2} + 2\eta^{(k)}(g_2^{\vec{v}_2})\left(xy + \binom{y}{2}\right) + \eta^{(k)}(g_2^{\vec{v}_3})\left(xy^2 + \binom{y^2}{2}\right).\end{aligned}$$

We now carry on. By looking at the coefficient of $\binom{x}{2}$ and $xy + \binom{y}{2}$, we see that $\eta^{(k)}(g_2^{\vec{v}_1})$ and $\eta^{(k)}(g_2^{\vec{v}_2})$ are both integers, and so $\eta^{(k)}$ vanishes on all elements of the form $h_2^{\vec{v}_1}$ and $h_2^{\vec{v}_2}$ with $h_2 \in G_2$. By looking at the coefficients of x and y, we can similarly show that $\eta^{(k)}$ vanishes on all elements of the form $h_1^{\vec{v}_1}$ and $h_1^{\vec{v}_2}$ with $h_1 \in G_1$. We are thus left with

$$\eta^{(k)} \circ g^{(k)}(x,y) = \eta^{(k)}(h^{(k)}(y)) + \eta^{(k)}(g_2^{\vec{v}_3})\left(xy^2 + \binom{y^2}{2}\right).$$

We first deal with the last term. Since $H^{(k)} = G_1^{\vec{v}_3} \mod G_2^{\vec{v}_3}$, we have $[H^{(k)}, H^{(k)}] = [G_1^{\vec{v}_3}, G_1^{\vec{v}_3}] \mod G_3^4$. Using the fact that $\eta^{(k)}$ vanishes on both

G_3^4 and $[H^{(k)}, H^{(k)}]$, we deduce that it also vanishes on $[G_1^{\vec{v}_3}, G_1^{\vec{v}_3}]$. Hence the function $\xi_{2,3} : G_2 \to \mathbb{R}$ given by $\xi_{2,3}(h) = \eta^{(k)}(h^{\vec{v}_3})$ is a 2^{nd} level character. By irrationality of g_2, it follows that $\xi_{2,3}$ is trivial, and so $\eta^{(k)}$ vanishes on $G_2^{\vec{v}_3}$. We have thus proved that $\eta^{(k)}$ vanishes on all of $G^{(k)}$ except $H^{(k)}$, and consequently that $\eta^{(k)} \circ g^{(k)} = \eta^{(k)} \circ h^{(k)}$.

We now show that

$$H^{(k+1)} = G_2^{\vec{v}_4} \mod G_1^{\vec{v}_3} \tag{4.42}$$

Suppose not; let U be a proper rational subgroup of $G_2^{\vec{v}_4}$ such that

$$H^{(k+1)} = U \mod G_1^{\vec{v}_3}.$$

Then

$$H^{(k+1)} \leqslant UG_1^{\vec{v}_3} \cap H^{(k)} \leqslant H^{(k)}.$$

We know from the rank-nullity theorem that $\dim H^{(k+1)} = \dim H^{(k)} - 1$, and we have $H^{(k)} = G_2^{\vec{v}_4} \mod G_1^{\vec{v}_3}$ from the inductive hypothesis. These two facts, together with the assumption that U is a proper rational subgroup of $G_2^{(0,0,1,0)}$, imply that $H^{(k+1)} = UG_1^{\vec{v}_3} \cap H^{(k)}$. It follows that

$$\eta^{(k)} \circ g^{(k)}(x, y) = \eta^{(k)}(a^{(k)}(y)^{\vec{v}_4}) + \eta^{(k)}(b^{(k)}(y)^{\vec{v}_3}) = \eta^{(k)}(a^{(k)}(y)^{\vec{v}_4})$$

We have already shown that $\eta^{(k)}$ vanishes on G_3^4. From the facts that $a^{(k)}(y) = \prod_{i=1}^s a_i^{(k)\binom{y}{i}}$ with $a_i^{(k)} \in G_i$, we deduce that $\eta^{(k)}(a^{(k)}(y)^{\vec{v}_4}) = \eta^{(k)}(a_1^{(k)})y + \eta^{(k)}(a_2^{(k)})\binom{y}{2}$. The map $\xi_{2,4}(h_2) = \eta^{(k)}(h_2^{\vec{v}_4})$ is a continuous group homomorphism on G_2 that vanishes on G_3 and sends Γ_2 to $\mathbb{Z}$. Since $\vec{v}_4 = (\vec{v}_2 \cdot \vec{v}_2 - \vec{v}_2)/2$, we also have

$$\xi_{2,4}([h_1, h_1']) = \frac{1}{2}\eta^{(k)}([h_1^{\vec{v}_2}, h_1'^{\vec{v}_2}]) - \frac{1}{2}\eta^{(k)}([h_1, h_1']^{\vec{v}_2}),$$

for any $h_1, h_1' \in G_1$, and so $\xi_{2,4}$ vanishes on $[G_1, G_1]$. Thus $\xi_{2,4}$ is a 2^{nd} level character on G_2 with respect to the filtration $G_\bullet$ on G, and since $a_2^{(k)}$ is an irrational element of G_2 with respect to this filtration, it follows that $\eta^{(k)}$ is trivial, a contradiction; hence (4.42) holds. The argument that

$$H^{(k+1)} = G_1^{\vec{v}_3} \mod G_2^{\vec{v}_4}$$

is similar.

Finally, we factorize $h^{(k)} = h^{(k+1)}\gamma^{(k+1)}$, where $\gamma^{(k+1)}$ is periodic and $h^{(k+1)}$ takes values in $H^{(k+1)} = \ker \eta^{(k+1)}$. It remains to show that $a_2^{(k+1)}$ and $b_2^{(k+1)}$ are irrational elements of G_2 and G_1 with respect to the filtration $G_\bullet$ on G. We observe that

$$a^{(k)} = a^{(k+1)}\gamma_a^{(k+1)} \quad \text{and} \quad b^{(k)} = b^{(k+1)}\gamma_b^{(k+1)}$$

for some periodic sequences γ_a and γ_b taking values in G_2 and G_1 respectively. Suppose that $\xi : G_2 \to \mathbb{R}$ is a 2^{nd} level character with respect to the filtration $G_\bullet$, for which $\xi(a_2^{(k+1)}) \in \mathbb{Z}$. The sequence $\gamma_a^{(k+1)}$ is periodic, hence $\xi \circ \gamma_a^{(k+1)}$ is $\mathbb{Q}$-valued, and so it follows that $\xi(a_2^{(k)}) \in \mathbb{Q}$ as well. Therefore there exists an integer $l > 0$ such that $l\xi(a_2^{(k)}) \in \mathbb{Z}$. Since $\xi' := l \cdot \xi$ is also a 2^{nd} level character, it follows from the irrationality of $a_2^{(k)}$ that ξ' is trivial. This implies that ξ is trivial as well, and hence $a_2^{(k+1)}$ is irrational. The argument showing that $b_2^{(k+1)}$ is irrational is identical.

We have thus shows inductively that $g^{(k)}$, $h^{(k)}$, $G^{(k)}$ and $H^{(k)}$ satisfy all the properties we want them to satisfy for all $k \geqslant 1$. Since $0 \leqslant \dim G^{(k+1)} < \dim G^{(k)}$, the procedure eventually terminates, at which point the sequence $g^{(k)}$ takes values in $G^{(k)}$ and is equidistributed on $G^{(k)}/\Gamma^{(k)}$. Letting $\tilde{G} = G^{(k)}$ for this value of k and $\gamma = \gamma^{(k)}...\gamma^{(1)}$, and observing that a product of periodic sequences is periodic, we finish the proof. $\square$

4.11 True complexity of $(x,\ x+y,\ ...,\ x+(t-1)y,\ x+y^d)$

Throughout this section, we let

$$\vec{Q}(x,y) = (x,\ x+y,\ ...,\ x+(t-1)y,\ x+y^d).$$

We present an alternative argument that allows us to prove Theorem 4.1.16, which gives true complexity for the progression $\vec{Q}$ whenever $2 \leqslant d \leqslant t-1$. If $d \geqslant t$, then $\vec{Q}$ is homogeneous, and this has been handled quantitatively in [Kuc21a]. A special case includes the already discussed progression $(x,\ x+y,\ x+2y,\ x+y^2)$. The method discussed in this section comes down to making the progression more homogeneous by replacing it with a longer progressions involving a higher number of variables using several applications of the Cauchy-Schwarz inequality. The content of this section has been adapted from Section 12 of [Kuc21b].

We start by proving true complexity for the nonlinear term at index t.

Proposition 4.11.1. *Let $t, d \in \mathbb{N}_+$ satisfy $t \geqslant 3$ and $d \geqslant 2$. Then*

$$\mathcal{A}_t(\vec{Q}) = \mathcal{T}_t(\vec{Q}) = \left\lfloor \frac{t-1}{d} \right\rfloor = \left\lceil \frac{t}{d} \right\rceil - 1$$

We note that one cannot get a control by a lower-degree Gowers norm here; this follows from the fact that the the space of polynomials in x and y of degree at most $t-1$ is spanned by polynomials in x, $x+y$, ..., $x+(t-1)y$ of degree at most $t-1$, so in particular it contains the $\left\lfloor \frac{t-1}{d} \right\rfloor$-th power of $x+y^d$.

Proof. We only prove the case $2 \leqslant d \leqslant t-1$, as the case $d \geqslant t$ follows from the results of [Kuc21a].

We apply the Cauchy-Schwarz inequality and translate $x \mapsto x-y$ exactly t times to remove $f_0, f_1, ..., f_{t-1}$, so that

$$\begin{aligned} &\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x) f_1(x+y) \cdots f_{t-1}(x+(t-1)y) f_t(x+y^d) \right|^{2^t} \\ &\leqslant \mathop{\mathbb{E}}_{x,y,h_1,\dots,h_t\in\mathbb{Z}/N\mathbb{Z}} \prod_{w\in\{0,1\}^t} \mathcal{C}^{|w|} f_t(\varepsilon_w), \end{aligned} \tag{4.43}$$

where

$$\varepsilon_w(x,y,h_1,...,h_t) = x + \left(y + \sum_{i=1}^t w_i h_i\right)^d - \sum_{i=1}^t (i-1) w_i h_i$$

for each $w \in \{0,1\}^t$. We moreover let

$$\vec{P}(x,y,h_1,...,h_t) = (\varepsilon_w(x,y,h_1,...,h_t))_{w\in\{0,1\}^t}.$$

The crux of the proof is to show that although the original progression $\vec{Q}$ is not homogeneous, the progression $\vec{P}$ is homogeneous and has algebraic complexity $\mathcal{A}_w(\vec{P}) \leqslant \left\lceil \frac{t}{d} \right\rceil - 1 = \left\lfloor \frac{t-1}{d} \right\rfloor$ at every index $w \in \{0,1\}^t$. One then applies the multiparameter version of Theorem 4.7.2 to deduce that the $\mathcal{T}_w(\vec{P}) = \left\lceil \frac{t}{d} \right\rceil - 1 = \left\lfloor \frac{t-1}{d} \right\rfloor$ at every index $w \in \{0,1\}^t$, and hence $\mathcal{T}_t(\vec{Q}) = \left\lceil \frac{t}{d} \right\rceil - 1 = \left\lfloor \frac{t-1}{d} \right\rfloor$ as well.

Given $w \in \{0,1\}^t$, we let $\vec{e}_w$ denote the basis vector in $\mathbb{R}^{\{0,1\}^t}$ of the form

$$\vec{e}_w(w') = \begin{cases} 1, & w' = w \\ 0, & w' \neq w. \end{cases}$$

The next lemma gives the structure of the polynomial spaces $\mathcal{P}_i$.

Lemma 4.11.2. *For each $i \in \mathbb{N}_+$, the space $\mathcal{P}_i$ is spanned by the set*

$$X_i = \left\{ \sum_{w: w_{k_1} = \ldots = w_{k_m} = 1} \vec{e}_w : \{k_1, ..., k_m\} \subseteq \{1, ..., t\},\ k_1 < ... < k_m,\ 0 \leqslant m \leqslant n \right\},$$

where $n = \min(id, t)$.

For instance, for $d = 2$, $t = 3$, we have

$$\begin{aligned}\vec{P}(x, y, h_1, h_2, h_3) = &\sum_w \vec{e}_w(x + y^2) + \sum_{w: w_1 = 1} \vec{e}_w(2h_1 y + h_1^2) \\ &+ \sum_{w: w_2 = 1} \vec{e}_w(2h_2 y + h_2^2 - h_2) + \sum_{w: w_3 = 1} \vec{e}_w(2h_3 y + h_3^2 - 2h_3) \\ &+ 2 \sum_{w: w_1 = w_2 = 1} \vec{e}_w h_1 h_2 + 2 \sum_{w: w_1 = w_3 = 1} \vec{e}_w h_1 h_3 + 2 \sum_{w: w_2 = w_3 = 1} \vec{e}_w h_2 h_3\end{aligned}$$

and an even longer expression for $\vec{P}(x, y, h_1, h_2, h_3)^2$.

Proof. To see that each vector of the form $\sum_{\substack{w: \\ w_{k_1} = \ldots = w_{k_m} = 1}} \vec{e}_w$ is contained in $\mathcal{P}_i$ for $0 \leqslant m \leqslant n$ and $k_1, ..., k_m \in \{1, ..., t\}$, we observe that the coefficient of

$$y^{id - m} h_{k_1} \cdots h_{k_m}$$

in $\vec{P}(x, y, h_1, ..., h_t)^i$ is a nonzero multiple of $\sum_{\substack{w: \\ w_{k_1} = \ldots = w_{k_m} = 1}} \vec{e}_w$. To deduce the converse, we note that the coefficient of a monomial of $\vec{P}(x, y, h_1, ..., h_t)^i$ is a nonzero multiple of $\sum_{w: w_{k_1} = \ldots = w_{k_m} = 1} \vec{e}_w$ for $0 \leqslant m \leqslant n$ if and only if the monomial contains the variables $h_{k_1}, ..., h_{k_m}$ but does not contain h_k for $k \in \{1, ..., t\} \setminus \{k_1, ..., k_m\}$. $\square$

Corollary 4.11.3. *If $i \geqslant \frac{t}{d}$, then $\mathcal{P}_i = \mathbb{R}^{\{0,1\}^t}$.*

Proof. The set X_i, which spans $\mathcal{P}_i$ by 4.11.2, consists of linearly independent vectors. If $id \geqslant t$, then X has 2^t elements, implying that $\mathcal{P}_i = \mathbb{R}^{\{0,1\}^t}$, as required. $\square$

Although the progression $\vec{Q}$ is not homogeneous, $\vec{P}$ is.

Lemma 4.11.4. *The progression $\vec{P}$ is homogeneous.*

Proof. It follows trivially from the definition of a homogeneous relation that every relation of degree 1 is homogeneous. We show that every algebraic relation

$$0 = \sum_{w} Q_w(\varepsilon_w(x, y, h_1, ..., h_t)) \tag{4.44}$$

of degree $i > 1$ satisfied by the terms of $\vec{P}$ is a sum of a homogeneous relation of degree i and an algebraic relation of degree at most $i-1$, and then use the induction hypothesis to deduce that (4.44) is a sum of homogeneous relations.

Let a_w be the coefficient of Q_w of degree i. We observe that for distinct $k_1, ..., k_m \subseteq \{1, ..., t\}$, the coefficient of $y^{id-m} h_{k_1} \cdots h_{k_m}$ in the polynomial on the right of (4.44) is a nonzero multiple of $\sum\limits_{w: w_{k_1} = ... = w_{k_m} = 1} a_w$. Hence

$$L\left(\sum_{w: w_{k_1} = ... = w_{k_m} = 1} \vec{e}_w\right) = 0,$$

where $L((x_w)_w) = \sum\limits_{w} a_w x_w$. By Lemma 4.11.2, L vanishes on all of $\mathcal{P}_i$, and as a consequence, we have

$$0 = \sum_{w} a_w \varepsilon_w^i,$$

as claimed. $\square$

We infer from the homogeneity of $\vec{P}$ and Corollary 4.11.3 that $\mathcal{A}_w(\vec{P}) \leqslant \left\lceil \frac{t}{d} \right\rceil - 1 = \left\lfloor \frac{t-1}{d} \right\rfloor$ for any $w \in \{0,1\}^t$. Since this is also a lower bound on the algebraic complexity of $\vec{P}$, we deduce that $\mathcal{A}_w(\vec{P}) = \left\lceil \frac{t}{d} \right\rceil - 1 = \left\lfloor \frac{t-1}{d} \right\rfloor$. An analogous statement for true complexity follows from the homogeneity of $\vec{P}$ and a version of Theorem 4.7.2 for multiparameter homogeneous progressions

$$(x,\ x + P_1(y_1, ..., y_r),\ ...,\ x + P_t(y_1, ..., y_r)).$$

The proof of the multiparameter version of Theorem 4.7.2 follows the same sequence of steps as the proof of the single-parameter version and encounters no additional difficulties; our decision to give the single-parameter proof of Theorem 4.7.2 has been motivated purely by the desire to enhance the clarity of the argument which would deteriorate by taking multivariable polynomials instead of polynomials of single variable in all the arguments leading up to Theorem 4.7.2. We could just conclude the proof of Proposition 4.11.1 here, but to satisfy the curious reader, we briefly summarise how one would use the homogeneity of $\vec{P}$ directly to prove that $\mathcal{T}_w(\vec{P}) = \mathcal{A}_w(\vec{P})$ for every $w \in \{0,1\}^t$.

First, a multiparameter version of Proposition 4.7.3 implies that there exists $s \in \mathbb{N}$ and $c > 0$ such that

$$\left| \mathop{\mathbb{E}}_{x,y,h_1,\ldots,h_t \in \mathbb{Z}/N\mathbb{Z}} \prod_{w \in \{0,1\}^t} \mathcal{C}^{|w|} f_t(\varepsilon_w) \right| \leqslant \|f_t\|_{U^{s+1}}^c + O(p^{-c})$$

for every 1-bounded function $f_t : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$.

Second, we use the arithmetic regularity lemma (Lemma 4.7.4) to decompose all the instances of f_t in

$$\mathop{\mathbb{E}}_{x,y,h_1,\ldots,h_t \in \mathbb{Z}/N\mathbb{Z}} \prod_{w \in \{0,1\}^t} \mathcal{C}^{|w|} f_t(\varepsilon_w).$$

For a growth function $\mathcal{F} : \mathbb{R}_+ \to \mathbb{R}_+$ to be fixed later, there exist $M = O_{\varepsilon,\mathcal{F}}(1)$, a filtered manifold G/Γ of degree s and complexity at most M, and an N-periodic, $\mathcal{F}(M)$-irrational sequence $g \in \text{poly}(\mathbb{Z}, G_\bullet)$ with $g(0) = 1$, for which there exists a decomposition

$$f_t = f_{nil} + f_{sml} + f_{unf}$$

such that $f_{nil}(n) = F(g(n)\Gamma)$ for an M-Lipschitz function $F : G/\Gamma \to \mathbb{C}$, $\|f_{sml}\|_2 \leqslant \varepsilon$ and $\|f_{unf}\|_{U^{s+1}} \leqslant \frac{1}{\mathcal{F}(M)}$. By picking $\mathcal{F}$ to be growing sufficiently fast, we can assume that $\|f_{unf}\|_{U^{s+1}} \leqslant \varepsilon^{\frac{1}{c}}$. Assuming that N is large enough with respect to ε, we have

$$\mathop{\mathbb{E}}_{x,y,h_1,\ldots,h_t \in \mathbb{Z}/N\mathbb{Z}} \prod_{w \in \{0,1\}^t} \mathcal{C}^{|w|} f(\varepsilon_w) = \mathop{\mathbb{E}}_{x,y,h_1,\ldots,h_t \in \mathbb{Z}/N\mathbb{Z}} \prod_{w \in \{0,1\}^t} \mathcal{C}^{|w|} F(g(\varepsilon_w)\Gamma) + O(\varepsilon).$$

Third, the homogeneity of $\vec{P}$ and a multiparameter version of Proposition 4.7.5 imply that the sequence $g^P(x, y, h_1, ..., h_t) = (g(\varepsilon_w)(x, y, h_1, ..., h_t))_{w \in \{0,1\}^t}$ is $O_M(\mathcal{F}(M)^{-c_M})$-equidistributed on the nilmanifold G^P/Γ^P for some $c_M > 0$, where G^P is the Leibman group associated with g^P. The proof of a multiparameter version of Theorem 4.6.5, from which the multiparameter version of Proposition 4.7.5 would be deduced, goes completely analogously to the proof of the single parameter version - it is the homogeneity that matters, not the number of parameters.

Fourth, we employ the fact that g^P is well equidistributed on G^P/Γ^P together with the same argument as in the proof of Theorem 4.7.2 to conclude the proof of Proposition 4.11.1.

□

The control by a low-degree Gowers norm of the nonlinear term $x+y^d$ is useful in that when combined with the regularity lemma (Lemma 4.7.4), it allows us to replace the function f_t by a low-degree nilsequence ψ. Proposition 4.11.5 shows how we can deal with ψ if it has sufficiently low degree.

Proposition 4.11.5 (Twisted generalized von Neumann's lemma)**.** *Let $2 \leqslant t \leqslant M$. There exists $c_M > 0$ such that for any $\delta_1, \delta_2 > 0$ and any 1-bounded functions $f_0, ..., f_{t-1} : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ satisfying*

$$\min_{0\leqslant i\leqslant t-1} \|f_i\|_{U^{t-1}} \leqslant \delta_1 \quad \text{and} \quad \min_{0\leqslant i\leqslant t-1} \|f_i\|_{U^t} \leqslant \delta_2,$$

the following holds:

1. *if $\psi(x,y) = F(g(x,y)\Gamma)$ is a 1-bounded, N-periodic nilsequence of complexity M and degree $t-2$, then*

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)\cdots f_{t-1}(x+(t-1)y)\psi(x,y) \ll_M \delta_1^{c_M};$$

2. *if $\psi(x,y) = F(g(x,y)\Gamma)$ is a 1-bounded, N-periodic nilsequence of complexity M and degree $t-1$, then*

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)\cdots f_{t-1}(x+(t-1)y)\psi(x,y) \ll_M \delta_2^{c_M}.$$

Before we prove Proposition 4.11.5, we state several definitions and lemmas from Section 3 of [GT12] that we need in the proof. We start with the notions of vertical characters and vertical oscillations.

Definition 4.11.6 (Vertical characters)**.** *Let G/Γ be a filtered nilmanifold of degree s. A* vertical character *is a continuous group homomorphism $\xi : G_s \to \mathbb{R}$ that sends $\xi(\Gamma_s) \subseteq \mathbb{Z}$.*

Definition 4.11.7 (Vertical oscillation)**.** *Let G/Γ be a filtered nilmanifold of degree s. A function $F : G/\Gamma \to \mathbb{C}$ has* vertical oscillation *if there exists a vertical character $\xi : G_s \to \mathbb{R}$ such that $F(g_s x\Gamma) = F(x\Gamma)\xi(g_s)$ for every $g_s \in G_s$.*

We observe that when $s = 1$, the nilmanifold is a torus, and vertical characters correspond to usual characters on the torus whereas functions with vertical oscillations are scaled versions thereof. For higher values of s, one can approximate an arbitrary Lipschitz function on nilmanifolds by linear combinations of Lipschitz functions with vertical oscillations in much the same way as one approximates Lipschitz functions on tori by trigonometric polynomials. This is made precise by the following lemma, the proof of which forms a part of the proofs of Lemmas 3.1 and 3.7 of [GT12], to which the lemma is applied.

Lemma 4.11.8. *Let G/Γ be a filtered nilmanifold of degree s and complexity M, and $F : G/\Gamma \to \mathbb{C}$ be an M-Lipschitz function. There exists $c_M > 0$ such that for every $\delta > 0$, there exists $Q = O_M(\delta^{-c_M})$ and a function*

$$\tilde{F}(x\Gamma) = \sum_{i=1}^{Q} a_i F_i(x\Gamma)$$

such that $\|F - \tilde{F}\|_\infty \leqslant \delta$, and for each $1 \leqslant i \leqslant Q$, the coefficients a_i satisfy $|a_i| \leqslant 1$ while F_i is an M-Lipschitz function with a vertical oscillation and $\|F_i\|_\infty \leqslant \|F\|_\infty$.

Finally, we state the following lemma which is used in the inductive step in the proof of Proposition 4.11.5.

Lemma 4.11.9. *Let $\psi(x, y) = F(g(x, y)\Gamma)$ be a nilsequence of degree s and complexity M, where $F : G/\Gamma \to \mathbb{C}$ has a vertical oscillation. For every h, the function $\tilde{\psi}_h(x, y) = \psi(x, y + h)\overline{\psi(x, y)}$ is a nilsequence of complexity $O_M(1)$ and degree $s - 1$, where the complexity bound is uniform in h.*

A statement similar to Lemma 4.11.9 has been proved in Sections 5 and 7 of [GT12], and it played a crucial role in the argument of that paper. Our proof uses very closely the ideas of [GT12].

Proof. Since the proof uses some of the most technical lemmas of [GT12], we will at times cite relevant lemmas from that paper without stating them here in full form. Without loss of generality, we can assume that $\dim G = M$.

First, we reduce to the case when $g(0, 0) = 1$ at the cost of worsening the Lipschitz norm of F from M to $O_M(1)$. This can be done as follows. By Lemma A.14 of [GT12], each element of G can be split as $x = \{x\}[x]$, where the Mal'cev coordinates of $\{x\}$ lie in $[0, 1)^M$, and $[x] \in \Gamma$ (this decomposition generalises the decomposition of each real number into its fractional and integer part). We then define $\tilde{F}(x\Gamma) = F(\{g(0, 0)\}x\Gamma)$ and

$$\tilde{g}(x, y) = \{g(0, 0)\}^{-1} g(x, y) g(0, 0)^{-1} \{g(0, 0)\} = \{g(0, 0)\}^{-1} g(x, y) [g(0, 0)]^{-1}.$$

We observe that $\tilde{g}(0, 0) = 1$ and $\tilde{F}(\tilde{g}(x, y)\Gamma) = F(g(x, y)\Gamma)$. The last thing to check is the Lipschitz norm of $\tilde{F}$. By Lemma A.5 of [GT12], the metric d on

G/Γ is approximately left-invariant[7], which in our case implies that

$$\begin{aligned}|\tilde{F}(x\Gamma) - \tilde{F}(y\Gamma)| &= |F(\{g(0,0)\}x\Gamma) - F(\{g(0,0)\}y\Gamma)| \\ &\leqslant Md(\{g(0,0)\}x\Gamma, \{g(0,0)\}y\Gamma) \ll_M d(x\Gamma, y\Gamma),\end{aligned}$$

the last step being precisely the consequence of the approximate left-invariance of d and the bound on the size of the Mal'cev coordinates of $\{g(0,0)\}$.

We thus assume from now on that $g(0,0) = 1$, so that g can be expressed as

$$g(x,y) = \prod_{\substack{i,j \geqslant 0,\\ 1 \leqslant i+j \leqslant s}} g_{ij}^{\binom{x}{i}\binom{y}{j}}.$$

We define $g_2(x,y) = g(x,y)g_{01}^{-y}g_{10}^{-x}$ and observe that g_2 takes values in G_2 because

$$g_2(x,y) = g_{10}^{x}g_{01}^{y}g_{01}^{-y}g_{10}^{-x} \mod G_2.$$

We rewrite

$$\begin{aligned}g(x,y+h) &= \{g_{01}^h\}\{g_{01}^h\}^{-1}g_2(x,y+h)g_{10}^{x}g_{01}^{y}\{g_{01}^h\}[g_{01}^h] \\ g(x,y) &= g_2(x,y)g_{10}^{x}g_{01}^{y}.\end{aligned}$$

Defining $\tilde{F}_h(x\Gamma, y\Gamma) = F(\{g_{01}^h\}x\Gamma)\overline{F(y\Gamma)}$ and

$$\tilde{g}_h(x,y) = (\{g_{01}^h\}^{-1}g_2(x,y+h)g_{10}^{x}g_{01}^{y}\{g_{01}^h\},\ g_2(x,y)g_{10}^{x}g_{01}^{y}),$$

we observe that

$$\tilde{\psi}_h(x,y) = \tilde{F}_h(\tilde{g}_h(x,y)\Gamma^2).$$

The bound on the complexity of ψ, approximate left-invariance of the metric on G/Γ (Lemma A.5 of [GT12]) and the fact that the Mal'cev coordinates of $\{g_{01}^h\}$ lie in $[0,1)^M$ imply that $\tilde{F}_h$ is $O_M(1)$-Lipschitz. Since the size of the Mal'cev coordinates of $\{g_{01}^h\}$ is bounded uniformly in h, the bound on the Lipschitz norm of $\tilde{F}_h$ is uniform in h, too.

The crucial observation that makes $\tilde{\psi}_h$ a nilsequence of degree $s-1$ is that

[7]Meaning that if G/Γ has complexity at most M and the Mal'cev coordinates on u are bounded by M, then $d(ux\Gamma, uy\Gamma) \ll_M d(x\Gamma, y\Gamma)$.

the sequence $\tilde{g}_h$ takes values in the subgroup

$$G^{\square} := G \times_{G_2} G = \{(u,v) \in G \times G : uv^{-1} \in G_2\}.$$

This follows from the identity

$$\begin{aligned}&\{g_{01}^h\}^{-1} g_2(x, y+h) g_{10}^x g_{01}^y \{g_{01}^h\} (g_2(x,y) g_{10}^x g_{01}^y)^{-1} \\ &= \{g_{01}^h\}^{-1} g_{10}^x g_{01}^y \{g_{01}^h\} g_{01}^{-y} g_{10}^{-x} \mod G_2 = 1 \mod G_2\end{aligned}$$

which uses the facts that g_2 takes values in G_2 and that G/G_2 is abelian; the latter allows us to commute the terms in the expansion above modulo G_2.

Following the notational convention of [GT12], we set $F_h^{\square} = \tilde{F}_h|_{G^{\square}}$. From the fact that F has a vertical character ξ it follows that for any element $g_s \in G_s$, we have

$$F_h^{\square}(g_s x\Gamma, g_s y\Gamma) = \xi(g_s) F(\{g_{01}^h\} x\Gamma) \overline{\xi(g_s) F(y\Gamma)} = F(\{g_{01}^h\} x\Gamma) \overline{F(y\Gamma)} = F_h^{\square}(x\Gamma, y\Gamma).$$

The function $F_h^{\square}$ is thus invariant under $G_s^{\triangle} = \{(g_s, g_s) \in G_s^2\}$, and so it induces a function $\overline{F_h^{\square}}$ on $\overline{G^{\square}} = G^{\square}/G_s^{\triangle}$. We set $\overline{\Gamma^{\square}} = \Gamma^2/(G_s^{\triangle} \cap \Gamma^2)$. It is the content of Proposition 7.2 of [GT12] that $\overline{G^{\square}}/\overline{\Gamma^{\square}}$ is a filtered nilmanifold of degree $s-1$ with filtration $\overline{G^{\square}}_{\bullet}$ given by $\overline{G_i^{\square}} = G_i \times_{G_{i+1}} G_i / G_s^{\triangle}$. Lemma 7.4 of [GT12] further states that the nilmanifold $\overline{G^{\square}}/\overline{\Gamma^{\square}}$ has complexity $O_M(1)$ and the function $\overline{F_h^{\square}}$ is $O_M(1)$-Lipschitz. Letting $\overline{g_h^{\square}}$ be the projection of $\tilde{g}_h$ onto $\overline{G^{\square}}/\overline{\Gamma^{\square}}$, we conclude that $\tilde{\psi}_h(x,y) = \overline{F_h^{\square}}(\overline{g_h^{\square}}\,\overline{\Gamma^{\square}})$ is indeed a nilsequence of degree $s-1$ and complexity $O_M(1)$ uniformly in h. $\square$

We now have all the tools necessary to prove Proposition 4.11.5.

Proof of Proposition 4.11.5. Proposition 4.11.5 is a variation of Lemma 4.2 of [GT10], and our proof follows very closely the proof of that resuls. We proceed by induction on t. For $t = 2$, the first statement of Proposition 4.11.5 is trivial since ψ, being a 0-step nilsequence, is just a constant.

To prove the second part for $t = 2$, we let $\delta > 0$ be a parameter to be fixed later. The function ψ takes the form $\psi(x,y) = F(\alpha x + \beta y)$ for some 1-bounded, M-Lipschitz function $F : \mathbb{R}^M/\mathbb{Z}^M \to \mathbb{C}$ and $\alpha, \beta \in \left(\frac{1}{N}\mathbb{Z}/\mathbb{Z}\right)^M$. Using Lemma 4.11.8, 1-boundedness of F and the fact that functions with vertical oscillations on the tori are scalar multiples of characters, we can find a 1-bounded trigonometric polynomial $\tilde{F} : \mathbb{R}^M/\mathbb{Z}^M \to \mathbb{C}$ of degree $O_M(\delta^{-C_M})$ satisfying $\|F - \tilde{F}\|_\infty \leqslant \delta$. It then follows from the triangle inequality and the pigeonhole principle that

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)\psi(x,y)\right| \ll_M \delta^{-C_M}\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)e_N(ax+by)\right| + \delta$$

for some $a,b\in\mathbb{Z}$. Applying the Cauchy-Schwarz inequality in x allows us to remove $f_0(x)e_N(ax)$

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)e_N(ax+by)\right|^2 &\leqslant \mathop{\mathbb{E}}_{x\in\mathbb{Z}/N\mathbb{Z}}\left|\mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} f_1(x+y)e_N(by)\right|^2\\ &= \mathop{\mathbb{E}}_{x,y,h_1\in\mathbb{Z}/N\mathbb{Z}} \Delta_{h_1}f_1(x+y)e_N(-bh_1),\end{aligned}$$

and translating $x\mapsto x-y$ followed by an application of the Cauchy-Schwarz inequality in h_1 allows us to bound

$$\left|\mathop{\mathbb{E}}_{h_1\in\mathbb{Z}/N\mathbb{Z}} e_N(-bh_1)\mathop{\mathbb{E}}_{x\in\mathbb{Z}/N\mathbb{Z}}\Delta_{h_1}f_1(x)\right|^2 \leqslant \mathop{\mathbb{E}}_{x,h_1,h_2}\Delta_{h_1,h_2}f(x);$$

thus,

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)e_N(ax+by)\right| \leqslant \|f_1\|_{U^2},$$

and analogues maneuvers also give a bound by $\|f_0\|_{U^2}$. It follows that

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)\psi(x,y)\right| \ll_M \delta^{-C_M}\delta_2+\delta.$$

Letting $\delta=\delta_2^{c_M}$ for a sufficiently small $0<c_M<1$, we obtain the claim.

We now assume $t>2$, and we let $\psi(x,y)=F(g(x,y)\Gamma)$ be a nilsequence of complexity M and degree $s=t-2$ (the argument for $s=t-1$ goes identically). Let $\delta>0$. Using Lemma 4.11.8, we can find a 1-bounded function $\tilde{F}:G/\Gamma\to\mathbb{C}$ that is a linear combination of $O_M(\delta^{-C_M})$ functions with vertical oscillations and satisfies $\|F-\tilde{F}\|_\infty$. Using the triangle inequality and pigeonhole principle, we can find a 1-bounded, M-Lipschitz function $F_1:G/\Gamma\to\mathbb{C}$ with a vertical character $\xi:G_s/\Gamma_s\to\mathbb{C}$ satisfying

$$\begin{aligned}&\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)\cdots f_{t-1}(x+(t-1)y)\psi(x,y)\right|\\ &\ll_M \delta^{-C_M}\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)\cdots f_{t-1}(x+(t-1)y)\psi_1(x,y)\right|+\delta,\end{aligned}$$

where $\psi_1(x,y) = F_1(g(x,y)\Gamma)$.

By the Cauchy-Schwarz inequality and change of variables, we have

$$\begin{aligned}&\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)\cdots f_{t-1}(x+(t-1)y)\psi_1(x,y)\right|^2\\ &\leqslant \left|\mathop{\mathbb{E}}_{x,y,h\in\mathbb{Z}/N\mathbb{Z}} \Delta_h f_1(x+y)\cdots \Delta_{(t-1)h} f_{t-1}(x+(t-1)y)\psi_1(x,y+h)\overline{\psi_1(x,y)}\right|,\end{aligned}$$

where we recall that $\Delta_h f(x) := f(x+h)\overline{f(x)}$. By Lemma 4.11.9, for every $h\in\mathbb{Z}/N\mathbb{Z}$, the function $\tilde{\psi}_h(x,y) = \psi_1(x,y+h)\overline{\psi_1(x,y)}$ is a 1-bounded, N-periodic nilsequence of complexity $O_M(1)$ and degree $s-1$, where the complexity bound is uniform in h. Picking $\delta = \delta_1^{c_M}$ for an appropriate value of $0 < c_M < 1$ and applying the inductive hypothesis, we obtain

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)\cdots f_{t-1}(x+(t-1)y)\psi(x,y)\right| \ll_M \min_{1\leqslant i\leqslant t-1} \mathop{\mathbb{E}}_{h\in\mathbb{Z}/N\mathbb{Z}} \|\Delta_{ih} f_i\|_{U^s}^{c_M}.$$

An application of the Hölder inequality and the recursive definition of Gowers norms give

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)\cdots f_{t-1}(x+(t-1)y)\psi(x,y)\right| \ll_M \min_{1\leqslant i\leqslant t-1} \|f_i\|_{U^{s+1}}^{c'_M}.$$

for some $0 < c'_M < 1$. A slight modification of the argument gives the same bound in terms of $\|f_0\|_{U^{s+1}}$, completing the proof of the lemma. □

Proposition 4.11.5 tells us how to proceed when f_t is a nilsequence. We now prove the general case.

Proposition 4.11.10. *Let $t,d\in\mathbb{N}_+$ satisfy $2\leqslant d\leqslant t-1$. Given any $\varepsilon>0$, there exists $\delta>0$ and $N_0\in\mathbb{N}$ s.t. for all primes $N>N_0$, we have*

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)\cdots f_{t-1}(x+(t-1)y)f_t(x+y^d)\right| \ll \varepsilon$$

uniformly for all 1-bounded functions $f_0,\ldots,f_t:\mathbb{Z}/N\mathbb{Z}\to\mathbb{C}$ such that $\|f_i\|_{U^s}\leqslant\delta$ for some $i\in\{0,\ldots,t-1\}$, where

$$s = \begin{cases} t, \text{ if} & d\mid t-1\\ t-1, \text{ if} & d\nmid t-1.\end{cases}$$

Proof. We fix $\varepsilon>0$, and we let $\delta>0$, $N_0\in\mathbb{N}_+$ and a growth function

$\mathcal{F} : \mathbb{R}_+ \to \mathbb{R}_+$ be chosen later. Suppose that $\min_{0\leqslant i\leqslant t-1} \|f_i\|_{U^s} \leqslant \delta$. By Lemma 4.7.4, there exist $M = O_{\varepsilon,\mathcal{F}}(1)$, a filtered nilmanifold G/Γ of degree $s_0 = \left\lceil \frac{t}{d} \right\rceil - 1$ and complexity at most M, and an N-periodic sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ with $g(0) = 1$, for which there exists a decomposition

$$f_t = f_{nil} + f_{sml} + f_{unf}$$

such that $f_{nil}(n) = F(g(n)\Gamma)$ for an M-Lipschitz function $F : G/\Gamma \to \mathbb{C}$, $\|f_{sml}\|_2 \leqslant \varepsilon$ and $\|f_{unf}\|_{U^{s_0+1}} \leqslant \frac{1}{\mathcal{F}(M)}$. Using the bound on f_{sml}, we crudely evaluate its contribution by

$$\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)\cdots f_{t-1}(x+(t-1)y)f_{sml}(x+y^d) \right| \leqslant \varepsilon. \tag{4.45}$$

To bound the contribution of f_{unf}, we choose $\delta' > 0$ and N_0 that work for ε as in Proposition 4.11.1. We then pick $\mathcal{F}$ to be growing sufficiently fast so that $\|f_{unf}\|_{U^{s_0+1}} \leqslant \delta'$. Assuming that $N > N_0$ and applying Proposition 4.11.1, we have

$$\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)\cdots f_{t-1}(x+(t-1)y)f_{unf}(x+y^d) \right| \ll \varepsilon. \tag{4.46}$$

Finally, we observe that $f_{nil}(x+y^d)$ is an N-periodic nilsequence of complexity M and degree $d\left\lfloor \frac{t-1}{d} \right\rfloor = s-1$. Using Proposition 4.11.5, we choose $\delta > 0$ in such a way as to guarantee that

$$\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)\cdots f_{t-1}(x+(t-1)y)f_{nil}(x+y^d) \right| \leqslant \varepsilon. \tag{4.47}$$

The Proposition follows from combining (4.45), (4.46) and (4.47). □

In the case of $(x,\ x+y,\ x+2y,\ x+y^2)$, Proposition 4.11.10 gives us control by the U^3 norm of the first three functions. It turns out, however, that for this specific example we can get control by the u^3 norm instead. This curious fact shows that Gowers norms need not be the "smalles" norms controlling a given configuration.

Proposition 4.11.11. *Given any $\varepsilon > 0$, there exists $\delta > 0$ and $N_0 \in \mathbb{N}$ s.t. for all primes $N > N_0$, we have*

$$\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)f_2(x+2y)f_3(x+y^2) \right| \ll \varepsilon$$

uniformly for all 1-bounded functions $f_0, f_1, f_2, f_3 : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ *satisfying* $\|f_i\|_{u^3} \leqslant \delta$ *for some* $i \in \{0, 1, 2\}$.

Proof. Let

$$F(x) = \mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} f_0(x-y^2)f_1(x+y-y^2)f_2(x+y-y^2)$$

be the dual function corresponding to this progression. Then

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)f_2(x+2y)f_3(x+y^2)\right|^2 = \left|\mathop{\mathbb{E}}_{x\in\mathbb{Z}/N\mathbb{Z}} F(x)f_3(x)\right|^2$$
$$\leqslant \|F\|_{L^2}^2 = \left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)f_2(x+2y)F(x+y^2)\right|.$$

This trick has allowed us to replace f_3 by the more structured dual function F. From the U^2 inverse theorem and the definition of F, it follows that

$$\|F\|_{U^2}^2 \leqslant \|\hat{F}\|_\infty = \left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)f_2(x+2y)e_N(\beta(x+y^2))\right|$$

for some $\beta \in \mathbb{Z}$. Using the algebraic relation (4.25), we can rewrite the expression above as

$$\|F\|_{U^2}^2 \leqslant \left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} \tilde{f}_0(x)\tilde{f}_1(x+y)\tilde{f}_2(x+2y)\right|,$$

where $\tilde{f}_i(x) = f_i(x)e_N(Q_i(x))$ for some quadratic polynomial Q_i. Since the counting operator for the 3-term arithmetic progression is bounded by the maximum Fourier coefficient of the weights, we deduce that

$$\|F\|_{U^2}^2 \leqslant \min_{i=0,1,2} \|\widehat{\tilde{f}_i}\|_\infty \leqslant \min_{i=0,1,2} \|f_i\|_{u^3},$$

where the last inequality follows from the fact that

$$\|\widehat{\tilde{f}_i}\|_\infty = \max_{\alpha\in\mathbb{Z}} \left|\mathop{\mathbb{E}}_{x\in\mathbb{Z}/N\mathbb{Z}} \tilde{f}_i(x)e_N(\alpha x)\right| \leqslant \max_{\alpha,\beta\in\mathbb{Z}} \left|\mathop{\mathbb{E}}_{x\in\mathbb{Z}/N\mathbb{Z}} f_i(x)e_N(\alpha x+\beta x^2)\right| = \|f_i\|_{u^3}.$$

Let $\varepsilon > 0$. Using Proposition 4.11.1, we choose $\delta > 0$ and a threshold N_0 such that for all primes $N > N_0$, we have

$$\left|\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} f_0(x)f_1(x+y)f_2(x+2y)F(x+y^2)\right| \ll \varepsilon^2$$

whenever $\|F\|_{U^2} \leqslant \delta^2$. Combining everything above, we deduce that if $\min_{i=0,1,2} \|f_i\|_{u^3} \leqslant \delta$, then $\|F\|_{U^2}^2 \leqslant \delta^2$, hence

$$\left| \mathop{\mathbb{E}}_{x,y \in \mathbb{Z}/N\mathbb{Z}} f_0(x) f_1(x+y) f_2(x+2y) F(x+y^2) \right| \leqslant \varepsilon^2,$$

and therefore

$$\left| \mathop{\mathbb{E}}_{x,y \in \mathbb{Z}/N\mathbb{Z}} f_0(x) f_1(x+y) f_2(x+2y) f_3(x+y^2) \right| \ll \varepsilon.$$

□

4.12 The equivalence of Weyl and algebraic complexity

While we are not able to show that Host-Kra and true complexities equal algebraic complexity for inhomogeneous progression, we can show the equivalence of Weyl and algebraic complexities for all integral progressions. We recall that an integral polynomial progression $\vec{P} \in \mathbb{R}[x, y]^{t+1}$ has Weyl complexity s at $0 \leqslant i \leqslant t$ if s the smallest natural number for which the factor $\mathcal{Z}_s$ is characteristic for the weak convergence of $\vec{P}$ at i for any Weyl system. The relative easy with which we can study Weyl complexity as compared to Host-Kra complexity has to do with the rich algebraic structure possessed by Weyl systems, especially the fact that the underlying spaces are abelian Lie groups.

Every disconnected Weyl system can be written as a finite union of isomorphic tori that are cyclically permuted by the transformation T, much the same way as each disconnected nilsystem is a union of connected nilsystems (cf. Proposition 4.2.2 and the remark below Theorem 3.5 of [BLL07]). Therefore we can restrict our attention to connected Weyl systems. These can in turn be reduced to standard Weyl systems, which are totally ergodic by Proposition 4.2.2. Throughout this section, we let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

Definition 4.12.1 (Standard Weyl system of order s). *Let $s \in \mathbb{N}_+$ and $X = \mathbb{T}^s$. A* standard Weyl system of order s *is a system $(X, \mathcal{X}, \mu, T)$, where $\mathcal{X}$ is the Borel σ-algebra on X, μ is the Lebesgue measure, and*

$$T(a_1, ..., a_s) = (a_1 + a_0, a_2 + a_1, ..., a_s + a_{s-1})$$

for some irrational a_0.

Proposition 4.12.2 (Lemma 4.1 of [FK05]). *Each connected Weyl system is* *a factor of a product of several standard Weyl systems.*

Determining Weyl complexity therefore amounts to analysing standard Weyl systems. Since each standard Weyl system is totally ergodic, we immediately deduce the following.

Proposition 4.12.3. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. Then $\mathcal{W}_i(\vec{P}) \leqslant \mathcal{HK}_i(\vec{P})$ for all $0 \leqslant i \leqslant t$.*

We now fix a standard Weyl system $(X, \mathcal{X}, \mu, T)$ of order s with some irrational a_0. Then

$$T^n(a_1, ..., a_s) = \left(a_1 + na_0, a_2 + na_1 + \binom{n}{2}a_0, ..., a_s + na_{s-1} + ... + \binom{n}{s}a_0\right) \tag{4.48}$$

$$= g_0 + g_1 n + ... + g_s \binom{n}{s},$$

where $g_i = (a_{1-i}, ..., a_{s-i})$ and $a_{-k} = 0$ for $k > 0$. For almost all points $a = (a_1, ..., a_s) \in \mathbb{R}^s$, the numbers $1, a_0, ..., a_s$ are rationally independent, and we fix a point $a \in \mathbb{R}^s$ for which this is the case. The sequence $g(n) = T^n a$ is adapted to the filtration $G_i = \{0\}^{i-1} \times \mathbb{R}^{s-i+1}$ for $1 \leqslant i \leqslant s$ and $G_i = 0$ for $i > s$ on $G = G_0 = \mathbb{R}^s$, and it is irrational due to the irrationality of a_0. Since the $\mathcal{Z}_i$ factor of X consists of all the functions whose values depend only on the first i coordinates, we have $Z_i = G/G_{i+1}\Gamma = \mathbb{T}^i \times \{0\}^{s-i}$, where $\Gamma = \mathbb{Z}^s$.

What we aim to show is therefore the following.

Proposition 4.12.4. *Let $t \in \mathbb{N}_+$, $(X, \mathcal{X}, \mu, T)$ be a standard Weyl system of order s and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. Fix $0 \leqslant i \leqslant t$ and suppose that $\mathcal{A}_i(\vec{P}) = s'$. Then the image of the group $\{0\}^i \times G_{s'+1} \times \{0\}^{t-i}$ is contained in the closure of g^P inside $(G/\Gamma)^{t+1}$.*

If $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ is a homogeneous progression, then the sequence g^P is equidistributed in G^P/Γ^P by Theorem 4.5.3, and Proposition 4.12.4 follows immediately; we want to say something about the closure of g^P in the general case. We fix an integral progression $\vec{P}$ for the rest of this section. For each $1 \leqslant i \leqslant s$, we pick linearly independent integral polynomials $Q_{i,1}, ..., Q_{i,t_i'}$ that form a basis for W_i'. We also let $\{R_1, ..., R_r\}$ be a basis for W^c consisting of integral polynomials. Thus,

$$\binom{\vec{P}}{i} = \sum_{j=1}^{t_i'} \vec{v}_{i,j} Q_{i,j} + \sum_{j=1}^{r} \vec{w}_{i,j} R_j$$

for some vectors $\vec{v}_{i,j}, \vec{w}_{i,j} \in \mathbb{Z}^{t+1}$, which follows from (4.28). Consequently,

$$g^P = g_0\vec{1} + \sum_{i=1}^{s} g_i \sum_{j=1}^{t'_i} \vec{v}_{i,j} Q_{i,j} + \sum_{j=1}^{r} \left(\sum_{i=1}^{s} g_i \vec{w}_{i,j} \right) R_j \tag{4.49}$$

We explain the notation used in (4.49). For $h \in G$ and $\vec{v} \in \mathbb{R}^{t+1}$, we interpret $h\vec{v}$ as the element of $(\mathbb{R}^s)^{t+1}$ of the form $(hv(0), ..., hv(t))$, where $hv(i) = (h_1 v(i), ..., h_s v(i))$ is an element of $\mathbb{R}^s$ for each $h = (h_1, ..., h_s) \in \mathbb{R}^s$ and $\vec{v} = (v(0), ..., v(t))$. Thus, $h\vec{v}$ is the same as what we previously called $h^{\vec{v}}$. We use the additive notation $h\vec{v}$ now since we are working in an abelian setting. We also denote $\vec{1} = (1, ..., 1)$.

We let $A_{i,j} = \mathrm{Span}_{\mathbb{R}}\{g_i \vec{v}_{i,j}\}$ and $B_j = \mathrm{Span}_{\mathbb{R}}\{\sum_{i=1}^{s} g_i \vec{w}_{i,j}\}$, and we denote the closure of their images in $(G/\Gamma)^{t+1}$ as $\overline{A}_{i,j}$ and $\overline{B}_j$ respectively. From the rational independence of a_i and the rationality of the entries of $\vec{v}_{i,j}$ and $\vec{w}_{i,j}$, we deduce that nonzero entries of $g_i\vec{v}_{i,j}$ and $\sum_{i=1}^{s} g_i \vec{w}_{i,j}$ are irrational; therefore the sequences $(x, y) \mapsto g_i \vec{v}_{i,j} Q_{i,j}(x, y)$ and $(x, y) \mapsto \sum_{i=1}^{s} g_i \vec{w}_{i,j} R_j(x, y)$ are equidistributed on $\overline{A}_{i,j}$ and $\overline{B}_j$ respectively. The linear independence of $Q_{i,j}, R_j$ then implies the following.

Proposition 4.12.5. *The closure of g^P is the image of $g_0\vec{1} + \tilde{G}$ inside $(G/\Gamma)^{t+1}$, where*

$$\tilde{G} = \sum_{i=1}^{s} \sum_{j=1}^{t'_i} A_{i,j} + \sum_{j=1}^{r} B_j.$$

In particular, the group $\tilde{G}$ contains

$$K = \sum_{i=1}^{s} \sum_{j=1}^{t'_i} A_{i,j} = \mathrm{Span}_{\mathbb{R}}\{h_i \vec{v}_{i,j} : h_i \in G_i, 1 \leqslant i \leqslant s, 1 \leqslant j \leqslant t'_i\}.$$

We observe that $K = \tilde{G} = G^P$ whenever $\vec{P}$ is homogeneous.

Corollary 4.12.6. *Fix $0 \leqslant i \leqslant t$ and let $\mathcal{A}_i(\vec{P}) < s$. For $k \leqslant s$, we have $\{0\}^i \times G_k \times \{0\}^{t-i} \leqslant K$ if and only if $k > \mathcal{A}_i(\vec{P})$.*

Proof. For each $1 \leqslant k \leqslant s$, we let $\mathcal{P}'_k = \mathrm{Span}_{\mathbb{R}}\{\vec{v}_{k,1}, ..., \vec{v}_{k,t'_k}\}$. Thus

$$K = \mathrm{Span}_{\mathbb{R}}\{h_k \vec{u}_k : \ 1 \leqslant k \leqslant s, \ h_k \in G_k, \ \vec{u}_k \in \mathcal{P}'_k\},$$

and so for $k \leqslant s$, we have the inclusion $\{0\}^i \times G_k \times \{0\}^{t-i} \leqslant K$ if and only if the vector $\vec{e}_i$ with 1 in the i-th position and 0 elsewhere is contained in $\mathcal{P}'_k$. The statement $\vec{e}_i \in \mathcal{P}'_k$ is equivalent to the inclusion $\binom{x+P_i(y)}{k} \in W'_k$. This is in turn equivalent to the statement that there are no algebraic relations of the form (4.6) with $\deg Q_i = k$, which is precisely the condition that $k > \mathcal{A}_i(\vec{P})$. $\square$

Corollary 4.12.7. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression. Then $\mathcal{W}_i(\vec{P}) \leqslant \mathcal{A}_i(\vec{P})$ for each $0 \leqslant i \leqslant t$.*

We finish this section by showing the converse.

Proposition 4.12.8. *Let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral polynomial progression for which $\mathcal{A}_i(\vec{P}) = s$ for some $0 \leqslant i \leqslant t$. Then for any standard Weyl system $(X, \mathcal{X}, \mu, T)$ of order s there exist smooth functions $f_0, ..., f_t : X \to \mathbb{C}$ such that $\mathbb{E}(f_i|\mathcal{Z}_{s-1}) = 0$ but the expression (4.29) is 1. In particular, $\mathcal{W}_i(\vec{P}) \geqslant s$.*

Before we prove Proposition 4.12.8, we define $\partial Q(x) = Q(x+1) - Q(x)$ for $Q \in \mathbb{R}[x]$. From the identity $\partial\binom{x}{k} = \binom{x+1}{k} - \binom{x}{k} = \binom{x}{k-1}$ we deduce that

$$\partial\left(a_0 + a_1\binom{x}{1} + ... + a_d\binom{x}{d}\right) = a_1 + a_2\binom{x}{1} + ... + a_d\binom{x}{d-1}.$$

Proof of Proposition 4.12.8. Let T be as in (4.48) for some irrational a_0. From $\mathcal{A}_i(\vec{P}) = s$ it follows that $\vec{P}$ satisfies an algebraic relation (4.6) with $\deg Q_i = s$. For each $0 \leqslant k \leqslant t$, we let $Q_k(u) = b_{k,1}u + ... + b_{k,s}\binom{u}{s}$. We define $\xi(u) = e(\alpha u)$ for some irrational α, and we let

$$f_k(a_1, ..., a_s) = \xi(b_{k,1}a_1 + ... + b_{k,s}a_s).$$

Thus, we have

$$f_k(T^{x+P_k(y)}a) = \xi(a_0 Q_k(x+P_k(y)) + a_1\partial Q_k(x+P_k(y)) + ... + a_s\partial^s Q_k(x+P_k(y))),$$

and so

$$\prod_{i=0}^{t} f_i(T^{x+P_i(y)}a) = \xi\left(\sum_{j=0}^{s} a_j\partial^j \sum_{k=0}^{t} Q_k(x+P_k(y))\right) = 1.$$

On the other hand, we have

$$|\,\mathbb{E}(f_i|\mathcal{Z}_{s-1})(a_1, ..., a_s)| = \left|\int_{\mathbb{T}} f_i(a_1, ..., a_s)da_s\right| = \left|\int_{\mathbb{T}} \xi(b_{i,s}a_s)da_s\right| = 0$$

for Lebesgue-a.e. a_s. $\square$

4.13 The proof of Theorem 4.1.18

We conclude the paper with the proof of Theorem 4.1.18; this is the statement to the effect that all progressions of complexity at most 1 satisfying the algebraic condition (4.10) possess many popular common differences and have good lower bounds for multiple recurrence. Throughout this section, we let $t \in \mathbb{N}_+$ and $\vec{P} \in \mathbb{R}[x,y]^{t+1}$ be an integral progression of algebraic complexity at most 1. We also let $Q_1, ..., Q_k$ be integral polynomials as in the statement of Theorem 4.1.18. Thus, $P_i = \sum_j a_{ij} Q_j$ and $Q_i = \sum_j a'_{ij} P_j$ for $a_{ij}, a'_{ij} \in \mathbb{Z}$. The second part of the theorem follows from the first part and the Furstenberg Correspondence Principle. We therefore proceed to prove part (i), followed by part (iii). Our argument for part (i) follows closely the proof of Theorem C of [Fra08].

Proof of Theorem 4.1.18(i). Suppose that $(X, \mathcal{X}, \mu, T)$ is a totally ergodic system with the Kronecker factor $(Z_1, \mathcal{Z}_1, \nu, S)$. The space Z_1 can be assumed to be a connected compact abelian group with an ergodic translation $Sx = x + b$. For each $\delta > 0$, let B_δ be the δ-neighbourhood of the identity in Z_1, and let

$$\tilde{B}_\delta = \{n \in \mathbb{N} : Q_1(n)b, ..., Q_k(n)b \in B_\delta\}.$$

It follows from the ergodicity of S and linear independence of $Q_1, ..., Q_k$ that

$$\lim_{N-M\to\infty} \frac{|\tilde{B}_\delta \cap [M,N)|}{N-M} = \nu(B_\delta)^k > 0$$

for any $\delta > 0$. In particular, $\tilde{B}_\delta$ is syndetic for any $\delta > 0$, otherwise we would have $\liminf_{N-M\to\infty} \frac{|\tilde{B}_\delta \cap [M,N)|}{N-M} = 0$.

We aim to show that for any $A \in \mathcal{X}$ with $\mu(A) > 0$ and any $\varepsilon > 0$, we have

$$\lim_{N-M\to\infty} \mathop{\mathbb{E}}_{n \in \tilde{B}_\delta \cap [M,N)} \mu(A \cap T^{P_1(n)}A \cap \cdots \cap T^{P_t(n)}A) \geqslant \mu(A)^{t+1} - \varepsilon \qquad (4.50)$$

for all sufficiently small $\delta > 0$. This implies part (i) of Theorem 4.1.18 as follows: if there is a sequence K_N of intervals in $\mathbb{N}$ of length converging to infinity, with the property that

$$\mu(A \cap T^{P_1(n)}A \cap \cdots \cap T^{P_t(n)}A) < \mu(A)^{t+1} - \varepsilon \qquad (4.51)$$

for all $n \in \bigcup_{N\in\mathbb{N}} K_N$, then the sets $\tilde{K}_N = K_N \cap \tilde{B}_\delta$ are nonempty for all sufficiently large N due to the syndecticity of B_δ (in fact, their cardinalities also converge to infinity). Since (4.51) holds for all $n \in \bigcup_{N\in\mathbb{N}} \tilde{K}_N$, the inequality (4.50) fails, leading to a contradiction.

We first show that if $\mathbb{E}(f_i|\mathcal{Z}_1) = 0$, then

$$\lim_{N-M\to\infty} \mathop{\mathbb{E}}_{n\in[M,N)} 1_{\tilde{B}_\delta}(n) \prod_{i=1}^{t} T^{P_i(n)} f_i = 0 \tag{4.52}$$

in L^2 for any $f_1, ..., f_t \in L^\infty(\mu)$. From the measurability of B_δ it follows that we can approximate $1_{\tilde{B}_\delta}(n) = \prod_{i=1}^k 1_{B_\delta}(Q_i(n)b)$ arbitrarily well by linear combinations of $\prod_{i=1}^k \xi_i(Q_i(n)b)$ for some characters $\xi_1, ..., \xi_k$ on Z_1. Using the fact that each Q_i is an integral linear combination of $P_1, ..., P_t$, we can rewrite $\prod_{i=1}^k \xi_i(Q_i(n)b) = \prod_{i=1}^t \tilde{\xi}_i(P_i(n)b)$ for some characters $\tilde{\xi}_1, ..., \tilde{\xi}_t$.

In effect, it suffices to show that

$$\lim_{N-M\to\infty} \mathop{\mathbb{E}}_{n\in[M,N)} \prod_{i=1}^{t} \tilde{\xi}_i(P_i(n)b) \prod_{i=1}^{t} T^{P_i(n)} f_i = 0. \tag{4.53}$$

We can rephrase the limit in (4.53) as

$$\lim_{N-M\to\infty} \prod_{i=1}^{t} \tilde{\xi}_i(-y) \mathop{\mathbb{E}}_{n\in[M,N)} \prod_{i=1}^{t} R^{P_i(n)}(f_i(x)\tilde{\xi}_i(y)), \tag{4.54}$$

where $R = T\times S$. Let $(R_t)_t$ be the ergodic components of R and $(f_i\otimes\xi_i)(x,y) = f_i(x)\xi_i(y)$; then $\mathbb{E}(f_i\otimes\xi_i|\mathcal{Z}_1(R_t)) = 0$ whenever $\mathbb{E}(f_i|\mathcal{Z}_1(T)) = 0$ for $(\mu\times\nu)$-a.e. t. It thus follows from Corollary 4.1.15 that if $\mathbb{E}(f_i|\mathcal{Z}_1) = 0$ for some i, then the limit in (4.54) is 0, which proves the claim.

We therefore deduce that

$$\begin{aligned}\lim_{N-M\to\infty} \mathop{\mathbb{E}}_{n\in\tilde{B}_\delta\cap[M,N)} \int_X \prod_{i=0}^{t} T^{P_i(n)} 1_A d\mu &= \lim_{N-M\to\infty} \mathop{\mathbb{E}}_{n\in\tilde{B}_\delta\cap[M,N)} \int_{Z_1} \prod_{i=0}^{t} S^{P_i(n)} \tilde{1}_A d\nu \\ &= \lim_{N-M\to\infty} \mathop{\mathbb{E}}_{n\in\tilde{B}_\delta\cap[M,N)} \int_{Z_1} \prod_{i=0}^{t} S^{\sum_j a_{ij}Q_j(n)} \tilde{1}_A d\nu,\end{aligned} \tag{4.55}$$

where $\tilde{1}_A = \mathbb{E}(1_A|\mathcal{Z}_1)$. Due to the ergodicity of S and the linear independence of $Q_1, ..., Q_k$, the limit in (4.55) equals

$$\frac{1}{\nu(B_\delta)^k} \int_{B_\delta^k} \int_{Z_1} \prod_{i=0}^{t} \tilde{1}_A(x + \sum_j a_{ij}y_j) d\nu(x) d\nu^k(y). \tag{4.56}$$

In the limit $\delta \to 0$, the expression in (4.56) converges to $\int_{Z_1} (\tilde{1}_A)^{t+1}$; hence for every $\varepsilon > 0$ and sufficiently small $\delta > 0$, we have

$$\frac{1}{\nu(B_\delta)^k} \int_{B_\delta^k} \int_{Z_1} \prod_{i=0}^{t} \tilde{1}_A(x + \sum_j a_{ij} y_j) d\nu(x) d\nu^k(y) \geqslant \int_{Z_1} (\tilde{1}_A)^{t+1} - \varepsilon. \tag{4.57}$$

Using Hölder inequality, we obtain that $\int_{Z_1} (\tilde{1}_A)^{t+1} \geqslant (\int_{Z_1} \tilde{1}_A)^{t+1} = \mu(A)^{t+1}$, which implies (4.50).

This finishes the totally ergodic case; the derivation of the ergodic case from the totally ergodic one proceeds in the same way as in the proof of Theorem C of [Fra08]. The identity (4.52) still holds as its proof only assumes the ergodicity of the map. We therefore still have

$$\lim_{N-M\to\infty} \mathop{\mathbb{E}}_{n\in\tilde{B}_\delta\cap[M,N)} \int_X \prod_{i=0}^{t} T^{P_i(n)} 1_A d\mu = \lim_{N-M\to\infty} \mathop{\mathbb{E}}_{n\in\tilde{B}_\delta\cap[M,N)} \int_{Z_1} \prod_{i=0}^{t} S^{P_i(n)} \tilde{1}_A d\nu$$

like in (4.55), except that now the Kronecker factor $(Z_1, \mathcal{Z}_1, \nu, S)$ is assumed to be ergodic rather than totally ergodic. Since a Kronecker factor is isomorphic to a rotation on a compact abelian Lie group, we can take $Z_1 = Z \times \mathbb{Z}/r\mathbb{Z}$ for some $r \in \mathbb{N}_+$, where Z is a connected compact abelian Lie group. Since the map S^r is a totally ergodic rotation on each connected component of $Z \times \mathbb{Z}/r\mathbb{Z}$, we consider every connected component $Z \times \{j\}$ separately, with the polynomials P_i replaced on each $Z \times \{j\}$ by $\tilde{P}_{i,j}(n) = \frac{P_i(r(n-1)+j)-P_i(j)}{r}$. An easy-to-verify fact that the progression

$$\vec{\tilde{P}}_j(x, y) = (x,\ x + \tilde{P}_{1,j}(y),\ \ldots,\ x + \tilde{P}_{t,j}(y)),$$

still satisfies the condition (4.10) ensures that we can now employ the argument in the totally ergodic case to each $\vec{\tilde{P}}_j$ in place of $\vec{P}$, and this finishes the proof of the ergodic case.

$\square$

We now proceed to the proof of part (iii) of Theorem 4.1.18. The argument can below can be seen as a finitary version of the argument above, with all the necessary modifications coming from working in the finitary setting. It follows the proof of the 3-term arithmetic progression case in Theorem 1.12 of [GT10].

Proof of Theorem 4.1.18(iii). Let $\alpha, \varepsilon > 0$, and suppose that $A \subseteq \mathbb{Z}/N\mathbb{Z}$ has size $|A| \geqslant \alpha N$ for a prime $N > N_0(\alpha, \varepsilon)$. Let $\mathcal{F} : \mathbb{R}_+ \to \mathbb{R}_+$ be a growth function to be specified later. By Theorem 5.1 of [CS12], the irrational and periodic version of the celebrated arithmetic regularity lemma of Green and

Tao (Theorem 1.2 of [GT10]), there exists a positive number $M = O_{\varepsilon,\mathcal{F}}(1)$ and a decomposition

$$1_A = f_{nil} + f_{sml} + f_{unf} \tag{4.58}$$

into 1-bounded functions such that

1. $f_{nil} = F(g(n)\Gamma)$ is an $\mathcal{F}(M)$-irrational, N-periodic nilsequence of degree 1 and complexity M;
2. $||f_{sml}||_1 \leqslant \varepsilon$;
3. $||f_{unf}||_{U^2} \leqslant \frac{1}{\mathcal{F}(M)}$.

Moreover, f_{nil} takes values in $[0,1]$. Unpacking the definition of f_{nil}, we see that $F : (\mathbb{R}/\mathbb{Z})^m \to [0,1]$ is M-Lipschitz, $1 \leqslant m \leqslant M$, and $g(n) = bn$ for some $\mathcal{F}(M)$-irrational element $b \in (\frac{1}{N}\mathbb{Z}/\mathbb{Z})^m$.

Our strategy is as follows. We shall define a weight $\tilde{\mu} : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}_{\geqslant 0}$ which satisfies

$$\mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) = 1 + O(\varepsilon) \tag{4.59}$$

and

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) \prod_{i=0}^{t} 1_A(x + P_i(y)) \geqslant \alpha^{t+1} - O(\varepsilon). \tag{4.60}$$

Using pigeonhole principle and (4.59), it can be deduced from (4.60) that for $\Omega_{\alpha,\varepsilon}(N)$ values of y, we have

$$\mathop{\mathbb{E}}_{x\in\mathbb{Z}/N\mathbb{Z}} \prod_{i=0}^{t} 1_A(x + P_i(y)) \geqslant \alpha^{t+1} - O(\varepsilon),$$

which proves part (iii) of Theorem 4.1.18.

We shall prove (4.60) by splitting each 1_A using (4.58) and showing that terms involving f_{sml} or f_{unf} have contributions at most $O(\varepsilon)$ while the term

$$\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) \prod_{i=0}^{t} f_{nil}(x + P_i(y)) \tag{4.61}$$

has size at least $\alpha^{t+1} - O(\varepsilon)$. Showing that the terms involving f_{sml} or f_{unf} make negligible contributions to (4.60) is akin to showing (4.52) for all functions with $\mathbb{E}(f_i|\mathcal{Z}_1) = 0$ in the proof of part (i) of Theorem 4.1.18. In doing so, we shall use the idea that while we fix $\varepsilon > 0$, we have control over how fast we choose $\mathcal{F}$ to grow - and we choose it to grow fast enough depending on α and ε to ensure that all the estimates work.

Let $\delta > 0$ be fixed later. We define $\psi : (\mathbb{R}/\mathbb{Z})^m \to \mathbb{R}_+$ to be a nonnegative, 1-bounded, $O_M(\delta^{-1})$-Lipschitz function that is 1 on $[-\frac{1}{4}\delta, \frac{1}{4}\delta]^m$ and 0 outside $[-\frac{1}{2}\delta, \frac{1}{2}\delta]^m$. We let $c = \int_{(\mathbb{R}/\mathbb{Z})^m} \psi$; thus $(\frac{1}{2}\delta)^m \leqslant c \leqslant \delta^m$. We then let $\mu(y) = \frac{\psi(by)}{c}$. Since b can be picked without the loss of generality from $[0,1]^m$, the function μ is $O_M(\delta^{-M-1})$-Lipschitz.

We let $\tilde{\mu}(y) = \mu(Q_1(y)) \cdots \mu(Q_k(y))$. It is a weight that picks out all the values y for which $Q_1(y)b$, ..., $Q_k(y)b$ are close to being an integer, and it plays a similar role as the function $1_{\tilde{B}_\delta}$ in the proof of part (i) of Theorem 4.1.18, except that it is constructed using a Lipschitz function rather than an indicator function. To show (4.59), we observe that

$$\mathop{\mathbb{E}}_{y \in \mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) = \frac{1}{c^k} \mathop{\mathbb{E}}_{y \in [N]} \prod_{i=1}^{k} \psi(bQ_i(y)). \tag{4.62}$$

Using the $\mathcal{F}(M)$-irrationality of g, linear independence of Q_1, ..., Q_k as well as Theorem 4.2.5, we deduce that (4.62) equals

$$\frac{1}{c^k}\left(\left(\int \psi\right)^k + O_M(\delta^{-1}\mathcal{F}(M)^{-c_M})\right) = 1 + O_M(\delta^{-M-1}\mathcal{F}(M)^{-c_M})$$

for some $c_M > 0$. The estimate (4.59) follows from choosing $\mathcal{F}$ growing fast enough depending on δ and picking $\delta = c'_M\varepsilon$ for an appropriately chosen $c'_M > 0$.

We decompose each 1_A in (4.60) using (4.58) and split (4.60) into 3^t accordingly using multilinearity. We first estimate (4.61), and subsequently we bound contributions of f_{sml} and f_{unf}.

Taking $\mathcal{F}$ growing fast enough, we assume that $||f_{unf}||_{U^2} \leqslant \varepsilon$, and thus $|\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_{unf}(x)| = ||f_{unf}||_{U^1} \leqslant ||f_{unf}||_{U^2} \leqslant \varepsilon$. From Hölder inequality and the bound on the L^1 norm of f_{sml}, we obtain a bound $|\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_{sml}| \leqslant \varepsilon$. From these bounds and (4.58) we deduce that $\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_{nil}(x) \geqslant \alpha - 2\varepsilon$.

We observe that by M-Lipschitzness of F and the definitions of μ, $\tilde{\mu}$ and Q_j, we have $f_{nil}(x + P_i(y)) = f_{nil}(x + \sum_j a_{ij}Q_j(y)) = f_{nil}(x) + O_M(\delta) =$

$f_{nil}(x) + O(\varepsilon)$ whenever $\tilde{\mu}(y) > 0$. It follows from this that

$$\begin{aligned} &\mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) \prod_{i=0}^{t} f_{nil}(x + \sum_j a_{ij} Q_j(y)) \\ &= \left(\mathop{\mathbb{E}}_{x\in\mathbb{Z}/N\mathbb{Z}} f_{nil}(x)^{t+1} + O(\varepsilon) \right) \mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y). \end{aligned} \tag{4.63}$$

Using the estimate for (4.59) and Hölder inequality, we deduce that (4.63) is bounded from below by

$$\left(\mathop{\mathbb{E}}_{x\in\mathbb{Z}/N\mathbb{Z}} f_{nil}(x) \right)^{t+1} - O(\varepsilon) \geqslant \alpha^{t+1} - O(\varepsilon),$$

where the last inequality follows from Hölder inequality.

We now bound terms involving f_{sml}. Suppose without loss of generality that f_{sml} is in the $i = 0$ position, and let $f_1, ..., f_t \in \{f_{nil}, f_{sml}, f_{unf}\}$. Then

$$\left| \mathop{\mathbb{E}}_{x,y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) f_{sml}(x) \prod_{i=1}^{t} f_i(x + P_i(y)) \right| \leqslant ||f_{sml}||_1 \mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) \leqslant \varepsilon, \tag{4.64}$$

where the first inequality follows from Hölder inequality, positivity of $\tilde{\mu}$ and 1-boundedness of $f_1, ..., f_t$.

It remains to bound the contributions of f_{unf}. Using a standard argument (see e.g. the proof of Proposition 3.1 of [GT12]), we want to approximate f_{unf} by a trigonometric polynomial, which allows us to essentially replace f_{unf} by additive characters. Let $K \in \mathbb{N}_+$ be fixed later. Since μ is an $O_M(\varepsilon^{-M})$-Lipschitz function, there exists a trigonometric polynomial $\mu_1 : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ such that $||\mu - \mu_1||_\infty \ll_M \varepsilon^{-C_M^{(1)}} K^{-c}$ for some $0 < c, C_M^{(1)}$. Moreover, μ_1 has degree at most K^M and its coefficients satisfy $||\widehat{\mu_1}||_\infty \leqslant ||\mu||_\infty \ll_M \varepsilon^{-M}$.

Let $f_0, ..., f_t \in \{f_{nil}, f_{sml}, f_{unf}\}$, with at least one of them being f_{unf}. We then bound

$$\left| \mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) \prod_{i=0}^{t} f_i(x + P_i(y)) \right| = \left| \mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} \prod_{i=1}^{k} \mu(Q_i(y)) \prod_{i=0}^{t} f_i(x + P_i(y)) \right| \tag{4.65}$$

$$\leqslant k \max(||\mu||_\infty, ||\mu_1||_\infty)^{k-1} ||\mu - \mu_1||_\infty + \left| \mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} \prod_{i=1}^{k} \mu_1(Q_i(y)) \prod_{i=0}^{t} f_i(x + P_i(y)) \right|.$$

The first term has size at most $C_M^{(2)} \varepsilon^{-C_M^{(2)}} K^{-c}$ for some $C_M^{(2)} > 0$. The second term is bounded by

$$K^M ||\widehat{\mu_1}||_\infty \cdot \left| \mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} \prod_{i=1}^{k} \xi_i(Q_i(y)) \prod_{i=1}^{t} f_i(x+P_i(y)) \right| \tag{4.66}$$

for some characters ξ_i on $\mathbb{Z}/N\mathbb{Z}$. Since each Q_i is an integral linear combination of P_i's, we can rewrite $\prod_{i=1}^{k} \xi_i(Q_i(y)) = \prod_{i=1}^{t} \tilde{\xi}_i(x+P_i(y))$. We let $\tilde{f}_i = f_i\tilde{\xi}_i$. Since each $\tilde{\xi}_i$ is a linear character, we have $||f_i||_{U^2} = ||\tilde{f}_i||_{U^2}$ for each i.

We recall from Theorem 4.1.12 that $\vec{P}$ has true complexity 1. Combining this fact with (4.65), (4.66) and the bound $||\tilde{f}_i||_{U^2} \leqslant 1/\mathcal{F}(M)$ for some i, we deduce that there is some increasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$, depending only on $\vec{P}$, such that

$$\left| \mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) \prod_{i=0}^{t} f_i(x+P_i(y)) \right| \leqslant C_M^{(2)} \varepsilon^{-C_M^{(2)}} K^{-c} + C_M^{(2)} \varepsilon^{-M} K^M \omega(1/\mathcal{F}(M)), \tag{4.67}$$

increasing the constant $C_M^{(2)}$ if necessary. We note that the existence of ω is equivalent to the statement that $\vec{P}$ is controlled by U^2 at i. We now show that we can choose K large enough and $\mathcal{F}$ growing fast enough so that the right-hand side of (4.67) is bounded by $O(\varepsilon)$.

For any given M, we find a constant $C_M^{(3)}$ such that $(C_M^{(3)})^c \geqslant C_M^{(2)}$ and $cC_M^{(3)} - C_M^{(2)} \geqslant 1$. We then let $K_M = C_m^{(3)} \varepsilon^{-C_M^{(3)}}$, so that

$$C_M^{(2)} \varepsilon^{-C_M^{(2)}} K_M^{-c} = C_M^{(2)} {C_M^{(3)}}^{-c} \varepsilon^{cC_M^{(3)} - C_M^{(2)}} \leqslant \varepsilon.$$

Picking $\mathcal{F}$ growing sufficiently fast depending on ε, we can ensure that

$$C_M^{(2)} \varepsilon^{-M} K_M^M \omega(1/\mathcal{F}(M)) \leqslant \varepsilon.$$

We thus set $K = K_M$ for the value of M induced by ε and $\mathcal{F}$, and so

$$\left| \mathop{\mathbb{E}}_{y\in\mathbb{Z}/N\mathbb{Z}} \tilde{\mu}(y) \prod_{i=0}^{t} f_i(x+P_i(y)) \right| \leqslant 2\varepsilon.$$

□

References

[Alt21] D. Altman. "On a conjecture of Gowers and Wolf". In: *ArXiv e-prints* (2021). arXiv: 2106.15437.

[BHK05] V. Bergelson, B. Host, and B. Kra. “Multiple recurrence and nilsequences”. In: *Invent. Math.* **160** (2005). With an appendix by I. Ruzsa, pp. 261–303.

[BL96] V. Bergelson and A. Leibman. “Polynomial extensions of van der Waerden’s and Szemerédi’s theorems”. In: *J. Amer. Math. Soc.* **9** (1996), pp. 725–753.

[BLL07] V. Bergelson, A. Leibman, and E. Lesigne. “Complexities of finite families of polynomials, Weyl systems, and constructions in combinatorial number theory”. In: *J. Anal. Math.* **103** (2007), pp. 47–92.

[CS12] P. Candela and O. Sisask. “Convergence results for systems of linear forms on cyclic groups and periodic nilsequences”. In: *SIAM J. Discrete Math.* **28** (2012), pp. 786–810.

[FK05] N. Frantzikinakis and B. Kra. “Polynomial averages converge to the product of integrals”. In: *Israel J. Math.* **148**.1 (2005), pp. 267–276.

[FK06] N. Frantzikinakis and B. Kra. “Ergodic averages for independent polynomials and applications”. In: *J. Lond. Math. Soc.* **74** (2006), pp. 131–142.

[Fra08] N. Frantzikinakis. “Multiple ergodic averages for three polynomials and applications”. In: *Trans. Amer. Math. Soc.* **360**.10 (2008), pp. 5435–5475.

[Fra16] N. Frantzikinakis. “Some open problems on multiple ergodic averages”. In: *Bull. Hellenic Math. Soc.* **60** (2016), pp. 41–90.

[GT10] B. Green and T. Tao. “An arithmetic regularity lemma, an associated counting lemma, and applications”. In: *An irregular mind. Szemerédi is 70*. Vol. **21**. Bolyai Soc. Math. Stud., 2010, pp. 261–334.

[GT12] B. Green and T. Tao. “The quantitative behaviour of polynomial orbits on nilmanifolds”. In: *Ann. of Math.* **175** (2 2012), pp. 465–540.

[GTZ11] B. Green, T. Tao, and T. Ziegler. “An inverse theorem for the Gowers U^4 norm”. In: *Glasg. Math. J.* **53**.1 (2011), pp. 1–50.

[GTZ12] B. Green, T. Tao, and T. Ziegler. “An inverse theorem for the Gowers $U^{s+1}[N]$-norm”. In: *Ann. of Math.* **176**.2 (2012), pp. 1231–1372.

[GW10] W. T. Gowers and J. Wolf. “The true complexity of a system of linear equations”. In: *Proc. Lond. Math. Soc.* **100**.1 (2010), pp. 155–176.

[GW11a] W. T. Gowers and J. Wolf. “Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$”. In: *Geom. Funct. Anal.* **21** (2011), pp. 36–69.

[GW11b] W. T. Gowers and J. Wolf. “Linear forms and quadratic uniformity for functions on $\mathbb{F}_p^n$”. In: *Mathematika* **57** (2 2011), pp. 215–237.

[GW11c] W. T. Gowers and J. Wolf. “Linear forms and quadratic uniformity for functions on $\mathbb{Z}_N$”. In: *J. Anal. Math.* **115**.1 (2011), pp. 121–186.

[HK05a] B. Host and B. Kra. “Convergence of polynomial ergodic averages”. In: *Israel J. Math.* **149**.1 (2005), pp. 1–19.

[HK05b] B. Host and B. Kra. “Nonconventional ergodic averages and nilmanifolds”. In: *Ann. of Math.* **161**.1 (2005), pp. 397–488.

[HK18] B. Host and B. Kra. *Nilpotent structures in ergodic theory.* AMS, 2018.

[Kuc21a] B. Kuca. “Further bounds in the polynomial Szemerédi theorem over finite fields”. In: *Acta Arith.* **198** (2021), pp. 77–108.

[Kuc21b] B. Kuca. “True complexity of polynomial progressions in finite fields”. In: *Proc. Edinb. Math. Soc.* (2021), pp. 1–53.

[Lei05a] A. Leibman. “Convergence of multiple ergodic averages along polynomials of several variables”. In: *Israel J. Math.* **146** (2005), pp. 303–315.

[Lei05b] A. Leibman. “Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold”. In: *Ergodic Theory Dynam. Systems* **25**.1 (2005), pp. 201–213.

[Lei09] A. Leibman. “Orbit of the diagonal in the power of a nilmanifold”. In: *Trans. Amer. Math. Soc.* **362**.03 (2009), pp. 1619–1658.

[Man18] F. Manners. “Good bounds in certain systems of true complexity 1”. In: *Discrete Anal.* **21** (2018), 40 pp.

[Man21] F. Manners. “True complexity and iterated Cauchy-Schwarz”. In: *ArXiv e-prints* (2021). arXiv: `2109.05731`.

[Pel19] S. Peluse. “On the polynomial Szemerédi theorem in finite fields”. In: *Duke Math. J.* **168**.5 (2019), pp. 749–774.

5 Conclusion

In Chapter 1, we have stated known results for questions listed in Chapter 1; in Chapters 2-4, we presented in detail our own contributions. Yet these questions are far from being resolved. In this concluding chapter, we indicate possible future research directions together with some obstacles that need to be overcome for future advances.

A significant portion of our results are composed of new bounds in the (multidimensional) polynomial Szemerédi theorem over finite fields. All known results for nonlinear configuration [Pel19; Pel20; PP19; PP20; Kuc21a; Kuc21b] make essential use of the fact that progressions studied in these papers have complexity 0, i.e. at least some polynomials in these progressions are linearly independent from others. Obtaining any reasonable bounds for an arbitrary polynomial progression, whether over finite fields or over natural numbers, is well outside the scope of existing methods and likely requires significant new ideas. Given that higher-degree Gowers norms are closely linked with higher-step nilsequences, it is rather plausible that to get anywhere close to the general case, one has to generalise arguments from [Pel19; Pel20; PP19; PP20; Pre; Kuc21a] by replacing purely Fourier analytic tools with more robust methods based on the higher order Fourier analysis. This seems unavoidable even for simply-looking configurations such as the progression

$$(x,\ x+y,\ x+2y,\ x+y^2),$$

discussed on so many occasions in this thesis. This progression is a natural starting point for such efforts, and it is one of the projects to which we hope to return in the future.

Conjecture 1.3.7 is another challenging problem towards which partial answers have been provided in this thesis. While trying to resolve this conjecture for inhomogeneous progressions, it has become clear to us that one either has to come up with a very different method of handling such questions, or one needs to obtain a much better understanding of the algebraic relations of the form

$$Q_0(x) + Q_1(x + P_1(y)) + ... + Q_t(x + P_t(y)) = 0$$

for inhomogeneous progressions than we currently possess. Because of the many intricate ways in which these relations show up in the problem, we find it likely that any attempt at resolving this conjecture might require a significant amount of algebraic geometry; this is something that we have successfully avoided for homogeneous progressions.

Finally, there remain the open problems related to the multidimensional version of the polynomial Szemerédi theorem: finding bounds in the multidimensional polynomial Szemerédi theorem on the combinatorial side of things (Question 1.1.19) and understanding the multiple recurrence properties and convergence of ergodic averages of several transformations on the ergodic side (Question 1.1.23). There are a number of technical issues that arise when dealing with multidimensional configurations but do not appear in the single-dimensional case. We present one of them here. We recall that our bounds in Chapter 3 hold under the assumption of polynomials having distinct degrees, whereas the analogous single-dimensional result of Peluse [Pel19] only requires polynomials to be linearly independent. This has to do with the fact that whereas single-dimensional progressions are controlled by Gowers norms, multidimensional progressions are in general only controlled by a broader class of *box norms*, such as the norm

$$\|f\| = \left(\mathop{\mathbb{E}}_{\substack{x_1, x_2,\\ h_1, h_2 \in \mathbb{F}_p}} f(x_1, x_2)\overline{f(x_1+h_1, x_2) f(x_1, x_2+h_2)} f(x_1+h_1, x_2+h_2)\right)^{\frac{1}{4}}.$$

More generally, a box norm of $f : \mathbb{F}_p^D \to \mathbb{C}$ in the direction of vectors $\mathbf{v}_1, ..., \mathbf{v}_s \in \mathbb{Z}^D$ is given by

$$\|f\| = \left(\mathop{\mathbb{E}}_{\substack{\mathbf{x} \in \mathbb{F}_p^D,\\ h_1, ..., h_s \in \mathbb{F}_p}} \prod_{w \in \{0,1\}^s} \mathcal{C}^{|w|} f(\mathbf{x} + h_1\mathbf{v}_1 + ... + h_s\mathbf{v}_s)\right)^{\frac{1}{2^s}}.$$

If $\mathbf{v}_1 = ... = \mathbf{v}_s = \mathbf{v}$, then this is a directional Gowers norm along $\mathbf{v}$ defined in Chapter 3, but if not all the vectors are identical up to scaling, then we get a rather different object.

Multidimensional polynomial progressions with polynomials of distinct degrees can still be controlled by (directional) Gowers norms, but if some of the polynomials have the same degrees, then one likely cannot avoid dealing with box norms. Whereas the inverse theory of Gowers norms is relatively well developed in the many works of Gowers [Gow01], Green, Tao, Ziegler [GT08; GTZ11; GTZ12] and Manners [Man18], only few results are known for box norms [Mil21], and therefore the necessity to deal with box norms is a difficulty in its own right.

A similar obstacle appears while trying to find a "nice" characteristic factor for the convergence of ergodic averages of several transformations. The main idea behind the study of ergodic averages of single transformations is that one can reduce the question of convergence of an average over an arbitrary ergodic system to an equidistribution problem over nilsystems. This is a consequence of two facts thoroughly explained in previous chapters: that Host-Kra factors are characteristic for the convergence of such averages, and that Host-Kra factors are inverse limits of nilsystems. We have indicated how this reduction is done in Chapter 4 when deriving the Host-Kra complexity of homogeneous progressions. The reduction-to-nilsystems argument is however in general not possible in the study of ergodic averages of several commuting transformations. This has to do with the fact that Host-Kra factors are in general not characteristic for the convergence of averages of several transformations, which is closely connected with the fact that Gowers norms do not control arbitrary multidimensional polynomial progressions.

Our knowledge is suboptimal even for the averages of several transformations for which we expect Host-Kra factors to be characteristic. We briefly consider averages of the form

$$\mathop{\mathbb{E}}_{n\in[N]} T_1^{P_1(n)}f_1\cdots T_t^{P_t(n)}f_t, \tag{5.1}$$

where $T_1, ..., T_t$ are commuting transformations on a probability space $(X, \mathcal{X}, \mu)$, the polynomials $P_1, ..., P_t \in \mathbb{Z}[n]$ have distinct degrees, and $f_1, ..., f_t \in L^\infty(\mu)$. We expect the L^2 limit of (5.1) to be 0 whenever $\mathbb{E}(f_i|\mathcal{K}_{rat}(T_i)) = 0$ for any $1 \leqslant i \leqslant t$; yet only weaker versions of this statement have been proved before [CFH11]. Proving that $\mathcal{K}_{rat}(T_i)$ is characteristic for the L^2 convergence of (5.1) at the index i seems to be the most accessible open problem in the study of convergence of averages of several commuting transformations, and we hope to attack it in the future.

The abovementioned list by no means exhausts research directions that one can take to further our understanding of the polynomial Szemerédi theorem; nor does it exhaust all the pitfalls that await a mathematician willing to take up the challenge. But it does give a taste of stumbling blocks that prevent us from making significant progress towards answering the questions listed in Chapter 1, and it gives examples of problems that should be within the reach of existing methods. We hope to tackle at least some of these problems in the future.

References

[CFH11] Q. Chu, N. Frantzikinakis, and B. Host. "Ergodic averages of commuting transformations with distinct degree polynomial iterates". In: *Proc. Lond. Math. Soc.* **102** (2011), pp. 801–842.

[Gow01] W. T. Gowers. "A new proof of Szemerédi's theorem". In: *Geom. Funct. Anal.* **11**.3 (2001), pp. 465–588.

[GT08] B. Green and T. Tao. "An inverse theorem for the Gowers $U^3(G)$ norm". In: *Proc. Edinb. Math. Soc.* **51**.1 (2008), pp. 73–153.

[GTZ11] B. Green, T. Tao, and T. Ziegler. "An inverse theorem for the Gowers U^4 norm". In: *Glasg. Math. J.* **53**.1 (2011), pp. 1–50.

[GTZ12] B. Green, T. Tao, and T. Ziegler. "An inverse theorem for the Gowers $U^{s+1}[N]$-norm". In: *Ann. of Math.* **176**.2 (2012), pp. 1231–1372.

[Kuc21a] B. Kuca. "Further bounds in the polynomial Szemerédi theorem over finite fields". In: *Acta Arith.* **198** (2021), pp. 77–108.

[Kuc21b] B. Kuca. "Multidimensional polynomial Szemerédi theorem in finite fields for distinct-degree polynomials". In: *ArXiv e-prints* (2021). arXiv: `2103.12606`.

[Man18] F. Manners. "Good bounds in certain systems of true complexity 1". In: *Discrete Anal.* **21** (2018), 40 pp.

[Mil21] L. Milićević. "An inverse theorem for certain directional Gowers uniformity norms". In: *ArXiv e-prints* (2021). arXiv: `2103.06354`.

[Pel19] S. Peluse. "On the polynomial Szemerédi theorem in finite fields". In: *Duke Math. J.* **168**.5 (2019), pp. 749–774.

[Pel20] S. Peluse. "Bounds for sets with no polynomial progressions". In: *Forum Math. Pi* **8** (e16 2020).

[PP19] S. Peluse and S. Prendiville. "Quantitative bounds in the non-linear Roth theorem". In: *ArXiv e-prints* (2019). arXiv: `1903.02592`.

[PP20] S. Peluse and S. Prendiville. "A polylogarithmic bound in the nonlinear Roth theorem". In: *Int. Math. Res. Nov. IMRN* (2020). rnaa261.

[Pre] S. Prendiville. *The inverse theorem for the nonlinear Roth configuration: an exposition*. arXiv: `2003.04121`.