MULTIPLE ERGODIC AVERAGES FOR VARIABLE POLYNOMIALS

ANDREAS KOUTSOGIANNIS

Abstract. In this paper we study multiple ergodic averages for "good" variable polynomials. In particular, under an additional assumption, we show that these averages converge to the expected limit, making progress related to an open problem posted by Frantzikinakis ([14, Problem 10]). Corresponding averages along prime numbers are studied too. These general convergence results imply various variable extensions of classical recurrence, combinatorial and number theoretical results which are presented as well.

Dedicated to the loving memory of Aris Deligiannis, a great mentor.

Contents

2010 *Mathematics Subject Classification.* Primary: 37A45; Secondary: 37A05, 05D10, 11B25, 11B83.
Key words and phrases. Variable polynomial sequences, multiple ergodic averages, multiple recurrence, characteristic factors, equidistribution, nilmanifolds, prime numbers, sublinear functions.

1. Introduction

The study of multiple ergodic averages along polynomials dates back to 1977, where Furstenberg, exploiting the (L^2) limiting behavior,[1] as $N \to \infty$, of

$$\frac{1}{N}\sum_{n=1}^{N} T^n f_1 \cdot T^{2n} f_2 \cdot \ldots \cdot T^{\ell n} f_\ell, \tag{1}$$

where $\ell \in \mathbb{N}$, $(X, \mathcal{B}, \mu, T)$ is a measure preserving system,[2] and $f_1, \ldots, f_\ell \in L^\infty(\mu)$, provided (in [20]) a purely ergodic theoretic proof of Szemerédi's theorem (i.e., every subset of natural numbers of positive upper density[3] contains arbitrarily long arithmetic progressions–result that can be immediately obtained by combining Theorem 2.3 with Theorem 2.5 below).

It was Bergelson who first visualized the iterates $n, 2n, \ldots, \ell n$ in (1) as "distinct enough" polynomials and studied (initially in [4]), via the use of van der Corput's lemma which is a crucial tool in "reducing the complexity" of the iterates (see Lemma 5.5 below for a variable variation of it), averages of the form

$$\frac{1}{N}\sum_{n=1}^{N} T^{p_1(n)} f_1 \cdot \ldots \cdot T^{p_\ell(n)} f_\ell, \tag{2}$$

for essentially distinct integer polynomials $p_1, \ldots, p_\ell$,[4] which, eventually, led to multidimensional polynomial extensions of Szemerédi's theorem (see [7]).

Bergelson and Leibman conjectured (in [6]) that multiple ergodic averages of the form

$$\frac{1}{N}\sum_{n=1}^{N} T_1^{p_1(n)} f_1 \cdot \ldots \cdot T_\ell^{p_\ell(n)} f_\ell, \tag{3}$$

in any system, for multiple commuting T_i's (i.e., $T_iT_j = T_jT_i$ for all i, j) and arbitrary integer polynomials p_i, always have limit (as $N \to \infty$). This conjecture was answered in the positive by Walsh (in [36]), who showed it in greater generality, namely, for products of transformations (i.e., $\mathbb{Z}^\ell$-actions) which generate a nilpotent group, averaging along Følner sequences in $\mathbb{Z}$.[5] Alas, no specific expression of the limit was provided by the method.

One of the questions that someone is called upon to answer is under which conditions on the polynomials and/ or the system we can explicitly find the limit of the aforementioned expressions. In this direction we have only a few recent results; we present some of them here. For a weakly mixing T[6] Furstenberg ([20]) showed that (1), and later Bergelson ([4]) that (2), converges to $\prod_{i=1}^{\ell} \int f_i \, d\mu$; we will refer to this limit as the "expected" one. Frantzikinakis and Kra (in [19]) proved that for a totally ergodic T (i.e., T^n is ergodic for all $n \in \mathbb{N}$) and independent integer polynomials $p_1, \ldots, p_\ell$ (i.e., every non-trivial linear

---

[1] All the limits in this article are taken with respect to the L^2 norm, except otherwise stated.

[2] I.e., $T : X \to X$ is an invertible measure preserving transformation on the probability space $(X, \mathcal{B}, \mu)$.

[3] For a set $A \subseteq \mathbb{N}$ we define its *upper density*, $\bar{d}(A)$, as $\bar{d}(A) := \limsup_{N\to\infty} \frac{|A\cap\{1,\ldots,N\}|}{N}$.

[4] A $p \in \mathbb{Q}[t]$ is called *integer polynomial* if $p(\mathbb{Z}) \subseteq \mathbb{Z}$. The non-constant integer polynomials $p_1, \ldots, p_\ell$, $\ell \in \mathbb{N}$, are called *essentially distinct* if $p_i - p_j$ is non-constant for all $i \neq j$.

[5] A *Følner sequence in* $\mathbb{Z}$ is a sequence $(\Phi_n)_n$ of finite subsets of $\mathbb{Z}$ such that for any $m \in \mathbb{Z}$ we have $\lim_{n\to\infty} \frac{|(\Phi_n+m)\cap\Phi_n|}{|\Phi_n|} = 1$. A special such sequence, that we will deal with, is $\Phi_n = \{1, \ldots, n\}$, $n \in \mathbb{N}$.

[6] T is *ergodic* if $T^{-1}A = A$ implies $\mu(A) \in \{0, 1\}$. T is *weakly mixing* if $T \times T$ is ergodic.

combination of the p_i's with scalars from $\mathbb{Z}$ is non-constant), (2) converges to the expected limit as well. For the multiple transformations case, when all T_i's are weakly mixing and the polynomials are of positive distinct degrees, Chu-Frantzikinakis-Host (in [8]) showed that (3) converges to the expected limit too, result that we extended in [30] for iterates of the form $([p_i(n)])_n$, for $p_i \in \mathbb{R}[t]$, where $[\cdot]$ denotes the floor function.[7]

Host and Kra (in [26]), developing the theory of characteristic factors, obtained an explicit expression of the limit of (1) in a general system. Analogously to that, Austin (in [1, 2]), studying (3), found precise characteristic factors for some specific cases of quadratic polynomials for $\ell = 2$ (and linear polynomials for $\ell = 3$). Together with Donoso and Sun, exploiting a result by Tao and Ziegler ([35]) on concatenation of factors, we studied (in [10]) expressions (even more general than (3), namely for $\mathbb{Z}^\ell$-actions) with essentially distinct integer polynomials (of multiple variables).

Showing that the characteristic factor coincides with the nilfactor of the system and exploiting the equidistribution property of the corresponding polynomial sequence in nilmanifolds, Frantzikinakis proved (in [13]) that the expression

$$\frac{1}{N}\sum_{n=1}^{N} T^{[p(n)]}f_1 \cdot T^{2[p(n)]}f_2 \cdot \ldots \cdot T^{\ell[p(n)]}f_\ell, \tag{4}$$

where $p \in \mathbb{R}[t]$ with $p(t) \neq cq(t) + d$, $c, d \in \mathbb{R}$, $q \in \mathbb{Q}[t]$, has in any system, the same limit (as $N \to \infty$) as (1); obtaining a refinement of Szemerédi's theorem. Generalizing the condition to multiple polynomials, following Frantzikinakis' approach, we showed with Karageorgos (in [27]) that for strongly independent real polynomials $p_1, \ldots, p_\ell$ (i.e., any non-trivial linear combination of the p_i's with scalars from $\mathbb{R}$ has at least one non-constant irrational coefficient) the expression

$$\frac{1}{N}\sum_{n=1}^{N} T^{[p_1(n)]}f_1 \cdot \ldots \cdot T^{[p_\ell(n)]}f_\ell, \tag{5}$$

has the expected limit.[8]

Generalizing the previous definition for sequences of real variable polynomials, i.e., of the form $(p_N)_N \subseteq \mathbb{R}[t]$, we have the following:

Definition 1.1 ([14])**.** The sequence $(p_N)_N$, where $p_N \in \mathbb{R}[t], N \in \mathbb{N}$, is *good* if the polynomials have bounded degree and for every non-zero $\alpha \in \mathbb{R}$ we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} e^{ip_N(n)\alpha} = 0. \tag{6}$$

[7]Letting $\lceil x \rceil$ and $[[x]]$ denote the ceiling and the closest integer functions respectively, using the relations $\lceil x \rceil = -[-x]$ and $[[x]] = [x + 1/2]$, we see that this last result remains true if the $[\cdot]$'s are individually and independently replaced by any rounding function. Similar results for Hardy field and tempered functions can be found in [11, 13] and [31] respectively.

[8]Here we mean, as it was stated before, in case T is ergodic, that the limit is equal to $\prod_{i=1}^{\ell} \int f_i \, d\mu$, whereas, in the general case, it's equal to $\prod_{i=1}^{\ell} \mathbb{E}(f_i|\mathcal{I}(T))$, where $\mathbb{E}(f_i|\mathcal{I}(T))$ is the conditional expectation of f_i with respect to the σ-algebra of the T-invariant sets. Actually, in (5) one can use any combination of the rounding functions $[\cdot]$ and $[[\cdot]]$, or even replace each $[\cdot]$ by $\lceil \cdot \rceil$.

The sequence of ℓ-tuples of variable polynomials $(p_{1,N},\ldots,p_{\ell,N})_N$, where $p_{i,N}\in\mathbb{R}[t]$, $N\in\mathbb{N}$, $1\leq i\leq \ell$, is *good* if every non-trivial linear combination of the sequences $(p_{1,N})_N,\ldots,(p_{\ell,N})_N$ is good.[9]

For this class of polynomial sequences, Frantzikinakis stated the following problem:

Problem 1 (Problem 10, [14])**.** *Let $(p_{1,N},\ldots,p_{\ell,N})_N$ be a good ℓ-tuple of variable polynomials. Then, for every ergodic system $(X,\mathcal{B},\mu,T)$ and functions $f_1,\ldots,f_\ell\in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{[p_{1,N}(n)]}f_1\cdot\ldots\cdot T^{[p_{\ell,N}(n)]}f_\ell=\prod_{i=1}^{\ell}\int f_i\,d\mu,$$

where the convergence takes place in $L^2(\mu)$.

Remark 1.2 ([14])**.** *If $p_N(t)=\sum_{k=1}^{d}c_{k,N}t^k$, $c_{k,N}\in\mathbb{R}$, $N\in\mathbb{N}$, then $(p_N)_N$ is good iff for any $a\in\mathbb{R}\setminus\{0\}$ we have*

$$\lim_{N\to\infty}N^j\left\|c_{j,N}a\right\|=\infty\ \textit{ for some }\ 1\leq j\leq d, \tag{7}$$

where $\|\cdot\|$ denotes the distance to the closest integer, i.e., $\|x\|:=d(x,\mathbb{Z})$.

As it is stated in [14], Problem 1 is interesting even in the following special cases (for a proof that the following sequences are good, see Lemma 8.1):

Example 1 ([14])**.** For $\ell=2$, the pair $(p_{1,N},p_{2,N})_N$, where $p_{1,N}(n)=n/N^a$, $p_{2,N}(n)=n/N^b$, $N,n\in\mathbb{N}$, $0<a<b<1$, is good.

Example 2 ([14])**.** For $\ell\in\mathbb{N}$, the ℓ-tuple $(p_{1,N},\ldots,p_{\ell,N})_N$, where $p_{i,N}(n)=n^i/N^a$, $1\leq i\leq\ell$, $N,n\in\mathbb{N}$, $0<a<1$, is good.

Showing that (4) has the same limit as (1) for $p\in\mathbb{R}[t]$ with $p(t)\neq cq(t)+d$, $c,d\in\mathbb{R}$, $q\in\mathbb{Q}[t]$, which follows from [13, Theorem 2.2], one naturally states the following problem, which is the good-variable-polynomial-version of Frantzikinakis' result:

Problem 2. *Let $(p_N)_N$ be a good polynomial sequence. Then, for every $\ell\in\mathbb{N}$, system $(X,\mathcal{B},\mu,T)$ and $f_1,\ldots,f_\ell\in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{[p_N(n)]}f_1\cdot T^{2[p_N(n)]}f_2\cdot\ldots\cdot T^{\ell[p_N(n)]}f_\ell=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{n}f_1\cdot T^{2n}f_2\cdot\ldots\cdot T^{\ell n}f_\ell,$$

where the convergence takes place in $L^2(\mu)$.

After stating Problem 1, Frantzikinakis mentions that its $\ell=1$ case (which also coincides with the $\ell=1$ case of Problem 2) can be obtained by using the spectral theorem. As this case can be proved without postulating any additional assumptions on the coefficients of the good variable polynomial sequence $(p_N)_N$, and it reflects its strong equidistribution behavior, we present, for reasons of completeness, its proof. In particular, in Section 3 we show:

[9] Notice that this definition, via Weyl's equidistribution theorem, for independent of N polynomials, characterizes the "strongly independence" notion from [27].

Theorem 1.3. *If $(p_N)_N$ is a good sequence of polynomials, $(X,\mathcal{B},\mu,T)$ an ergodic system and $f\in L^2(\mu)$, then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{[p_N(n)]}f=\int f\,d\mu,$$

where the convergence takes place in $L^2(\mu)$.[10]

For general $\ell\in\mathbb{N}$, we also make progress towards the validation of both Problems 1 and 2. In particular, under some additional assumptions on the coefficients of the good variable polynomials, we show two general results, Theorem 2.1 and Theorem 2.2 (see next section). In this introductory section, we will present an easier application of each of them, which still covers both Examples 1 and 2.

To this end, we first recall the set of sublinear logarithmico-exponential Hardy field functions (of polynomial degree 0) which converge to ($\pm$) infinity:[11]

$$\mathcal{SLE}:=\{g\in\mathcal{LE}:\ 1\prec g(x)\prec x\}.^{12}$$

Next, we define an appropriate set of functions, from which we will get our variable sequences of coefficients:

$$\mathcal{C}:=\left\{\sum_{i=1}^{l}\frac{\rho_i}{g_i}:\ \rho_i\in\mathbb{R},\ g_i\in\mathcal{SLE},\ 1\leq i\leq l,\ l\in\mathbb{N},\ \text{with } g_l\prec\ldots\prec g_1\right\}.$$

Extending the definition from [27], we say that the sequence of ℓ-tuple of variable polynomials $(p_{1,N},\ldots,p_{\ell,N})_N$, where for each $1\leq i\leq\ell$, $p_{i,N}$ has the form:

$$p_{i,N}(n)=a_{i,d_i,N}n^{d_i}+\cdots+a_{i,1,N}n+a_{i,0,N},\tag{8}$$

with $(a_{i,0,N})_N$ bounded, and $a_{i,j,\cdot}\in\mathcal{C}$, $1\leq j\leq d_i$, is *strongly independent* if for any $(\lambda_1,\ldots,\lambda_\ell)\in\mathbb{R}^\ell\setminus\{\vec{0}\}$ we have that $\sum_{i=1}^{\ell}\lambda_i p_{i,N}(n)$ is a non-constant polynomial in n.

For example, the following tuple of variable polynomials

$$\left(\left(\frac{\sqrt{2}}{N^{1/2}}+\frac{1}{\log N\cdot\log\log N}\right)n^3-\frac{31}{\log N}n+1,\frac{1}{N^{1/2}}n^3,\left(\frac{\sqrt{3}}{N^{e/3}}-\frac{17}{N^{\pi/4}\log^{3/2}N}\right)n^2\right)_N$$

is strongly independent.

For Problem 1, i.e., multiple variable polynomial sequences, we have:

[10]Of course here someone can use $\lceil\cdot\rceil$ or $[[\cdot]]$ instead of $[\cdot]$.

[11]Let R be the collection of equivalence classes of real valued functions defined on some halfline (c,∞), $c\geq 0$, where two functions that agree eventually are identified. These equivalence classes are called *germs* of functions. A *Hardy field* is a subfield of the ring $(R,+,\cdot)$ that is closed under differentiation. Here, we use the word *function* when we refer to elements of R (understanding that all the operations defined and statements made for elements of R are considered only for sufficiently large values of $x\in\mathbb{R}$). We say that g is a *logarithmico-exponential Hardy field function*, and we write $g\in\mathcal{LE}$, if it belongs to a Hardy field of real valued functions and it's defined on some $(c,+\infty)$, $c\geq 0$, by a finite combination of symbols $+,-,\times,\div,\sqrt[n]{\cdot},\exp,\log$ acting on the real variable x and on real constants (for more on Hardy field functions, and in particular for logarithmico-exponential ones, one can check [11] and [13]).

[12]We write $g_2\prec g_1$ if $|g_1(x)|/|g_2(x)|\to\infty$ as $x\to\infty$.

Theorem 1.4. *For $\ell \in \mathbb{N}$, let $(p_{1,N},\dots,p_{\ell,N})_N$ be a strongly independent ℓ-tuple of polynomials of the form* (8)*. Then, for every ergodic system $(X,\mathcal{B},\mu,T)$ and $f_1,\dots,f_\ell \in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} T^{[p_{1,N}(n)]}f_1\cdot\ldots\cdot T^{[p_{\ell,N}(n)]}f_\ell=\prod_{i=1}^{\ell}\int f_i\;d\mu,$$

where the convergence takes place in $L^2(\mu)$.

It is true that strongly independent sequences of variable polynomials as in (8) are good (for this, see Lemma 8.1). It is also easy to check that the sequences $(p_{1,N},p_{2,N})_N$, with $p_{1,N}(n)=n/N^a$, $p_{2,N}(n)=n/N^b$, $0<a<b<1$, and $(q_{1,N},\dots,q_{\ell,N})_N$, with $q_{i,N}(n)=n^i/N^a$, $1\leq i\leq \ell$, and $0<a<1$, are strongly independent (here $g_1(N):=N^a\prec g_2(N):=N^b\in\mathcal{SLE}$), hence Theorem 1.4 indeed covers both Examples 1 and 2.

For Problem 2, i.e., a single variable polynomial sequence, we have the following:

Theorem 1.5. *Let $(p_N)_N$ be a non-constant polynomial sequence of the form* (8)*. Then, for every $\ell\in\mathbb{N}$, system $(X,\mathcal{B},\mu,T)$ and $f_1,\dots,f_\ell\in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{[p_N(n)]}f_1\cdot T^{2[p_N(n)]}f_2\cdot\ldots\cdot T^{\ell[p_N(n)]}f_\ell=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{n}f_1\cdot T^{2n}f_2\cdot\ldots\cdot T^{\ell n}f_\ell,$$

where the convergence takes place in $L^2(\mu)$.

Following arguments from [18] and [29], we get the corresponding to Theorems 1.4 and 1.5 results along prime numbers (see also the more general Theorems 2.16 and 2.17, and Subsection 2.4 for recurrence results along shifted primes). To the best of our knowledge, these are the very first results in the literature of this nature (i.e., for variable iterates).

Theorem 1.6. *For $\ell\in\mathbb{N}$, let $(q_{1,N},\dots,q_{\ell,N})_N$ be a strongly independent ℓ-tuple of polynomials of the form* (8)*. Then, for every ergodic system $(X,\mathcal{B},\mu,T)$ and $f_1,\dots,f_\ell\in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty}\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\cap[1,N]}T^{[q_{1,N}(p)]}f_1\cdot\ldots\cdot T^{[q_{\ell,N}(p)]}f_\ell=\prod_{i=1}^{\ell}\int f_i\;d\mu,$$

where the convergence takes place in $L^2(\mu)$ and $\pi(N)=\mathbb{P}\cap[1,N]$ denotes the prime numbers up to N.

Theorem 1.7. *Let $(q_N)_N$ be a non-constant polynomial sequence of the form* (8)*. Then, for every $\ell\in\mathbb{N}$, system $(X,\mathcal{B},\mu,T)$ and $f_1,\dots,f_\ell\in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty}\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\cap[1,N]}T^{[q_N(p)]}f_1\cdot T^{2[q_N(p)]}f_2\cdot\ldots\cdot T^{\ell[q_N(p)]}f_\ell=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{\ell}T^{in}f_i,$$

where the convergence takes place in $L^2(\mu)$.

As we previously highlighted, very few convergence results for ergodic averages with polynomial iterates, in which we can explicitly find the limit, exist. Generally, results for variable polynomials are extremely sparse. We will conclude this introduction, mentioning some of them. Kifer (in [28]) studied multiple averages for variable polynomials of the form

$p_{i,N}(n) = p_i(n) + q_i(N)$, with p_i's essentially distinct and $q_i(\mathbb{Z}) \subseteq \mathbb{Z}$ for T weakly mixing, and, for more general polynomials, for T strongly mixing "enough". Finally, Frantzikinakis (in [13]) found characteristic factors for averages with variable polynomial iterates $p_{i,N}$, with, independent of N, leading coefficients. It is the arguments from this last part that we will adapt, in order to find characteristic factors for our averages as well, which is one of the main two ingredients of the proof (with the second one being the equidistribution of particular sequences for which we adapt arguments from [12]).

Notation. With $\mathbb{P}$, $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ we denote the set of prime, natural, integer, rational and real numbers respectively. For a measurable function f on a measure space X with a transformation $T : X \to X$, we denote with Tf the composition $f \circ T$. For $s \in \mathbb{N}$, $\mathbb{T}^s = \mathbb{R}^s/\mathbb{Z}^s$ denotes the s dimensional torus, and $(a(n))_n$ denotes a sequence indexed over the natural numbers (i.e., $(a(n))_{n\in\mathbb{N}}$). Finally, for two non-negative quantities a, b we write $a \ll b$, if there exists a positive constant C such that $a \leq Cb$.

2. General results and applications

In this section we will state our most general results together with some applications of them, which are variable extensions of classical results. For the proofs of these implications, we follow [11] and [27], adapting the corresponding arguments to the variable polynomial case.

We first cover Problem 1 for a subclass of good polynomial sequences:

Theorem 2.1. *For $\ell \in \mathbb{N}$, let $(p_{1,N}, \ldots, p_{\ell,N})_N$ be a good and super nice ℓ-tuple of polynomials.*[13] *Then, for every ergodic system $(X, \mathcal{B}, \mu, T)$ and $f_1, \ldots, f_\ell \in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{[p_{1,N}(n)]} f_1 \cdot \ldots \cdot T^{[p_{\ell,N}(n)]} f_\ell = \prod_{i=1}^{\ell} \int f_i \, d\mu, \tag{9}$$

where the convergence takes place in $L^2(\mu)$.

For a single good variable polynomial sequence, we also cover the following case of Problem 2:

Theorem 2.2. *Let $(p_N)_N \subseteq \mathbb{R}[t]$ be a good polynomial sequence such that, for all $\ell \in \mathbb{N}$, $(p_N, 2p_N, \ldots, \ell p_N)_N$ is super nice. Then, for every $\ell \in \mathbb{N}$, system $(X, \mathcal{B}, \mu, T)$ and $f_1, \ldots, f_\ell \in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{[p_N(n)]} f_1 \cdot T^{2[p_N(n)]} f_2 \cdot \ldots \cdot T^{\ell[p_N(n)]} f_\ell = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{\ell} T^{in} f_i, \tag{10}$$

where the convergence takes place in $L^2(\mu)$.

We will show that Theorem 2.1 implies Theorem 1.4 (resp. Theorem 2.2 implies Theorem 1.5), hence it covers Examples 1 and 2, and that also holds for any polynomial family $\{p_1, \ldots, p_\ell\}$ (resp. $\{p, 2p, \ldots, \ell p\}$ for Theorem 2.2) which is independent of N and non-trivial linear combinations of its members satisfy (6). In particular it implies [27, Theorem 2.1] for strongly independent polynomials (and of course the same is true for

[13]The "super niceness" property is a technical one which is defined in Definition 5.12 (via Definition 5.6).

Theorem 2.2 for a single polynomial $p \in \mathbb{R}[t]$, any non-zero scalar multiple of which has at least one non-constant irrational coefficient).[14]

The approach we follow to show these results is similar to the one in [13] and [27], with a few extra twists as the variable case is trickier to deal with. Namely, one has to find the characteristic factor of (9) and (10) (which turns out in both cases to be equal to the nilfactor of the system–see Proposition 5.14 and 5.15) and show some equidistribution results in nilmanifolds (mainly see Lemma 6.5 and Theorem 6.2). The "super niceness" property will be introduced so we can deal with the characteristic factor part, while the "goodness" property implies the equidistribution one.

As it was mentioned in the previous section, the ergodicity assumption in Theorem 2.1 can be dropped.[15] Hence, our main theorems hold for any system. The strong nature of these results is also reflected in the fact that they have immediate recurrence and combinatorial implications.

To have some specific example in mind, one can imagine the following results under the assumptions of Theorems 1.4 and 1.5, i.e., for variable polynomials of the form (8).

2.1. **Single sequence.** We first deal with a single variable polynomial sequence, assuming the validity of Theorem 2.2.

2.1.1. *Recurrence.* It is Furstenberg Multiple Recurrence Theorem that will help us obtain recurrence results:

Theorem 2.3 (Furstenberg Multiple Recurrence Theorem, [20])**.** *Let $(X, \mathcal{B}, \mu, T)$ be a system. Then, for any $\ell \in \mathbb{N}$ and $A \in \mathcal{B}$ with $\mu(A) > 0$, we have*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-\ell n}A) > 0.$$ [16]

Theorem 2.2, via Theorem 2.3, implies the following:

Corollary 2.4. *Let $(p_N)_N \subseteq \mathbb{R}[t]$ as in Theorem 2.2. Then, for every $\ell \in \mathbb{N}$, system $(X, \mathcal{B}, \mu, T)$, and $A \in \mathcal{B}$ with $\mu(A) > 0$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(A \cap T^{-[p_N(n)]}A \cap T^{-2[p_N(n)]}A \cap \ldots \cap T^{-\ell[p_N(n)]}A\right) > 0.$$

Proof. Using Theorem 2.2 with $f_i = \mathbf{1}_A$, $1 \leq i \leq \ell$ (i.e., the characteristic function of A), together with Theorem 2.3, implies

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigcap_{i=0}^{\ell} T^{-i[p_N(n)]}A\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigcap_{i=0}^{\ell} T^{-in}A\right) > 0,$$

[14] This will be justified later via the van der Corput operation, defined in Subsection 5.2. Also, both Theorems 2.1 and 2.2 remain true under the use of any combination of the rounding functions $[\cdot]$ and $[[\cdot]]$, or even if we replace each $[\cdot]$ by $\lceil\cdot\rceil$, as addition by 1/2 or multiplication by -1 respectively (using in the second case T^{-1} instead of T) doesn't alter the properties of the (variable) polynomials.

[15] The limit in this case is equal to $\prod_{i=1}^{\ell} \mathbb{E}(f_i|\mathcal{I}(T))$. Indeed, if $\mu = \int \mu_t \; d\lambda(t)$ denotes the ergodic decomposition of μ, it suffices to show that if $\mathbb{E}(f_i|\mathcal{I}(T)) = 0$ for some i then the averages converge to 0. Since $\mathbb{E}(f_i|\mathcal{I}(T)) = 0$, we have that $\int f_i \; d\mu_t = 0$ for λ-a.e. t. By (9), we have that the averages go to 0 in $L^2(\mu_t)$ for λ-a.e. t, hence the limit is equal to 0 in $L^2(\mu)$.

[16] This lim inf, as we mentioned before, is actually a limit.

which is exactly what we wanted to show. □

2.1.2. *Combinatorics.* Via Furstenberg Correspondence Principle, one gets combinatorial results from recurrence ones. We present here a reformulation of this principle from [3].

Theorem 2.5 (Furstenberg Correspondence Principle, [20], [3])**.** *Let E be a subset of integers. There exists a system $(X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{B}$ with $\mu(A) = \bar{d}(E)$ such that*

$$\bar{d}(E \cap (E - n_1) \cap \ldots \cap (E - n_\ell)) \geq \mu(A \cap T^{-n_1} A \cap \ldots \cap T^{-n_\ell} A) \tag{11}$$

for every $\ell \in \mathbb{N}$ and $n_1, \ldots, n_\ell \in \mathbb{Z}$.

Using Corollary 2.4 and Theorem 2.5, we have the following variable combinatorial result:

Corollary 2.6. *Let $(p_N)_N \subseteq \mathbb{R}[t]$ as in Theorem 2.2. Then, for every $\ell \in \mathbb{N}$ and $E \subseteq \mathbb{N}$ with $\bar{d}(E) > 0$, we have*

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \bar{d}\left(E \cap (E - [p_N(n)]) \cap (E - 2[p_N(n)]) \cap \ldots \cap (E - \ell[p_N(n)])\right) > 0.$$

Proof. For $E \subseteq \mathbb{N}$ with $\bar{d}(E) > 0$, let $(X, \mathcal{B}, \mu, T)$ system and $A \in \mathcal{B}$ with $\mu(A) = \bar{d}(E) > 0$ that satisfies (11). Using Corollary 2.4 we get

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \bar{d}\left(\bigcap_{i=0}^{\ell} (E - i[p_N(n)])\right) \geq \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigcap_{i=0}^{\ell} T^{-i[p_N(n)]} A\right) > 0,$$

as was to be shown. □

Hence, we immediately get the following refinement of Szemerédi's theorem:

Corollary 2.7. *Let $(p_N)_N \subseteq \mathbb{R}[t]$ as in Theorem 2.2. Then, for every $\ell \in \mathbb{N}$, any $E \subseteq \mathbb{N}$ with $\bar{d}(E) > 0$ contains arithmetic progressions of the form:*

$$\{m, m + [p_N(n)], m + 2[p_N(n)], \ldots, m + \ell[p_N(n)]\},$$

for some $m \in \mathbb{Z}$, $N \in \mathbb{N}$, and $1 \leq n \leq N$, with $[p_N(n)] \neq 0$.

2.2. **Multiple sequences.** Analogously to the previous results, assuming the validity of Theorems 2.1, we have various implications for multiple variable polynomial sequences.

2.2.1. *Recurrence.* A first recurrence result is the following (we skip the proof as the argument is the same to the one in [12, Theorem 2.8]):

Theorem 2.8. *For $\ell \in \mathbb{N}$, let $(p_{1,N}, \ldots, p_{\ell,N})_N$ as in Theorem 2.1. If $(X, \mathcal{B}, \mu, T)$ is a system and $A_0, A_1, \ldots, A_\ell \in \mathcal{B}$ such that*

$$\mu\left(A_0 \cap T^{k_1} A_1 \cap \ldots \cap T^{k_\ell} A_\ell\right) = \alpha > 0$$

for some $k_1, \ldots, k_\ell \in \mathbb{Z}$, then

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(A_0 \cap T^{-[p_{1,N}(n)]} A_1 \cap \ldots \cap T^{-[p_{\ell,N}(n)]} A_\ell\right) \geq \alpha^{\ell+1}.$$

Setting $A_i = A$ and $k_i = 0$ we immediately get the following (optimal lower bound):

Corollary 2.9. *For $\ell \in \mathbb{N}$, let $(p_{1,N}, \ldots, p_{\ell,N})_N$ as in Theorem 2.1. Then, for every system $(X, \mathcal{B}, \mu, T)$ and $A \in \mathcal{B}$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(A \cap T^{-[p_{1,N}(n)]}A \cap \ldots \cap T^{-[p_{\ell,N}(n)]}A\right) \geq (\mu(A))^{\ell+1}.$$

2.2.2. *Combinatorics.* Theorem 2.8, via [15, Proposition 3.3], which is a variant of Theorem 2.5 for several sets, implies the following (we are skipping the routine details):

Theorem 2.10. *For $\ell \in \mathbb{N}$, let $(p_{1,N}, \ldots, p_{\ell,N})_N$ as in Theorem 2.1. If $E_0, E_1, \ldots, E_\ell \subseteq \mathbb{N}$ such that*

$$\bar{d}(E_0 \cap (E_1 + k_1) \cap \ldots \cap (E_\ell + k_\ell)) = \alpha > 0$$

for some $k_1, \ldots, k_\ell \in \mathbb{Z}$, then

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \bar{d}(E_0 \cap (E_1 - [p_{1,N}(n)]) \cap \ldots \cap (E_\ell - [p_{\ell,N}(n)])) \geq \alpha^{\ell+1}.$$

Setting $E_i = E$ and $k_i = 0$ in the previous result, we get:

Corollary 2.11. *For $\ell \in \mathbb{N}$, let $(p_{1,N}, \ldots, p_{\ell,N})_N$ as in Theorem 2.1. Then, for every $E \subseteq \mathbb{N}$, we have*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \bar{d}(E \cap (E - [p_{1,N}(n)]) \cap \ldots \cap (E - [p_{\ell,N}(n)])) \geq (\bar{d}(E))^{\ell+1}.$$

So, we immediately obtain the following combinatorial result:

Corollary 2.12. *For $\ell \in \mathbb{N}$, let $(p_{1,N}, \ldots, p_{\ell,N})_N$ as in Theorem 2.1. Then any $E \subseteq \mathbb{N}$ with $\bar{d}(E) > 0$ contains arithmetic configurations of the form*

$$\{m, m + [p_{1,N}(n)], m + [p_{2,N}(n)], \ldots, m + [p_{\ell,N}(n)]\},$$

for some $m \in \mathbb{Z}$, $N \in \mathbb{N}$, and $1 \leq n \leq N$, with $[p_{i,N}(n)] \neq 0$, for all $1 \leq i \leq \ell$.

Applying Theorem 2.10 to syndetic sets $E_0, E_1, \ldots, E_\ell$[17] and $\alpha = (\prod_{i=1}^{\ell} r_i)^{-1}$, where r_i is the syndeticity constant of E_i, $1 \leq i \leq \ell$, we have:

Corollary 2.13. *For $\ell \in \mathbb{N}$, let $(p_{1,N}, \ldots, p_{\ell,N})_N$ as in Theorem 2.1. If $E_0, E_1, \ldots, E_\ell \subseteq \mathbb{N}$ are syndetic sets, then there exists $m \in \mathbb{Z}$, $N \in \mathbb{N}$, and $1 \leq n \leq N$, with $[p_{i,N}(n)] \neq 0$, for all $1 \leq i \leq \ell$, such that*

$$m \in E_0, m + [p_{1,N}(n)] \in E_1, \ldots, m + [p_{\ell,N}(n)] \in E_\ell.$$

The previous result, setting $E_i = c_i E$,[18] $0 \leq i \leq \ell$, where $E \subseteq \mathbb{N}$ is a syndetic set and $c_0, c_1, \ldots, c_\ell \in \mathbb{N}$, implies that we can find $x_0, x_1, \ldots, x_\ell \in E$, $N \in \mathbb{N}$, and $1 \leq n \leq N$,

---

[17] A set $E \subseteq \mathbb{N}$ is called *syndetic* if finite translations of it are covering $\mathbb{N}$. The cardinality of such a set of translations is a *syndeticity constant* of E.

[18] Where $cE := \{cn : n \in E\}$.

solution to the following system of equations:

$$\begin{aligned} c_1x_1 - c_0x_0 &= [p_{1,N}(n)] \\ c_2x_2 - c_0x_0 &= [p_{2,N}(n)] \\ &\vdots \\ c_\ell x_\ell - c_0x_0 &= [p_{\ell,N}(n)]. \end{aligned}$$

2.2.3. *Topological dynamics.* Let (X,T) be a (topological) dynamical system, whereby (X,d) is a compact metric space and $T : X \to X$ an invertible continuous transformation. Suppose T is minimal (i.e., $\overline{\{T^n x:\ n \in \mathbb{N}\}} = X$ for all $x \in X$, hence, for every $x \in X$ and non-empty open set U the set $\{n \in \mathbb{N}:\ T^n x \in U\}$ is syndetic). There exists a T-invariant Borel measure which gives positive value to every non-empty open set. So, due to syndeticity, for every $x \in X$ and every non-empty open set U we have

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_U(T^n x) > 0 \tag{12}$$

(this limit actually exists). Since $\mathbb{E}(f_i|\mathcal{I}(T)) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} T^n f_i$, combining (12) with the result from Theorem 2.1, we get for almost every $x \in X$ (and hence for a dense set) and every $U_1, \dots, U_\ell$ from a given countable basis of non-empty open sets that

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{U_1}(T^{[p_{1,N}(n)]}x) \cdot \ldots \cdot \mathbf{1}_{U_\ell}(T^{[p_{\ell,N}(n)]}x) > 0. \tag{13}$$

Using this we get:

Theorem 2.14. *For $\ell \in \mathbb{N}$, let $(p_{1,N}, \dots, p_{\ell,N})_N$ as in Theorem 2.1. If (X,T) is a minimal dynamical system, then, for a residual and T-invariant set of $x \in X$, we have*

$$\overline{\left\{(T^{[p_{1,N}(n)]}x, \dots, T^{[p_{\ell,N}(n)]}x):\ N \in \mathbb{N},\ 1 \le n \le N\right\}} = X \times \cdots \times X. \tag{14}$$

Proof. Relation (13) immediately implies that the set of points that satisfy (14), say R, is dense. To see that R is G_δ, take $\ell = 1$ (the general case is analogous). Then

$$R = \left\{x \in X :\ \forall\, m, r \in \mathbb{N},\ \exists\, N \in \mathbb{N} \text{ and } 1 \le n \le N \text{ with } T^{[p_{1,N}(n)]}x \in B(x_m, 1/r)\right\},$$

where $\{x_m:\ m \in \mathbb{N}\}$ is a countable, dense subset of X and $B(x_m, 1/r)$ denotes the open ball centered at x_m with radius $1/r$. The claim now follows since

$$R = \bigcap_{m,r\in\mathbb{N}} \bigcup_{\substack{N\in\mathbb{N}\\ 1\le n\le N}} T^{-[p_{1,N}(n)]}B(x_m, 1/r).$$

By the fact that

$$(T^{[p_{1,N}(n)]}(Tx), \dots, T^{[p_{\ell,N}(n)]}(Tx)) = (T \times \ldots \times T)(T^{[p_{1,N}(n)]}x, \dots, T^{[p_{\ell,N}(n)]}x)$$

we also get the T-invariance of R. □

Using Zorn's lemma, we know that every dynamical system has a minimal subsystem. Using this and Theorem 2.14 we get:

Corollary 2.15. *For $\ell \in \mathbb{N}$, let $(p_{1,N},\ldots,p_{\ell,N})_N$ as in Theorem 2.1. If (X,T) is a dynamical system, then, for a non-empty and T-invariant set of $x \in X$, we have*

$$\overline{\left\{(T^{[p_1(n)]}x,\ldots,T^{[p_\ell(n)]}x):\ N\in\mathbb{N},\ 1\leq n\leq N\right\}}=\overline{\{T^n x:\ n\in\mathbb{N}\}}\times\cdots\times\overline{\{T^n x:\ n\in\mathbb{N}\}}.$$

2.3. **Convergence along primes.** For averages along prime numbers, we will also show some more general, comparing to Theorems 1.6 and 1.7, results. Once again, for ease, one can imagine the following results under the assumptions of Theorems 1.6 and 1.7.

Theorem 2.16. *For $\ell \in \mathbb{N}$, let $(q_{1,N},\ldots,q_{\ell,N})_N$ be a sequence of ℓ-tuple of polynomials such that, for all $W \in \mathbb{N}$ and $1 \leq r \leq W$, the sequence $(q_{W,r,1,N},\ldots,q_{W,r,\ell,N})_N$ is good and super nice, where $q_{W,r,i,N}(n) = q_{i,N}(Wn+r)$, $1 \leq i \leq \ell$.*[19] *Then, for every ergodic system $(X,\mathcal{B},\mu,T)$ and $f_1,\ldots,f_\ell \in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty}\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\cap[1,N]}T^{[q_{1,N}(p)]}f_1\cdot\ldots\cdot T^{[q_{\ell,N}(p)]}f_\ell=\prod_{i=1}^{\ell}\int f_i\,d\mu,$$

where the convergence takes place in $L^2(\mu)$.

Theorem 2.17. *Let $(q_N)_N \subseteq \mathbb{R}[t]$ be a polynomial sequence such that, for every $W \in \mathbb{N}$, $1 \leq r \leq W$, the sequence $(q_{W,r,N})_N$ is good and for all $\ell \in \mathbb{N}$, $(q_{W,r,N}, 2q_{W,r,N},\ldots,\ell q_{W,r,N})_N$ is super nice, where $q_{W,r,N}(n) = q_N(Wn+r)$.*[20] *Then, for every $\ell \in \mathbb{N}$, system $(X,\mathcal{B},\mu,T)$, and $f_1,\ldots,f_\ell \in L^\infty(\mu)$, we have*

$$\lim_{N\to\infty}\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\cap[1,N]}T^{[q_N(p)]}f_1\cdot T^{2[q_N(p)]}f_2\cdot\ldots\cdot T^{\ell[q_N(p)]}f_\ell=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{\ell}T^{in}f_i,$$

where the convergence takes place in $L^2(\mu)$.

Theorems 2.16 and 2.17 imply Theorems 1.6 and 1.7 but also [27, Theorem 2.14] and [27, Theorem 2.12], respectively, which are about constant (with respect to N) polynomial sequences.

As with Theorem 2.1, we can drop the ergodicity assumption in Theorem 2.16 as well. Assuming the validity of Theorems 2.16 and 2.17 we present some applications.

For a single sequence, Theorem 2.17, via Theorem 2.3, implies the following:

Corollary 2.18. *Let $(q_N)_N \subseteq \mathbb{R}[t]$ as in Theorem 2.17. Then, for every $\ell \in \mathbb{N}$, system $(X,\mathcal{B},\mu,T)$, and $A \in \mathcal{B}$ with $\mu(A) > 0$, we have*

$$\lim_{N\to\infty}\frac{1}{\pi(N)}\sum_{p\in\mathbb{P}\cap[1,N]}\mu\left(A\cap T^{-[q_N(p)]}A\cap T^{-2[q_N(p)]}A\cap\ldots\cap T^{-\ell[q_N(p)]}A\right)>0.$$

Corollary 2.18, immediately implies the following Szemerédy-type theorem:

[19]Letting $W = r = 1$, it is easy to see that this condition implies that $(q_{1,N},\ldots,q_{\ell,N})_N$ is good and super nice as well.

[20]As in the previous result, for $W = r = 1$, we have that $(q_N)_N$ is good and, for all $\ell \in \mathbb{N}$, $(q_N, 2q_N,\ldots,\ell q_N)_N$ is super nice.

Corollary 2.19. *Let $(q_N)_N \subseteq \mathbb{R}[t]$ as in Theorem 2.17. Then, for every $\ell \in \mathbb{N}$, any $E \subseteq \mathbb{N}$ with $\bar{d}(E) > 0$ contains arithmetic progressions of the form:*

$$\{m, m + [q_N(p)], m + 2[q_N(p)], \dots, m + \ell[q_N(p)]\},$$

for some $m \in \mathbb{Z}$, $N \in \mathbb{N}$, and $p \in \mathbb{P} \cap [1, N]$, with $[q_N(p)] \neq 0$.

For multiple sequences, Theorem 2.16 implies:

Corollary 2.20. *For $\ell \in \mathbb{N}$, let $(q_{1,N}, \dots, q_{\ell,N})_N$ as in Theorem 2.16. Then, for every system $(X, \mathcal{B}, \mu, T)$ and $A \in \mathcal{B}$, we have*

$$\lim_{N\to\infty} \frac{1}{\pi(N)} \sum_{p\in\mathbb{P}\cap[1,N]} \mu\left(A \cap T^{-[q_{1,N}(p)]}A \cap \ldots \cap T^{-[q_{\ell,N}(p)]}A\right) \geq (\mu(A))^{\ell+1}.$$

Via the corresponding to Corollary 2.20 combinatorial result, which we get by using the correspondence principle, we finally obtain:

Corollary 2.21. *For $\ell \in \mathbb{N}$, let $(q_{1,N}, \dots, q_{\ell,N})_N$ as in Theorem 2.16. Then any $E \subseteq \mathbb{N}$ with $\bar{d}(E) > 0$ contains arithmetic configurations of the form*

$$\{m, m + [q_{1,N}(p)], m + [q_{2,N}(p)], \dots, m + [q_{\ell,N}(p)]\},$$

for some $m \in \mathbb{Z}$, $N \in \mathbb{N}$, and $p \in \mathbb{P} \cap [1, N]$, with $[q_{i,N}(n)] \neq 0$, for all $1 \leq i \leq \ell$.

2.4. **Recurrence along shifted primes.** Our method also implies recurrence results along shifted primes. More specifically, we get the following:

Theorem 2.22. *Let $(q_N)_N$ be a polynomial sequence such that, for every $W \in \mathbb{N}$, the sequence $(q_{W,N})_N$ is good and for all $\ell \in \mathbb{N}$, $(q_{W,N}, 2q_{W,N}, \dots, \ell q_{W,N})_N$ is super nice, where $q_{W,N}(n) = q_N(Wn)$. Then, for every $\ell \in \mathbb{N}$, system $(X, \mathcal{B}, \mu, T)$, and $A \in \mathcal{B}$ with $\mu(A) > 0$, the set*

$$\bigcup_{N\in\mathbb{N}} \left\{n \in [1, N] :\ \mu\left(A \cap T^{-[q_N(n)]}A \cap T^{-2[q_N(n)]}A \cap \ldots \cap T^{-\ell[q_N(n)]}A\right) > 0\right\}$$

has non-empty intersection with $\mathbb{P} - 1$ and $\mathbb{P} + 1$.

Theorem 2.23. *For $\ell \in \mathbb{N}$, let $(q_{1,N}, \dots, q_{\ell,N})_N$ be a sequence of ℓ-tuple of polynomials such that, for all $W \in \mathbb{N}$ the sequence $(q_{W,1,N}, \dots, q_{W,\ell,N})_N$ is good and super nice, where $q_{W,i,N}(n) = q_{i,N}(Wn)$, $1 \leq i \leq \ell$. Then, for every system $(X, \mathcal{B}, \mu, T)$ and $A \in \mathcal{B}$ with $\mu(A) > 0$, the set*

$$\bigcup_{N\in\mathbb{N}} \left\{n \in [1, N] :\ \mu\left(A \cap T^{-[q_{1,N}(n)]}A \cap \ldots \cap T^{-[q_{\ell,N}(n)]}A\right) > 0\right\}$$

has non-empty intersection with $\mathbb{P} - 1$ and $\mathbb{P} + 1$.

Via Furstenberg's correspondence principle the previous results imply:

Corollary 2.24. *Under the assumptions of Theorem 2.22, for every $\ell \in \mathbb{N}$ and $E \subseteq \mathbb{N}$ with $\bar{d}(E) > 0$, we have that the set*

$$\bigcup_{N\in\mathbb{N}} \left\{n \in [1, N] :\ \bar{d}(E \cap (E - [q_N(n)]) \cap (E - 2[q_N(n)]) \cap \ldots \cap (E - \ell[q_N(n)])) > 0\right\}$$

has non-empty intersection with $\mathbb{P} - 1$ and $\mathbb{P} + 1$.

Corollary 2.25. *Under the assumptions of Theorem 2.23, for any* $E \subseteq \mathbb{N}$ *with* $\bar{d}(E) > 0$*, we have that the set*

$$\bigcup_{N\in\mathbb{N}} \big\{n \in [1,N] : \ \bar{d}(E \cap (E - [q_{1,N}(n)]) \cap \ldots \cap (E - [q_{\ell,N}(n)])) > 0\big\}$$

has non-empty intersection with $\mathbb{P} - 1$ *and* $\mathbb{P} + 1$.

3. The $\ell = 1$ case

This subsection is dedicated to proving Theorem 1.3.[21] Someone who is familiar with Weyl's equidistribution theorem, can see that there is a strong equidistribution nature behind the definition of good sequences of variable polynomials, i.e., property (6).[22]

We start with some intermediate lemmas.[23]

Lemma 3.1. *Let* $(p_N)_N$ *be a good sequence of polynomials. Then, for every complex-valued continuous function* f *on* $\mathbb{R}$ *with period* 1, *we have*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(p_N(n)) = \int_0^1 f(x)\, dx. \tag{15}$$

Proof. We follow the arguments of the proof of [33, Theorem 2.1, Page 7]. For $\varepsilon > 0$, by the Stone-Weierstrauss theorem, we find a trigonometric polynomial q (i.e., $q(x)$ is a finite linear combination of functions of the form $e^{2\pi i h x}$, $h \in \mathbb{Z}$, with complex coefficients) such that $\sup_{0\leq x\leq 1} |f(x) - q(x)| < \varepsilon$. We have

$$\begin{aligned}
\left|\int_0^1 f(x)\, dx - \frac{1}{N}\sum_{n=1}^{N} f(p_N(n))\right| \ &\leq \ \left|\int_0^1 (f(x) - q(x))\, dx\right| \\
&+ \ \left|\int_0^1 q(x)\, dx - \frac{1}{N}\sum_{n=1}^{N} q(p_N(n))\right| \\
&+ \ \left|\frac{1}{N}\sum_{n=1}^{N}(q(p_N(n)) - f(p_N(n)))\right| \\
&< \ 2\varepsilon + \left|\int_0^1 q(x)\, dx - \frac{1}{N}\sum_{n=1}^{N} q(p_N(n))\right|,
\end{aligned}$$

where the last quantity goes to 0 as $N \to \infty$ by the goodness property of $(p_N)_N$. The conclusion follows as ε was chosen to be arbitrarily small. □

Using a standard argument, the previous statement can be upgraded to the following:

[21] We chose to present this proof at this point, before going deeper into characteristic factors and equidistribution of (variable) sequences in general nilmanifolds, as we will only make use of the definition of a good polynomial sequence and Herglotz's theorem.

[22] Note though that (6) is strictly stronger comparing to the one from Weyl's criterion, as, for example, $(\sqrt{2}n^2)_n$ is equiditributed but $(n^2 = 1/\sqrt{2}(\sqrt{2}n^2))_n$ is not.

[23] The following lemmas, in case of independent of N polynomial sequences, are characterizations of the equidistribution notion.

Lemma 3.2. *Let $(p_N)_N$ be a good sequence of polynomials. The conclusion of Lemma 3.1 holds for every complex-valued Riemann integrable function f on $\mathbb{R}$ with period* 1.

Proof. Without loss of generality assume that f take real values (otherwise we use the conclusion for the real and imaginary part of f). We follow [33, Theorem 1.1, Page 2].

Given $\varepsilon > 0$ let f_1, f_2 be continuous functions with $f_1(0) = f_1(1)$, $f_2(0) = f_2(1)$, $f_1(x) \leq f(x) \leq f_2(x)$ for all $x \in [0,1]$ and $\int_0^1 (f_2(x) - f_1(x))\, dx < \varepsilon$. Using Lemma 3.1 for both f_1, f_2 we get

$$\begin{aligned}
\int_0^1 f(x)\, dx - \varepsilon &\leq \int_0^1 f_1(x)\, dx = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N f_1(p_N(n)) \\
&\leq \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^N f(p_N(n)) \leq \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^N f(p_N(n)) \\
&\leq \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N f_2(p_N(n)) = \int_0^1 f_2(x)\, dx \leq \int_0^1 f(x)\, dx + \varepsilon.
\end{aligned}$$

The conclusion follows as ε was chosen to be arbitrarily small. □

Noting that for a good sequence of polynomials $(p_N)_N$ and $\gamma \in \mathbb{R}\setminus\mathbb{Q}$ we have that $((h_1\gamma + h_2)p_N)_N$ is good for all $(h_1, h_2) \in \mathbb{Z}^2 \setminus \{(0,0)\}$, we have the following:

Lemma 3.3. *Let $(p_N)_N$ be a good sequence of polynomials. Then, for every complex-valued continuous periodic mod 1 function f on $\mathbb{R}^2$ and $\gamma \in \mathbb{R}\setminus\mathbb{Q}$, we have*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N f(p_N(n)\gamma, p_N(n)) = \int_0^1\int_0^1 f(x,y)\, dxdy. \tag{16}$$

Proof. All continuous complex-valued periodic mod 1 (on each coordinate) functions on $\mathbb{R}^2$ can be approximated (with respect to the uniform norm) by finite linear combinations with complex coefficients of functions $e^{2\pi i(h_1 x + h_2 y)}$, $(h_1, h_2) \in \mathbb{Z}^2$. The result now follows, as in the proof of Lemma 3.1, as (6) implies $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N e^{2\pi i(h_1\gamma+h_2)p_N(n)} = 0$ for all $(h_1, h_2) \in \mathbb{Z}^2 \setminus \{(0,0)\}$. □

Analogously to the proof of Lemma 3.2, using the previous result, we get:

Lemma 3.4. *Let $(p_N)_N$ be a good sequence of polynomials. Then, for every complex-valued Riemann integrable periodic mod 1 function f on $\mathbb{R}^2$ and $\gamma \in \mathbb{R}\setminus\mathbb{Q}$, we have*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N f(p_N(n)\gamma, p_N(n)) = \int_0^1\int_0^1 f(x,y)\, dxdy.$$

The next result is the last ingredient that we need in order to prove Theorem 1.3:

Lemma 3.5. *If $(p_N)_N$ is a good sequence of polynomials and $\gamma \in \mathbb{R}\setminus\mathbb{Z}$, then*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N e^{2\pi i [p_N(n)]\gamma} = 0.$$

Proof. We split the proof into two cases.

Case 1: $\gamma \in \mathbb{Q} \setminus \mathbb{Z}$.

For $m \in \mathbb{N}$, $m > 1$, it suffices to show, for all $1 \leq h \leq m-1$, that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i [p_N(n)] \frac{h}{m}} = 0.$$

For $0 \leq j \leq m-1$, let $A(j,m,N)$ denote the number of terms $p_N(1), \dots, p_N(N)$ that satisfy $[p_N(n)] \equiv j (\text{mod } m)$. This is equivalent to $j/m \leq \{p_N(n)/m\} < (j+1)/m$.[24] Hence, since $(p_N/m)_N$ is good, using Lemma 3.2 for the characteristic function of the interval $[j/m, (j+1)/m)$, we get

$$\lim_{N\to\infty} \frac{A(j,m,N)}{N} = \frac{1}{m}.$$

This last relation, together with

$$\sum_{h=1}^{m-1} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i [p_N(n)] \frac{h}{m}} \right|^2 = m \sum_{j=0}^{m-1} \left(\frac{A(j,m,N)}{N} - \frac{1}{m} \right)^2,$$

(see [33, Exercise 1.5, Page 318]) implies the claim.

Case 2: $\gamma \in \mathbb{R} \setminus \mathbb{Q}$.

We write

$$\frac{1}{N} \sum_{n=1}^{N} e^{2\pi i [p_N(n)]\gamma} = \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i (p_N(n) - \{p_N(n)\})\gamma} = \frac{1}{N} \sum_{n=1}^{N} f(p_N(n)\gamma, p_N(n)),$$

where $f(x,y) = e^{2\pi i (x - \{y\}\gamma)}$ is a complex-valued Riemann integrable periodic mod 1 function on $\mathbb{R}^2$. By Lemma 3.4 we have that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i [p_N(n)]\gamma} = \int_0^1 \int_0^1 f(x,y)\, dxdy = 0,$$

as was to be shown. □

Using a standard argument we are now ready to prove Theorem 1.3:

[24]Where $\{x\} = x - [x]$ is the fractional part of the real number x.

Proof of Theorem 1.3. We can assume without loss of generality that $\int f \, d\mu = 0$. With Herglotz's theorem we have

$$\begin{aligned}\left\| \frac{1}{N}\sum_{n=1}^{N} T^{[p_N(n)]}f \right\|_2^2 &= \frac{1}{N^2}\sum_{n,m=1}^{N}\int T^{[p_N(n)]}fT^{[p_N(m)]}f \, d\mu \\ &= \frac{1}{N^2}\sum_{n,m=1}^{N}\int fT^{[p_N(n)]-[p_N(m)]}f \, d\mu \\ &= \frac{1}{N^2}\sum_{n,m=1}^{N}\int e^{2\pi i([p_N(n)]-[p_N(m)])\gamma} \, d\nu_f(\gamma) \\ &= \int \left| \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i [p_N(n)]\gamma} \right|^2 \, d\nu_f(\gamma),\end{aligned}$$

where ν_f is the spectral measure with $\nu_f(\{0\}) = 0$. The conclusion now follows since Lemma 3.5 implies that $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} e^{2\pi i [p_N(n)]\gamma} = 0$ for all $\gamma \in (0,1)$. □

Remark 3.6. *Someone can get numerous applications from Theorem 1.3 as the ones in Section 2 (in the prime numbers cases one has to postulate some additional "good" assumptions for the sequences but no "super nice" ones). As these implications are clear from our arguments, we leave the formulations and proofs of them to the interested reader.*

4. Some background material

In this section we list some materials that will be used for the multiple average case.

4.1. **Factors.** A *homomorphism* from a system $(X, \mathcal{B}, \mu, T)$ onto a system $(Y, \mathcal{Y}, \nu, S)$ is a measurable map $\pi : X' \to Y'$, where X' is T-invariant subset of X and Y' is an S-invariant subset of Y, both of full measure, such that $\mu \circ \pi^{-1} = \nu$ and $S \circ \pi(x) = \pi \circ T(x)$ for $x \in X'$. When we have such a homomorphism we say that the system $(Y, \mathcal{Y}, \nu, S)$ is a *factor* of the system $(X, \mathcal{B}, \mu, T)$. If the factor map $\pi : X' \to Y'$ can be chosen to be injective, then we say that the systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ are *isomorphic*. A factor can also be characterised by $\pi^{-1}(\mathcal{Y})$ which is a T-invariant sub-σ-algebra of $\mathcal{B}$, and, conversely, any T-invariant sub-σ-algebra of $\mathcal{B}$ defines a factor. By abusing the terminology, we denote by the same letter the σ-algebra $\mathcal{Y}$ and its inverse image by π, so, if $(Y, \mathcal{Y}, \nu, S)$ is a factor of $(X, \mathcal{B}, \mu, T)$, we think of $\mathcal{Y}$ as a sub-σ-algebra of $\mathcal{B}$.

4.1.1. *Seminorms.* We follow [26] and [8] for the inductive definition of the seminorms $\|\!| \cdot |\!\|_k$. More specifically, the definition that we use here follows from [26] (in the ergodic case), [8] (in the general case) and the use of von Neumann's ergodic theorem.

Let $(X, \mathcal{B}, \mu, T)$ be a system and $f \in L^\infty(\mu)$. We define inductively the seminorms $\|\!| f |\!\|_{k,\mu,T}$ (or just $\|\!| f |\!\|_k$ if there is no confusion) as follows: For $k = 1$ we set

$$\|\!| f |\!\|_1 := \|\mathbb{E}(f|\mathcal{I}(T))\|_2 \, ,$$

where $\mathcal{I}(T)$ is the σ-algebra of T-invariant sets and $\mathbb{E}(f|\mathcal{I}(T))$ the conditional expectation of f with respect to $\mathcal{I}(T)$, satisfying $\int \mathbb{E}(f|\mathcal{I}(T)) \, d\mu = \int f \, d\mu$ and $T\mathbb{E}(f|\mathcal{I}(T)) = \mathbb{E}(Tf|\mathcal{I}(T))$.

For $k \geq 1$ we let

$$|||f|||_{k+1}^{2^{k+1}} := \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} |||\bar{f}\cdot T^n f|||_k^{2^k}.$$

All these limits exist and $|||\cdot|||_k$ define seminorms on $L^\infty(\mu)$ ([26]). Also, we remark that for all $k \in \mathbb{N}$ we have $|||f|||_k \leq |||f|||_{k+1}$ and $|||f\otimes \bar{f}|||_{k,\mu\times\mu,T\times T} \leq |||f|||_{k+1,\mu,T}^2$.

4.1.2. *Nilfactors.* Using the seminorms we defined above we can construct factors $\mathcal{Z}_k = \mathcal{Z}_k(T)$ of X characterized by:

$$\text{for } f\in L^\infty(\mu),\quad \mathbb{E}(f|\mathcal{Z}_{k-1}) = 0 \text{ if and only if } |||f|||_k = 0.$$

The following profound fact from [26] shows that for every $k \in \mathbb{N}$ the factor $\mathcal{Z}_k$ has a purely algebraic structure; in particular for all practical reasons we can assume that it is a k-step nilsystem (see Subsection 4.2 below for the definitions):

Theorem 4.1 (Host & Kra, [26])**.** *Let $(X,\mathcal{B},\mu,T)$ be an ergodic system and $k\in\mathbb{N}$. Then the factor $\mathcal{Z}_k(T)$ is an inverse limit of k-step nilsystems.*[25]

Because of this result (also known as the "Structure Theorem") we call $\mathcal{Z}_k$ the *k-step nilfactor* of the system. The smallest factor that is an extension of all finite step nilfactors is denoted by $\mathcal{Z} = \mathcal{Z}(T)$, meaning, $\mathcal{Z} = \bigvee_{k\in\mathbb{N}} \mathcal{Z}_k$, and is called the *nilfactor* of the system. The nilfactor $\mathcal{Z}$ is of particular interest because it controls the limiting behaviour in $L^2(\mu)$ of the averages in (9) and (10).

4.2. **Nilmanifolds.** Let G be a k-step nilpotent Lie group, meaning $G_{k+1} = \{e\}$ for some $k\in\mathbb{N}$, where $G_k = [G, G_{k-1}]$ denotes the k-th commutator subgroup, and Γ a discrete cocompact subgroup of G. The compact homogeneous space $X = G/\Gamma$ is called *k-step nilmanifold* (or *nilmanifold*). The group G acts on G/Γ by left translations, where the translation by an element $b\in G$ is given by $T_b(g\Gamma) = (bg)\Gamma$. We denote by m_X the normalized *Haar measure* on X, i.e., the unique probability measure that is invariant under the action of G, and by $\mathcal{G}/\Gamma$ the Borel σ-algebra of G/Γ. If $b\in G$, we call the system $(G/\Gamma, \mathcal{G}/\Gamma, m_X, T_b)$ *k-step nilsystem* (or *nilsystem*) and the elements of G *nilrotations.*

4.2.1. *Equidistribution.* For a connected and simply connected Lie group G, let $\exp : \mathfrak{g}\to G$ be the exponential map, where $\mathfrak{g}$ is the Lie algebra of G. For $b\in G$ and $s\in\mathbb{R}$ we define the element b^s of G as follows: If $X\in\mathfrak{g}$ is such that $\exp(X) = b$, then $b^s = \exp(sX)$ (this is well defined since under the aforementioned assumptions exp is a bijection).

If $(a(n))_n$ is a sequence of real numbers and $X = G/\Gamma$ is a nilmanifold with G connected and simply connected, we say that the sequence $(b^{a(n)}x)_n$, $b\in G$, is *equidistributed* in a subnilmanifold Y of X, if for every $F\in C(X)$ we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} F(b^{a(n)}x) = \int F\, dm_Y. \tag{17}$$[26]

A nilrotation $b\in G$ is *ergodic* (or *acts ergodically*) on X, if the sequence $(b^n\Gamma)_n$ is dense in X. If $b\in G$ is ergodic, then for every $x\in X$ the sequence $(b^n x)_n$ is equidistributed

[25]By this we mean that there exist T-invariant sub-σ-algebras $\mathcal{Z}_{k,i}$, $i\in\mathbb{N}$, of $\mathcal{B}$ such that $\mathcal{Z}_k = \bigcup_{i\in\mathbb{N}}\mathcal{Z}_{k,i}$ and for every $i\in\mathbb{N}$, the factors induced by the σ-algebras $\mathcal{Z}_{k,i}$ are isomorphic to k-step nilsystems.

[26]If $(a(n))_n\subseteq\mathbb{Z}$, we can drop the assumptions that G is connected and simply connected.

in X (this is a non-trivial fact which follows by unique ergodicity). The orbit closure $\overline{(b^n\Gamma)_n}$ of $b \in G$ has the structure of a nilmanifold with $(b^n\Gamma)_n$ being equidistributed in it. Analogously, if G is connected and simply connected, then $\overline{(b^s\Gamma)_{s\in\mathbb{R}}}$ is a nilmanifold with $(b^s\Gamma)_{s\in\mathbb{R}}$ being equidistributed in it.

4.2.2. *Change of base point formula.* Let $X = G/\Gamma$ be a nilmanifold. As mentioned before, for every $b \in G$ the sequence $(b^n\Gamma)_n$ is equidistributed in $X_b = \overline{\{b^n\Gamma : n \in \mathbb{N}\}}$. Using the identity $b^n g = g(g^{-1}bg)^n$ we see that the nil-orbit $(b^n g\Gamma)_n$ is equidistributed in the set $gX_{g^{-1}bg}$. A similar formula holds when G is connected and simply connected, where we replace the integer parameter n with the real parameter s and the nilmanifold X_b with $Y_b = \overline{\{b^s\Gamma :\ s \in \mathbb{R}\}}$.

4.2.3. *Lifting argument.* Giving a topological group G, we denote the connected component of its identity element, e, by G_0. In several instances it will be convenient for us to assume that a nilmanifold has a representation G/Γ with G connected and simply connected (to this end, one can follow for example [34]). Since all our results deal with the action on X of finitely many elements of G we can and will assume that the discrete group G/G_0 is finitely generated. In this case one can show that $X = G/\Gamma$ is isomorphic to a sub-nilmanifold of a nilmanifold $\tilde{X} = \tilde{G}/\tilde{\Gamma}$, where $\tilde{G}$ is a connected and simply connected nilpotent Lie group, with all translations from G "represented" in $\tilde{G}$.[27] We caution the reader that such a construction is only helpful when our working assumptions impose no restrictions on a nilrotation. Any assumption made about $b \in G$, which acts on a nilmanifold X, is typically lost when passing to the lifted nilmanifold $\tilde{X}$.

5. Finding the characteristic factor

In this technical section we find characteristic factors for the required expressions that appear in Theorems 2.1 and 2.2. Actually, in both cases, we will show that the nilfactor is the characteristic one (Proposition 5.14 and Proposition 5.15 respectively).

We first start with the degree 1 case and then move on to the general one. At this point we recall (adapted to our study) the notion of a characteristic factor:

Definition 5.1. For $\ell \in \mathbb{N}$ let $(X, \mathcal{B}, \mu, T)$ be a system. The sub-σ-algebra $\mathcal{Y}$ of $\mathcal{B}$ is a *characteristic factor* for the variable tuple of integer-valued sequences $(a_{1,N}, \dots, a_{\ell,N})_N$ if it is T-invariant and

$$\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=1}^{N} T^{a_{1,N}(n)}f_1 \cdot \ldots \cdot T^{a_{\ell,N}(n)}f_\ell - \frac{1}{N}\sum_{n=1}^{N} T^{a_{1,N}(n)}\tilde{f}_1 \cdot \ldots \cdot T^{a_{\ell,N}(n)}\tilde{f}_\ell \right\|_2 = 0,$$

for all $f_i \in L^\infty(\mu)$, where $\tilde{f}_i = \mathbb{E}(f_i|\mathcal{Y})$, $1 \leq i \leq \ell$.[28]

[27] Practically this means that for every $F \in C(X)$, $b \in G$ and $x \in X$, there exists $\tilde{F} \in C(\tilde{X})$, $\tilde{b} \in \tilde{G}$ and $\tilde{x} \in \tilde{X}$, such that $F(b^n x) = \tilde{F}(\tilde{b}^n\tilde{x})$ for every $n \in \mathbb{N}$.

[28] Equivalently, $\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=1}^{N} T^{a_{1,N}(n)}f_1 \cdot \ldots \cdot T^{a_{\ell,N}(n)}f_\ell \right\|_2 = 0$ if $\mathbb{E}(f_i|\mathcal{Y}) = 0$ for some $1 \leq i \leq \ell$.

5.1. **The base case.** The following crucial lemma, which can be understood as a "change of variables" procedure, will be used in the base $\ell = 1$ case for $\deg p_N = 1$, i.e., $p_N(n) = a_N n + b_N$. We will assume that $(b_N)_N$ is bounded, so, as such error terms don't affect our averages, we mainly have to deal with the expression $\frac{1}{N}\sum_{n=1}^N T^{[a_N n]}f$.

Lemma 5.2. *Let $(a_N)_N \subseteq (0, +\infty)$ bounded with $(a_N \cdot N)_N$ tending increasingly to ∞. For any sequence $(c_N(n))_{n,N} \subseteq [0,\infty)$ we have*

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^N c_{[a_N N]}([a_N n]) \ll \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^N c_N(n).$$

Proof. For a fixed $N \in \mathbb{N}$, since $(a_N)_N$ is bounded, we have the relation

$$\frac{1}{N}\sum_{n=1}^N c_{[a_N N]}([a_N n]) \leq \left(\left[\frac{1}{a_N}\right] + 1\right) \cdot a_N \cdot \frac{[a_N N]}{a_N N} \cdot \frac{1}{[a_N N]}\sum_{n=0}^{[a_N N]} c_{[a_N N]}(n).$$

Since $a_N N \to \infty$, we get that $[a_N N]/a_N N \to 1$. Finally, using yet again that $(a_N)_N$ is bounded, we have

$$\left(\left[\frac{1}{a_N}\right] + 1\right) \cdot a_N \leq a_N + 1 \ll 1.$$

The result now follows by taking $\limsup$. □

Using Lemma 5.2, following the argument of [13, Lemma 5.2] we get:

Lemma 5.3. *Let $(p_N)_N$ be a sequence of polynomials of degree 1 of the form*

$$p_N(n) = a_N n + b_N,\ n, N \in \mathbb{N},$$

where $(a_N)_N, (b_N)_N$ are bounded with $(a_N)_N \subseteq (0, +\infty)$ and $(a_N \cdot N)_N$ tending increasingly to ∞. Then, for any system $(X, \mathcal{B}, \mu, T)$ and $f_1 \in L^\infty(\mu)$, we have

$$\limsup_{N\to\infty} \sup_{\|f_0\|_\infty \leq 1} \frac{1}{N}\sum_{n=1}^N \left|\int f_0 \cdot T^{[p_N(n)]}f_1 \, d\mu\right| \ll |||f_1|||_2. \tag{18}$$

Proof. For every $N \in \mathbb{N}$ we choose functions $f_{0,N}$ with $\|f_{0,N}\|_\infty \leq 1$ so that the corresponding average in (18) is $1/N$ close to its $\sup_{\|f_0\|_\infty \leq 1}$. To show (18), it suffices to show that

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^N \left|\int f_{0,[a_N N]} \cdot T^{[p_N(n)]}f_1 \, d\mu\right| \ll |||f_1|||_2. \tag{19}$$

We write

$$[p_N(n)] = [a_N n + b_N] = [a_N n] + [b_N] + e(n,N),\quad e(n,N) \in \{0,1\}.$$

If $([b_N] + e(n,N))_{n,N}$ takes values in the finite set of integers E, we have that

$$\frac{1}{N}\sum_{n=1}^N \left|\int f_{0,[a_N N]} \cdot T^{[p_N(n)]}f_1 \, d\mu\right| \quad \ll \quad \max_{e\in E} \frac{1}{N}\sum_{n=1}^N \left|\int f_{0,[a_N N]} \cdot T^{[a_N n]+e}f_1 \, d\mu\right|.$$

Taking squares and using the Cauchy-Schwarz inequality, the right-hand side of the previous relation is bounded by

$$\max_{e\in E}\frac{1}{N}\sum_{n=1}^{N}\left|\int f_{0,[a_N N]}\cdot T^{[a_N n]+e}f_1\,d\mu\right|^2=\max_{e\in E}\frac{1}{N}\sum_{n=1}^{N}\int F_{0,[a_N N]}\cdot S^{[a_N n]+e}F_1\,d\tilde{\mu},$$

where $S=T\times T$, $F_{0,[a_N N]}=f_{0,[a_N N]}\otimes\bar{f}_{0,[a_N N]}$, $F_1=f_1\otimes\bar{f}_1$, and $\tilde{\mu}=\mu\times\mu$. For every $e\in E$, using Lemma 5.2, the average on the right-hand side of the previous relation is bounded by a constant multiple of

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int F_{0,N}\cdot S^{n+e}F_1\,d\tilde{\mu}\leq\limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}S^{n+e}F_1\right\|_{L^2(\tilde{\mu})},$$

where the last inequality follows by Cauchy-Schwarz and the fact that $\|F_{0,N}\|_\infty\leq 1$. Using von Neymann's mean ergodic theorem, the last term is equal to

$$\|\mathbb{E}(S^e F_1|\mathcal{I}(S))\|_{L^2(\tilde{\mu})}=\|\mathbb{E}(F_1|\mathcal{I}(S))\|_{L^2(\tilde{\mu})}\leq\|\!|f_1|\!\|_2^2,$$

where we used the fact that S is measure preserving and the definition of the seminorms $\|\!|\cdot|\!\|$. (19) now follows by removing the squares. □

Remark 5.4. *Lemma 5.3 holds also for $(a_N)_N\subseteq(-\infty,0)$ with $(a_N\cdot N)_N$ tending decreasingly to $-\infty$.*

Indeed, In this case we write

$$[p_N(n)]=-[-a_N n]+[b_N]+e(n,N),\quad e(n,N)\in\{-1,0\},$$

so,

$$\frac{1}{N}\sum_{n=1}^{N}\left|\int f_0\cdot T^{[p_N(n)]}f_1\,d\mu\right|\ll\max_{e\in E}\frac{1}{N}\sum_{n=1}^{N}\left|\int f_0\cdot T^{-[-a_N n]}(T^e f_1)\,d\mu\right|,$$

where E is a finite subset of integers. Since $-a_N>0$ and $(-a_N\cdot N)_N$ tends increasingly to ∞, we get the conclusion by the previous lemma (working with T^{-1} instead of T).[29]

To extend Lemma 5.3 for multiple terms, we use the following variant of the classical van der Corput trick, the main tool for reducing the complexity of our iterates (see next subsection for more details):

Lemma 5.5 (Lemma 4.6, [13]). *Let $(v_{N,n})_{N,n}$ be a bounded sequence in a Hilbert space. Then*

$$\limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}v_{N,n}\right\|^2\leq 4\limsup_{H\to\infty}\frac{1}{H}\sum_{h=1}^{H}\limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N}\langle v_{N,n+h},v_{N,n}\rangle\right|.$$

We will now demonstrate the main idea behind the generalization of Lemma 5.3, for which we follow [13, Proposition 5.3, Case 1]. In that statement, to show

$$\limsup_{N\to\infty}\sup_{\|f_0\|_\infty,\|f_i\|_\infty\leq 1}\frac{1}{N}\sum_{n=1}^{N}\left|\int f_0\cdot T^{[a_1 n]}f_1\cdot T^{[a_2 n]}f_2\right|\ll\|\!|f_j|\!\|_4,$$

[29]We note that the seminorms that are taken with respect to the commuting transformations T or T^{-1}, since in Theorems 2.1 and 2.2 we are working under the ergodicity assumption, coincide.

where $(i,j)=(1,2)$ or $(2,1)$, one uses Lemma 5.5, compose with, say, $-[a_1n]$, and gets the terms (notice that we keep the h-term in the first difference even though it's bounded)

$$[a_1(n+h)]-[a_1n]\approx[a_1h],$$
$$[a_2(n+h)]-[a_1n]\approx[(a_2-a_1)n], \text{ and}$$
$$[a_2n]-[a_1n]\approx[(a_2-a_1)n],^{30}$$

so, after grouping the last two terms together, using the first one as constant (since it only depends on h–the average along which is taken at the very end), one can use the base $\ell=1$ case; this is actually how the inductive step works in the proof of the general $\ell\in\mathbb{N}$ case.

The variable case is more complicated to deal with. We demonstrate the main idea behind it by considering Example 1, i.e., $p_{1,N}(n)=a_{1,N}n$ and $p_{2,N}(n)=a_{2,N}n$, where $a_{1,N}=1/N^a$ and $a_{2,N}=1/N^b$ for $0<a<b<1$. The previous approach cannot be imitated, as, for example,

$$[a_{1,N}(n+h)]-[a_{1,N}n]\approx[a_{1,N}h]$$

is in general a variable term and we cannot proceed with the same argument. What we do instead is to transform the iterates in the initial sum to the following:

$$(0,[a_{1,N}n],[a_{2,N}n])\approx\left(0,[a_{1,N}n],\left[\frac{a_{2,N}}{a_{1,N}}[a_{1,N}n]\right]\right)\xrightarrow[\text{change of variables}]{\text{Lemma 5.2}}\left(0,n,\left[\frac{a_{2,N}}{a_{1,N}}n\right]\right),$$

and then we use Lemma 5.5 to bound, eventually, everything by $\|\!|f_1|\!\|_4$. (To use Lemma 5.2 note the crucial fact that $(a_{2,N}/a_{1,N})_N$ is bounded.) Additionally, to bound our expression by $\|\!|f_2|\!\|_4$, the previous argument needs an additional twist to work since the quantity $(a_{1,N}/a_{2,N})_N$ is unbounded. What we do in this case is to compose with $-[a_{2,N}n]$ to get

$$\begin{aligned}(0,[a_{1,N}n],[a_{2,N}n]) &\approx ([-a_{2,N}n],[(a_{1,N}-a_{2,N})n],0)\\ &\approx \left(\left[\frac{-a_{2,N}}{a_{1,N}-a_{2,N}}[(a_{1,N}-a_{2,N})n]\right],[(a_{1,N}-a_{2,N})n],0\right)\\ &\xrightarrow[\text{variables}]{\text{change of}} \left(\left[\frac{-a_{2,N}}{a_{1,N}-a_{2,N}}n\right],n,0\right).\end{aligned}$$

As $(a_{2,N}/(a_{1,N}-a_{2,N}))_N$ is bounded, we can now finish the argument as before.

The previous discussion, naturally leads to the following assumption on the leading coefficients of the linear (variable) polynomials:

Definition 5.6. A sequence of real numbers $(a_N)_N$ has the R_1*-property* if

(i) it is bounded; and

(ii) $(a_N)_N\subseteq(0,+\infty)$ or $(-\infty,0)$ and $(|a_N|\cdot N)_N$ tends increasingly to $+\infty$.

For $\ell\in\mathbb{N}$ we say that the sequences of real numbers $\{(a_{i,N})_N:\ 1\leq i\leq\ell\}$ have the R_ℓ*-property* if for all $1\leq i\leq\ell$:

(i) $(a_{i,N})_N$ has the R_1-property; and

[30]Here, by "$\approx$", we mean "modulo bounded error terms".

(ii) at least one of the following three properties holds:

(a) there exists $1 \leq j_0 \neq i \leq \ell$ such that the sequences $\left\{ \left(\frac{a_{j_0,N}-a_{j,N}}{a_{i,N}}\right)_N : 1 \leq j \neq j_0 \leq \ell \right\}$ have the $R_{\ell-1}$-property.

(b) there exists $1 \leq j_0 \neq i \leq \ell$ such that the sequence $(a_{i,N} - a_{j_0,N})_N$ has the R_1-property and the sequences $\left\{ \left(\frac{a_{j,N}}{a_{i,N}-a_{j_0,N}}\right)_N : 1 \leq j \neq j_0 \leq \ell \right\}$ have the $R_{\ell-1}$-property.

(c) there exist $1 \leq j_0 \neq i \leq \ell$ such that the sequence $(a_{i,N} - a_{j_0,N})_N$ has the R_1-property and $1 \leq k_0 \neq j_0, i \leq \ell$ such that the sequences $\left\{ \left(-\frac{a_{k_0,N}}{a_{i,N}-a_{j_0,N}}\right)_N, \left(\frac{a_{j,N}-a_{k_0,N}}{a_{i,N}-a_{j_0,N}}\right)_N : 1 \leq j \neq k_0, j_0 \leq \ell \right\}$ have the $R_{\ell-1}$-property.

Remark 5.7. The polynomial family of Example 1, i.e., $p_{1,N}(n) = n/N^a$, $p_{2,N}(n) = n/N^b$, $n, N \in \mathbb{N}$, where $0 < a < b < 1$, has the R_2-property.

Indeed, skipping the trivial calculations, both sequences $(1/N^a)_N$, $(1/N^b)_N$ have the R_1-property and for $i = 1$ we have the (ii) (a) case, while for $i = 2$ the (ii) (b) case.

We are now ready to extend Lemma 5.3 to multiple terms along polynomials of degree 1, following the main idea of [13, Proposition 5.3, Case 1]:

Proposition 5.8. *Let $(p_{1,N})_N, \ldots, (p_{\ell,N})_N$ be polynomial sequences of degree 1 of the form*

$$p_{i,N}(n) = a_{i,N}n + b_{i,N}, \quad n, N \in \mathbb{N}, 1 \leq i \leq \ell,$$

where the sequences $(a_{i,N})_N$, $1 \leq i \leq \ell$, have the R_ℓ-property and $(b_{i,N})_N$, $1 \leq i \leq \ell$, are bounded. Then, for every $f_1 \in L^\infty(\mu)$, we have

(20)

$$\limsup_{N\to\infty} \sup_{\|f_0\|_\infty, \|f_2\|_\infty, \ldots, \|f_\ell\|_\infty \leq 1} \frac{1}{N}\sum_{n=1}^{N} \left| \int f_0 \cdot T^{[p_{1,N}(n)]}f_1 \cdot \ldots \cdot T^{[p_{\ell,N}(n)]}f_\ell \, d\mu \right| \ll \vert\vert\vert f_1 \vert\vert\vert_{2\ell}.$$ [31]

Proof. We use induction on ℓ. The base case, $\ell = 1$, follows from Lemma 5.3. We assume that $\ell \geq 2$ and that the statement holds for $\ell - 1$.

Case 1: For $i = 1$, the property (ii) (a) from the Definition 5.6 holds.

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N} \left| \int f_0 \cdot T^{[a_{1,N}n+b_{1,N}]}f_1 \cdot \ldots \cdot T^{[a_{\ell,N}n+b_{\ell,N}]}f_\ell \, d\mu \right| \\
&= \frac{1}{N}\sum_{n=1}^{N} \left| \int f_0 \cdot T^{[a_{1,N}n]+e_1(n,N)}f_1 \cdot \prod_{i=2}^{\ell} T^{\left[\frac{a_{i,N}}{a_{1,N}}[a_{1,N}n]\right]+e_i(n,N)} f_i \, d\mu \right| \qquad (21)\\
&\ll \max_{e_1,\ldots,e_\ell \in E} \frac{1}{N}\sum_{n=1}^{N} \left| \int f_0 \cdot T^{[a_{1,N}n]}(T^{e_1}f_1) \cdot \prod_{i=2}^{\ell} T^{\left[\frac{a_{i,N}}{a_{1,N}}[a_{1,N}n]\right]}(T^{e_i}f_i) \, d\mu \right|,
\end{aligned}$$

[31] The constant here depends on the bounds of the coefficients and the number of transformations ℓ. Note that, because of symmetry, we also have the respective estimates for $2 \leq i \leq \ell$ with f_i in place of f_1.

where E is a finite subset of integers (the error terms $e_i(n,N)$, as $(b_{i,N})_N$, and $(a_{i,N}/a_{1,N})_N$ are bounded for $1\leq i\leq \ell$, take finitely many values).

For every $N\in\mathbb{N}$ we now choose functions $f_{i,N}$ with $\|f_{i,N}\|_\infty\leq 1$ for $i\in\{0,2,\ldots,\ell\}$, so that the last relation in (21) is $1/N$ close to the corresponding $\sup_{\|f_0\|_\infty,\|f_2\|_\infty,\ldots,\|f_\ell\|_\infty\leq 1}\cdot$ Using the Cauchy-Schwarz inequality and that $|a_{1,N}|\cdot N\to\infty$, we have that (20) follows if we show, for each choice of $e_1,\ldots,e_\ell\in E$, that

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left|\int f_{0,[a_{1,N}N]}\cdot T^{[a_{1,N}n]}(T^{e_1}f_1)\cdot\prod_{i=2}^{\ell}T^{\left[\frac{a_{i,N}}{a_{1,N}}[a_{1,N}n]\right]}(T^{e_i}f_{i,[a_{1,N}N]})\,d\mu\right|^2 \tag{22}$$

is bounded above by a constant multiple of $\vert\vert\vert f_1\vert\vert\vert_{2\ell}^2$. Using Lemma 5.2 it suffices to show

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left|\int f_{0,N}\cdot T^{n}(T^{e_1}f_1)\cdot\prod_{i=2}^{\ell}T^{\left[\frac{a_{i,N}}{a_{1,N}}n\right]}(T^{e_i}f_{i,N})\,d\mu\right|^2\ll \vert\vert\vert f_1\vert\vert\vert_{2\ell}^2. \tag{23}$$

The left-hand side of (23) is equal to

$$A:=\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int F_{0,N}\cdot S^{n}(S^{e_1}F_1)\cdot\prod_{i=2}^{\ell}S^{\left[\frac{a_{i,N}}{a_{1,N}}n\right]}(S^{e_i}F_{i,N})\,d\tilde{\mu},$$

where $S=T\times T$, $F_1=f_1\otimes\bar{f}_1$, $F_{i,N}=f_{i,N}\otimes\bar{f}_{i,N}$, $i=0,2,\ldots,\ell$, and $\tilde{\mu}=\mu\times\mu$. Using Cauchy-Schwarz, and then Lemma 5.5, we have that

$$|A|^2\ll\limsup_{H\to\infty}\frac{1}{H}\sum_{h=1}^{H}A_h,$$

where

$$A_h:=\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\Bigg|\int S^{n+h}(S^{e_1}F_1)\cdot\prod_{i=2}^{\ell}S^{\left[\frac{a_{i,N}}{a_{1,N}}(n+h)\right]}(S^{e_i}F_{i,N})\cdot S^{n}(S^{e_1}\overline{F}_1)\cdot\prod_{i=2}^{\ell}S^{\left[\frac{a_{i,N}}{a_{1,N}}n\right]}(S^{e_i}\overline{F}_{i,N})\,d\tilde{\mu}\Bigg|.$$

Factoring out the term S^n we get

$$\begin{aligned}A_h&=\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\Bigg|\int S^{h}(S^{e_1}F_1)\cdot S^{e_1}\overline{F}_1\cdot\prod_{i=2}^{\ell}S^{\left[\left(\frac{a_{i,N}}{a_{1,N}}-1\right)n\right]+\left[\frac{a_{i,N}}{a_{1,N}}h\right]+e_j(n,h,N)}(S^{e_i}F_{i,N})\\&\qquad\cdot\prod_{i=2}^{\ell}S^{\left[\left(\frac{a_{i,N}}{a_{1,N}}-1\right)n\right]+\tilde{e}_j(n,h,N)}(S^{e_i}\overline{F}_{i,N})\,d\tilde{\mu}\Bigg|\\&=\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left|\int F_{1,h}\cdot\prod_{i=2}^{\ell}S^{\left[\left(\frac{a_{i,N}}{a_{1,N}}-1\right)n\right]}F_{i,h,n,N}\,d\tilde{\mu}\right|,\end{aligned} \tag{24}$$

where $F_{1,h}=S^h(S^{e_1}F_1)\cdot S^{e_1}\overline{F}_1$, $F_{i,h,n,N}=S^{\left[\frac{a_{i,N}}{a_{1,N}}h\right]+e_j(n,h,N)}(S^{e_i}F_{i,N})\cdot S^{\tilde{e}_j(n,h,N)}(S^{e_i}\overline{F}_{i,N})$, $2\leq i\leq\ell$. Using the hypothesis, for $i=1$, there exists $2\leq j_0\leq\ell$ such that the sequences

$\left\{\left(\frac{a_{j_0,N}-a_{j,N}}{a_{1,N}}\right)_N : 1\leq j\neq j_0\leq \ell\right\}$ have the $R_{\ell-1}$-property. Factoring out $S^{\left[\left(\frac{a_{j_0,N}}{a_{1,N}}-1\right)n\right]}$ in the previous relation we have that

$$\begin{aligned} A_h &= \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left|\int F_{1,h}\cdot\prod_{i=2}^{\ell} S^{\left[\left(\frac{a_{i,N}}{a_{1,N}}-1\right)n\right]}F_{i,h,n,N}\,d\tilde{\mu}\right| \\ &= \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left|\int F_{j_0,h,n,N}\cdot S^{\left[-\left(\frac{a_{j_0,N}}{a_{1,N}}-1\right)n\right]+e_1'(n,N)}F_{1,h}\right. \\ &\qquad\qquad \left.\cdot\prod_{2\leq i\neq j_0\leq \ell} S^{\left[-\left(\frac{a_{j_0,N}-a_{i,N}}{a_{1,N}}\right)n\right]}\tilde{F}_{i,h,n,N}\,d\tilde{\mu}\right|, \end{aligned}$$

where $\tilde{F}_{i,h,n,N}=S^{e_i'(n,N)}F_{i,h,n,N}$ for some error terms $e_i'(n,N)\in\{0,1\}$.

As we previously highlighted, for every fixed N, we can partition the set of integers so that $e_1'(n,N)$ is constant. So, fixing $e_1'\in\{0,1\}$, using the induction hypothesis, we have

$$\begin{aligned} A_h &\ll \limsup_{N\to\infty}\sup_{\|F_0\|_\infty,\|F_2\|_\infty,\dots,\|F_\ell\|_\infty\leq 1}\frac{1}{N}\sum_{n=1}^{N}\left|\int F_0\cdot S^{\left[-\left(\frac{a_{j_0,N}}{a_{1,N}}-1\right)n\right]}(S^{e_1'}F_{1,h})\right. \\ &\qquad\qquad\left.\cdot\prod_{2\leq i\neq j_0\leq\ell} S^{\left[-\left(\frac{a_{j_0,N}-a_{i,N}}{a_{1,N}}\right)n\right]}F_i\,d\tilde{\mu}\right| \\ &\ll |||S^{e_1'}F_{1,h}|||_{2(\ell-1)} = |||F_{1,h}|||_{2(\ell-1)} = |||S^h(S^{e_1}F_1)\cdot S^{e_1}\overline{F}_1|||_{2(\ell-1)} \\ &= |||(T^{h+e_1}f_1\cdot T^{e_1}\overline{f}_1)\otimes\overline{(T^{h+e_1}f_1\cdot T^{e_1}\overline{f}_1)}|||_{2(\ell-1)} \leq |||T^{h+e_1}f_1\cdot T^{e_1}\overline{f}_1|||_{2\ell-1}^2. \end{aligned}$$

So, using Hölder inequality and the definition of the seminorms $|||\cdot|||$, we have

$$\begin{aligned} |A|^2 &\ll \limsup_{H\to\infty}\frac{1}{H}\sum_{h=1}^{H}A_h \ll \limsup_{H\to\infty}\frac{1}{H}\sum_{h=1}^{H}|||T^{h+e_1}f_1\cdot T^{e_1}\overline{f}_1|||_{2\ell-1}^2 \\ &\leq \limsup_{H\to\infty}\left(\frac{1}{H}\sum_{h=1}^{H}|||T^{h+e_1}f_1\cdot T^{e_1}\overline{f}_1|||_{2\ell-1}^{2^{2\ell-1}}\right)^{1/2^{2(\ell-1)}} = |||T^{e_1}f_1|||_{2\ell}^4 = |||f_1|||_{2\ell}^4, \end{aligned}$$

hence, (22) is bounded above by a constant multiple of $|||f_1|||_{2\ell}^2$ as was to be shown.

Cases 2 & 3: For $i=1$, we either have (ii) (b) or (ii) (c) in Definition 5.6.

Here we will skip the details already outlined in Case 1. If $2\leq j_0\leq\ell$ is the one guaranteed by Definition 5.6, the averaging term in the last part of (21) will become (setting, without loss, $e_i=0$)

$$f_{j_0}\cdot T^{[(a_{1,N}-a_{j_0,N})n]}f_1\cdot T^{\left[-\frac{a_{j_0,N}}{a_{1,N}-a_{j_0,N}}[(a_{1,N}-a_{j_0,N})n]\right]}f_0\cdot\prod_{2\leq j\neq j_0\leq\ell}T^{\left[\frac{a_{j,N}-a_{j_0,N}}{a_{1,N}-a_{j_0,N}}[(a_{1,N}-a_{j_0,N})n]\right]}f_j$$

and the one in (24)

$$\int F_{1,h} \cdot S^{\left[\left(\frac{-a_{j_0,N}}{a_{1,N}-a_{j_0,N}}-1\right)n\right]} F_{0,h,n,N} \cdot \prod_{2\leq j\neq j_0\leq \ell} S^{\left[\left(\frac{a_{j,N}-a_{j_0,N}}{a_{1,N}-a_{j_0,N}}-1\right)n\right]} F_{i,h,n,N}\, d\tilde{\mu}.$$

Factoring out $S^{\left[\left(\frac{-a_{j_0,N}}{a_{1,N}-a_{j_0,N}}-1\right)n\right]} = S^{\left[\frac{a_{1,N}}{a_{1,N}-a_{j_0,N}}n\right]}$ (for Case 2) and $S^{\left[\left(\frac{a_{k_0,N}-a_{j_0,N}}{a_{1,N}-a_{j_0,N}}-1\right)n\right]} = S^{\left[\frac{a_{k_0,N}-a_{1,N}}{a_{1,N}-a_{j_0,N}}n\right]}$ (for Case 3–where $2 \leq k_0 \neq j_0 \leq \ell$ is the one guaranteed by Definition 5.6), we can continue (using the induction hypothesis) and finish the argument as in Case 1. The proof of the statement is now complete. □

Remark 5.9. *To the best of our knowledge, when we deal with norm convergence of averages with (non-variable) polynomial iterates, we can always replace the conventional Cesàro averages, i.e.,* $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}$, *with uniform ones, i.e.,* $\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N-1}$. *Our method though, exactly because of the choice of functions* $f_{i,[a_{1,N}N]}$ *(to go from Equation* (18) *to* (19) *and from* (21) *to* (22)*), cannot give the corresponding uniform results.*

5.2. **The general case.** We start by recalling (see, for example, [4] and [13]) the definition of the degree and type of a polynomial family that we will adapt in our study:

Definition 5.10. For $\ell \in \mathbb{N}$ let $\mathcal{P} = \{p_1, \ldots, p_\ell\}$ be a family of non-constant real polynomials. We denote with $\deg(\mathcal{P})$ the maximum degree of the p_i's and we call it *degree* of $\mathcal{P}$. If w_i denotes the number of distinct leading coefficients of polynomials from $\mathcal{P}$ of degree i and $d = \deg(\mathcal{P})$, then the vector $(d, w_d, \ldots, w_1)$ is the *type* of $\mathcal{P}$. We order all the possible type vectors lexicographically.[32]

In order to reduce the complexity (i.e., the type) of a polynomial family, one has to use the classic PET induction,[33] which was introduced in [4].

At this point we remind the reader that the real polynomials $p_1, \ldots, p_\ell$ are called *essentially distinct* if they are, together with their pairwise differences, non-constant. Given such a family of polynomials $\mathcal{P} = \{p_1, \ldots, p_\ell\}$, $p \in \mathcal{P}$ and $h \in \mathbb{N}$, the *van der Corput operation (vdC-operation)*, acting on $\mathcal{P}$, gives the family

$$\mathcal{P}(p, h) := \{p_1(t+h) - p(t), \ldots, p_\ell(t+h) - p(t), p_1(t) - p(t), \ldots, p_\ell(t) - p(t)\},^{34}$$

whereby we remove all the terms that are bounded[35] and we group the ones of degree 1 with bounded difference (i.e., of the same leading coefficient), thus obtaining a new family of essentially distinct polynomials.

The following lemma acknowledges that there exists a choice of a polynomial in a family of essentially distinct polynomials, via which the vdC-operation reduces its type:

Lemma 5.11 (Lemma 4.5, [13]). *Let $\ell \in \mathbb{N}$ and $\mathcal{P} = \{p_1, \ldots, p_\ell\}$ be a family of essentially distinct polynomials with* $\deg(\mathcal{P}) = \deg(p_1) \geq 2$. *Then there exists $p \in \mathcal{P}$ (of minimum*

[32] I.e., $(d, w_d, \ldots, w_1) > (d', w'_d, \ldots, w'_1)$ iff, reading from left to right, the first instance where the two vectors disagree the coordinate of the first vector is greater than that of the second one.

[33] I.e., Polynomial Exhaustion Technique.

[34] Notice that if $\mathcal{P}$ lists the polynomial iterates in the expression $T^{[p_1(n)]}f_1 \cdot \ldots \cdot T^{[p_\ell(n)]}f_\ell$, then $\mathcal{P}(p,h)$ lists the respective iterates (modulo error terms) after using Lemma 5.5 and factoring out the iterate $-p(n)$.

[35] This is justified with the use of the Cauchy-Schwarz inequality.

degree in the polynomial family) such that for every large h the family $\mathcal{P}(p,h)$ has type smaller than that of $\mathcal{P}$, and $\deg(\mathcal{P}(p,h)) = \deg(p_1(t+h) - p(t))$.

What is crucial for us is that every decreasing sequence of types is eventually (after finitely many steps) stationary and that, by using the previous lemma, there is a point that all the polynomials have degree 1 (hence, one is able to use the base case, finishing the inductive scheme). Also, by its definition, the vdC-operation preserves the essential distinctness property.

Switching to variable polynomials, notice that, for all $N \in \mathbb{N}$, the type in Examples 1 and 2 is $(1,2)$ and $(\ell, 1, \dots, 1)$ respectively, i.e., both independent of N. This is not always the case though.

Indeed, consider the pairs $(p_{1,N}, p_{2,N})_N$, and $(q_{1,N}, q_{2,N})_N$, where, for $N \in \mathbb{N}$ we have

$$p_{1,N}(t) = (1/N)t^4 + \sqrt{2}t^3 + t^2, \text{ and } p_{2,N}(t) = ((-1)^N/N)t^4 + \sqrt{3}t^3 - t$$

and

$$q_{1,N}(t) = (1/N^2)t^4 + \sqrt{2}t^3 + t^2, \text{ and } q_{2,N}(t) = (1/N)t^4 + \sqrt{3}t^3 - t.$$

In the first case the type of the pair is $(4,1,0,0,0)$ for N even and $(4,2,0,0,0)$ otherwise, while the type of the second one equals to $(4,1,0,0,0)$ for $N = 1$ and $(4,2,0,0,0)$ for $N > 1$, i.e., has the property that it eventually (for $N > 1$) becomes constant.

We will deal with sequences of families of real polynomials, $(\mathcal{P}_N)_N$, where $\mathcal{P}_N = \{p_{1,N}, \dots, p_{\ell,N}\}$, $N \in \mathbb{N}$,[36] that, for large N,[37] has type independent of N, to be able to use the facts that we just mentioned.

Next we will define the subclass of variable polynomials that we will deal with.

Definition 5.12. For $\ell \in \mathbb{N}$ let $(\mathcal{P}_N)_N = (p_{1,N}, \dots, p_{\ell,N})_N$ be a sequence of ℓ-tuples of real polynomials with bounded coefficients. We say that $(\mathcal{P}_N)_N$ is *super nice* if, for every (large enough) $N \in \mathbb{N}$:

(i) the polynomials $p_{i,N}$ and, for all $i \neq j$, $p_{i,N} - p_{j,N}$ are non-constant and their degrees are independent of N;

(ii) after performing, if needed, (finitely many) vdC-operations to $(\mathcal{P}_N)_N$ to end up only with polynomials of degree 1, say $k \equiv k((\mathcal{P}_N)_N)$ many, the leading coefficients (for large enough h_i's–from the vdC-operations) have the R_k-property; and

(ii)$'$ if $\deg((p_{i_0,N})_N) = \deg((\mathcal{P}_N)_N)$, then (ii) holds for the polynomial sequence $(\mathcal{P}'_N)_N := (p_{1,N} - p_{i_0,N}, \dots, p_{i_0-1,N} - p_{i_0,N}, -p_{i_0,N}, p_{i_0+1,N} - p_{i_0,N}, \dots, p_{\ell,N} - p_{i_0,N})_N$.[38]

In the following remark we list interesting facts on the super niceness property, some of which will be used in what follows.

[36]Abusing the notation we will refer to such sequences as (variable) sequences of ℓ-tuples of real polynomials and we will denote them with $(\mathcal{P}_N)_N = (p_{1,N}, \dots, p_{\ell,N})_N$.

[37]From now on whatever we write about variable polynomials we understand it for "large" enough N.

[38]It is not clear whether (ii) implies $(ii)'$. Indeed, consider for example the polynomials $p_{1,N}(n) = -a_N n^2 - b_N n, p_{2,N}(n) = (a_N - b_N)n$ and $(\mathcal{P}_N)_N = (p_{1,N}, p_{2,N})_N$. After performing two vdC-operations we get the triple $\{-2a_N(h'+h)n, -2a_N hn, -2a_N h'n\}$, while for the $(\mathcal{P}'_N)_N = (-p_{1,N}, p_{2,N} - p_{1,N})_N$, after a single vdC-operation we get $\{(a_N - b_N)n, 2a_N hn, ((2h+1)a_N - b_N)n\}$. So, in the second case we have to impose assumptions on both $(a_N)_N$, $(b_N)_N$, while in the first one only on $(a_N)_N$.

Remark 5.13. (1) *The degree and type of every super nice sequence, together with the integer k in (ii) (and, analogously, in $(ii)'$ as well), are independent of N.*

(2) *Every independent of N family $\mathcal{P}=\{p_1,\ldots,p_\ell\}$ of essentially distinct polynomials is super nice. (This shows that the set of super nice families of non-variable polynomials is not empty.)*

Indeed, since (i) is immediate, we are showing (ii) ($(ii)'$ follows with exactly the very same argument). As it was mentioned before, the vdC-operation preserves the essential distinctness property, hence, all the k linear polynomial will have distinct leading coefficients, which, as they are independent of N, will have the $\mathcal{R}_k$-property.

(3) *The set of super nice variable polynomial sequences is non-empty. Actually, the ℓ-tuple $(p_{1,N},\ldots,p_{\ell,N})_N$, where $p_{i,N}(n)=n^i/N^a$, $1\leq i\leq \ell$, $N,n\in\mathbb{N}$, and $0<a<1$, from Example 2, is super nice (see Lemma 8.2 below for a more general statement).*

Indeed, as the variable part of the coefficients of the polynomials, after applying vdC-operations, is the same for all terms (and equal to $1/N^a$), at each step we have that the ratios of the coefficients are independent of N, hence we have all the properties.

(4) *Even though the number k of degree 1 terms (that appears in (ii) and $(ii)'$) is not a priori known, when we have a single variable polynomial sequence*

$$p_N(n)=a_{d,N}n^d+\ldots+a_{1,N}n+a_{0,N},$$

where $(a_{d,N})_N$ has the R_1-property and all $(a_{i,N})_N$, $0\leq i\leq d$, are bounded, we have that for all $\ell\in\mathbb{N}$, $(\mathcal{P}_N)_N=(p_N,2p_N,\ldots,\ell p_N)_N$ is super nice (without having to know the corresponding k).

Given (i) is immediate, it suffices to show (ii) ((ii)$'$ follows with exactly the same argument–the important part here is that $(\mathcal{P}_N)_N$ consists of distinct non-zero multiples of the same polynomial). We start with $(\ell p_N(n),\ldots,\ p_N(n))$ (we write it in this order for convenience) and use the vdC-operation which leads to differences of polynomials (derivatives).[39] Factoring out $p_N(n)$ we get

$$(\ell-1)p_N(n+h_1)+p_N(n+h_1)-p_N(n),\ldots,p_N(n+h_1)-p_N(n),(\ell-1)p_N(n),\ldots,p_N(n).$$

In the next iteration of the vdC-operation we factor out $p_N(n+h_1)-p_N(n)$, and then $p_N(n)-(p_N(n+h_1)-p_N(n))=p_N(n+h_1)$ (i.e., polynomials of minimum degree at each step), reducing by 1 the appearance of each p_N in the corresponding expression. Continuing the procedure, we eventually arrive at, say k many, degree 1 iterates with distinct leading coefficients (by the essential distinctness property of the initial polynomials), which are all multiples of $d!\cdot a_{d,N}$. As all the iterated ratios of these coefficients are independent of $a_{d,N}$ and non-zero, we get that they satisfy the R_k-property, completing the claim.

(5) *If $(\mathcal{P}_N)_N$ is super nice, then $(\mathcal{P}'_N)_N$ is super nice too.*

[39]Note that these differences lead to reduction of degrees, as for example $\Delta_1(p(n);h_1):=p(n+h_1)-p(n)$, and $\Delta_2(p(n);h_1,h_2):=\Delta_1(\Delta_1(p(n);h_1);h_2)=p(n+h_1+h_2)-p(n+h_2)-(p(n+h_1)-p(n))$ reduce the degree of p by 1 and 2 respectively.

Looking at Property (i) for $(\mathcal{P}_N)_N$, we have that each polynomial (sequence) in $(\mathcal{P}'_N)_N$ is non-constant and has degree independent of N, equal to $\deg(p_{i_0,N} - p_{1,N})$. If we let

$$q_{i,N} := \begin{cases} p_{i,N} - p_{i_0,N} & i \neq i_0 \\ -p_{i_0,N} & i = i_0 \end{cases}, \text{ then, we have } q_{i,N} - q_{j,N} = \begin{cases} p_{i,N} - p_{j,N} & i,j \neq i_0 \\ p_{i,N} & j = i_0 \\ -p_{j,N} & i = i_0 \end{cases},$$

and (i) follows for $(\mathcal{P}'_N)_N$. (ii) and $(ii)'$ follow by the fact that $((\mathcal{P}')'_N)_N = (\mathcal{P}_N)_N$.

(6) *Property (i) is invariant under the vdC-operation.*[40]

Indeed, if $p_{i_0,N}$ is the polynomial guaranteed by Lemma 5.11, then we have the iterates:

$$p_{1,N}(n+h) - p_{i_0,N}(n), \dots, p_{\ell,N}(n+h) - p_{i_0,N}(n), \text{ and}$$

$$p_{1,N}(n) - p_{i_0,N}(n), \dots, p_{i_0-1,N}(n) - p_{i_0,N}(n), p_{i_0+1,N}(n) - p_{i_0,N}(n), \dots, p_{\ell,N}(n) - p_{i_0,N}(n).$$

The degrees of these polynomials satisfy

$$\deg(p_{i,N}(n+h) - p_{i_0,N}(n)) = \begin{cases} \deg(p_{i,N}(n) - p_{i_0,N}(n)) & i \neq i_0 \\ \deg(p_{i_0,N}) - 1 & i = i_0 \end{cases}.^{41}$$

For the pairwise differences part, for $i \neq j$, we have

$$p_{i,N}(n+h) - p_{i_0,N}(n) - (p_{j,N}(n+h) - p_{i_0,N}(n)) = p_{i,N}(n+h) - p_{j,N}(n+h),$$

$$p_{i,N}(n) - p_{i_0,N}(n) - (p_{j,N}(n) - p_{i_0,N}(n)) = p_{i,N}(n) - p_{j,N}(n),$$

and, finally,

$$p_{i,N}(n+h) - p_{i_0,N}(n) - (p_{j,N}(n) - p_{i_0,N}(n)) = p_{i,N}(n+h) - p_{j,N}(n),$$

so, everything follows by Property (i) for $(\mathcal{P}_N)_N$.[42]

Notice that Remark 5.13 (4) implies that Theorem 1.5, via Theorem 2.2, holds for a larger class of variable polynomial sequences; even with coefficients that oscillate.

A real-valued function g which is continuously differentiable on $[c, \infty)$, where $c \geq 0$, is called *Fejér* if the following hold: $g'(x)$ tends monotonically to 0 as $x \to \infty$; and $\lim_{x\to\infty} x|g'(x)| = \infty$.[43] Any such function is eventually monotonic and satisfies the growth conditions $\log x \prec g(x) \prec x$, hence $(1/g(N))_N$ has the R_1-property. So, modulo the goodness property, Theorem 2.2 will also hold for polynomial sequences of the form:

$$\left(\frac{\sqrt{5}}{N^{1/2}(2+\cos\sqrt{\log N})}n^3 + p_{1,N}(n)\right)_N, \text{ or } \left(\frac{7}{N^{1/40}(1/10+\sin\log N)^3}n^{17} + p_{2,N}(n)\right)_N,$$

where $p_{1,N}$, $p_{2,N}$ are polynomials of degrees less than 3 and 17 respectively with bounded coefficients, while the functions $g_1(x) = x^{1/2}(2+\cos\sqrt{\log x})$, $g_2(x) = x^{1/40}(1/10+\sin\log x)^3$ are Fejér but, because of oscillation, don't belong to $\mathcal{SLE}$.

[40] So, for sequences $(\mathcal{P}_N)_N$ with degree ≥ 2, the vdC-operation preserves the super niceness property.

[41] Notice that the vdC-operation will remove the $p_{i_0,N}(n+h) - p_{i_0,N}(n)$ iterate in case $\deg(p_{i_0,N}) = 1$.

[42] Recall here that if, in the case where $i = j$, it happens $\deg(p_{i,N}) = 1$, then $\deg(p_{i_0,N}) = 1$ (as a non-constant polynomial of minimum degree in $(\mathcal{P}_N)_N$), so the vdC-operation will group the terms $p_{i,N}(n+h) - p_{i_0,N}(n)$ and $p_{i,N}(n) - p_{i_0,N}(n)$ together, being of degree 1 with bounded difference.

[43] For a deep study of averages with general sublinear iterates one is referred to [9], while for more general such functions, i.e., tempered ones, to [5] and [31].

The following result shows that the nilfactor $\mathcal{Z}$ is characteristic for a super nice collection of polynomial sequences $(p_{1,N},\dots,p_{\ell,N})_N$.

Proposition 5.14. *For $\ell\in\mathbb{N}$ let $(p_{1,N},\dots,p_{\ell,N})_N$ be a super nice sequence of polynomials, $(X,\mathcal{B},\mu,T)$ a system, and suppose that at least one of the functions $f_1,\dots,f_\ell\in L^\infty(\mu)$ is orthogonal to the nilfactor $\mathcal{Z}$. Then, we have*

$$\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^N T^{[p_{1,N}(n)]}f_1\cdot\ldots\cdot T^{[p_{\ell,N}(n)]}f_\ell\right\|_2=0. \tag{25}$$

Proof. We assume without loss of generality that f_1 is orthogonal to $\mathcal{Z}$. As in [13, Lemma 4.7], to show (25), it suffices to show:

$$\lim_{N\to\infty}\sup_{\|f_0\|_\infty,\|f_2\|_\infty,\dots,\|f_\ell\|_\infty\le 1}\frac{1}{N}\sum_{n=1}^N\left|\int f_0\cdot T^{[p_{1,N}(n)]}f_1\cdot\ldots\cdot T^{[p_{\ell,N}(n)]}f_\ell\;d\mu\right|=0. \tag{26}$$

We claim next that we can further assume that $\deg(p_{1,N})=\deg(\mathcal{P}_N)$. If this is not the case and $\deg(p_{1,N})<\deg(p_{i_0,N})=\deg(\mathcal{P}_N)$, then, factoring out $T^{[p_{i_0,N}(n)]}$, (26) becomes

$$\lim_{N\to\infty}\sup_{\|f_0\|_\infty,\|f_2\|_\infty,\dots,\|f_\ell\|_\infty\le 1}\frac{1}{N}\sum_{n=1}^N\Bigg|\int f_{i_0}\cdot T^{[-p_{i_0,N}(n)]}f_{0,n,N}\cdot\prod_{1\le i\neq i_0\le\ell}T^{[p_{i,N}(n)-p_{i_0,N}(n)]}f_{i,n,N}\;d\mu\Bigg|=0,$$

where $f_{i,n,N}=T^{e_i(n,N)}f_i$ for some error terms $e_i(n,N)\in\{0,1\}$. So, it suffices to show, for any choice of $e\in\{0,1\}$, that

$$\lim_{N\to\infty}\sup_{\|f_0\|_\infty,\|f_2\|_\infty,\dots,\|f_\ell\|_\infty\le 1}\frac{1}{N}\sum_{n=1}^N\Bigg|\int f_{i_0}\cdot T^{[-p_{i_0,N}(n)]}f_0\cdot T^{[p_{1,N}(n)-p_{i_0,N}(n)]}(T^ef_1)\cdot\prod_{2\le i\neq i_0\le\ell}T^{[p_{i,N}(n)-p_{i_0,N}(n)]}f_i\;d\mu\Bigg|=0.$$

With Remark 5.13 (5), we have that the polynomial family $(p_{i_0,N}-p_{1,N},\dots,p_{i_0,N}-p_{i_0-1,N},$ $p_{i_0,N},p_{i_0,N}-p_{i_0+1,N},\dots,p_{i_0,N}-p_{\ell,N})_N$ is super nice with degree$=\deg(p_{i_0,N}-p_{1,N})$, from which the claim follows.

In case all the $p_{i,N}$'s are of degree 1, the result follows from Proposition 5.8.
Assume now that $\deg(p_{1,N})\ge 2$. We will use induction on the type of the polynomial family of ℓ-tuple of sequences.

For every $N\in\mathbb{N}$ we choose functions $f_{i,N}$ with $\|f_{i,N}\|_\infty\le 1$ for $i\in\{0,2,\dots,\ell\}$, so that the average in (26) is $1/N$ close to the corresponding $\sup_{\|f_0\|_\infty,\|f_2\|_\infty,\dots,\|f_\ell\|_\infty\le 1}$. Using the Cauchy-Schwarz inequality, we have that (26) follows if we show

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N\left|\int f_{0,N}\cdot T^{[p_{1,N}(n)]}f_1\cdot\prod_{i=2}^\ell T^{[p_{i,N}(n)]}f_{i,N}\;d\mu\right|^2=0. \tag{27}$$

Setting $S = T \times T$, $F_1 = f_1 \otimes \overline{f}_1$, $F_{i,N} = f_{i,N} \otimes \overline{f}_{i,N}$, $i = 0, 2, \dots, \ell$, and $\tilde{\mu} = \mu \times \mu$, (27) can be rewritten as

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int F_{0,N} \cdot S^{[p_{1,N}(n)]} F_1 \cdot \prod_{i=2}^{\ell} S^{[p_{i,N}(n)]} F_{i,N} \, d\tilde{\mu} = 0.$$

So, after using Cauchy-Schwarz, it suffices to show

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} S^{[p_{1,N}(n)]} F_1 \cdot \prod_{i=2}^{\ell} S^{[p_{i,N}(n)]} F_{i,N} \right\|_{L^2(\tilde{\mu})} = 0. \tag{28}$$

Using Lemma 5.5, (28) follows if, for large enough h, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Big| \int S^{[p_{1,N}(n+h)]} F_1 \cdot \prod_{i=2}^{\ell} S^{[p_{i,N}(n+h)]} F_{i,N} \cdot S^{[p_{1,N}(n)]} \overline{F}_1 \cdot \prod_{i=2}^{\ell} S^{[p_{i,N}(n)]} \overline{F}_{i,N} \, d\tilde{\mu} \Big| = 0.$$

Picking $p_{j_0,N}$ as guaranteed by Lemma 5.11 (note that since the degrees of the $p_{i,N}$'s are fixed, the choice of j_0 is independent of N), factoring out $S^{[p_{j_0,N}(n)]}$, the previous relation becomes

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \Big| \int \overline{F}_{j_0,N} \cdot S^{[p_{1,N}(n+h)-p_{j_0,N}(n)]+e_1(n,h,N)} F_1 \cdot \prod_{i=2}^{\ell} S^{[p_{i,N}(n+h)-p_{j_0,N}(n)]+e_i(n,h,N)} F_{i,N} \cdot S^{[p_{1,N}(n)-p_{j_0,N}(n)]+\tilde{e}_1(n,N)} \overline{F}_1 \cdot \prod_{2\leq i\neq j_0\leq \ell} S^{[p_{i,N}(n)-p_{j_0,N}(n)]+\tilde{e}_i(n,N)} \overline{F}_{i,N} \, d\tilde{\mu} \Big| = 0, \tag{29}$$

for some error terms $e_i(n,h,N), \tilde{e}_i(n,N) \in \{0,1\}$.

Next, we group the degree 1 iterates. More specifically, if $\deg(p_{i,N}) = 1$ for some $2 \leq i \leq \ell$ (recall that $\deg(p_{1,N}) \geq 2$), then $p_{i,N}(n+h) = p_{i,N}(n) + c_{i,N}h$, so $[p_{i,N}(n+h)] = [p_{i,N}(n)] + [c_{i,N}h] + e_i'(n,h,N)$, for some error terms in $\{0,1\}$. Hence, in such a case, we have

$$\begin{aligned} &S^{[p_{i,N}(n+h)-p_{j_0,N}(n)]+e_i(n,h,N)} F_{i,N} \cdot S^{[p_{i,N}(n)-p_{j_0,N}(n)]+\tilde{e}_i(n,h,N)} \overline{F}_{i,N} \\ &= S^{[p_{i,N}(n)-p_{j_0,N}(n)]} \big(S^{[c_{i,N}h]+e_i(n,h,N)+e_i'(n,h,N)} F_{i,N} \cdot S^{\tilde{e}_i(n,h,N)} \overline{F}_{i,N} \big), \end{aligned}$$

and we treat this product as one iterate. After this grouping, assuming that the remaining terms in (29) are r many, it suffices to show, for large enough h, and every choice of

$e \in \{0,1\}$ that

$$\lim_{N\to\infty} \sup_{\|F_0\|_\infty, \|F_2\|_\infty, \dots, \|F_r\|_\infty \le 1} \Big| \frac{1}{N}\sum_{n=1}^N \int F_0 \cdot S^{[p_{1,h,N}(n)]}(S^e F_1) \cdot \prod_{i=2}^r S^{[p_{i,h,N}(n)]} F_i \, d\tilde{\mu} \Big| = 0, \tag{30}$$

where the polynomial sequences $(p_{i,h,N})_N$ form the polynomial family $(\mathcal{P}_N(p_{j_0,N},h))_N$, with $p_{1,h,N}(n) = p_{1,N}(n+h) - p_{j_0,N}(n)$ and $\deg(p_{1,h,N}) = \deg(\mathcal{P}_N(p_{j_0,N},h))$.

(30) is as (26), with the polynomial family of the latter having type strictly less than the former (from Lemma 5.11). Using Remark 5.13 (6), we are done by induction. □

For the expression of Theorem 2.2, using Proposition 5.14, we get the following result:

Proposition 5.15. *For $\ell \in \mathbb{N}$ let $(p_N, 2p_N, \dots, \ell p_N)_N$ be a super nice sequence of polynomials, $(X, \mathcal{B}, \mu, T)$ a system, and suppose that at least one of the functions $f_1, \dots, f_\ell \in L^\infty(\mu)$ is orthogonal to the nilfactor $\mathcal{Z}$. Then, we have*

$$\lim_{N\to\infty} \Big\| \frac{1}{N}\sum_{n=1}^N T^{[p_N(n)]} f_1 \cdot T^{2[p_N(n)]} f_2 \cdot \ldots \cdot T^{\ell[p_N(n)]} f_\ell \Big\|_2 = 0. \tag{31}$$

Proof. We can rewrite (31) as

$$\lim_{N\to\infty} \Big\| \frac{1}{N}\sum_{n=1}^N T^{[p_N(n)]} f_1 \cdot T^{[2p_N(n)]+e_{2,N}(n)} f_2 \cdot \ldots \cdot T^{[\ell p_N(n)]+e_{\ell,N}(n)} f_\ell \Big\|_2 = 0,$$

for some bounded error terms $e_{i,N}(n) \in \{-i, \dots, -1, 0\}$, $2 \le i \le \ell$. The result now follows by the proof of Proposition 5.14 as $(p_N, 2p_N, \dots, \ell p_N)_N$ is a super nice sequence. □

6. Equidistribution

In order to prove our main equidistribution result (Theorem 6.2), we start with some definitions and facts, following [12] (see [12, Subsubsection 2.3.2] for more details).

If G is a nilpotent group, then a sequence $g : \mathbb{N} \to G$ of the form $g(n) = b_1^{p_1(n)} \cdots b_k^{p_k(n)}$, where $b_i \in G$, and p_i are integer polynomials, is called a *polynomial sequence in G*. If the maximum degree of the p_i's is at most d we say that the *degree* of $g(n)$ is at most d.

Given a nilmanifold $X = G/\Gamma$ the *horizontal torus* is defined to be the compact abelian group $Z = G/([G,G]\Gamma)$. If X is connected, then Z is isomorphic to some finite dimensional torus $\mathbb{T}^s$. A *horizontal character* $\chi : G \to \mathbb{C}$ is a continuous homomorphism that satisfies $\chi(g\gamma) = \chi(g)$ for every $\gamma \in \Gamma$ and can be thought of as a character of $\mathbb{T}^s$, in which case there exists a unique $\kappa \in \mathbb{Z}^s$ such that $\chi(t\mathbb{Z}^s) = e(\kappa \cdot t)$, where "$\cdot$" denotes the inner product operation, and $e(x) := e^{2\pi i x}$.

Let $p : \mathbb{Z} \to \mathbb{R}$ be a polynomial sequence of degree d of the form $p(n) = \sum_{i=0}^{d} a_i n^i$, where $a_i \in \mathbb{R}$, $1 \le i \le d$. Recalling that $\|\cdot\| = d(\cdot, \mathbb{Z})$, we define the *smoothness norm*

$$\|e(p(n))\|_{C^\infty[N]} := \max_{1\le i\le d}(N^i \|a_i\|).^{44} \tag{32}$$

Given $N \in \mathbb{N}$, a finite sequence $(g(n)\Gamma)_{1\le n\le N}$ is said to be *δ-equidistributed in X*, if

$$\left|\frac{1}{N}\sum_{n=1}^{N} F(g(n)\Gamma) - \int_X F d_{m_X}\right| \le \delta \|F\|_{\mathrm{Lip}(X)}$$

for every Lipschitz function $F : X \to \mathbb{C}$, where

$$\|F\|_{\mathrm{Lip}(X)} = \|F\|_\infty + \sup_{x,y\in X, x\ne y} \frac{|F(x) - F(y)|}{d_X(x, y)}$$

for some appropriate metric d_X.

At this point we quote [12, Theorem 2.9], a direct consequence of [22, Theorem 2.9]:

Theorem 6.1 (Green & Tao, [22]). *Let $X = G/\Gamma$ be a nilmanifold with G connected and simply connected, and $d \in \mathbb{N}$. Then for every small enough $\delta > 0$ there exist a positive constant $M \equiv M(X, d, \delta)$ with the following property: For every $N \in \mathbb{N}$, if $g : \mathbb{Z} \to G$ is a polynomial sequence of degree at most d such that the finite sequence $(g(n)\Gamma)_{1\le n\le N}$ is not δ-equidistributed, then for some non-trivial horizontal character χ with $\|\chi\| \le M$ we have*

$$\|\chi(g(n))\|_{C^\infty[N]} \le M$$

(χ here is thought of as a character of the horizontal torus $Z = \mathbb{T}^s$ and $g(n)$ as a polynomial sequence in $\mathbb{T}^s$).

Adapting the notion of equidistribution of a sequence in a nilmanifold (recall (17)) to our case, abusing the notation, we say that $(b^{a_N(n)}x)_{1\le n\le N}$, where $(a_N(n))_{1\le n\le N}$ is a variable sequence of real numbers and $X = G/\Gamma$ is a nilmanifold with G connected and simply connected, is *equidistributed* in a subnilmanifold Y of X, if for every $F \in C(X)$ we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} F(b^{a_N(n)}x) = \int F \, d_{m_Y}.$$

In order for us to prove Theorems 2.1 and 2.2, we prove the following equidistribution theorem, which is the main result of this section:

Theorem 6.2. *Let $(p_{1,N}, \dots, p_{\ell,N})_N$ be a good sequence of ℓ-tuples of polynomials.*

(i) *If $X_i = G_i/\Gamma_i$, $1 \le i \le \ell$, are nilmanifolds with G_i connected and simply connected, then for every $b_i \in G_i$ and $x_i \in X_i$ the sequence*

$$(b_1^{p_{1,N}(n)}x_1, \dots, b_\ell^{p_{\ell,N}(n)}x_\ell)_{1\le n\le N}$$

is equidistributed in the nilmanifold

$$\overline{(b_1^s x_1)}_{s\in\mathbb{R}} \times \cdots \times \overline{(b_\ell^s x_\ell)}_{s\in\mathbb{R}}.$$

[44] There is an alternative, and equivalent, definition for this norm (see [22, Definition 2.7]). We can rewrite $p(n) = \sum_{i=0}^{d} a_i' \binom{n}{i}$ and define $\|e(p(n))\|'_{C^\infty[N]} := \max_{1\le i\le k}(N^i \|a_i'\|)$. By ([16, Section 5]), there are positive constants $c \equiv c(d)$ and $C \equiv C(d)$ with $c\|e(p(n))\|'_{C^\infty[N]} \le \|e(p(n))\|_{C^\infty[N]} \le C\|e(p(n))\|'_{C^\infty[N]}$.

(ii) *If $X_i = G_i/\Gamma_i$, $1 \leq i \leq \ell$, are nilmanifolds, then for every $b_i \in G_i$ and $x_i \in X_i$ the sequence*
$$(b_1^{[p_{1,N}(n)]}x_1, \ldots, b_\ell^{[p_{\ell,N}(n)]}x_\ell)_{1\leq n\leq N}$$
is equidistributed in the nilmanifold
$$\overline{(b_1^n x_1)}_n \times \cdots \times \overline{(b_\ell^n x_\ell)}_n.$$

Remark 6.3. *In order to prove Theorem 6.2, we can assume that $X_1 = \ldots = X_\ell = X$.*

Indeed, in the general case we consider the nilmanifold $\tilde{X} = X_1 \times \cdots \times X_\ell$. Then $\tilde{X} = \tilde{G}/\tilde{\Gamma}$, where $\tilde{G} = G_1 \times \cdots \times G_\ell$ is connected and simply connected and $\tilde{\Gamma} = \Gamma_1 \times \cdots \times \Gamma_\ell$ is a discrete cocompact subgroup of $\tilde{G}$. Each b_i can be considered as an element of $\tilde{G}$ and each x_i as an element of $\tilde{X}$. Changing the base point we can also assume that $x = \Gamma$.

The following lemma, which is analogous to [12, Lemma 5.1], shows that Part (ii) of the previous result follows from Part (i).

Lemma 6.4. *Let $\ell \in \mathbb{N}$ and $(a_{1,N}, \ldots, a_{\ell,N})_N$ be sequence of ℓ-tuples of real numbers. Suppose that for every nilmanifold $X = G/\Gamma$, with G connected and simply connected, and every $b_1, \ldots, b_\ell \in G$ the sequence*
$$(b_1^{a_{1,N}(n)}\Gamma, \ldots, b_\ell^{a_{\ell,N}(n)}\Gamma)_{1\leq n\leq N}$$
is equidistributed in the nilmanifold
$$\overline{(b_1^s\Gamma)}_{s\in\mathbb{R}} \times \cdots \times \overline{(b_\ell^s\Gamma)}_{s\in\mathbb{R}}.$$
Then, for every nilmanifold $X = G/\Gamma$, $b_1, \ldots, b_\ell \in G$ and $x_1, \ldots, x_\ell \in X$, the sequence
$$(b_1^{[a_{1,N}(n)]}x_1, \ldots, b_\ell^{[a_{\ell,N}(n)]}x_\ell)_{1\leq n\leq N}$$
is equidistributed in the nilmanifold
$$\overline{(b_1^n x_1)}_n \times \cdots \times \overline{(b_\ell^n x_\ell)}_n.$$

Sketch of the proof. Following [12, Lemma 4.1], we show the $\ell = 1$ case, as the general one follows by straightforward modifications.

Let $X = G/\Gamma$ be a nilmanifold, $b \in G$ and $x \in X$. Using some standard reductions (namely, the lifting argument and the change of base point formula), we can and will assume that G is connected and simply connected and that $x = \Gamma$.

Letting $X_b := \overline{(b^n\Gamma)}_n$ and m_{X_b} the corresponding normalized Haar measure, we will show that for every $F \in C(X)$ we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N F(b^{[a_N(n)]}\Gamma) = \int_{X_b} F\; dm_{X_b}. \tag{33}$$

Using our assumption for the case $\tilde{X} := \tilde{G}/\tilde{\Gamma}$, where $\tilde{G} := \mathbb{R} \times G$ is connected and simply connected, $\tilde{\Gamma} := \mathbb{Z} \times \Gamma$ and $\tilde{b} := (1, b)$, for every $H \in C(\tilde{X})$ we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N H(\tilde{b}^{a_N(n)}\tilde{\Gamma}) = \int_{\tilde{X}_{\tilde{b}}} H\; dm_{\tilde{X}_{\tilde{b}}}, \tag{34}$$

where $\tilde{X}_{\tilde{b}} := \overline{(s\mathbb{Z}, b^s\Gamma)}_{s\in\mathbb{R}}$ and $m_{\tilde{X}_{\tilde{b}}}$ the corresponding normalized Haar measure.[45]

Let $F \in C(X)$, and define $\tilde{F} : \tilde{X} \to \mathbb{C}$ with $\tilde{F}(t\mathbb{Z}, g\Gamma) := F(b^{-\{t\}}g\Gamma)$. While $\tilde{F}$ may be discontinuous, for every $0 < \delta < 1/2$ there exists $\tilde{F}_\delta \in C(\tilde{X})$ that equals to $\tilde{F}$ on $\tilde{X}_\delta = I_\delta \times X$, where $I_\delta = \{t\mathbb{Z} : \ \|t\| \geq \delta\}$, and it is uniformly bounded by $2\|F\|_\infty$.

Since $\tilde{b}^{a_N(n)} = (a_N(n), b^{a_N(n)})$, our assumption implies (see also the argument that we used to get Lemma 3.2 from Lemma 3.1) that $a_N(n)\mathbb{Z} \in I_\delta$, and so $\tilde{b}^{a_N(n)}\tilde{\Gamma} \in \tilde{X}_\delta$, for a set of n's with density $1 - 2\delta$.[46] So,

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{\infty} |\tilde{F}(\tilde{b}^{a_N(n)}\tilde{\Gamma}) - \tilde{F}_\delta(\tilde{b}^{a_N(n)}\tilde{\Gamma})| \leq 4\delta \|F\|_\infty ,$$

hence, since (34) holds for every $\tilde{F}_\delta$, it also holds for $\tilde{F}$.

The map $(s\mathbb{Z}, g\Gamma) \mapsto b^{-\{s\}}g\Gamma$ sends $\tilde{X}_{\tilde{b}}$ onto X_b. Defining the measure m on X_b as

$$\int_{X_b} F \ dm := \int_{\tilde{X}_{\tilde{b}}} F(b^{-\{s\}}g\Gamma) \ dm_{\tilde{X}_{\tilde{b}}}(s\mathbb{Z}, g\Gamma),$$

we have (see [12, Lemma 4.1]) that $m = m_{X_b}$. Thus, since

$$\tilde{F}(\tilde{b}^{a_N(n)}\tilde{\Gamma}) = \tilde{F}(a_N(n)\mathbb{Z}, b^{a_N(n)}\Gamma) = F(b^{-\{a_N(n)\}}b^{a_N(n)}\Gamma) = F(b^{[a_N(n)]}\Gamma),$$

using (34) for the function $\tilde{F}$, we get

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} F(b^{[a_N(n)]}\Gamma) = \int_{\tilde{X}_{\tilde{b}}} \tilde{F} \ dm_{\tilde{X}_{\tilde{b}}} = \int_{\tilde{X}_{\tilde{b}}} F(b^{-\{s\}}g\Gamma) \ dm_{\tilde{X}_{\tilde{b}}}(s\mathbb{Z}, g\Gamma) = \int_{X_b} F \ dm_{X_b},$$

so we have (33). □

Recalling that a sequence of ℓ-tuples of variable polynomials $(p_{1,N}, \dots, p_{\ell,N})_N$ is good if every non-trivial linear combination of $(p_{1,N})_N, \dots, (p_{\ell,N})_N$ is good, we have the following:

Lemma 6.5. *Let $(p_{1,N}, \dots, p_{\ell,N})_N$ be a good sequence of ℓ-tuples of polynomials, $X_i = G_i/\Gamma_i$ nilmanifolds, with G_i connected and simply connected, and suppose that $b_i \in G_i$ acts ergodically on X_i, $1 \leq i \leq \ell$. Then the sequence $(b_1^{p_{1,N}(n)}\Gamma_1, \dots, b_\ell^{p_{\ell,N}(n)}\Gamma_\ell)_{1\leq n\leq N}$ is equidistributed in $X_1 \times \cdots \times X_\ell$.*

Proof. We follow [12, Lemma 5.3]. As the general case is similar, we assume that $X_1 = \ldots = X_\ell = X$. Arguing by contradiction, we will also assume that for some $\delta > 0$, $(b_1^{p_{1,N}(n)}\Gamma, \dots, b_\ell^{p_{\ell,N}(n)}\Gamma)_{1\leq n\leq N}$ is not δ-equidistributed in X^ℓ.

If $p_{i,N}(t) = \sum_{k=0}^{d_i} c_{i,k,N}t^k$, then

$$b_i^{p_{i,N}(n)} = b_{i,0,N} \cdot b_{i,1,N}^n \cdots b_{i,d_i,N}^{n^{d_i}},$$

where $b_{i,j,N} = b^{c_{i,j,N}}$, $0 \leq j \leq d_i$, $1 \leq i \leq \ell$, so, for every $N \in \mathbb{N}$, $(b_1^{p_{1,N}(n)}, \dots, b_\ell^{p_{\ell,N}(n)})_n$ is a polynomial sequence in G^ℓ.

[45] Here we adapt the notation $z\mathbb{Z}$ which is more convenient than $z(\text{mod}1)$.

[46] By this we mean $\lim_{N\to\infty} \frac{|\{1\leq n\leq N : \ a_N(n)\mathbb{Z}\in I_\delta\}|}{N} = 1 - 2\delta$.

Applying Theorem 6.1, we have a constant $M \equiv M(\delta, X, d_1, \ldots, d_\ell)$ and a horizontal character χ of X^ℓ with $\|\chi\| \leq M$ such that

$$\left\| \chi(b_1^{p_{1,N}(n)}, \ldots, b_\ell^{p_{\ell,N}(n)}) \right\|_{C^\infty[N]} \leq M.$$

Let $\pi(b_i) = (\beta_{i,1}\mathbb{Z}, \ldots, \beta_{i,s}\mathbb{Z})$, $1 \leq i \leq \ell$, where $\beta_{i,j} \in \mathbb{R}$, be the projection of b_i on the horizontal torus $\mathbb{T}^s$ (this s is bounded by the dimension of X). Using the ergodicity assumption on the b_i's, for all $1 \leq i \leq \ell$, the set $\{1, \beta_{i,1}, \ldots, \beta_{i,s}\}$ consists of rationally independent elements. For $t \in \mathbb{R}$ we have $\pi(b_i^t) = (t\tilde{\beta}_{i,1}\mathbb{Z}, \ldots, t\tilde{\beta}_{i,s}\mathbb{Z})$ for some $\tilde{\beta}_{i,j} \in \mathbb{R}$ with $\tilde{\beta}_{i,j}\mathbb{Z} = \beta_{i,j}\mathbb{Z}$,[47] so, we have that

$$\chi(b_1^{p_{1,N}(n)}, \ldots, b_\ell^{p_{\ell,N}(n)}) = e\left(\sum_{i=1}^{\ell} p_{i,N}(n) \sum_{j=1}^{s} \lambda_{i,j}\tilde{\beta}_{i,j}\right)$$

for some $\lambda_i \in \mathbb{Z}$ with $|\lambda_i| \leq M$.

If, for $n \in \mathbb{N}$, we set

$$p_N(n) := \sum_{i=1}^{\ell} p_{i,N}(n) \sum_{j=1}^{s} \lambda_{i,j}\tilde{\beta}_{i,j} = \sum_{k=0}^{d} c_{k,N} n^k,$$

we have that the sequence $(p_N)_N$, being a non-trivial (as χ is non-trivial and $\tilde{\beta}_{i,j}$'s are rationally independent) linear combination of the $(p_{i,N})_N$'s, is good. Combining the last three relations, we have a contradiction to (7), as

$$M \geq \|e(p_N(n))\|_{C^\infty[N]} \geq \max_{1\leq j\leq d} \left(N^j \|c_{j,N}\|\right).$$

So, the sequence $(b_1^{p_{1,N}(n)}\Gamma, \ldots, b_\ell^{p_{\ell,N}(n)}\Gamma)_{1\leq n\leq N}$ is δ-equidistributed for all $\delta > 0$, hence it is equidistributed. □

The last ingredient in proving Part (i) of Theorem 6.2 is the following lemma:

Lemma 6.6 (Lemma 5.2, [12])**.** *Let $X = G/\Gamma$ be a nilmanifold with G connected and simply connected. Then, for every $b_1, \ldots, b_\ell \in G$, there exists an $s_0 \in \mathbb{R}$ such that for all $1 \leq i \leq \ell$ the element $b_i^{s_0}$ acts ergodically on the nilmanifold $\overline{(b_i^s\Gamma)}_{s\in\mathbb{R}}$.*

We are now ready to prove Theorem 6.2.

Proof of Theorem 6.2. Using Lemma 6.4 we see that Part (ii) of Theorem 6.2 follows from Part (i). To establish Part (i) let $b_1, \ldots, b_\ell \in G$. By Lemma 6.6 there exists a non-zero $s_0 \in \mathbb{R}$ such that for every $1 \leq i \leq \ell$ the element $b_i^{s_0}$ acts ergodically on the nilmanifold $\overline{(b_i^s\Gamma)}_{s\in\mathbb{R}}$. Using Lemma 6.5 for the elements $b_i^{s_0}$ and the polynomials $p_{i,N}/s_0$ (which are still forming a good sequence of ℓ-tuples of polynomials) we get that the sequence $(b_1^{p_{1,N}(n)}\Gamma, \ldots, b_\ell^{p_{\ell,N}(n)}\Gamma)_{1\leq n\leq N}$ is equidistributed in the nilmanifold $\overline{(b_1^s\Gamma)}_{s\in\mathbb{R}} \times \cdots \times \overline{(b_\ell^s\Gamma)}_{s\in\mathbb{R}}$, hence we get the conclusion. □

[47] Note here that, for all $1 \leq i \leq \ell$, the $\tilde{\beta}_{i,j}$'s are also rationally independent.

7. From natural to prime numbers

In this section we will provide the steps we need to follow in order to study averages along prime numbers. To do so, we follow [17], [18] and [29].

To state the first lemma which compares averages along primes to averages along natural numbers, we need to recall some standard notation:

We start with the definition of the *von Mangoldt function*, $\Lambda : \mathbb{N} \to \mathbb{R}$, where $\Lambda(n) = \begin{cases} \log(p) & \text{, if } n = p^k \text{ for some } p \in \mathbb{P} \text{ and some } k \in \mathbb{N} \\ 0 & \text{, elsewhere} \end{cases}.$

Instead of Λ, it is more convenient for us to deal with the function $\Lambda' : \mathbb{N} \to \mathbb{R}$, where $\Lambda'(n) = \mathbf{1}_{\mathbb{P}}(n) \cdot \Lambda(n) = \mathbf{1}_{\mathbb{P}}(n) \cdot \log(n)$.[48]

The argument of the following lemma, using the fact that the sequence is bounded, is exactly the same as the one in [17, Lemma 1].

Lemma 7.1 (Lemma 1, [17])**.** *If $(a_N(n))_{N,n} \subseteq \mathbb{C}$ is a bounded sequence, then*

$$\left| \frac{1}{\pi(N)} \sum_{p \in \mathbb{P} \cap [1,N]} a_N(p) - \frac{1}{N} \sum_{n=1}^{N} \Lambda'(n) \cdot a_N(n) \right| = o_N(1).^{49}$$

Next, we recall the definition of Gowers norms.

Definition 7.2. If $a : \mathbb{Z}_N \to \mathbb{C}$, we inductively define:

$$\|a\|_{U_1(\mathbb{Z}_N)} = \left| \mathbb{E}_{n \in \mathbb{Z}_N} a(n) \right|;$$

and

$$\|a\|_{U_{d+1}(\mathbb{Z}_N)} = \left(\mathbb{E}_{h \in \mathbb{Z}_N} \|a_h \cdot \overline{a}\|_{U_d(\mathbb{Z}_N)}^{2^d} \right)^{1/2^{d+1}},$$

where $a_h(n) = a(n+h)$. As Gowers showed in [21], $\|\cdot\|_{U_d(\mathbb{Z}_N)}$ defines a norm on $\mathbb{Z}_N$ for $d \geq 2$.

In what follows, we will use the, standard by now, "W-trick". For $w > 2$, let

$$W = \prod_{p \in \mathbb{P} \cap [1, w-1]} p$$

be the product of primes bounded above by w. For $r \in \mathbb{N}$, let

$$\Lambda'_{w,r}(n) = \frac{\phi(W)}{W} \cdot \Lambda'(Wn + r),$$

where ϕ is the Euler function, be the *modified von Mangoldt function.*

The next result shows the Gowers uniformity of the modified von Mangoldt function and can be derived by [23], [24] and [25]. We present here a formulation of it from [18].

Theorem 7.3 (Theorem 7.2, [23])**.** *For every $d \in \mathbb{N}$ we have that*

$$\lim_{w \to \infty} \left(\lim_{N \to \infty} \left(\max_{1 \leq r \leq W,\, (r,W)=1} \left\| (\Lambda'_{w,r} - 1) \cdot \mathbf{1}_{[1,N]} \right\|_{U_d(\mathbb{Z}_{dN})} \right) \right) = 0.$$

---

[48] The difference of the averages $\frac{1}{N} \sum_{n=1}^{N} (\Lambda(n) - \Lambda'(n))$ is of the order of $1/\sqrt{N}$.

[49] By $o_N(1)$ we mean that the quantity goes to 0 as $N \to \infty$.

The following uniformity estimate, which can be shown by iterating a variant of van der Corput's lemma, is an important step towards our results.

Lemma 7.4 (Lemma 3.5, [18])**.** *Let $\ell, m \in \mathbb{N}$, $(X, \mathcal{B}, \mu, T_1, \dots, T_\ell)$ be a system,*[50] *$q_{i,j} \in \mathbb{Z}[t]$ polynomials, $1 \leq i \leq \ell$, $1 \leq j \leq m$, $f_1, \dots, f_m \in L^\infty(\mu)$ and $a : \mathbb{N} \to \mathbb{C}$ be a sequence satisfying $a(n)/n^c \to 0$ for every $c > 0$. Then there exists $d \in \mathbb{N}$, depending only on the maximum degree of the polynomials $q_{i,j}$ and the integers ℓ, m and a constant C_d depending on d, such that*

$$\left\| \frac{1}{N} \sum_{n=1}^{N} a(n) \cdot \prod_{j=1}^{m} \prod_{i=1}^{\ell} T_i^{q_{i,j}(n)} f_j \right\|_{L^2(\mu)} \leq C_d \left(\left\| a \cdot \mathbf{1}_{[1,N]} \right\|_{U_d(\mathbb{Z}_{dN})} + o_N(1) \right).$$

Furthermore, the constant C_d is independent of the sequence $(a(n))_n$ and the $o_N(1)$ term depends only on the integer d and on the sequence $(a(n))_n$.

The crucial fact of the previous result for our study is that, because of the dependence of the constants, it can also be used for variable polynomials of bounded degree, i.e., the ones that we deal with, of the form $q_N(n) = a_{d,N} n^d + \ldots + a_{1,N} n + a_{0,N}$.

Let $r \in \mathbb{N}$ and $(X, \mathcal{B}, \mu)$ be a probability space. We call a jointly measurable family $(T_t)_{t \in \mathbb{R}^r}$ of measure preserving transformations $T_t \colon X \to X$, a *measure preserving flow*, if for all $s, t \in \mathbb{R}^r$, we have $T_{s+t} = T_s \circ T_t$.

We will show a similar estimation to [29, Theorem 3.1], via the use of a multidimensional variant of the special flow above a system under the constant ceiling function 1:

Theorem 7.5. *For $\ell, m \in \mathbb{N}$, let $(X, \mathcal{B}, \mu, T_1, \dots, T_\ell)$ be a system, $(q_{i,j,N})_N \in \mathbb{R}[t]$ good polynomials, $1 \leq i \leq \ell$, $1 \leq j \leq m$, and $f_1, \dots, f_\ell \in L^\infty(\mu)$.*

For $N \in \mathbb{N}$, for the functions

$$b_N(n) := \Big(\prod_{i=1}^{\ell} T_i^{[q_{i,1,N}(n)]}\Big) f_1 \cdot \ldots \cdot \Big(\prod_{i=1}^{\ell} T_i^{[q_{i,m,N}(n)]}\Big) f_m$$

there exists $d \in \mathbb{N}$, depending only on the maximum degree of the polynomials $q_{i,j,N}$ and the integers ℓ and m, such that for every $0 < \delta < 1$ there exists a constant $C_{d,\delta}$ depending on d and δ, such that

$$\begin{aligned} \left\| \frac{1}{N} \sum_{n=1}^{N} (\Lambda'_{w,r}(n) - 1) b_N(n) \right\|_{L^2(\mu)} &\leq C_{d,\delta} \left(\left\| (\Lambda'_{w,r} - 1) \cdot \mathbf{1}_{[1,N]} \right\|_{U_d(\mathbb{Z}_{dN})} + o_N(1) \right) \\ &+ c_\delta \left(1 + o_{N \to \infty; w}(1) \right), \end{aligned}$$

for all $r \in \mathbb{N}$, where $c_\delta \to 0$ as $\delta \to 0^+$ and the term $o_N(1)$ depends on the integer d.[51]

Proof. We follow [29, Theorem 3.1]. Let $0 < \delta < 1$ and $w, r \in \mathbb{N}$. For the given transformations on X, we define the $\mathbb{R}^{\ell m}$-action $\prod_{i=1}^{\ell} T_{i,s_{i,1}} \cdot \ldots \cdot \prod_{i=1}^{\ell} T_{i,s_{i,m}}$ on the probability space

[50]Here we mean that $T_1, \dots, T_\ell : X \to X$ are invertible, commuting, measure preserving transformations on the probability space $(X, \mathcal{B}, \mu)$.

[51]The quantity $o_{N \to \infty; w}(1)$ goes to 0 as $N \to \infty$ and then $w \to \infty$.

$Y := X \times [0,1)^{\ell m}$, endowed with the measure $\nu := \mu \times \lambda^{\ell m}$ (λ is the Lebesgue measure on $[0,1)$), by

$$\prod_{j=1}^{m}\prod_{i=1}^{\ell} T_{i,s_{i,j}}(x, a_{1,1}, \dots, a_{\ell,1}, a_{1,2}, \dots, a_{\ell,2}, \dots, a_{1,m}, \dots, a_{\ell,m}) =$$

$$\left(\prod_{j=1}^{m}\prod_{i=1}^{\ell} T_i^{[s_{i,j}+a_{i,j}]}x, \{s_{1,1}+a_{1,1}\}, \dots, \{s_{\ell,1}+a_{\ell,1}\}, \dots, \{s_{1,m}+a_{1,m}\}, \dots, \{s_{\ell,m}+a_{\ell,m}\}\right).$$

One can routinely verify that the above action defines a measure preserving flow on the product probability space Y.

If $f_1, \dots, f_m$ are bounded functions on X, we define the Y-extensions of f_j, setting for every element $(a_{1,1}, \dots, a_{\ell,1}, a_{1,2}, \dots, a_{\ell,2}, \dots, a_{1,m}, \dots, a_{\ell,m}) \in [0,1)^{\ell m}$:

$$\hat{f}_j(x, a_{1,1}, \dots, a_{\ell,1}, a_{1,2}, \dots, a_{\ell,2}, \dots, a_{1,m}, \dots, a_{\ell,m}) = f_j(x), \quad 1 \le j \le m; \quad \text{and}$$

$$\hat{f}_0(x, a_{1,1}, \dots, a_{\ell,1}, a_{1,2}, \dots, a_{\ell,m}) = 1_{[0,\delta]^{\ell m}}(a_{1,1}, \dots, a_{\ell,1}, a_{1,2}, \dots, a_{\ell,m}).$$

For every $1 \le n \le N$, let

$$\tilde{b}_N(n) := \hat{f}_0 \cdot (\prod_{j=1}^{m}\prod_{i=1}^{\ell} T_{i,\delta_{j1}\cdot q_{i,1,N}(n)})\hat{f}_1 \cdot \dots \cdot (\prod_{j=1}^{m}\prod_{i=1}^{\ell} T_{i,\delta_{jm}\cdot q_{i,m,N}(n)})\hat{f}_m;$$

and for every $x \in X$

$$b'_N(n)(x) := \int_{[0,1)^{\ell m}} \tilde{b}_N(n)(x, a_{1,1}, \dots, a_{\ell,1}, a_{1,2}, \dots, a_{\ell,2}, \dots, a_{1,m}, \dots, a_{\ell,m})\, d\lambda^{\ell m},$$

where the integration is with respect to the variables $a_{i,j}$.

By using the triangle and the Cauchy-Schwarz inequality, if $a(n) = \Lambda'_{w,r}(n) - 1$, we have

$$\begin{aligned}\delta^{\ell m}\left\|\frac{1}{N}\sum_{n=1}^{N} a(n)b_N(n)\right\|_{L^2(\mu)} &\le \left\|\frac{1}{N}\sum_{n=1}^{N} a(n)\cdot(\delta^{\ell m}b_N(n) - b'_N(n))\right\|_{L^2(\mu)} \\ &+ \left\|\frac{1}{N}\sum_{n=1}^{N} a(n)\tilde{b}_N(n)\right\|_{L^2(\nu)}.\end{aligned}$$

Using Lemma 7.4, we find an integer $d \in \mathbb{N}$, depending only on the maximum degree of the polynomials $q_{i,j,N}$ and the integers ℓ, m and a constant C_d depending on d, such that

$$\left\|\frac{1}{N}\sum_{n=1}^{N} a(n)\tilde{b}_N(n)\right\|_{L^2(\nu)} \le C_d\left(\left\|a \cdot \mathbf{1}_{[1,N]}\right\|_{U_d(\mathbb{Z}_{dN})} + o_N(1)\right),$$

where the $o_N(1)$ term depends only on the integer d and the sequence $(a(n))_n$.

Next we will study the term $\left\| \frac{1}{N}\sum_{n=1}^N a(n)\cdot (\delta^{\ell m}b_N(n)-b'_N(n))\right\|_{L^2(\mu)}$. For every $x\in X$ and $1\leq n\leq N$, we have

$$\left|\delta^{\ell m}b_N(n)(x)-b'_N(n)(x)\right| = \left|\int_{[0,\delta]^{\ell m}}\left(\prod_{j=1}^m f_j(\prod_{i=1}^{\ell}T_i^{[q_{i,j,N}(n)]}x)-\prod_{j=1}^m f_j(\prod_{i=1}^{\ell}T_i^{[q_{i,j,N}(n)+a_{i,j}]}x)\right)\,d\lambda^{\ell m}\right|.$$

Since all the relevant $a_{i,j}$ in the integrand are less or equal than δ, if $\{q_{i,j,N}(n)\}<1-\delta$, we have $T_i^{[q_{i,j,N}(n)+a_{i,j}]}=T_i^{[q_{i,j,N}(n)]}$ for all $1\leq i\leq \ell$, $1\leq j\leq m$. We deal with the case where $\{q_{i,j,N}(n)\}\geq 1-\delta$.

For every $1\leq i\leq\ell$, $1\leq j\leq m$, let $E_{\delta,N}^{i,j}:=\{1\leq n\leq N\colon \{q_{i,j,N}(n)\}\in[1-\delta,1)\}$. Then, by using the fact that $\mathbf{1}_{E_{\delta,N}^{1,1}\cup\ldots\cup E_{\delta,N}^{1,m}\cup E_{\delta,N}^{2,1}\cup\ldots\cup E_{\delta,N}^{\ell,m}}\leq \sum_{(i,j)\in[1,\ell]\times[1,m]}\mathbf{1}_{E_{\delta,N}^{i,j}}$ and that $\mathbf{1}_{E_{\delta,N}^{i,j}}(n)=\mathbf{1}_{[1-\delta,1)}(\{q_{i,j,N}(n)\})$, for every $x\in X$ we have

$$\left|\delta^{\ell m}b_N(n)(x)-b'_N(n)(x)\right|\leq 2\delta^{\ell m}\cdot\sum_{(i,j)\in[1,\ell]\times[1,m]}\mathbf{1}_{[1-\delta,1)}(\{q_{i,j,N}(n)\}).$$

So, recalling that $a(n)=\Lambda'_{w,r}(n)-1$,

$$\frac{1}{N}\sum_{n=1}^N|a(n)|\cdot\mathbf{1}_{[1-\delta,1)}(\{q_{i,j,N}(n)\})\leq\frac{1}{N}\sum_{n=1}^N\Lambda'_{w,r}(n)\cdot\mathbf{1}_{[1-\delta,1)}(\{q_{i,j,N}(n)\})+\frac{|E_{\delta,N}^{i,j}|}{N}.$$

Since each $(q_{i,j,N})_N$ is good, for large N and small enough δ, the term (and the sum of finitely many terms of this form) $\frac{|E_{\delta,N}^{i,j}|}{N}$ is as small as we want.

It remains to show that the term $\frac{1}{N}\sum_{n=1}^N\Lambda'_{w,r}(n)\cdot\mathbf{1}_{[1-\delta,1)}(\{q_{i,j,N}(n)\})$ goes to zero as $N\to\infty$, then $w\to\infty$ and finally $\delta\to 0^+$. To this end, it suffices to show that $\frac{1}{N}\sum_{n=1}^N\Lambda'_{w,r}(n)e^{2\pi ikq_{i,j,N}(n)}\to 0$ as $N\to\infty$ and then $w\to\infty$ for all $k\in\mathbb{Z}\setminus\{0\}$.[52] Writing

$$\frac{1}{N}\sum_{n=1}^N\Lambda'_{w,r}(n)e^{2\pi ikq_{i,j,N}(n)}=\frac{1}{N}\sum_{n=1}^N(\Lambda'_{w,r}(n)-1)e^{2\pi ikq_{i,j,N}(n)}+\frac{1}{N}\sum_{n=1}^N e^{2\pi ikq_{i,j,N}(n)},$$

we have that the second term goes to 0 for large N from (6). The first term goes to zero as $N\to\infty$ and then $w\to\infty$ as in [18],[53] The result now follows. □

The following implication of Theorems 7.5 and 7.3 is the main result that we will use to obtain the convergence and recurrence results along primes, stated in Section 2.

[52]This follows by the fact that if f is Riemann integrable on $[0,1)$ with $\int_{[0,1)}f(x)\,dx=c$, then, for every $\varepsilon>0$, we can find trigonometric polynomials q_1, q_2, with no constant terms, with $q_1(t)+c-\varepsilon\leq f(t)\leq q_2(t)+c+\varepsilon$. We use this for the function $f=\mathbf{1}_{[1-\delta,1)}$.

[53]One can imitate here the proof of [18, Lemma 3.5], since, after finitely many iterates of the van der Corput lemma, the polynomial $q_{i,j,N}$ becomes constant.

Proposition 7.6. *For $\ell, m \in \mathbb{N}$, let $(X, \mathcal{B}, \mu, T_1, \ldots, T_\ell)$ be a system, $(p_{i,j,N})_N$ polynomials such that $(p_{W,r,i,j,N})_N$ is good for every $W \in \mathbb{N}$, $1 \leq r \leq W$, where $p_{W,r,i,j,N}(n) = p_{i,j,N}(Wn + r)$, $1 \leq i \leq \ell$, $1 \leq j \leq m$, and $f_1, \ldots, f_\ell \in L^\infty(\mu)$. Then,*

$$\max_{1\leq r\leq W,\, (r,W)=1} \left\| \frac{1}{N}\sum_{n=1}^{N} (\Lambda'_{w,r}(n) - 1) \cdot \prod_{j=1}^{m}\prod_{i=1}^{\ell} T_i^{[p_{i,j,N}(Wn+r)]} f_j \right\|_{L^2(\mu)} = o_{N\to\infty;w}(1).$$

Proof. Using Theorem 7.5, for the polynomials $p_{W,r,i,j,N}$, we get that for every $0 < \delta < 1$, there exists $d \in \mathbb{N}$, depending only on the maximum degree of the polynomials $q_{i,j,N}$ and the integers ℓ and m, and a constant $C_{d,\delta}$ depending on d and δ, such that

$$\max_{1\leq r\leq W,\, (r,W)=1} \left\| \frac{1}{N}\sum_{n=1}^{N} (\Lambda'_{w,r}(n) - 1) \cdot \prod_{j=1}^{m}\prod_{i=1}^{\ell} T_i^{[p_{i,j,N}(Wn+r)]} f_j \right\|_{L^2(\mu)}$$

$$\leq C_{d,\delta}\left(\max_{1\leq r\leq W,\, (r,W)=1} \left\| (\Lambda'_{w,r} - 1) \cdot \mathbf{1}_{[1,N]} \right\|_{U_d(\mathbb{Z}_{dN})} + o_N(1) \right) + c_\delta \left(1 + o_{N\to\infty;w}(1)\right),$$

where $c_\delta \to 0$ as $\delta \to 0^+$. Taking first $N \to \infty$ and then $w \to \infty$ in this expression, by Theorem 7.3, we have that the required limit is bounded above by c_δ. Taking $\delta \to 0^+$, we get the result. □

8. Proof of main results

In this final section we will prove of our main results.

8.1. Averages along natural numbers.

We will show Theorems 2.1 and 2.2. First we show that the polynomial sequences from Theorems 1.4 and 1.5 are good and super nice.

We remind the reader at this point that every Hardy field function is eventually monotone and that limits of ratios of such functions always exist (in the extended real line).

Recall that we write $g_2 \prec g_1$ if $|g_1(x)|/|g_2(x)| \to \infty$ as $x \to \infty$. We also introduce some new notation: We write $g_2 \sim g_1$, in case $|g_1(x)|/|g_2(x)|$ converges to a non-zero real number as $x \to \infty$, and $g_2 \precsim g_1$ if either $g_2 \prec g_1$ or $g_2 \sim g_1$.

Lemma 8.1. *The polynomial sequences from Theorems 1.4 and 1.5 are good.*

Proof. Let $\lambda_1 p_{1,N} + \ldots + \lambda_\ell p_{\ell,N}$ be a non-trivial linear combination of strongly independent variable polynomials as in (8). In case this combination is a polynomial of degree 1, factoring its constant term out, without loss, we can assume that it is of the form $h(N)n$, where $h(N) \sim 1/g(N)$, with $1 \prec g(N) \prec N$ (hence $h(N) \to 0$ and $|h(N)|N \to \infty$ monotonically as $N \to \infty$). For any $\alpha \neq 0$, we have that

$$\frac{1}{N}\sum_{n=0}^{N-1} e^{i\alpha h(N)n} = \frac{1}{N} \cdot \frac{1 - e^{i\alpha h(N)N}}{1 - e^{i\alpha h(N)}} = \frac{h(N)}{1 - e^{i\alpha h(N)}} \cdot \frac{1 - e^{i\alpha h(N)N}}{h(N)N} \to \frac{i}{\alpha} \cdot 0 = 0, \text{ as } N \to \infty.$$

In case the combination is a polynomial of degree d, after using Lemma 5.5 $(d-1)$ times, we are getting a polynomial of degree 1, hence the result follows from the previous step. □

Recall that when we want to check that a k-tuple, for $k > 1$, has the R_k-property, we have to check that for every $1 \leq i \leq k$ the corresponding (according to Definition 5.6) $(k-1)$-tuple has the R_{k-1}-property. If a $(k-1)$-tuple corresponds to the index i_0, we say that it is *descending* from the i_0 term of the previous step.

Lemma 8.2. *The polynomial sequences from Theorems 1.4 and 1.5 are super nice.*

Proof. For a single polynomial sequence as in (8), the result follows immediately from Remark 5.13 (4) and the properties of Hardy field functions.

For multiple sequences, (i) follows by the form (8) that the variable polynomial sequences have. As $(\mathcal{P}_N)_N$ and $(\mathcal{P}'_N)_N$ consist of polynomials of the same form, (ii) and $(ii)'$ will both follow by the same argument.

Assuming that we have k many essentially distinct terms of the form $a_{i,N}n$, $1 \leq i \leq k$, we have $a_{i,\cdot} \in \mathcal{C}$ and $a_{k,N} \precsim \ldots \precsim a_{2,N} \precsim a_{1,N}$.[54] We present an algorithmic way of finding the coefficients in order to check that they satisfy the R_k-property:

Step 1: For $i = 1$ we pick $j_0 = k$ (i.e., the largest index). In this case we will show that we have (ii) (a) (of Definition 5.6). The coefficients become:

$$\frac{a_{k,N} - a_{j,N}}{a_{1,N}} \sim \frac{a_{j,N}}{a_{1,N}}, \quad 1 \leq j \leq k-1.$$

For $i > 1$, we pick $j_0 = 1$ (i.e., the smallest index). In this case we will show that we have (ii) (b). The coefficients become:

$$\frac{a_{j,N}}{a_{i,N} - a_{1,N}} \sim \frac{a_{j,N}}{a_{1,N}}, \quad 2 \leq j \leq k.$$

Step λ: After we order them from largest to smallest growth, we denote the j-th term at the λ-th step with $a_{\lambda,j,N}$. We have two cases:

- The sequence of coefficients is descending from the $i = 1$ term of the $(\lambda-1)$-th step.

For $i = 1$ we pick $j_0 = k - \lambda + 1$ and show (ii) (a) (always for the $i = 1$ case we pick the largest index j_0 and show (ii) (a)). For $1 \leq j \leq j_0 - 1$, we have

$$\frac{a_{\lambda,j_0,N} - a_{\lambda,j,N}}{a_{\lambda,1,N}} = \frac{a_{\lambda-1,j,N} - a_{\lambda-1,j_0,N}}{a_{\lambda-1,j_0+1,N} - a_{\lambda-1,1,N}} \sim \frac{a_{\lambda-1,j,N}}{a_{\lambda-1,1,N}},$$

where the numerator comes from the difference $(a_{\lambda-1,j_0+1,N} - a_{\lambda-1,j_0,N}) - (a_{\lambda-1,j_0+1,N} - a_{\lambda-1,j,N})$, and the (common) denominators are canceled.

For $i > 1$ we pick $j_0 = 1$ and show (ii) (b) (always for the $i > 1$ case we pick $j_0 = 1$ and show (ii) (b)). For $2 \leq j \leq k - \lambda + 1$ we have

$$\frac{a_{\lambda,j,N}}{a_{\lambda,i,N} - a_{\lambda,1,N}} = \frac{a_{\lambda-1,k-\lambda+2,N} - a_{\lambda-1,j,N}}{a_{\lambda-1,1,N} - a_{\lambda-1,i,N}} \sim \frac{a_{\lambda-1,j,N}}{a_{\lambda-1,1,N}},$$

where the denominator comes from the difference $(a_{\lambda-1,k-\lambda+2,N} - a_{\lambda-1,i,N}) - (a_{\lambda-1,k-\lambda+2,N} - a_{\lambda-1,1,N})$, and, as in the previous case, the (common) denominators are canceled.

- The sequence of coefficients is descending from the $i > 1$ term of the $(\lambda-1)$-th step.

[54] This happens because vdC-operations preserve the essential distinctness property of the polynomials and at each step the coefficient functions belong to $\mathcal{C}$.

For $i = 1$, we choose $j_0 = k - \lambda + 1$. For all $1 \leq j \leq j_0 - 1$ we have:

$$\frac{a_{\lambda,j_0,N} - a_{\lambda,j,N}}{a_{\lambda,1,N}} = \frac{a_{\lambda-1,j_0+1,N} - a_{\lambda-1,j+1,N}}{a_{\lambda-1,2,N}} \sim \frac{a_{\lambda-1,j+1,N}}{a_{\lambda-1,2,N}}.$$

For $i > 1$, we choose $j_0 = 1$. For all $2 \leq j \leq k - \lambda + 1$, we have

$$\frac{a_{\lambda,j,N}}{a_{\lambda,i,N} - a_{\lambda,1,N}} = \frac{a_{\lambda-1,j+1,N}}{a_{\lambda-1,i+1,N} - a_{\lambda-1,2,N}} \sim \frac{a_{\lambda-1,j+1,N}}{a_{\lambda-1,2,N}}.$$

Note that each of the aforementioned terms, at each step, is (up to a sign) of the form

$$\frac{a_{i,N} - a_{j,N}}{a_{s,N} - a_{t,N}},\ s > i > j \geq t, \text{ or, } \frac{a_{i,N} - a_{j,N}}{a_{t,N}},\ i > j \geq t, \text{ or, } \frac{a_{j,N}}{a_{i,N} - a_{t,N}},\ t < \min\{i, j\},$$

i.e., combinations of terms from the initial sequence (because of the cancellations that we mentioned before) which are all $\sim$ to $a_{j,N}/a_{t,N}$. The claim now follows by the properties of elements from $\mathcal{LE}$ as each coefficient is a logarithmico-exponential Hardy function, hence eventually monotone, which is either ~ 1 or $\sim 1/g(N)$ with $1 \prec g(N) \prec N$ by the construction. □

We now prove Theorem 2.1 (which of course implies Theorem 1.4):

Proof of Theorem 2.1. We start by using Proposition 5.14 in order to get that the nilfactor $\mathcal{Z}$ is characteristic for the multiple average in (9). Via Theorem 4.1 we can assume without loss of generality that our system is an inverse limit of nilsystems. By a standard approximation argument, we can further assume that our system is a nilsystem.

Let $(X = G/\Gamma, \mathcal{G}/\Gamma, m_X, T_b)$ be a nilsystem, where $b \in G$ is ergodic, and $F_1, \ldots, F_\ell \in L^\infty(m_X)$. Our objective now is show that if $(p_{1,N}, \ldots, p_{\ell,N})_N$ is a super nice sequence of ℓ-tuples of polynomials, then

$$\lim_{N\to\infty} \sum_{n=1}^{N} F_1(b^{[p_{1,N}(n)]}x) \cdot \ldots \cdot F_\ell(b^{[p_{\ell,N}(n)]}x) = \int F_1\, dm_X \cdot \ldots \cdot \int F_\ell\, dm_X, \tag{35}$$

where the convergence takes place in $L^2(m_X)$. By density, we can assume that the functions $F_1, \ldots, F_\ell$ are continuous. Then, applying Theorem 6.2 to the nilmanifold X^ℓ, the nilrotation $\tilde{b} = (b, \ldots, b) \in G^\ell$, the point $\tilde{x} = (x, \ldots, x) \in X^\ell$, and the continuous function $\tilde{F}(x_1, \ldots, x_\ell) = F_1(x_1) \cdot \ldots \cdot F_\ell(x_\ell)$, we get that

$$\lim_{N\to\infty} \sum_{n=1}^{N} \tilde{F}(b^{[p_{1,N}(n)]}x, \ldots, b^{[p_{\ell,N}(n)]}x) = \int \tilde{F}\, dm_{X^\ell};$$

which gives the desired limit in (35), completing the proof. □

Next, we show Theorem 2.2 (which in turn implies Theorem 1.5):

Proof of Theorem 2.2. From Proposition 5.15 the nilfactor $\mathcal{Z}$ is characteristic for the multiple ergodic averages in (10). Via Theorem 4.1, using the ergodic decomposition, we can assume without loss of generality that our system is an inverse limit of nilsystems. As in the previous proof, we can further assume without loss of generality that our system is a nilsystem. Let $(X = G/\Gamma, \mathcal{G}/\Gamma, m_X, T_b)$ be a nilsystem and $F_1, \ldots, F_\ell \in L^\infty(m_X)$.

Our objective now is to show that if $(p_N)_N \subseteq \mathbb{R}[t]$ is a good polynomial sequence with $(p_N, 2p_N, \ldots, \ell p_N)_N$ being super nice, then the limit

$$\lim_{N\to\infty} \sum_{n=1}^{N} F_1(b^{[p_N(n)]}x) \cdot F_2(b^{2[p_N(n)]}x) \cdot \ldots \cdot F_\ell(b^{\ell[p_N(n)]}x) \tag{36}$$

exists in $L^2(m_X)$ and it's equal to the limit

$$\lim_{N\to\infty} \sum_{n=1}^{N} F_1(b^{n}x) \cdot F_2(b^{2n}x) \cdot \ldots \cdot F_\ell(b^{\ell n}x). \tag{37}$$

Note that by density, we can assume that every F_i is continuous. Then, for every $x \in X$, applying Theorem 6.2 (for the single good polynomial sequence $(p_N)_N$) to the nilmanifold X^ℓ, the nilrotation $\tilde{b} = (b, b^2, \ldots, b^\ell)$, the point $\tilde{x} = (x, x, \ldots, x)$, and the continuous function $\tilde{F}(x_1, \ldots, x_\ell) = F_1(x_1) \cdot F_2(x_2) \cdot \ldots \cdot F_\ell(x_\ell)$, as the sequences $(\tilde{b}^{[p_N(n)]}\tilde{x})_n$ and $(\tilde{b}^n\tilde{x})_n$ are equidistributed to the nilmanifold $\overline{(\tilde{b}^n\tilde{x})}_n$, we get that the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \tilde{F}(\tilde{b}^{[p_N(n)]}\tilde{x})$$

exists and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \tilde{F}(\tilde{b}^{[p_N(n)]}\tilde{x}) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \tilde{F}(\tilde{b}^{n}\tilde{x}).$$

This implies that the limits in (36) and (37) exist for every $x \in X$ and are equal, completing the proof. □

8.2. **Averages along prime numbers.** We will now show Theorems 2.17 and 2.16, which imply Theorems 1.7 and 1.6 (it is easy to see that polynomials of the form (8) are satisfying the required assumptions of these results). We will also show Theorems 2.22 and 2.23. To do so we follow the arguments from [18] (see also [29] and [27]).

Proof of Theorem 2.17. By Lemma 7.1 it suffices to show that the sequence

$$A(N) := \frac{1}{N} \sum_{n=1}^{N} \Lambda'(n) \cdot T^{[q_N(n)]}f_1 \cdot T^{2[q_N(n)]}f_2 \cdot \ldots \cdot T^{\ell[q_N(n)]}f_\ell$$

converges in $L^2(\mu)$ to the same limit as the sequence $\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{\ell} T^{in}f_i$ as $N \to \infty$. For w (which gives a corresponding W), $r \in \mathbb{N}$, we define

$$B_{w,r}(N) := \frac{1}{N} \sum_{n=1}^{N} T^{[q_N(Wn+r)]}f_1 \cdot T^{2[q_N(Wn+r)]}f_2 \cdot \ldots \cdot T^{\ell[q_N(Wn+r)]}f_\ell.$$

For any $\varepsilon > 0$, using Proposition 7.6 with $m = \ell$, $T_i = T$, $1 \leq i \leq \ell$, and $p_{i,j,N} = \begin{cases} 0 & , \text{ if } i \leq \ell - j \\ q_N & , \text{ elsewhere} \end{cases}$, for sufficiently large N and some w_0 we have

$$\left\| A(W_0N) - \frac{1}{\phi(W_0)} \sum_{1\leq r\leq W_0,\ (r,W_0)=1} B_{w_0,r}(N) \right\|_{L^2(\mu)} < \varepsilon.$$

By the assumption we have that $(q_{W,r,N})_N$, where $q_{W,r,N}(n) = q_N(Wn+r)$, is good and $(q_{W,r,N}, 2q_{W,r,N}, \ldots, \ell q_{W,r,N})_N$ is super nice.

By Theorem 2.2, we have that for any $1 \leq r \leq W_0$ the sequence $(B_{w_0,r}(N))_N$ converges, in $L^2(\mu)$, to the same limit as the sequence $\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{\ell} T^{in}f_i$, and since

$$\lim_{N\to\infty} \|A(W_0N+r) - A(W_0N)\|_{L^2(\mu)} = 0$$

for every $1 \leq r \leq W_0$, we get the result. □

Proof of Theorem 2.16. The argument is analogous to the one in the previous statement. We define $A(N) := \frac{1}{N}\sum_{n=1}^{N}\Lambda'(n)\cdot T^{[q_{1,N}(n)]}f_1\cdot\ldots\cdot T^{[q_{\ell,N}(n)]}f_\ell$ and for $w, r \in \mathbb{N}$, $B_{w,r}(N) := \frac{1}{N}\sum_{n=1}^{N} T^{[q_{1,N}(Wn+r)]}f_1\cdot\ldots\cdot T^{[q_{\ell,N}(Wn+r)]}f_\ell$. We use Proposition 7.6 with $m = \ell$, $T_i = T$, $1 \leq i \leq \ell$, $p_{i,j,N} = \begin{cases} 0 & , \text{ if } i \neq j \\ q_{i,N} & , \text{ if } i = j \end{cases}$, and we note that the family $(q_{W,r,1,N}, \ldots, q_{W,r,\ell,N})_N$ is good and super nice, where $q_{W,r,i,N}(n) = q_{i,N}(Wn+r)$. The result now follows similarly to the previous one since, by Theorem 2.1, we have that for any $1 \leq r \leq W_0$ the sequence $(B_{w_0,r}(N))_N$ converges, in $L^2(\mu)$, to $\prod_{i=1}^{\ell}\int f_i\, d\mu$. □

We now prove both Theorems 2.22 and 2.23 for the set $\mathbb{P}-1$; the corresponding results for $\mathbb{P}+1$ follow by the same arguments.

Proof of Theorem 2.22. Using Proposition 7.6 with $m = \ell$, $T_i = T$, $1 \leq i \leq \ell$, $r = 1$ and $p_{i,j,N}(n) = \begin{cases} 0 & , \text{ if } i \leq \ell - j \\ q_N(n-1) & , \text{ elsewhere} \end{cases}$, together with Corollary 2.4, we have, for sufficiently large $\omega \in \mathbb{N}$, that

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \Lambda'_{\omega,1}(n)\cdot\mu\left(A \cap \bigcap_{i=1}^{\ell} T^{-i[q_N(Wn)]}A\right) > 0,$$

from which we get the required non-empty intersection with $\mathbb{P}-1$. □

Proof of Theorem 2.23. The proof is analogous to the one of Theorem 2.22. More specifically, we use Proposition 7.6 with $m = \ell$, $T_i = T$, $1 \leq i \leq \ell$, $r = 1$ and $p_{i,j,N}(n) = \begin{cases} 0 & , \text{ if } i \neq j \\ q_{i,N}(n-1) & , \text{ if } i = j \end{cases}$, and Corollary 2.9 instead of Corollary 2.4 to get, for some sufficiently large $\omega \in \mathbb{N}$, that

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \Lambda'_{\omega,1}(n)\cdot\mu\left(A \cap T^{-[q_{1,N}(Wn)]}A \cap \ldots \cap T^{-[q_{\ell,N}(Wn)]}A\right) > 0,$$

from which we get the result. □

8.3. **Closing comments and problems.** In the generality that it's stated, Problem 1 (i.e., [14, Problem 10]), except in the $\ell = 1$ case, remains open. In this article, we first showed that the nilfactor of a system is characteristic for the corresponding sequence of iterates under the additional super niceness assumption. Second, for the equidistribution part, we proceeded to show that the goodness property alone was enough for us to display what we wanted in full generality. This comes as no surprise though for, as we have already mentioned, the goodness property is a strong equidistribution notion. Hence, to completely resolve the problem, someone has to show the following:

Problem 3. *For $\ell \in \mathbb{N}$ let $(\mathcal{P}_N)_N = (p_{1,N}, \dots, p_{\ell,N})_N$ be a good sequence of ℓ-tuples of polynomials. Then, for every system, its nilfactor $\mathcal{Z}$ is characteristic for $(\mathcal{P}_N)_N$.*

Analogously, to verify Problem 2, it suffices to show:

Problem 4. *Let $(p_N)_N$ be a good sequence of polynomials. Then, for every $\ell \in \mathbb{N}$ and every system, its nilfactor $\mathcal{Z}$ is characteristic for $(\mathcal{P}_N)_N = (p_N, 2p_N \dots, \ell p_N)_N$.*

The resolution of these problems will of course lead to stronger variable-iterate results along prime (and shifted prime) numbers.

As in our results we have convergence to the expected limit, it is reasonable, under the recent developments (see [32]), for someone to study the corresponding pointwise results along natural numbers. So, we naturally close this article with the following problem:

Problem 5. *Find classes of good variable polynomial iterates (e.g., the ones in Theorems 2.1 and 2.2) for which we have the corresponding pointwise convergence results.*[55]

Acknowledgements. Thanks go to D. Karageorgos with whom I started discussing the problem; N. Frantzikinakis for his constant support and fruitful discussions during the writing of this article; and N. Kotsonis for his detailed corrections on the text.

References

[1] T. Austin. Pleasant extensions retaining algebraic structure, I. *J. Anal. Math.* **125** (2015), 1–36.
[2] T. Austin. Pleasant extensions retaining algebraic structure, II. *J. Anal. Math.* **126** (2015), 1–111.
[3] V. Bergelson. Ergodic Ramsey theory. *Logic and combinatorics (Arcata, Calif., 1985)*, 63–87, *Contemp. Math.*, **65**, *Amer. Math. Soc., Providence, RI*, 1987.
[4] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems* **7** (1987), no. 3, 337–349.
[5] V. Bergelson, I. Håland-Knutson. Weakly mixing implies mixing of higher orders along tempered functions. *Ergodic Theory Dynam. Systems* **29** (2009), no. 5, 1375–1416.
[6] V. Bergelson, A. Leibman. A nilpotent Roth theorem. *Invent. Math.* **147** (2002), 429–470.
[7] V. Bergelson, A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's Theorems. *Journal of the A. M. S.* **9** (1996), 725–753.
[8] Q. Chu, N. Frantzikinakis, B. Host. Ergodic averages of commuting transformations with distinct degree polynomial iterates. *Proc. of the London Math. Society.* (3), **102** (2011), 801–842.
[9] S. Donoso, A. Koutsogiannis, W. Sun. Pointwise multiple averages for sublinear functions. *Ergodic Theory Dynam. Systems* **40** (2020), 1594–1618.
[10] S. Donoso, A. Koutsogiannis, W. Sun. Seminorms for multiple averages along polynomials and applications to joint ergodicity. To appear in J. d'Analyse Math.
[11] N. Frantzikinakis. A multidimensional Szemerédi theorem for Hardy sequences of different growth. *Tran. of the A. M. S.* **367**, no. 8, (2015), 5653–5692.
[12] N. Frantzikinakis. Equidistribution of sparse sequences on nilmanifolds. *J. d'Analyse Math.* **109** (2009), 353–395.
[13] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial growth. *J. d'Analyse Math.* **112** (2010), 79–135.
[14] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bulletin of the Hellenic Mathematical Society.* **60** (2016), 41-90.
[15] N. Frantzikinakis. An averaged Chowla and Elliott conjecture along independent polynomials. *International Mathematics Research Notices* **2018** (2018), no. 12, 3721–3743.
[16] N. Frantzikinakis, B. Host. Higher order Fourier analysis of multiplicative functions and applications. *Journal of the A. M. S.* **30**, no. 1, (2017), 67-157.

[55]Of course someone can start by studying the pointwise convergence for the special cases of averages with iterates from Examples 1 and 2.

[17] N. Frantzikinakis, B. Host, B. Kra. Multiple recurrence and convergence for sequences related to the prime numbers. *J. Reine Angew. Math.* **611** (2007), 131–144.
[18] N. Frantzikinakis, B. Host, B. Kra. The polynomial multidimensional Szemerédi theorem along shifted primes. *Israel J. Math.* **194** (2013), no. 1, 331–348.
[19] N. Frantzikinakis, B. Kra. Polynomial averages converge to the product of integrals. *Israel. J. Math.* 148 (2005), 267–276.
[20] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.* **31** (1977), 204-–256.
[21] W. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** (2001), 465–588.
[22] B. Green, T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Annals of Math.* **175** (2012), 465–540.
[23] B. Green, T. Tao. Linear equations in primes. *Annals. Math.* **171** (2010), 1753–1850.
[24] B. Green, T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Annals. Math.* **175** (2012), no. 2, 541–566.
[25] B. Green, T. Tao, T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Annals. Math.* **176** (2012), no. 2, 1231–1372.
[26] B. Host, B. Kra. Nonconventional ergodic averages and nilmanifolds. *Annals of Math.* **161** (2005), no. 1, 397–488.
[27] D. Karageorgos, A. Koutsogiannis. Integer part independent polynomial averages and applications along primes. *Studia Mathematica* **249** (2019), no. 3, 233–257.
[28] Y. Kifer. Ergodic theorems for nonconventional arrays and an extension of the Szemerédi theorem. *Discrete & Continuous Dynamical Systems* **38** (2018), no. 6, 2687–2716.
[29] A. Koutsogiannis. Closest integer polynomial multiple recurrence along shifted primes. *Ergodic Theory Dynam. Systems* **38** (2018), no. 2, 666–685.
[30] A. Koutsogiannis. Integer part polynomial correlation sequences. *Ergodic Theory Dynam. Systems* **38** (2018), no. 4, 1525–1542.
[31] A Koutsogiannis. Multiple ergodic averages for tempered functions. To appear in Descrete & Continuous Dynamical Systems.
[32] B. Krause, M. Mirek, T. Tao. Pointwise ergodic theorems for non-conventional bilinear polynomial averages. Preprint, arXiv:2008.00857.
[33] L. Kuipers, H. Niederreiter. Uniform distribution of sequences. Pure and Applied Mathematics. *Wiley-Interscience*, New York-London-Sydney, 1974.
[34] A. Leibman. Pointwise Convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Ergodic Theory Dynam. Systems* **25** (2005) no. 1, 201–213.
[35] T. Tao, T. Ziegler. Concatenation Theorems for anti-Gowers-uniform functions and Host-Kra characteristic factors. *Discrete Anal.* 2016, Paper No. 13.
[36] M. Walsh. Norm convergence of nilpotent ergodic averages. *Annals of Math.* **175** (2012), no. 3, 1667–1688.

(Andreas Koutsogiannis) The Ohio State University, Department of mathematics, Columbus, Ohio, USA
Email address: `koutsogiannis.1@osu.edu`