

RESEARCH ARTICLE

Joint ergodicity of fractional powers of primes

Nikos Frantzikinakis 

Department of Mathematics, University of Crete, Voutes University Campus, Heraklion, 71003, Greece;
E-mail: frantzikinakis@gmail.com.

Received: 14 July 2021; **Revised:** 7 April 2022; **Accepted:** 24 April 2022

Keywords and phrases: Joint ergodicity, multiple recurrence, prime numbers, fractional powers

2020 Mathematics subject classification: *Primary* – 37A44; *Secondary* – 28D05, 05D10, 11B30

Abstract

We establish the mean convergence for multiple ergodic averages with iterates given by distinct fractional powers of primes and related multiple recurrence results. A consequence of our main result is that every set of integers with positive upper density contains patterns of the form $\{m, m + [p_n^a], m + [p_n^b]\}$, where a, b are positive nonintegers and p_n denotes the n th prime, a property that fails if a or b is a natural number. Our approach is based on a recent criterion for joint ergodicity of collections of sequences, and the bulk of the proof is devoted to obtaining good seminorm estimates for the related multiple ergodic averages. The input needed from number theory are upper bounds for the number of prime k -tuples that follow from elementary sieve theory estimates and equidistribution results of fractional powers of primes in the circle.

Contents

1	Introduction and main results	2
1.1	Introduction	2
1.2	Main results	3
1.3	Limitations of our techniques and open problems	4
1.4	Notation	5
2	Proof strategy	5
3	Seminorm estimates – some preparation	7
3.1	A more general statement	7
3.2	Feedback from number theory	9
3.3	Two elementary lemmas	12
4	Seminorm estimates – sublinear case	13
4.1	An example	13
4.2	Reduction to averages with bounded weights and change of variables	16
4.3	Reduction to averages with polynomial iterates	19
4.4	Averages with polynomial iterates	22
4.5	Proof of Theorem 3.1 in the sublinear case	23

5 Seminorm estimates – induction step **24**
 5.1 An example 24
 5.2 The van der Corput operation and reduction of type 25
 5.3 Proof of Theorem 3.1 28

1. Introduction and main results

1.1. Introduction

Given an ergodic measure preserving system (X, μ, T) and functions $f, g \in L^\infty(\mu)$, it was shown in [6] that for distinct $a, b \in \mathbb{R}_+ \setminus \mathbb{Z}$, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N T^{[na]} f \cdot T^{[nb]} g = \int f \, d\mu \cdot \int g \, d\mu \tag{1}$$

in $L^2(\mu)$.¹ An immediate consequence of this limit formula is that for every (not necessarily ergodic) measure preserving system and measurable set A , we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-[na]} A \cap T^{-[nb]} A) \geq \mu(A)^3. \tag{2}$$

Examples of periodic systems show that equations (1) and (2) fail if either a or b is an integer greater than 1. Using the Furstenberg correspondence principle [10, 11], it is easy to deduce from equation (2) that every set of integers with positive upper density contains patterns of the form

$$\{m, m + [n^a], m + [n^b]\}$$

for some $m, n \in \mathbb{N}$.

The main goal of this article is to establish similar convergence and multiple recurrence results, and deduce related combinatorial consequences, when in the previous statements we replace the variable n with the n th prime number p_n . For instance, we show in Theorem 1.1 that if $a, b \in \mathbb{R}_+$ are distinct nonintegers, then

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N T^{[p_n^a]} f \cdot T^{[p_n^b]} g = \int f \, d\mu \cdot \int g \, d\mu \tag{3}$$

in $L^2(\mu)$. We also prove more general statements of this sort involving two or more linearly independent polynomials with fractional exponents evaluated at primes (related results for fractional powers of integers were previously established in [4, 6, 26]).

If $a, b \in \mathbb{N}$ are natural numbers, then equation (3) fails because of obvious congruence obstructions. On the other hand, using the method in [9], it can be shown that if $a, b \in \mathbb{N}$ are distinct, then equation (3) does hold under the additional assumption that the system is totally ergodic; see also [19, 20] for related work regarding polynomials in $\mathbb{R}[t]$ evaluated at primes. The main idea in the proof of these results is to show that the difference of a modification of the averages in equation (3) and the averages equation (1) converges to 0 in $L^2(\mu)$. This comparison method works well when a, b are positive integers since, in this case, one can bound this difference by the Gowers uniformity norm of the modified von Mangoldt function $\tilde{\Lambda}_N$ (see [9, Lemma 3.5] for the precise statement), which is known by [14] to converge to 0 as $N \rightarrow \infty$. Unfortunately, if a, b are not integers, this comparison step breaks down, since it requires a uniformity property for $\tilde{\Lambda}_N$ in which some of the averaging parameters lie in very short intervals, a property that is currently not known. An alternative approach for establishing equation (3) is given

¹Throughout, with (X, μ, T) , we mean a probability space (X, \mathcal{X}, μ) together with an invertible, measurable and measure-preserving $T : X \rightarrow X$. The system is ergodic if the only T -invariant sets in \mathcal{X} have measure 0 or 1. If $f \in L^\infty(\mu)$, with $T^n f$, we denote the composition $f \circ T^n$, where $T^n := T \circ \dots \circ T$.

by the argument used in [6] to prove equation (1). It uses the theory of characteristic factors that originates from [16] and eventually reduces the problem to an equidistribution result on nilmanifolds. This method is also blocked since we are unable to establish the needed equidistribution properties on general nilmanifolds.²

Our approach is quite different and is based on a recent result of the author from [8] (see Theorem 2.1 below); it implies that in order to verify equation (3), it suffices to obtain suitable seminorm estimates and equidistribution results on the circle (versus the general nilmanifold that the method of characteristic factors requires). The needed equidistribution property follows from [2] (see Theorem 2.2 below), and the bulk of this article is devoted to the rather tricky proof of the seminorm estimates (see Theorem 1.4 below).

1.2. Main results

To facilitate discussion, we use the following definition from [8].

Definition. We say that the collection of sequences $b_1, \dots, b_\ell: \mathbb{N} \rightarrow \mathbb{Z}$ is *jointly ergodic* if, for every ergodic system (X, μ, T) and functions $f_1, \dots, f_\ell \in L^\infty(\mu)$, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N T^{b_1(n)} f_1 \cdot \dots \cdot T^{b_\ell(n)} f_\ell = \int f_1 d\mu \cdot \dots \cdot \int f_\ell d\mu$$

in $L^2(\mu)$.

For instance, the identities in equations (1) and (3) are equivalent to the joint ergodicity of the pairs of sequences $\{[n^a], [n^b]\}$ and $\{[p_n^a], [p_n^b]\}$ when $a, b \in \mathbb{R}_+$ are distinct nonintegers.

We will establish joint ergodicity properties involving the class of fractional polynomials that we define next.

Definition. A *polynomial with real exponents* is a function $a: \mathbb{R}_+ \rightarrow \mathbb{R}$ of the form $a(t) = \sum_{j=1}^r \alpha_j t^{d_j}$, where $\alpha_j \in \mathbb{R}$ and $d_j \in \mathbb{R}_+, j = 1, \dots, r$. If $d_1, \dots, d_r \in \mathbb{R}_+ \setminus \mathbb{Z}$, we call it a *fractional polynomial*.

The following is the main result of this article:

Theorem 1.1. *Let a_1, \dots, a_ℓ be linearly independent³ fractional polynomials. Then the collection of sequences $[a_1(p_n)], \dots, [a_\ell(p_n)]$ is jointly ergodic.*

In particular, this applies to the collection of sequences $[n^{c_1}], \dots, [n^{c_\ell}]$, where $c_1, \dots, c_\ell \in \mathbb{R}_+ \setminus \mathbb{Z}$ are distinct. We remark also that the linear independence assumption is necessary for joint ergodicity. Indeed, suppose that a_1, \dots, a_ℓ is a collection of linearly depended sequences. Then $c_1 a_1 + \dots + c_\ell a_\ell = 0$ for some $c_1, \dots, c_\ell \in \mathbb{R}$ not all of them 0. After multiplying by an appropriate constant, we can assume that at least one of the c_1, \dots, c_ℓ is not an integer and $\max_{i=1, \dots, \ell} |c_i| \leq 1/(10\ell)$. Then $c_1 [a_1(n)] + \dots + c_\ell [a_\ell(n)] \in [-1/10, 1/10]$ for all $n \in \mathbb{N}$, and this easily implies that the collection $[a_1(n)], \dots, [a_\ell(n)]$ is not good for equidistribution (see definition in Section 2) and hence not jointly ergodic.

Using standard methods, we immediately deduce from Theorem 1.1 the following multiple recurrence result:

Corollary 1.2. *Let a_1, \dots, a_ℓ be linearly independent fractional polynomials. Then for every system (X, μ, T) and measurable set A , we have*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-[a_1(p_n)]} A \cap \dots \cap T^{-[a_\ell(p_n)]} A) \geq (\mu(A))^{\ell+1}.$$

²For polynomials with integer degrees, the needed equidistribution property can be verified using a comparison method that again breaks down when the degrees are fractional.

³Henceforth, when we say ‘linearly independent’, we mean linearly independent over \mathbb{R} .

Using the Furstenberg correspondence principle [10, 11], we deduce the following combinatorial consequence:

Corollary 1.3. *Let a_1, \dots, a_ℓ be linearly independent fractional polynomials. Then for every subset Λ of \mathbb{N} , we have⁴*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \bar{d}(\Lambda \cap (\Lambda - [a_1(p_n)]) \cap \dots \cap (\Lambda - [a_\ell(p_n)])) \geq (\bar{d}(\Lambda))^{\ell+1}.$$

Hence, every set of integers with positive upper density contains patterns of the form $\{m, m + [a_1(p_n)], \dots, m + [a_\ell(p_n)]\}$ for some $m, n \in \mathbb{N}$.

An essential tool in the proof of our main result is the following statement that is of independent interest since it covers a larger class of collections of fractional polynomials (not necessarily linearly independent) evaluated at primes. See Section 2 for the definition of the seminorms $\|\cdot\|_s$.

Theorem 1.4. *Suppose that the fractional polynomials a_1, \dots, a_ℓ and their pairwise differences are nonzero. Then there exists $s \in \mathbb{N}$ such that for every ergodic system (X, μ, T) and functions $f_1, \dots, f_\ell \in L^\infty(\mu)$ with $\|f_i\|_s = 0$ for some $i \in \{1, \dots, \ell\}$, we have*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N T^{[a_1(p_n)]} f_1 \cdot \dots \cdot T^{[a_\ell(p_n)]} f_\ell = 0 \tag{4}$$

in $L^2(\mu)$.

Remark. It seems likely that with some additional effort the techniques of this article can cover the more general case of Hardy field functions a_1, \dots, a_ℓ such that the functions and their differences belong to the set $\{a: \mathbb{R}_+ \rightarrow \mathbb{R}: t^{k+\varepsilon} < a(t) < t^{k+1-\varepsilon} \text{ for some } k \in \mathbb{Z}_+ \text{ and } \varepsilon > 0\}$. Using the equidistribution result in [3] and the argument in Section 2, this would immediately give a corresponding strengthening of Theorem 1.1. We opted not to deal with these more general statements because the added technical complexity would obscure the main ideas of the proof of Theorem 1.4.

The proof of Theorem 1.4 crucially uses the fact that the iterates a_1, \dots, a_ℓ have ‘fractional power growth’, and our argument fails for iterates with ‘integer power growth’. Similar results that cover the case of polynomials with integer or real coefficients were obtained in [9, 29] and [19], respectively, and depend on deep properties of the von Mangoldt function from [13] and [14], but these results and their proofs do not appear to be useful for our purposes. Instead, we rely on some softer number theory input that follows from standard sieve theory techniques (see Section 3.2) and an argument that is fine-tuned for the case of fractional polynomials (but fails for polynomials with integer exponents). This argument eventually enables us to bound the averages in equation (4) with averages involving iterates given by multivariate polynomials with real coefficients evaluated at the integers, a case that was essentially handled in [23].

1.3. Limitations of our techniques and open problems

We expect that the following generalisation of Theorem 1.1 holds:

Problem. *Let a_1, \dots, a_ℓ be functions from a Hardy field with polynomial growth such that every nontrivial linear combination b of them satisfies $|b(t) - p(t)|/\log t \rightarrow \infty$ for all $p \in \mathbb{Z}[t]$. Then the collection of sequences $[a_1(p_n)], \dots, [a_\ell(p_n)]$ is jointly ergodic.*

⁴For $A \subset \mathbb{N}$, we let $\bar{d}(A) := \limsup_{N \rightarrow \infty} \frac{|A \cap [N]|}{N}$.

By Theorem 2.1, it suffices to show that the collection $[a_1(p_n)], \dots, [a_\ell(p_n)]$ is good for equidistribution and seminorm estimates. Although the needed equidistribution property has been proved in [3, Theorem 3.1], the seminorm estimates that extend Theorem 1.4 seem hard to establish. Our argument breaks down when some of the functions, or their differences, are close to integral powers of t : for example, when they are $t^k \log t$ or $t^k / \log \log t$ for some $k \in \mathbb{N}$. In both cases, the vdC-operation (see Section 5.2) leads to sequences of sublinear growth for which we can no longer establish Lemma 4.1, in the first case because the estimate equation (20) fails and in the second case because in equation (22), the length of the interval in the average is too short for Corollary 3.4 to be applicable.

Finally, we remark that although the reduction offered by Theorem 2.1 is very helpful when dealing with averages with independent iterates, as is the case in equation (3), it does not offer any help when the iterates are linearly dependent, which is the case for the averages

$$\frac{1}{N} \sum_{n=1}^N T^{\lfloor p_n^a \rfloor} f \cdot T^{2\lfloor p_n^a \rfloor} g, \tag{5}$$

where $a \in \mathbb{R}_+$ is not an integer. We do expect the $L^2(\mu)$ -limit of the averages in equation (5) to be equal to the $L^2(\mu)$ -limit of the averages $\frac{1}{N} \sum_{n=1}^N T^n f \cdot T^{2n} g$, but this remains a challenging open problem⁵; see Problem 27 in [7].

1.4. Notation

With \mathbb{N} , we denote the set of positive integers, and with \mathbb{Z}_+ , the set of nonnegative integers. With \mathbb{P} , we denote the set of prime numbers. With \mathbb{R}_+ , we denote the set of nonnegative real numbers. For $t \in \mathbb{R}$, we let $e(t) := e^{2\pi i t}$. If $x \in \mathbb{R}_+$, when there is no danger for confusion, with $[x]$, we denote both the integer part of x and the set $\{1, \dots, [x]\}$. We denote with $\Re(z)$ the real part of the complex number z .

Let $a: \mathbb{N} \rightarrow \mathbb{C}$ be a bounded sequence. If A is a nonempty finite subset of \mathbb{N} , we let

$$\mathbb{E}_{n \in A} a(n) := \frac{1}{|A|} \sum_{n \in A} a(n).$$

If $a, b: \mathbb{R}_+ \rightarrow \mathbb{R}$ are functions, we write

- $a(t) < b(t)$ if $\lim_{t \rightarrow +\infty} a(t)/b(t) = 0$;
- $a(t) \sim b(t)$ if $\lim_{t \rightarrow +\infty} a(t)/b(t)$ exists and is nonzero;
- $A_{c_1, \dots, c_\ell}(t) \ll_{c_1, \dots, c_\ell} B_{c_1, \dots, c_\ell}(t)$ if there exist $t_0 = t_0(c_1, \dots, c_\ell) \in \mathbb{R}_+$ and $C = C(c_1, \dots, c_\ell) > 0$ such that $|A_{c_1, \dots, c_\ell}(t)| \leq C|B_{c_1, \dots, c_\ell}(t)|$ for all $t \geq t_0$.

We use the same notation for sequences $a, b: \mathbb{N} \rightarrow \mathbb{R}$.

Throughout, we let $L_N := \lfloor e^{\sqrt{\log N}} \rfloor$, $N \in \mathbb{N}$.

We say that a sequence $(c_{N, \underline{h}}(n))$, where $\underline{h} \in [L_N]^k$, $n \in [N]$, $N \in \mathbb{N}$, is bounded if there exists $C > 0$ such that $|c_{N, \underline{h}}(n)| \leq C$ for all $\underline{h} \in [L_N]^k$, $n \in [N]$, $N \in \mathbb{N}$.

2. Proof strategy

Our argument depends upon a convenient criterion for joint ergodicity that was established recently in [8] (and was motivated by work in [24, 25]). To state it, we need to review the definition of the ergodic seminorms from [16].

⁵Although the method of Theorem 2.4 does give good seminorm bounds for the averages in equation (5), the needed equidistribution properties on nilmanifolds present serious difficulties.

Definition. For a given ergodic system (X, μ, T) and function $f \in L^\infty(\mu)$, we define $\|\cdot\|_s$ inductively as follows:

$$\|f\|_1 := \left| \int f \, d\mu \right|;$$

$$\|f\|_{s+1}^{2^{s+1}} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \|\bar{f} \cdot T^n f\|_s^{2^s}, \quad s \in \mathbb{N}.$$

It was shown in [16], via successive uses of the mean ergodic theorem, that for every $s \in \mathbb{N}$, the above limit exists, and $\|\cdot\|_s$ defines an increasing sequence of seminorms on $L^\infty(\mu)$.

Definition. We say that the collection of sequences $b_1, \dots, b_\ell: \mathbb{N} \rightarrow \mathbb{Z}$ is:

1. *Good for seminorm estimates*, if for every ergodic system (X, μ, T) , there exists $s \in \mathbb{N}$ such that if $f_1, \dots, f_\ell \in L^\infty(\mu)$ and $\|f_m\|_s = 0$ for some $m \in \{1, \dots, \ell\}$, then

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} T^{b_1(n)} f_1 \cdot \dots \cdot T^{b_m(n)} f_m = 0 \tag{6}$$

in $L^2(\mu)$.⁶

2. *Good for equidistribution*, if for all $t_1, \dots, t_\ell \in [0, 1)$, not all of them 0, we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} e(b_1(n)t_1 + \dots + b_\ell(n)t_\ell) = 0.$$

We remark that any collection of nonconstant integer polynomial sequences with pairwise nonconstant differences is known to be good for seminorm estimates [23], and examples of periodic systems show that no such collection is good for equidistribution (unless $\ell = 1$ and $b_1(t) = \pm t + k$). On the other hand, a collection of linearly independent fractional polynomials is known to be good both for seminorm estimates [6, Theorem 2.9] and equidistribution (follows from [22, Theorem 3.4] and [8, Lemma 6.2]).

A crucial ingredient used in the proof of our main result is the following result that gives convenient necessary and sufficient conditions for joint ergodicity of a collection of sequences (see also [5] for an extension of this result for sequences $b_1, \dots, b_\ell: \mathbb{N}^k \rightarrow \mathbb{Z}$).

Theorem 2.1 ([8]). *The sequences $b_1, \dots, b_\ell: \mathbb{N} \rightarrow \mathbb{Z}$ are jointly ergodic if and only if they are good for equidistribution and seminorm estimates.*

Remark. The proof of this result uses ‘soft’ tools from ergodic theory and avoids deeper tools like the Host-Kra theory of characteristic factors (see [17, Chapter 21] for a detailed description) and equidistribution results on nilmanifolds.

In view of this result, in order to establish Theorem 1.1, it suffices to show that a collection of linearly independent fractional polynomials evaluated at primes is good for seminorm estimates and equidistribution. The good equidistribution property is a consequence of the following result [2, Theorem 2.1]:

Theorem 2.2 ([2]). *If $a(t)$ is a nonzero fractional polynomial, then the sequence $(a(p_n))$ is equidistributed(mod 1).*

Using the previous result and [8, Lemma 6.2], we immediately deduce the following:

Corollary 2.3. *If a_1, \dots, a_ℓ are linearly independent fractional polynomials, then the collection of sequences $[a_1(p_n)], \dots, [a_\ell(p_n)]$ is good for equidistribution.*

We let $\Lambda': \mathbb{N} \rightarrow \mathbb{R}_+$ be the following slight modification of the von Mangoldt function: $\Lambda'(n) := \log(n)$ if n is a prime number and 0 otherwise. To establish that the collection $[a_1(p_n)], \dots, [a_\ell(p_n)]$

⁶In practice, s can often be chosen independently of the system, and equation (6) can be established with $m = \ell$ (note that Property (i) with $m = \ell$ in equation (6) is a stronger property than Property (i) as stated).

is good for seminorm estimates, it suffices to prove the following result (the case $w_N(n) := \Lambda'(n)$, $N, n \in \mathbb{N}$, implies Theorem 1.4 in a standard way; see, for example, [9, Lemma 2.1]):

Theorem 2.4. *Suppose that the fractional polynomials a_1, \dots, a_ℓ and their pairwise differences are nonzero. Then there exists $s \in \mathbb{N}$ such that the following holds: If (X, μ, T) is an ergodic system and $f_1, \dots, f_\ell \in L^\infty(\mu)$ are such that $\|f_i\|_s = 0$ for some $i \in \{1, \dots, \ell\}$, then for every 1-bounded sequence $(c_N(n))$, we have*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} w_N(n) \cdot T^{[a_1(n)]} f_1 \cdot \dots \cdot T^{[a_\ell(n)]} f_\ell = 0 \tag{7}$$

in $L^2(\mu)$, where $w_N(n) := \Lambda'(n) \cdot c_N(n)$, $n \in [N]$, $N \in \mathbb{N}$.

Remarks.

- The sequence $(c_N(n))$ is not essential in order to deduce Theorem 1.4. It is used because it helps us absorb error terms that often appear in our argument.
- Our proof shows that the place of the sequence $(\Lambda'(n))$ can take any nonnegative sequence $(b(n))$ that satisfies properties (i) and (ii) of Corollary 3.4 and the estimate $b(n) \ll n^\varepsilon$ for every $\varepsilon > 0$.

To prove Theorem 2.4, we use an induction argument, similar to the polynomial exhaustion technique (PET-induction) introduced in [1], which is based on variants of the van der Corput inequality stated immediately after Lemma 3.5. The fact that the weight sequence $(w_N(n))$ is unbounded forces us to apply Lemma 3.5 in the form given in equation (15) with $L_N \in \mathbb{N}$ that satisfy $L_N > (\log N)^A$ for every $A > 0$. On the other hand, since we have to take care of some error terms that are of the order L_N^B/N^a for arbitrary $a, B > 0$, we are also forced to take $L_N < N^a$ for every $a > 0$ in order for these errors to be negligible. These two estimates are satisfied for example when $L_N = \lfloor e^{\sqrt{\log N}} \rfloor$, $N \in \mathbb{N}$, which is the value of L_N that we use henceforth.

During the course of the PET-induction argument, we have to keep close track of the additional parameters h_1, \dots, h_k that arise after each application of Lemma 3.5 in the form given in equation (15). This is why we prove a more general variant of Theorem 2.4 that is stated in Theorem 3.1 and involves fractional polynomials with coefficients depending on finitely many parameters. It turns out that the most laborious part of its proof is the base case of the induction where all iterates have sublinear growth. This case is dealt with in three steps. First, in Lemma 4.1, we use a change of variables argument and the number theory input from Corollary 3.4 to reduce matters to the case where the weight sequence $(w_N(n))$ is bounded. Next, in Lemma 4.2, we use another change of variables argument and Lemma 3.5 to successively ‘eliminate’ the sequences a_1, \dots, a_ℓ , and, after ℓ -iterations, we get an upper bound that involves iterates given by the integer parts of polynomials in several variables with real coefficients. Finally, in Lemma 4.3, we show that averages with such iterates obey good seminorm bounds. This last step is carried out by adapting an argument from [23] to our setup; this is done by another PET-induction, which this time uses Lemma 3.5 in the form given in equation (16). In Sections 4.1 and 5.1, the reader will find examples that explain how these arguments work in some model cases that contain the essential ideas of the general arguments.

To conclude this section, we remark that to prove Theorem 1.1, it suffices to prove Theorem 2.4; the remaining sections are devoted to this task.

3. Seminorm estimates – some preparation

3.1. A more general statement

To prove Theorem 2.4, it will be convenient to establish a technically more complicated statement that is better suited for a PET-induction argument. We state it in this subsection.

Throughout, the sequence L_N is chosen to satisfy $(\log N)^A < L_N < N^a$ for all $A, a > 0$; so we can take, for example,

$$L_N := \lceil e^{\sqrt{\log N}} \rceil, \quad N \in \mathbb{N}.$$

With $\mathbb{R}[t_1, \dots, t_k]$, we denote the set of polynomials with real coefficients in k -variables.

Definition. We say that $a : \mathbb{Z}^k \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is a *polynomial with real exponents and k -parameters* if it has the form

$$a(\underline{h}, t) = \sum_{j=0}^r p_j(\underline{h}) t^{d_j}$$

for some $r \in \mathbb{N}, 0 = d_0 < d_1 < \dots < d_r \in \mathbb{R}_+$, and $p_0, \dots, p_r \in \mathbb{R}[t_1, \dots, t_k]$. If $d_1, \dots, d_r \in \mathbb{R}_+ \setminus \mathbb{Z}$, we call it a *fractional polynomial with k -parameters*. If p_j is nonzero for some $j \in \{1, \dots, r\}$, we say that $a(\underline{h}, t)$ is *nonconstant*. We define the *fractional degree* of $a(\underline{h}, t)$ to be the maximum exponent d_j for which the polynomial p_j is nonzero and denote it by $\text{f-deg}(a)$. We call the integer part of its fractional degree the *(integral) degree* of $a(\underline{h}, t)$ and denote it by $\text{deg}(a)$. We also let $\text{deg}(0) := -1$.

For example, the fractional polynomial with 1-parameter $h^2 t^{0.5} + (h^2 \sqrt{2} + h) t^{0.1}$ has fractional degree 0.5 and degree 0.

Definition. We say that a collection a_1, \dots, a_ℓ of polynomials with real exponents and k -parameters is *nice* if

1. $\text{f-deg}(a_i) \leq \text{f-deg}(a_1)$ for $i = 2, \dots, \ell$, and
2. the functions a_1, \dots, a_ℓ and the functions $a_1 - a_2, \dots, a_1 - a_\ell$ are nonconstant in the variable t (and as a consequence they have positive fractional degree).

Given a sequence $u : \mathbb{N} \rightarrow \mathbb{C}$, we let $(\Delta_h u)(n) := u(n+h) \cdot \overline{u(n)}$, $h, n \in \mathbb{N}$, and if $\underline{h} = (h_1, \dots, h_k)$, we let $(\Delta_{\underline{h}})(u(n)) := (\Delta_{h_k} \cdots \Delta_{h_1})(u(n))$, $h_1, \dots, h_k, n \in \mathbb{N}$. For example, $(\Delta_{(h_1, h_2)})(u(n)) = u(n+h_1+h_2) \cdot \overline{u(n+h_1)} \cdot \overline{u(n+h_2)} \cdot u(n)$, $h_1, h_2, n \in \mathbb{N}$.

Theorem 3.1. For $k \in \mathbb{Z}_+, \ell \in \mathbb{N}$, let $a_1, \dots, a_\ell : \mathbb{N}^k \times \mathbb{N} \rightarrow \mathbb{R}$ be a nice collection of fractional polynomials with k -parameters and $(c_{N, \underline{h}}(n))$ be a 1-bounded sequence. Then there exists $s \in \mathbb{N}$ such that the following holds: If (X, μ, T) is a system and $f_{N, \underline{h}, 1}, \dots, f_{N, \underline{h}, \ell} \in L^\infty(\mu)$, $\underline{h} \in [L_N]^k, N \in \mathbb{N}$, are 1-bounded functions with $f_{N, \underline{h}, 1} = f_1, \underline{h} \in \mathbb{N}^k, N \in \mathbb{N}$ and $\|f_1\|_s = 0$, then

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in [N]} w_{N, \underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h}, n)]} f_{N, \underline{h}, i} \right\|_{L^2(\mu)} = 0, \tag{8}$$

where $w_{N, \underline{h}}(n) := (\Delta_{\underline{h}} \Lambda')(n) \cdot c_{N, \underline{h}}(n)$, $\underline{h} \in [L_N]^k, n \in [N], N \in \mathbb{N}$.

Remark. Our argument also works if $\Delta_{\underline{h}} \Lambda'(n)$ is replaced by other expressions involving Λ' : for example, when $k = 0$, one can use the expression $\prod_{i=1}^m \Lambda'(n+c_i)$, where c_1, \dots, c_m are distinct integers.

If in Theorem 3.1, we take $k = 0$, then we get Theorem 2.4 using an argument that we describe next.

Proof of Theorem 2.4 assuming Theorem 3.1. Let a_1, \dots, a_ℓ and $w_N(n)$ be as in Theorem 2.4. Since the assumptions of Theorem 2.4 are symmetric with respect to the sequences a_1, \dots, a_ℓ , it suffices to show that there exists $s \in \mathbb{N}$ such that if $\|f_1\|_s = 0$, then equation (7) holds.

If a_1 has maximal fractional degree within the family a_1, \dots, a_ℓ , then if we take $k = 0$ and all functions to be independent of N in Theorem 3.1, we get that the conclusion of Theorem 2.4 holds.

Otherwise, we can assume that a_ℓ is the function with the highest fractional degree and, as a consequence, $f\text{-deg}(a_1) < f\text{-deg}(a_\ell)$. It suffices to show that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} w_N(n) \cdot \int f_{N,0} \cdot \prod_{i=1}^{\ell} T^{[a_i(n)]} f_i \, d\mu = 0,$$

where

$$f_{N,0} := \mathbb{E}_{n \in [N]} \bar{w}_N(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(n)]} \bar{f}_i, \quad N \in \mathbb{N}.$$

Note that since f_1, \dots, f_ℓ and c_N are 1-bounded, we have

$$\limsup_{N \rightarrow \infty} \|f_{N,0}\|_\infty \leq \lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \Lambda'(n) = 1,$$

(the last identity follows from the prime number theorem, but we only need the much simpler upper bound) hence, we can assume that $f_{N,0}$ is 1-bounded for every $N \in \mathbb{N}$.

After composing with $T^{-[a_\ell(n)]}$, using the Cauchy-Schwarz inequality, and the identity $[x] - [y] = [x - y] + e$ for some $e \in \{0, 1\}$, we are reduced to showing that

$$\lim_{N \rightarrow \infty} \left\| \mathbb{E}_{n \in [N]} w_N(n) \cdot \prod_{i=1}^{\ell-1} T^{[a_i(n) - a_\ell(n) + e_i(n)]} f_i \cdot T^{[-a_\ell(n) + e_\ell(n)]} f_{N,0} \right\|_{L^2(\mu)} = 0,$$

for some $e_1(n), \dots, e_{\ell-1}(n) \in \{0, 1\}$, $n \in \mathbb{N}$. Next, we would like to replace the error sequences $e_1, \dots, e_{\ell-1}$ with constant sequences. To this end, we use Lemma 3.6 for I a singleton, $J := [N]$, $X := L^\infty(\mu)$, $A_N(n_1, \dots, n_\ell) := \prod_{i=1}^{\ell-1} T^{n_i} f_i \cdot T^{-n_\ell} f_{N,0}$, $n_1, \dots, n_\ell \in \mathbb{Z}$, and $b_i := [a_i - a_\ell]$, $i = 1, \dots, \ell - 1$, $b_\ell := [-a_\ell]$. We get that it suffices to show that

$$\lim_{N \rightarrow \infty} \left\| \mathbb{E}_{n \in [N]} z_N(n) \cdot \prod_{i=1}^{\ell} T^{[a'_i(n)]} g_{N,i} \right\|_{L^2(\mu)} = 0, \tag{9}$$

where

$$a'_i := a_i - a_\ell, \quad i = 1, \dots, \ell - 1, \quad a'_\ell := -a_\ell$$

for some 1-bounded sequence $(z_N(n))$, where $g_{N,i} := T^{\epsilon_i} f_i$, $i = 1, \dots, \ell - 1$, $g_{N,\ell} := T^{\epsilon_\ell} f_{N,0}$, $N \in \mathbb{N}$, for some constants $\epsilon_1, \dots, \epsilon_\ell \in \{0, 1\}$. Note that the family a'_1, \dots, a'_ℓ is nice, and $g_{N,1} = T^{\epsilon_1} f_1$, $N \in \mathbb{N}$, so Theorem 3.1 applies (for $k = 0$ and all but one of the functions independent of N) and gives that there exists $s \in \mathbb{N}$ so that if $\|f_1\|_s = 0$, then equation (9) holds. This completes the proof. \square

We will prove Theorem 3.1 in Sections 4 and 5 using a PET-induction technique. The first section covers the base case of the induction where all the iterates have sublinear growth, and the subsequent section contains the proof of the induction step. Before moving into the details, we gather some basic tools that will be used in the argument.

3.2. Feedback from number theory

The next statement is well known and can be proved using elementary sieve theory methods (see, for example, [15, Theorem 5.7] or [18, Theorem 6.7]).

Theorem 3.2. *Let \mathbb{P} be the set of prime numbers. For all $k \in \mathbb{N}$, there exist $C_k > 0$ such that for all distinct $h_1, \dots, h_k \in \mathbb{N}$ and all $N \in \mathbb{N}$, we have*

$$|\{n \in [N] : n + h_1, \dots, n + h_k \in \mathbb{P}\}| \leq C_k \mathfrak{G}_k(h_1, \dots, h_k) \frac{N}{(\log N)^k},$$

where

$$\mathfrak{G}_k(h_1, \dots, h_k) := \prod_{p \in \mathbb{P}} \left(1 - \frac{1}{p}\right)^{-k} \left(1 - \frac{\nu_p(h_1, \dots, h_k)}{p}\right) \tag{10}$$

and $\nu_p(h_1, \dots, h_k)$ denotes the number of congruence classes mod p that are occupied by h_1, \dots, h_k .

We remark that although $\mathfrak{G}_1 = 1$, the expression $\mathfrak{G}_k(h_1, \dots, h_k)$ is not bounded in h_1, \dots, h_k if $k \geq 2$, and this causes some problems for us. Asymptotics for averages of powers of $\mathfrak{G}_k(h_1, \dots, h_k)$ are given in [12] and [21, Theorem 1.1] using elementary but somewhat elaborate arguments. These results are not immediately applicable for our purposes, since we need to understand the behavior of \mathfrak{G}_k on thin subsets of \mathbb{Z}^k : for instance, when $k = 4$, we need to understand the averages of $\mathfrak{G}_4(0, h_1, h_2, h_1 + h_2)$. Luckily, we only need to get upper bounds for these averages, and this can be done rather easily, as we will see shortly (a similar argument was used in [28] to handle averages over r of $\mathfrak{G}_k(0, r, 2r, \dots, (k - 1)r)$).

Definition. Let $\ell \in \mathbb{N}$, and for $\underline{h} \in \mathbb{N}^\ell$, let $\text{Cube}(\underline{h}) \in \mathbb{N}^{2^\ell}$ be defined by

$$\text{cube}(\underline{h}) := (\underline{\epsilon} \cdot \underline{h})_{\underline{\epsilon} \in \{0,1\}^\ell},$$

where $\underline{\epsilon} \cdot \underline{h}$ is the inner product of $\underline{\epsilon}$ and \underline{h} .

If S is a subset of \mathbb{N}^ℓ , we define

$$S^* := \{\underline{h} \in S : \text{cube}(\underline{h}) \text{ has distinct coordinates}\}.$$

For instance, when $\ell = 3$, we have

$$\text{cube}(h_1, h_2, h_3) = (0, h_1, h_2, h_3, h_1 + h_2, h_1 + h_3, h_2 + h_3, h_1 + h_2 + h_3),$$

and $([N]^\ell)^*$ consists of all triples $(h_1, h_2, h_3) \in [N]^3$ with distinct coordinates that in addition satisfy $h_i \neq h_j + h_k$ for all distinct $i, j, k \in \{1, 2, 3\}$. Since the complement of $([N]^\ell)^*$ in $[N]^\ell$ is contained on the zero set of finitely many (at most 3^ℓ) linear forms, we get that there exists $K_\ell > 0$ such that

$$|[N]^\ell \setminus ([N]^\ell)^*| \leq K_\ell N^{\ell-1} \tag{11}$$

for every $N \in \mathbb{N}$.

Proposition 3.3. *For every $\ell \in \mathbb{N}$, there exists $C_\ell > 0$ such that*

$$\mathbb{E}_{\underline{h} \in [N]^\ell} (\mathfrak{G}_{2^\ell}(\text{cube}(\underline{h})))^2 \leq C_\ell$$

for all $N \in \mathbb{N}$, where $\mathfrak{G}_{2^\ell}(\text{cube}(\underline{h}))$ is as in equation (10).

Remark. If we use k th powers instead of squares, we get similar upper bounds (which also depend on k), but we will not need this.

Proof. In the following argument, whenever we write p , we assume that p is a prime number.

Let $\underline{h} \in [N]^\ell$. Note that if $\nu_p(\text{cube}(\underline{h})) = 2^\ell$, then

$$\left(1 - \frac{1}{p}\right)^{-2^\ell} \left(1 - \frac{\nu_p(\text{cube}(\underline{h}))}{p}\right) \leq 1,$$

and if $v_p(\text{cube}(\underline{h})) < 2^\ell$, then for $a_\ell := 2^{\ell+1} - 2$, we have

$$\left(1 - \frac{1}{p}\right)^{-2^\ell} \left(1 - \frac{v_p(\text{cube}(\underline{h}))}{p}\right) \leq \left(1 - \frac{1}{p}\right)^{-(2^\ell-1)} \leq e^{\frac{a_\ell}{p}},$$

where we use that $\frac{1}{1-x} \leq e^{2x}$ for $x \in [0, \frac{1}{2}]$. Note also that if $v_p(\text{cube}(\underline{h})) < 2^\ell$, then there exist distinct $\underline{\epsilon}, \underline{\epsilon}' \in \{0, 1\}^\ell$ such that $p | (\underline{\epsilon} - \underline{\epsilon}') \cdot \underline{h}$, in which case we have that $p \in \mathcal{P}(\underline{h})$, where

$$\mathcal{P}(\underline{h}) := \bigcup_{\substack{\underline{\epsilon}, \underline{\epsilon}' \in \{0,1\}^\ell, \\ \underline{\epsilon}, \underline{\epsilon}' \neq \underline{\epsilon}'}} \{p \in \mathbb{P} : p | (\underline{\epsilon} - \underline{\epsilon}') \cdot \underline{h}\}, \quad \underline{h} \in \mathbb{N}^\ell.$$

We deduce from the above facts and equation (10) that

$$\mathfrak{G}_{2^\ell}(\text{cube}(\underline{h})) \leq e^{a_\ell \sum_{p \in \mathcal{P}(\underline{h})} \frac{1}{p}}. \tag{12}$$

By [27, Lemma E.1], we have for some $b_\ell, c_\ell > 0$ that

$$e^{a_\ell \sum_{p \in \mathcal{P}(\underline{h})} \frac{1}{p}} \leq b_\ell \sum_{p \in \mathcal{P}(\underline{h})} \frac{(\log p)^{c_\ell}}{p} = b_\ell \sum_{\substack{\underline{\epsilon}, \underline{\epsilon}' \in \{0,1\}^\ell, \\ \underline{\epsilon} \neq \underline{\epsilon}'}} \left(\sum_{p | (\underline{\epsilon} - \underline{\epsilon}') \cdot \underline{h}} \frac{(\log p)^{c_\ell}}{p} \right). \tag{13}$$

Moreover, we get for some $d_\ell, e_\ell > 0$ that

$$\sum_{\underline{h} \in [N]^\ell} \left(\sum_{p | (\underline{\epsilon} - \underline{\epsilon}') \cdot \underline{h}} \frac{(\log p)^{c_\ell}}{p} \right) \leq d_\ell \sum_p \frac{(\log p)^{c_\ell}}{p} \frac{N^\ell}{p} \leq e_\ell N^\ell, \tag{14}$$

for all $N \in \mathbb{N}$, where to get the first estimate, we used the fact that for some $d_\ell > 0$, we have

$$|\underline{h} \in [N]^\ell : p | (\underline{\epsilon} - \underline{\epsilon}') \cdot \underline{h}| \leq d_\ell \frac{N^\ell}{p}$$

for all $N \in \mathbb{N}$, and to get the second estimate, we used that $\sum_p \frac{(\log p)^{c_\ell}}{p^2} < \infty$.

If we take squares in equation (12), sum over all $\underline{h} \in [N]^\ell$ and then use equations (13) and (14), we get the asserted estimate. □

From this we deduce the following estimate that is a crucial ingredient used in the proof of Theorem 2.4:

Corollary 3.4. *Let $\ell \in \mathbb{N}$. Then for every $A \geq 1$, there exist $C_{A,\ell}(\underline{h}) > 0$, $\underline{h} \in \mathbb{N}^\ell$ and $D_{A,\ell} > 0$ such that*

1. *for all $N \in \mathbb{N}$, $\underline{h} = (h_1, \dots, h_\ell) \in (\mathbb{N}^\ell)^*$, $c \in \mathbb{N}$, such that $c + h_1 + \dots + h_\ell \leq N^A$, we have*

$$\mathbb{E}_{n \in [N]} (\Delta_{\underline{h}} \Lambda')(n + c) \leq C_{A,\ell}(\underline{h});$$

2. $\mathbb{E}_{\underline{h} \in [H]^\ell} (C_{A,\ell}(\underline{h}))^2 \leq D_{A,\ell}$ for every $H \in \mathbb{N}$.

Remark. We will use this result in the proof of Lemma 4.1 for values of c that are larger than N and smaller than N^A for some $A > 0$ (the choice of A depends on the situation).

Proof. Since Λ' is supported on primes and $c + h_1 + \dots + h_\ell \leq N^A$, we have that

$$\sum_{n \in [N]} (\Delta_{\underline{h}} \Lambda')(n + c) \leq |\{n \in [N] : \underline{n} + c + \text{cube}(\underline{h}) \in \mathbb{P}^{2^\ell}\}| \cdot (\log(N + N^A))^{2^\ell},$$

where $\underline{n+c}$ is a vector with 2^ℓ coordinates, all equal to $n+c$. Note that for $\underline{h} \in (\mathbb{N}^\ell)^*$, we can apply Theorem 3.2, and we get that there exists $D_{A,\ell} > 0$ such that for every $N \in \mathbb{N}$, the last expression is bounded by

$$D_{A,\ell} \mathfrak{G}_{2^\ell}(\text{cube}(\underline{h})) N.$$

If we let $C_{A,\ell}(\underline{h}) := D_{A,\ell} \mathfrak{G}_{2^\ell}(\text{cube}(\underline{h}))$ and $\underline{h} \in \mathbb{N}^\ell$ and use Proposition 3.3, we get that properties (i) and (ii) hold. □

3.3. Two elementary lemmas

We will use the following inner product space variant of a classical elementary estimate of van der Corput (see [22, Lemma 3.1]):

Lemma 3.5. *Let $N \in \mathbb{N}$ and $(u(n))_{n \in [N]}$ be vectors in some inner product space. Then for all $H \in [N]$, we have*

$$\|\mathbb{E}_{n \in [N]} u(n)\|^2 \leq \frac{2}{H} \mathbb{E}_{n \in [N]} \|u(n)\|^2 + 4 \mathbb{E}_{h \in [H]} \left(1 - \frac{h}{H}\right) \Re \left(\frac{1}{N} \sum_{n=1}^{N-h} \langle u(n+h), u(n) \rangle \right).$$

We will apply the previous lemma in the following two cases, depending on the range of the shift parameter h (the first case will be used when the relevant sequences are not necessarily bounded).

1. If $M_N := 1 + \max_{n \in [N]} \|u_N(n)\|^2$, $N \in \mathbb{N}$ and L_N are such that $M_N < L_N < \frac{N}{M_N}$, then for $H := L_N$, we have

$$\|\mathbb{E}_{n \in [N]} u_N(n)\|^2 \leq 4 \mathbb{E}_{h \in [L_N]} \left| \mathbb{E}_{n \in [N]} \langle u_N(n+h), u_N(n) \rangle \right| + o_N(1), \tag{15}$$

where for every fixed $N \in \mathbb{N}$, the sequence $(u_N(n))$ is either defined on the larger interval $[N + L_N]$ or extended to be zero outside the interval $[N]$. In all the cases where we will apply this estimate, we have $M_N \ll (\log N)^A$ for some $A > 0$, and we take $L_N = \lceil e^{\sqrt{\log N}} \rceil$, $N \in \mathbb{N}$.

2. If the sequence $(u_N(n))$ is bounded, then we have

$$\limsup_{N \rightarrow \infty} \|\mathbb{E}_{n \in [N]} u_N(n)\|^2 \leq 4 \limsup_{H \rightarrow \infty} \mathbb{E}_{h \in [H]} \limsup_{N \rightarrow \infty} \left| \mathbb{E}_{n \in [N]} \langle u_N(n+h), u_N(n) \rangle \right|, \tag{16}$$

where for every fixed $N \in \mathbb{N}$, the sequence $(u_N(n))$ is either defined on the larger interval $[N + H]$ or extended to be zero outside the interval $[N]$.

We will also make frequent use of the following simple lemma, or variants of it, to replace error sequences that take finitely many integer values with constant sequences.

Lemma 3.6. *For $f, \ell \in \mathbb{N}$, there exists $C_{f,\ell} > 0$ such that the following holds: Let $(X, \|\cdot\|)$ be a normed space and F be a finite subset of \mathbb{Z} with $|F| = f$, $k \in \mathbb{N}$, and $I \subset \mathbb{N}^k$, $J \subset \mathbb{N}$ be finite. For $\underline{h} \in I$, consider sequences $A_{\underline{h}}: \mathbb{Z}^\ell \rightarrow X$, $b_{1,\underline{h}}, \dots, b_{\ell,\underline{h}}: J \rightarrow \mathbb{Z}$, $w_{\underline{h}}: J \rightarrow \mathbb{C}$, and $e_{1,\underline{h}}, \dots, e_{\ell,\underline{h}}: J \rightarrow F$. Then there exist sequences $\tilde{w}_{\underline{h}}: J \rightarrow \mathbb{C}$, $\underline{h} \in I$, with $\|\tilde{w}_{\underline{h}}\|_{L^\infty(J)} \leq \|w_{\underline{h}}\|_{L^\infty(J)}$ and constants $\epsilon_1, \dots, \epsilon_\ell \in F$, such that*

$$\sum_{\underline{h} \in I} \left\| \sum_{n \in J} w_{\underline{h}}(n) \cdot A_{\underline{h}}(b_{1,\underline{h}}(n) + e_{1,\underline{h}}(n), \dots, b_{\ell,\underline{h}}(n) + e_{\ell,\underline{h}}(n)) \right\| \leq C_{f,\ell} \sum_{\underline{h} \in I} \left\| \sum_{n \in J} \tilde{w}_{\underline{h}}(n) \cdot A_{\underline{h}}(b_{1,\underline{h}}(n) + \epsilon_1, \dots, b_{\ell,\underline{h}}(n) + \epsilon_\ell) \right\|.$$

Remark. Often, when this estimate is used, the sequence $A_{\underline{h}}$ is defined only on a subset of \mathbb{Z}^ℓ , and we assume that it is extended to be zero at the elements where it is not defined.

Proof. The expression on the left-hand side is bounded by

$$\sum_{j=1}^t \sum_{\underline{h} \in I} \left\| \sum_{n \in J} w_{\underline{h}}(n) \cdot A_{\underline{h}}(b_{1,\underline{h}}(n) + e_{1,\underline{h}}(n), \dots, b_{\ell,\underline{h}}(n) + e_{\ell,\underline{h}}(n)) \cdot \mathbf{1}_{E_{j,\underline{h}}}(n) \right\|,$$

where for $t = f^\ell$, the sets $E_{1,\underline{h}}, \dots, E_{t,\underline{h}}$ form a partition of \mathbb{N} into sets (possibly empty) on which all the sequences $e_{1,\underline{h}}, \dots, e_{\ell,\underline{h}}$ are constant (and the constants do not depend on \underline{h}). If the maximum of the summands over j occurs for some $j_0 \in [t]$, then there exist $\epsilon_1, \dots, \epsilon_\ell \in F$ such that for all $n \in E_{j_0,\underline{h}}$, we have $e_{i,\underline{h}}(n) = \epsilon_i, i \in [\ell], \underline{h} \in I$. Hence, the last sum is bounded by

$$t \sum_{\underline{h} \in I} \left\| \sum_{n \in J} \tilde{w}_{\underline{h}}(n) \cdot A_{\underline{h}}(b_{1,\underline{h}}(n) + \epsilon_1, \dots, b_{\ell,\underline{h}}(n) + \epsilon_\ell) \right\|,$$

where $\tilde{w}_{\underline{h}}(n) := w_{\underline{h}}(n) \cdot \mathbf{1}_{E_{j_0,\underline{h}}}(n), n \in J, \underline{h} \in I$. □

We will use the previous lemma to handle some error sequences that occur when we use the Taylor expansion in order to perform some approximations and when we replace the sum (or the difference) of the integer parts of sequences with the corresponding integer part of their sum (or the difference), and vice versa. For instance, if $e_1(n), \dots, e_\ell(n) \in (-1, 1), n \in [N]$, we have

$$\left\| \mathbb{E}_{n \in [N]} w(n) \cdot A([a_1(n) + b_1(n) + e_1(n)], \dots, [a_\ell(n) + b_\ell(n) + e_\ell(n)]) \right\| \leq 4^\ell \left\| \mathbb{E}_{n \in [N]} \tilde{w}(n) \cdot A([a_1(n)] + [b_1(n)] + \epsilon_1, \dots, [a_\ell(n)] + [b_\ell(n)] + \epsilon_\ell) \right\|$$

for some $\epsilon_1, \dots, \epsilon_\ell \in \{-1, 0, 1, 2\}$ and $\tilde{w}: [N] \rightarrow \mathbb{C}$ with $\|\tilde{w}\|_{L^\infty[N]} \leq \|w\|_{L^\infty[N]}$. Often the constants $\epsilon_1, \dots, \epsilon_\ell$ make no difference for our argument and can be ignored.

4. Seminorm estimates – sublinear case

The goal of this section is to establish Theorem 3.1 in the case where all the iterates have fractional degree smaller than 1; see Proposition 4.4 below.

4.1. An example

We explain in some detail how the proof of Theorem 3.1 works when $k = 1, \ell = 2$ and $a_1(h, t) := p_1(h)t^{0.5} + q_1(h)t^{0.1}, a_2(h, t) := p_2(h)t^{0.5} + q_2(h)t^{0.1}, h \in \mathbb{N}, t \in \mathbb{R}_+$. We assume that $p_1 \neq 0$ and $a_1, a_2, a_1 - a_2$ are nonzero.

We also assume that the sequence of weights $(w_{N,h}(n))$ is defined by

$$w_{N,h}(n) := \Lambda'(n) \cdot \Lambda'(n + h) \cdot c_{N,h}(n), \quad h \in [L_N], n \in [N], N \in \mathbb{N},$$

where $(c_{N,h}(n))$ is a 1-bounded sequence.

Our aim is to show that there exists $s \in \mathbb{N}$ such that if $\|f_1\|_s = 0$, then

$$\lim_{N \rightarrow \infty} \mathbb{E}_{h \in [L_N]} \left\| \mathbb{E}_{n \in [N]} w_{N,h}(n) \cdot \prod_{i=1}^2 T^{[p_i(h)n^{0.5} + q_i(h)n^{0.1}]} f_i \right\|_{L^2(\mu)} = 0.$$

Step 1. Our first goal is to use the number theory feedback of Section 3.2 to reduce matters to showing mean convergence to zero for some other averages with bounded weights $w_{N,h}$ (this step corresponds to Lemma 4.1 below). We let

$$p(h) := [\max\{|p_1|(h), |p_2|(h), |q_1|(h), |q_2|(h)\}^{10}] + 1, \quad h \in \mathbb{N}.$$

Note that p is not a polynomial, but this will not bother us. After splitting the average over $[N]$ into subintervals, we see (this reduction will be explained in more detail in the proof of Lemma 4.1) that it suffices to show mean convergence to zero for

$$\mathbb{E}_{h \in [L_N]} \left\| \mathbb{E}_{n \in I_{N,h}} \mathbb{E}_{n_1 \in J_{n,h}} w_{N,h}(n_1) \cdot \prod_{i=1}^2 T^{[p_i(h)n_1^{0.5} + q_i(h)n_1^{0.1}]} f_i \right\|_{L^2(\mu)},$$

where

$$I_{N,h} := [N^{0.5}p(h)], \quad J_{n,h} := \left[\left(\frac{n-1}{p(h)} \right)^2, \left(\frac{n}{p(h)} \right)^2 \right), \quad n \in I_{N,h}, h \in [L_N], N \in \mathbb{N}.$$

For convenience, we write

$$J_{n,h} = (k_{n,h}, k_{n,h} + l_{n,h}], \quad n \in I_{N,h}, h \in [L_N], N \in \mathbb{N}$$

for some $k_{n,h}, l_{n,h} \in \mathbb{N}$.

Note that for fixed $n, h \in \mathbb{N}$, when n_1 ranges in $J_{n,h}$, the value of $p_1(h)n_1^{0.5}$ ranges in an interval of length at most 1; the same property holds for the values of $p_2(h)n_1^{0.5}, q_1(h)n_1^{0.1}, q_2(h)n_1^{0.1}$. Hence, for $n_1 \in J_{n,h}$, we have

$$p_i(h)n_1^{0.5} + q_i(h)n_1^{0.1} = \frac{p_i(h)}{p(h)}n + q_i(h)\left(\frac{n}{p(h)}\right)^{0.2} + e_i(h, n, n_1), \quad i = 1, 2,$$

where $e_1(h, n, n_1), e_2(h, n, n_1)$ are bounded by 2 for all $n_1 \in J_{n,h}, n \in I_{N,h}, h \in [L_N], N \in \mathbb{N}$. Using Lemma 3.6, and since replacing f_i with $T^{\epsilon_{i,N}} f_i, i = 1, 2$, where $\epsilon_{1,N}, \epsilon_{2,N}$ take finitely many values for $N \in \mathbb{N}$, does not introduce changes to our argument, we can ignore these error terms. We are thus left with showing convergence to zero for

$$\mathbb{E}_{h \in [L_N]} \left\| \mathbb{E}_{n \in I_{N,h}} \tilde{w}_{N,h}(n) \cdot \prod_{i=1}^2 T^{[\frac{p_i(h)}{p(h)}n + q_i(h)(\frac{n}{p(h)})^{0.2}]} f_i \right\|_{L^2(\mu)},$$

where for $n \in [I_{N,h}], h \in [L_N], N \in \mathbb{N}$, we let

$$\tilde{w}_{N,h}(n) := \mathbb{E}_{n_1 \in J_{n,h}} w_{N,h}(n_1) = \mathbb{E}_{n_1 \in [l_{n,h}]} \Lambda'(n_1 + k_{n,h}) \cdot \Lambda'(n_1 + k_{n,h} + h) \cdot c_{N,h}(n_1 + k_{n,h}). \quad (17)$$

From the definition of $k_{n,h}, l_{n,h}, L_N$, we get that there exists $N_0 = N_0(p) \in \mathbb{N}$ such that

$$k_{n,h} + h \leq l_{n,h}^3, \quad \text{for all } n \in [N^{0.4}, N^{0.5}p(h)], h \in [L_N], N \geq N_0.$$

Using Corollary 3.4 (with $\ell = 1, A = 3, c = k_{n,h}, N = l_{n,h}$), we see that there exist $D > 0$ and $C(h), h \in \mathbb{N}$, such that for the above-mentioned values of n, h, N , we can write

$$\tilde{w}_{N,h}(n) = C(h) \cdot z_{N,h}(n),$$

where $(z_{N,h}(n))$ is 1-bounded and

$$\mathbb{E}_{h \in [L_N]} (C(h))^2 \leq D$$

for every $N \in \mathbb{N}$.

We use this estimate, apply the Cauchy-Schwarz inequality and keep in mind that the part of the intervals $I_{N,h}$ that intersects the interval $[N^{0.4}]$ is negligible for our averages. We deduce that it suffices to show convergence to zero for

$$\mathbb{E}_{h \in [L_N]} \left\| \mathbb{E}_{n \in I_{N,h}} z_{N,h}(n) \cdot \prod_{i=1}^2 T \left[\frac{p_i(h)}{p(h)} n + q_i(h) \left(\frac{n}{p(h)} \right)^{0.2} \right] f_i \right\|_{L^2(\mu)}^2,$$

where the sequence $(z_{N,h}(n))$ is 1-bounded. We write $n = n'p(h) + s$ for some $n' \in [N^{0.5}]$ and $s \in [p(h)]$. For convenience, we also rename n' as n and use Lemma 3.6 to treat finite valued error sequences that are introduced when we approximate $q_i(h)(n + s/p(h))^{0.2}$ with $q_i(h)n^{0.2}$, $i = 1, 2$. We get that it suffices to show convergence to zero for

$$\mathbb{E}_{h \in [L_N]} \mathbb{E}_{s \in [p(h)]} \left\| \mathbb{E}_{n \in [N^{0.5}]} z_{N,h,s}(n) \cdot \prod_{i=1}^2 T^{[p_i(h)n + q_i(h)n^{0.2} + e_i(h,s)]} f_i \right\|_{L^2(\mu)}^2,$$

where $(z_{N,h,s}(n))$ is some other 1-bounded sequence and $e_i(h, s) := s \frac{p_i(h)}{p(h)}$, $i = 1, 2$. After replacing the average $\mathbb{E}_{s \in [p(h)]}$ with $\max_{s \in [p(h)]}$, we are left with dealing with the averages

$$\mathbb{E}_{h \in [L_N]} \left\| \mathbb{E}_{n \in [N^{0.5}]} z_{N,h}(n) \cdot \prod_{i=1}^2 T^{[p_i(h)n + q_i(h)n^{0.2} + e_{i,N}(h)]} f_i \right\|_{L^2(\mu)}^2$$

for some other 1-bounded sequence $(z_{N,h}(n))$ and arbitrary sequences of real numbers $(e_{1,N}(h))$, $(e_{2,N}(h))$ (which will be eliminated later, so their particular form is not important).

Step 2. Our next goal is to reduce matters to showing mean convergence to zero for averages with iterates given by polynomials in several variables and real coefficients (this step corresponds to Lemma 4.2 below). After using equation (15) for the average over n , we are left with showing convergence to zero for

$$\mathbb{E}_{h, h_1 \in [L_N]} \left| \mathbb{E}_{n \in [N^{0.5}]} c_{N,h,h_1}(n) \cdot \int \prod_{i=1}^2 T^{[p_i(h)(n+h_1) + q_i(h)(n+h_1)^{0.2} + e_{i,N}(h)]} f_i \cdot \prod_{i=1}^2 T^{[p_i(h)n + q_i(h)n^{0.2} + e_{i,N}(h)]} \bar{f}_i \, d\mu \right|,$$

where $(c_{N,h,h_1}(n))$ is a 1-bounded sequence. We compose with $T^{-[p_2(h)n + q_2(h)n^{0.2} + e_{2,N}(h)]}$ (and not with $T^{-[p_1(h)n + q_1(h)n^{0.2} + e_{1,N}(h)]}$ because we want the highest fractional degree iterate to be applied to the function f_1), use that $(n + h_1)^{0.2}$ can for our purposes be replaced with $n^{0.2}$, ignore errors that take finitely many values using Lemma 3.6 and use the Cauchy-Schwarz inequality. We are left with showing convergence to zero for

$$\mathbb{E}_{h, h_1 \in [L_N]} \left\| \mathbb{E}_{n \in [N^{0.5}]} c_{N,h,h_1}(n) \cdot T^{[(p_1 - p_2)(h)n + (q_1 - q_2)(h)n^{0.2} + e_{3,N}(h)]} (T^{[p_1(h)h_1]} f_1 \cdot \bar{f}_1) \right\|_{L^2(\mu)},$$

where $(c_{N,h,h_1}(n))$ is some other 1-bounded sequence and the sequence $(e_{3,N}(h))$ takes arbitrary real values.

We consider two cases. Suppose first that $p_1 = p_2$. Then by assumption, $q_1 - q_2 \neq 0$. Repeating the argument used in Step 1, we are left with showing convergence to zero for

$$\mathbb{E}_{h, h_1 \in [L_N]} \left\| \mathbb{E}_{n \in [N^{0.1}]} c_{N, h, h_1}(n) \cdot T^{[(q_1 - q_2)(h)n + e_{4,N}(h)]} (T^{[p_1(h)h_1]} f_1 \cdot \bar{f}_1) \right\|_{L^2(\mu)}$$

for some other 1-bounded sequence of complex numbers $(c_{N, h, h_1}(n))$ and $(e_{4,N}(h))$ arbitrary sequence of real numbers. Using as above equation (15) for the average over n , composing with $T^{-[(q_1 - q_2)(h)n + e_{4,N}(h)]}$ and then using the Cauchy-Schwarz inequality and Lemma 3.6 to treat errors, we are left with showing mean convergence to zero for

$$\mathbb{E}_{h, h_1, h_2 \in [L_N]} c_{N, h, h_1, h_2}(n) \cdot T^{[(q_1 - q_2)(h)h_2 + p_1(h)h_1]} f_1 \cdot T^{[(q_1 - q_2)(h)h_2]} \bar{f}_1 \cdot T^{[p_1(h)h_1]} \bar{f}_1$$

for some 1-bounded sequence of complex numbers $(c_{N, h, h_1, h_2}(n))$.

If $p_1 \neq p_2$, we apply equation (15) for the average over n , compose with the transformation $T^{-[(p_1 - p_2)(h)n + (q_1 - q_2)(h)n^{0.2} + e_{3,N}(h)]}$ and use the Cauchy-Schwarz inequality and Lemma 3.6 to treat errors. We are left with showing mean convergence to zero for

$$\mathbb{E}_{h, h_1, h_2 \in [L_N]} c_{N, h, h_1, h_2}(n) \cdot T^{[(p_1 - p_2)(h)h_2 + p_1(h)h_1]} f_1 \cdot T^{[(p_1 - p_2)(h)h_2]} \bar{f}_1 \cdot T^{[p_1(h)h_1]} \bar{f}_1$$

for some other 1-bounded sequence of complex numbers $(c_{N, h, h_1, h_2}(n))$.

Step 3. In Step 2, we were led to show mean convergence to zero for averages with iterates given by nonconstant polynomials with real coefficients in several variables that have pairwise nonconstant differences. For such averages, one can argue as in [23] to show that there exists $s \in \mathbb{N}$ such that if $\|f_1\|_s = 0$, then we have mean convergence to zero. For more details, see the proof of Lemma 4.3 below. This achieves our goal.

4.2. Reduction to averages with bounded weights and change of variables

Our first goal is to prove the following result that allows us to restrict to the case where the weights $w_{N, \underline{h}}$ are 1-bounded and also allows us to perform the substitution $n \mapsto n^{1/d}$.

Lemma 4.1. *For $k \in \mathbb{Z}_+, \ell \in \mathbb{N}$, let a_1, \dots, a_ℓ be a nice collection of fractional polynomials with k -parameters and suppose that $d := \text{f-deg}(a_1) \in (0, 1)$. Then the following holds: If (X, μ, T) is a system, $f_{N, \underline{h}, 1}, \dots, f_{N, \underline{h}, \ell} \in L^\infty(\mu)$, $\underline{h} \in \mathbb{N}^k, N \in \mathbb{N}$, are 1-bounded functions, $a > 0$, and*

$$w_{N, \underline{h}}(n) := (\Delta_{\underline{h}} \Lambda')(n) \cdot c_{N, \underline{h}}(n) \quad \text{or} \quad w_{N, \underline{h}}(n) := c_{N, \underline{h}}(n), \quad \underline{h} \in [L_N]^k, n \in [N^a], N \in \mathbb{N},$$

where $(c_{N, \underline{h}}(n))$ is a 1-bounded sequence, then there exist a 1-bounded sequence $(z_{N, \underline{h}}(n))$ and sequences of real numbers $(e_{1, N}(\underline{h})), \dots, (e_{\ell, N}(\underline{h}))$, such that

$$\begin{aligned} \mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in [N^a]} w_{N, \underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h}, n)]} f_{N, \underline{h}, i} \right\|_{L^2(\mu)} &\ll_{k, a_1, \dots, a_\ell} \\ \mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in [N^{ad}]} z_{N, \underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h}, n^{1/d}) + e_{i, N}(\underline{h})]} f_{N, \underline{h}, i} \right\|_{L^2(\mu)} &+ o_N(1), \end{aligned} \quad (18)$$

where $o_N(1)$ is a quantity that converges to 0 when $N \rightarrow \infty$ and all other parameters remain fixed.

Remark. It is important that the function a_1 has sublinear growth; our argument would not work if a_1 had linear or larger than linear growth.

Proof. We cover the case where $w_{N, \underline{h}}(n) = (\Delta_{\underline{h}} \Lambda')(n) \cdot c_{N, \underline{h}}(n)$; the case where $w_{N, \underline{h}}(n) = c_{N, \underline{h}}(n)$ is similar (in fact, easier).

By assumption, we have that $a_i(\underline{h}, t) := \sum_{j=0}^r p_{i,j}(\underline{h})t^{d_j}$, $i = 1, \dots, \ell$, where $0 = d_0 < d_1 < \dots < d_r = d < 1$ and $p_{i,j} \in \mathbb{R}[t_1, \dots, t_k]$ with $p_{1,r} \neq 0$. We let

$$p(\underline{h}) := \left[\max_{i,j} \{ |p_{i,j}(\underline{h})| \}^{\frac{1}{d_1}} \right] + 1, \quad \underline{h} \in \mathbb{N}^k.$$

For $\underline{h} \in [L_N]^k$, after partitioning $[N^a]$ into subintervals, we deduce that it suffices to get an upper bound for the averages

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in I_{N,\underline{h}}} \mathbb{E}_{n_1 \in J_{n,\underline{h}}}^* w_{N,\underline{h}}(n_1) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h}, n_1)]} f_{N,\underline{h},i} \right\|_{L^2(\mu)},$$

where

$$I_{N,\underline{h}} := [N^{ad} p(\underline{h})], \quad J_{n,\underline{h}} := \left[\left(\frac{n-1}{p(\underline{h})} \right)^{\frac{1}{d}}, \left(\frac{n}{p(\underline{h})} \right)^{\frac{1}{d}} \right), \quad n \in I_{N,\underline{h}}, \underline{h} \in [L_N]^k, N \in \mathbb{N},$$

and for $D: \mathbb{N} \rightarrow \mathbb{C}$ and fixed $N \in \mathbb{N}$, $n \in I_{N,\underline{h}}$, $\underline{h} \in [L_N]^k$, we let

$$\mathbb{E}_{n_1 \in J_{n,\underline{h}}}^* D(n_1) := \frac{1}{N^a / |I_{N,\underline{h}}|} \sum_{n_1 \in J_{n,\underline{h}}} D(n_1). \tag{19}$$

Note that an application of the mean value theorem gives

$$|J_{n,\underline{h}}| \leq \frac{1}{d} \frac{(N^{ad} p(\underline{h}))^{\frac{1}{d}-1}}{p(\underline{h})^{\frac{1}{d}}} = \frac{1}{d} \cdot \frac{N^a}{|I_{N,\underline{h}}|}, \quad n \in I_{N,\underline{h}}, \underline{h} \in [L_N]^k, N \in \mathbb{N}. \tag{20}$$

For⁷ convenience, we write

$$J_{n,\underline{h}} = (k_{n,\underline{h}}, k_{n,\underline{h}} + l_{n,\underline{h}}], \quad n \in I_{N,\underline{h}}, \underline{h} \in [L_N]^k, N \in \mathbb{N}$$

for some $k_{n,\underline{h}}, l_{n,\underline{h}} \in \mathbb{N}$. Note that for $i = 1, \dots, \ell$, $j = 1, \dots, r$ and fixed n, \underline{h} , when n_1 ranges on $J_{n,\underline{h}}$, the values of $p_{i,j}(\underline{h})n_1^{d_j}$ belong to an interval of length 1. Hence, for $i = 1, \dots, \ell$, we can write

$$a_i(\underline{h}, n_1) = a_i(\underline{h}, (n/p(\underline{h}))^{1/d}) + \epsilon_i(\underline{h}, n, n_1),$$

where $\epsilon_i(\underline{h}, n, n_1)$ is bounded by r for all $n_1 \in J_{n,\underline{h}}$, $n \in I_{N,\underline{h}}$, $\underline{h} \in [L_N]^k$, $N \in \mathbb{N}$.

The terms $\epsilon_i(\underline{h}, n, n_1)$ can be easily taken care by using Lemma 3.6 and appropriately modifying $(c_{N,\underline{h}}(n))$ to another bounded sequence of weights. We deduce that it suffices to get an upper bound for the averages

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in I_{N,\underline{h}}} \mathbf{1}'_{N,\underline{h}}(n) \tilde{w}_{N,\underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h}, (n/p(\underline{h}))^{1/d})] + \epsilon_{i,N}} f_{N,\underline{h},i} \right\|_{L^2(\mu)}, \tag{21}$$

where $\mathbf{1}'_{N,\underline{h}} := [N^{\frac{\epsilon_d}{2}}, N^{ad} p(\underline{h})]$, $N \in \mathbb{N}$ (the indicator introduces a negligible $o_N(1)$ term), $\epsilon_{1,N}, \dots, \epsilon_{\ell,N}$ take finitely many values for $N \in \mathbb{N}$, and for $n \in [I_{N,\underline{h}}]$, $\underline{h} \in [L_N]^k$, $N \in \mathbb{N}$, we let

$$\tilde{w}_{N,\underline{h}}(n) := \mathbb{E}_{n_1 \in J_{n,\underline{h}}}^* w_{N,\underline{h}}(n_1). \tag{22}$$

We used that $L_N, \Lambda'(N) < N^\epsilon$ for all $\epsilon > 0$ to justify that inserting the indicator $\mathbf{1}'_{N,\underline{h}}$ only introduces an $o_N(1)$ term, which is fine for our purposes.

⁷We crucially used here that fractional polynomials do not grow too slowly. The estimate would fail if, for example, for $\ell = 1$, we started with $a_1(t) := \log t$.

Using that $(c_{N,\underline{h}}(n))$ is 1-bounded, $((\Delta_{\underline{h}}\Lambda')(n))$ is nonnegative and equations (19) and (20), we deduce that

$$|\tilde{w}_{N,\underline{h}}(n)| \leq d^{-1} \cdot \mathbb{E}_{n_1 \in J_{n,\underline{h}}} (\Delta_{\underline{h}}\Lambda')(n_1) = d^{-1} \cdot \mathbb{E}_{n_1 \in [l_{n,\underline{h}}]} (\Delta_{\underline{h}}\Lambda')(n_1 + k_{n,\underline{h}}). \tag{23}$$

From the definition of $l_{n,\underline{h}}$ and the mean value theorem, we have that

$$l_{n,\underline{h}} \geq \frac{n^{\frac{1}{d}-1}}{d(p(\underline{h}))^{\frac{1}{d}}}, \quad n \in \mathbb{N}.$$

Since $L_N < N^\varepsilon$ for every $\varepsilon > 0$ and $k_{n,\underline{h}} \leq n^{1/d}$, it follows that if $A > \frac{1}{1-d}$ – for example, if $A := \frac{1}{1-d} + 1$ – then there exists $N_0 = N_0(d, p) \in \mathbb{N}$ such that for all $N \geq N_0$ and all $n \in I'_{N,\underline{h}}, \underline{h} \in [L_N]^k$, we have

$$k_{n,\underline{h}} \leq l_{n,\underline{h}}^A.$$

Hence,⁸ there exists $N_1 = N_1(d, k, p) \in \mathbb{N}$ such that for all $N \geq N_1$, we have for all $n \in I'_{N,\underline{h}}$ and $\underline{h} = (h_1, \dots, h_k) \in [L_N]^k$ that

$$k_{n,\underline{h}} + h_1 + \dots + h_k \leq l_{n,\underline{h}}^A.$$

We will combine this with the identity

$$(\Delta_{\underline{h}}\Lambda')(n) = \prod_{\epsilon \in \{0,1\}^k} \Lambda'(n + \epsilon \cdot \underline{h}),$$

the estimate equation (23) and Corollary 3.4 (with $\ell := k, c := k_{n,\underline{h}}, N := l_{n,\underline{h}}, A := \frac{1}{1-d} + 1$). We deduce that there exist $C = C(d, k) > 0$ and $C_{d,k}(\underline{h}) > 0, \underline{h} \in \mathbb{N}^k$, such that for all large enough N (depending only on d, k, p), for every $n \in I_{N,\underline{h}}, \underline{h} \in ([L_N]^k)^*$, we can write

$$\tilde{w}_{N,\underline{h}}(n) = C_{d,k}(\underline{h}) \cdot z_{N,\underline{h}}(n), \tag{24}$$

where $(z_{N,\underline{h}}(n))$ is 1-bounded and

$$\mathbb{E}_{\underline{h} \in [L_N]^k} (C_{d,k}(\underline{h}))^2 \leq C \tag{25}$$

for every $N \in \mathbb{N}$.

Note that since $L_N > (\log N)^K$ for every $K > 0$ and $\Lambda'(n) \leq \log n$, for every $n \in \mathbb{N}$, we have that $\max_{\underline{h} \in [L_N]^k, n \in [N]} (\tilde{w}_{N,\underline{h}}(n))^2 < L_N$. Using this, and since by equation (11), we have that $\frac{1}{L_N^k} |[L_N]^k \setminus ([L_N]^k)^*| \ll_k \frac{1}{L_N}$, we deduce that we can redefine $C(\underline{h})$ on the complement of $([L_N]^k)^*$ so that for all large enough N (depending on d, k, p), equation (24) holds for all $n \in I_{N,\underline{h}}, \underline{h} \in [L_N]^k$, and equation (25) also holds (for some larger constant C' in place of C).

We now use equations (24) and (25) and the Cauchy-Schwarz inequality to bound the averages in equation (21). We can also remove the indicator $\mathbf{1}_{I'_{N,\underline{h}}}(n)$ since it has a negligible effect on our averages. We deduce that it suffices to get an upper bound for the averages

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in I_{N,\underline{h}}} z_{N,\underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h}, (n/p(\underline{h}))^{1/d})] + \epsilon_{i,N}} f_{N,\underline{h},i} \right\|_{L^2(\mu)}^2.$$

⁸In the process of deriving this estimate, we crucially used that sublinear fractional polynomials are not too close to linear ones. The estimate would fail if, for example, for $\ell = 1$, we started with $a_1(t) := t/\log t$.

Note that since the weights and functions are bounded, it suffices to get an upper bound for the previous expression, ignoring the square. For $\underline{h} \in [L_N]^k$, we can express $n \in I_{N,\underline{h}}$ as $n = n'p(\underline{h}) + s$ for some $n' \in [N^{ad}]$ or $n' = 0$ and $s \in [p(\underline{h})]$. After renaming n' as n for convenience, we are led to upper-bounding the averages

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \mathbb{E}_{s \in [p(\underline{h})]} \left\| \mathbb{E}_{n \in [N^{ad}]} z_{N,\underline{h},s}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h},(n+s/p(\underline{h}))^{1/d})+\epsilon_{i,N}]} f_{N,\underline{h},i} \right\|_{L^2(\mu)} \tag{26}$$

for some 1-bounded sequence $(z_{N,\underline{h},s}(n))$. Note that if $u \in (0, 1)$ and $q \in \mathbb{R}[t_1, \dots, t_k]$, then an application of the mean value theorem shows that for every $\varepsilon > 0$, we have

$$\lim_{N \rightarrow \infty} \sup_{c \in [0,1], \underline{h} \in [L_N]^k, n \geq N^\varepsilon} |q(\underline{h})((n+c)^u - n^u)| = 0.$$

It follows that in equation (26), when computing $a_i(\underline{h}, (n+s/p(\underline{h}))^{1/d})$, we can replace $n+s/p(\underline{h})$ with n in the nonlinear monomials; this will lead to some error sequences that are 1-bounded for large enough N and can be handled by appealing to Lemma 3.6 (and redefining the sequence $z_{N,\underline{h}}(n)$). With this in mind, it follows that in equation (26), we can replace $a_i(\underline{h}, (n+s/p(\underline{h}))^{1/d})$ with $a_i(\underline{h}, n^{1/d}) + \frac{P_{i,r}(\underline{h})}{p(\underline{h})}s$. Hence, it suffices to get an upper bound for the averages

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \mathbb{E}_{s \in [p(\underline{h})]} \left\| \mathbb{E}_{n \in [N^{ad}]} z_{N,\underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h},n^{1/d})+e_{i,N}(\underline{h},s)]} f_{N,\underline{h},i} \right\|_{L^2(\mu)},$$

where $e_{i,N}(\underline{h}, s) := \frac{P_{i,r}(\underline{h})}{p(\underline{h})}s + \epsilon_{i,N}$, $i = 1, \dots, \ell$ and $\epsilon_{1,N}, \dots, \epsilon_{\ell,N}$ take finitely many values for $N \in \mathbb{N}$. After replacing the average $\mathbb{E}_{s \in [p(\underline{h})]}$ with $\max_{s \in [p(\underline{h})]}$, we are led to the asserted upper bound in equation (18). □

4.3. Reduction to averages with polynomial iterates

For the purposes of the next lemma, it will be convenient to slightly enlarge the class of polynomials with real exponents that we work with to include those with fractional degree equal to 1.

Lemma 4.2. *Let $k \in \mathbb{Z}_+, \ell \in \mathbb{N}$ and $a_1, \dots, a_\ell: \mathbb{N}^k \times \mathbb{N} \rightarrow \mathbb{R}$ be a nice collection of polynomials with real exponents and k -parameters of fractional degree at most 1. Then there exist $l, r \in \mathbb{N}$ and nonconstant polynomials $P_1, \dots, P_r \in \mathbb{R}[t_1, \dots, t_{k+l}]$, with pairwise nonconstant differences, such that the following holds: If (X, μ, T) is a system and $f_{N,\underline{h},1}, \dots, f_{N,\underline{h},\ell} \in L^\infty(\mu)$, $\underline{h} \in \mathbb{N}^k$, $N \in \mathbb{N}$, are 1-bounded functions, then for every $a > 0$, sequences of real numbers $(e_{1,N}(\underline{h})), \dots, (e_{\ell,N}(\underline{h}))$ and 1-bounded sequence of complex numbers $(c_{N,\underline{h}}(n))$, we have*

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in [N^a]} c_{N,\underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h},n)+e_{i,N}(\underline{h})]} f_{N,\underline{h},i} \right\|_{L^2(\mu)} \ll_{k,a_1,\dots,a_\ell} \mathbb{E}_{\underline{h}_1 \in [L_N]^k, \underline{h}_2 \in [L_N]^l} \left| \int \prod_{i=0}^r T^{[P_i(\underline{h}_1,\underline{h}_2)]+\epsilon_{i,N}} F_{N,\underline{h}_1,i} d\mu \right| + o_N(1), \tag{27}$$

where $P_0 := 0$, $F_{N,\underline{h}_1,i} \in \{f_{N,\underline{h}_1,1}, \overline{f_{N,\underline{h}_1,1}}\}$ for $i = 0, \dots, r$, $\underline{h}_1 \in [L_N]^k$, $N \in \mathbb{N}$, $\epsilon_{0,N}, \dots, \epsilon_{r,N}$ take finitely many values for $N \in \mathbb{N}$, and $o_N(1)$ is a quantity that converges to 0 when $N \rightarrow \infty$ and all other parameters remain fixed.

Proof. We first reduce to the case where $e_{i,N}(\underline{h}) = 0$ for $i = 1, \dots, \ell$. To do this, we replace $[a_i(\underline{h}, n) + e_{i,N}(\underline{h})]$ with $[a_i(\underline{h}, n)] + [e_{i,N}(\underline{h})]$; this introduces some error sequences on the exponents that take

finitely many values. To treat the error sequences, we use Lemma 3.6, redefine the weight $(c_{N,\underline{h}}(n))$ and introduce some sequences $\epsilon_{1,N}, \dots, \epsilon_{\ell,N}$ that take finitely many values for $N \in \mathbb{N}$. Next, we compose with $T^{-[e_{1,N}(\underline{h})] - \epsilon_{1,N}}$, and we are left with upper-bounding the expression

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in [N^a]} c_{N,\underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h},n)]} (T^{[e_{i,N}(\underline{h})] - [e_{1,N}(\underline{h})] + \epsilon_{i,N} - \epsilon_{1,N}} f_{N,\underline{h},i}) \right\|_{L^2(\mu)}.$$

If we rename for $i = 2, \dots, \ell$ the functions $T^{[e_{i,N}(\underline{h})] - [e_{1,N}(\underline{h})] + \epsilon_{i,N} - \epsilon_{1,N}} f_{N,\underline{h},i}$ as $f_{N,\underline{h},i}$, we are reduced to bounding equation (27) when $e_{i,N}(\underline{h}) = 0$ for $i = 1, \dots, \ell$.

We will prove the statement by induction on $\ell \in \mathbb{N}$. For $\ell = 1$, the argument is similar to the one used in the inductive step, so we only summarise it briefly (for more details; see Steps 1–3 below). We first use Lemma 4.1, and we are led to upper-bounding the averages

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in [N^a]} c_{N,\underline{h}}(n) \cdot T^{[p_1(\underline{h})n + q_1(\underline{h},n)]} f_{N,\underline{h},1} \right\|_{L^2(\mu)},$$

where $p_1 \neq 0$ and q_1 is a polynomial with real exponents and $\text{f-deg}(q_1) < 1$. We then apply equation (15) for the average over n , compose with $T^{-[p_1(\underline{h})n + q_1(\underline{h},n)]}$, use that $q_1(\underline{h}, n + h_{k+1}) - q_1(\underline{h}, n)$ is negligible for the range of parameters we are interested in and use Lemma 3.6 to treat the finite valued error sequences that arise. We get an upper bound by the averages

$$\mathbb{E}_{\underline{h} \in [L_N]^k, h_{k+1} \in [L_N]} \left| \int T^{[p_1(\underline{h})h_{k+1}] + \epsilon_N} f_{N,\underline{h},1} \cdot \bar{f}_{N,\underline{h},1} d\mu \right|,$$

where ϵ_N takes finitely many values for $N \in \mathbb{N}$. This proves equation (27) (with $\ell = r = 1$).

Suppose that $\ell \geq 2$ and the statement holds for all nice collections of $\ell - 1$ polynomials with real exponents and finitely many parameters.

We have that $a_i(\underline{h}, t) := \sum_{j=1}^r p_{i,j}(\underline{h}) t^{d_j}$, $i = 1, \dots, \ell$, where $0 \leq d_1 < \dots < d_r = d \leq 1$ and $p_{i,j} \in \mathbb{R}[t_1, \dots, t_k]$. Furthermore, we can assume that the polynomial $p_{1,r}$ is nonzero, and hence the fractional degree of a_1 is d .

Step 1 (Linearising the highest-order term). If the fractional degree of a_1 is 1, then we proceed to Step 2. If not, then Lemma 4.1 (for $w_{N,\underline{h}} := c_{N,\underline{h}}$) applies, and we get an estimate of the form equation (18). Hence, to get an estimate of the form equation (27), it suffices to get a similar estimate for the averages

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in [N^{ad}]} c_{N,\underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[\tilde{a}_i(\underline{h},n)] + e_{i,N}(\underline{h})} f_{N,\underline{h},i} \right\|_{L^2(\mu)},$$

where $(c_{N,\underline{h}}(n))$ is another 1-bounded sequence, $(e_{1,N}(\underline{h})), \dots, (e_{\ell,N}(\underline{h}))$ are sequences of real numbers and

$$\tilde{a}_i(\underline{h}, t) := p_{i,r}(\underline{h}) t + q_i(\underline{h}, t), \quad \text{where} \quad q_i(\underline{h}, t) := \sum_{j=1}^{r-1} p_{i,j}(\underline{h}) t^{\frac{d_j}{d}}, \quad i = 1, \dots, \ell. \quad (28)$$

After composing with $T^{-e_{1,N}(\underline{h})}$ and redefining the functions $f_{N,\underline{h},i}$, $i = 2, \dots, \ell$, we are reduced to the case where $e_{i,N}(\underline{h}) = 0$ for $i = 1, \dots, \ell$. So we only treat this case henceforth. We also remark that since the collection a_1, \dots, a_{ℓ} is nice, and $\tilde{a}_i(\underline{h}, t) = a_i(\underline{h}, t^{1/d})$, $i = 1, \dots, \ell$, the collection $\tilde{a}_1, \dots, \tilde{a}_{\ell}$ is also nice.

Step 2 (Reduction of ℓ via vdC). Applying equation (15) for the average over n , we get that it suffices to obtain an upper bound for the following averages:

$$\mathbb{E}_{(\underline{h}, h_{k+1}) \in [L_N]^{k+1}} \mathbb{E}_{n \in [N \text{ ad}]} \left| \int \prod_{i=1}^{\ell} T^{[\tilde{a}_i(\underline{h}, n+h_{k+1})]} f_{N, \underline{h}, i} \prod_{i=1}^{\ell} T^{[\tilde{a}_i(\underline{h}, n)]} \bar{f}_{N, \underline{h}, i} d\mu \right|.$$

We compose with $T^{-[\tilde{a}_1(\underline{h}, n)]}$, and for $i = 1, \dots, \ell$, we replace the differences $[\tilde{a}_i(\underline{h}, n+h_{k+1})] - [\tilde{a}_1(\underline{h}, n)]$, $[\tilde{a}_i(\underline{h}, n)] - [\tilde{a}_1(\underline{h}, n)]$ with $[\tilde{a}_i(\underline{h}, n+h_{k+1}) - \tilde{a}_1(\underline{h}, n)]$, $[\tilde{a}_i(\underline{h}, n) - \tilde{a}_1(\underline{h}, n)]$, respectively. To do so, we have to introduce some error sequences that take values on a finite subset of \mathbb{N} . We use Lemma 3.6 to treat the errors that arise, and we are left with upper-bounding averages of the form

$$\mathbb{E}_{(\underline{h}, h_{k+1}) \in [L_N]^{k+1}} \mathbb{E}_{n \in [N \text{ ad}]} \left| \int \prod_{i=1}^{\ell} T^{[\tilde{a}_i(\underline{h}, n+h_{k+1}) - \tilde{a}_1(\underline{h}, n)] + \epsilon_{i,N}} f_{N, \underline{h}, i} \prod_{i=1}^{\ell} T^{[\tilde{a}_i(\underline{h}, n) - \tilde{a}_1(\underline{h}, n)] + \epsilon'_{i,N}} \bar{f}_{N, \underline{h}, i} d\mu \right|,$$

where $\epsilon_{i,N}, \epsilon'_{i,N}, i = 1, \dots, \ell$ take finitely many values for $N \in \mathbb{N}$. Note that the fractional degree of q_1, \dots, q_{ℓ} is strictly smaller than 1. It follows from this and the mean value theorem that

$$\lim_{N \rightarrow \infty} \max_{(\underline{h}, h_k) \in [L_N]^{k+1}, i \in \{1, \dots, \ell\}} |q_i(\underline{h}, t+h_{k+1}) - q_i(\underline{h}, t)| = 0. \tag{29}$$

Using equations (28) and (29) and then Lemma 3.6, we get that it suffices to get an upper bound for the averages

$$\mathbb{E}_{(\underline{h}, h_{k+1}) \in [L_N]^{k+1}} \mathbb{E}_{n \in [N \text{ ad}]} c_{N, \underline{h}, h_{k+1}}(n) \cdot \int (T^{[p_{1,r}(\underline{h})h_{k+1}] + \epsilon_{1,N}} f_{N, \underline{h}, 1} \cdot \bar{f}_{N, \underline{h}, 1}) \prod_{i=2}^{\ell} T^{[b_i(\underline{h}, n)] + \epsilon_{i,N}} \bar{f}_{N, \underline{h}, h_{k+1}, i} d\mu,$$

where $\epsilon_{1,N}, \dots, \epsilon_{\ell,N}$ take finitely many values for $N \in \mathbb{N}$,

$$b_i(\underline{h}, t) := (p_{i,r} - p_{1,r})(\underline{h})t + (q_i - q_1)(\underline{h}, t), \quad i = 2, \dots, \ell,$$

$\tilde{f}_{N, \underline{h}, h_{k+1}, i} \in L^\infty(\mu), i = 2, \dots, \ell$ are 1-bounded functions and $(c_{N, \underline{h}, h_{k+1}}(n))$ is a 1-bounded sequence. Without loss of generality, we can assume that b_ℓ has maximal fractional degree within the collection b_2, \dots, b_ℓ (note that some of the polynomials $p_{i,r} - p_{1,r}$ may vanish). We compose with $T^{-[b_\ell(\underline{h}, n)]}$ and apply Lemma 3.6 to treat finite-valued error sequences that we get when we replace differences of integer parts with the integer part of the corresponding differences. After using the Cauchy-Schwarz inequality, we deduce that it suffices to get an upper bound for the following averages:

$$\mathbb{E}_{h_{k+1} \in [L_N]} \left(\mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in [N \text{ ad}]} c_{N, \underline{h}, h_{k+1}}(n) \prod_{i=1}^{\ell-1} T^{[\tilde{b}_i(\underline{h}, n)] + \epsilon'_{i,N}} \tilde{f}_{N, \underline{h}, h_{k+1}, i} \right\|_{L^2(\mu)} \right), \tag{30}$$

where $\epsilon'_{1,N}, \dots, \epsilon'_{\ell-1,N}$ take finitely many values for $N \in \mathbb{N}$,

$$\tilde{b}_i(\underline{h}, t) := (p_{i,r} - p_{\ell,r})(\underline{h})t + (q_i - q_\ell)(\underline{h}, t), \quad i = 1, \dots, \ell - 1,$$

and

$$\tilde{f}_{N,\underline{h},h_{k+1},1} := T^{[p_{1,r}(\underline{h})h_{k+1}] + \epsilon_{1,N}} f_{N,\underline{h},1} \cdot \overline{f}_{N,\underline{h},1}, \tag{31}$$

where $\epsilon_{1,N}$ takes finitely many values for $N \in \mathbb{N}$.

Note that our assumptions imply that $\tilde{b}_1, \dots, \tilde{b}_{\ell-1}$, thought of as a collection of polynomials with real exponents and $(k + 1)$ -parameters, is nice.

Step 3 (Applying the induction hypothesis). Using the induction hypothesis for the expression in equation (30) that is inside the parentheses, and the fact that $\tilde{b}_1, \dots, \tilde{b}_{\ell-1}$ do not depend on the parameter h_{k+1} , we get that there exist $l, r \in \mathbb{N}$ and nonconstant polynomials $P_1, \dots, P_r \in \mathbb{R}[t_1, \dots, t_{k+l}]$ with pairwise nonconstant differences, such that the averages in equation (30) are bounded by an $o_N(1)$ term plus a constant C_{k,a_1,\dots,a_ℓ} (note that $\tilde{b}_1, \dots, \tilde{b}_{\ell-1}$ are determined by a_1, \dots, a_ℓ) times the expression

$$\mathbb{E}_{(\underline{h}_1, h_{k+1}) \in [L_N]^{k+1}, \underline{h}_2 \in [L_N]^l} \left| \int \prod_{i=0}^r T^{[P_i(\underline{h}_1, \underline{h}_2)] + \epsilon'_{i,N}} F_{N,\underline{h}_1, h_{k+1}, i} d\mu \right|,$$

where $P_0 := 0$, $F_{N,\underline{h},h_{k+1},i} \in \{\tilde{f}_{N,\underline{h},h_{k+1},1}, \overline{f}_{N,\underline{h},h_{k+1},1}\}$ for $i = 0, \dots, r$, $\underline{h}_1 \in [L_N]^k$, $h_{k+1} \in [L_N]$, $N \in \mathbb{N}$ and $\epsilon'_{0,N}, \dots, \epsilon'_{r,N}$ take finitely many values for $N \in \mathbb{N}$.

Using equation (31) and Lemma 3.6, we can bound this expression by a constant C_r times the following average:

$$\mathbb{E}_{(\underline{h}_1, h_{k+1}) \in [L_N]^{k+1}, \underline{h}_2 \in [L_N]^l} \left| \int f_{N,\underline{h}_1,1} \cdot T^{[p_{1,r}(\underline{h}_1)h_{k+1}] + \epsilon'_{0,N}} \overline{f}_{N,\underline{h}_1,1} \cdot \prod_{i=1}^r (T^{[P_i(\underline{h}_1, \underline{h}_2) + p_{1,r}(\underline{h}_1)h_{k+1}] + \epsilon'_{i,N}} G_{N,\underline{h}_1, h_{k+1}, i} \cdot T^{[P_i(\underline{h}_1, \underline{h}_2)] + \epsilon'_{r+i,N}} G_{N,\underline{h}_1, h_{k+1}, r+i}) d\mu \right|,$$

where for $i = 1, \dots, 2r$, we have $G_{N,\underline{h}_1, h_{k+1}, i} \in \{f_{N,\underline{h}_1,1}, \overline{f}_{N,\underline{h}_1,1}\}$, $\underline{h}_1 \in [L_N]^k$, $h_{k+1} \in [L_N]$, $N \in \mathbb{N}$, and $\epsilon'_{i,N}, i = 0, \dots, 2r$ take finitely many values for $N \in \mathbb{N}$. Since the polynomial $p_{1,r}$ is nonzero and the polynomials P_1, \dots, P_r with $k+l$ variables are nonconstant and have nonconstant pairwise differences, the same holds for the $2r+1$ polynomials with $k+l+1$ variables $p_{1,r}(\underline{h}_1)h_{k+1}, P_i(\underline{h}_1, \underline{h}_2) + p_{1,r}(\underline{h}_1)h_{k+1}, P_i(\underline{h}_1, \underline{h}_2), i = 1, \dots, r$. This completes the proof. \square

4.4. Averages with polynomial iterates

Lemma 4.1 and Lemma 4.2 show that in the case of iterates with sublinear growth, to get good seminorm estimates for the averages in Theorem 3.1, it suffices to study averages with iterates given by polynomials in $\mathbb{R}[t_1, \dots, t_k]$ for some $k \in \mathbb{N}$. This is the context of the next result.

Lemma 4.3. *Let $k, r \in \mathbb{N}$ and $P_1, \dots, P_r \in \mathbb{R}[t_1, \dots, t_k]$ be nonconstant polynomials with pairwise nonconstant differences. Then there exists $s \in \mathbb{N}$ such that the following holds: If (X, μ, T) is an ergodic system and $f_1, \dots, f_r \in L^\infty(\mu)$ are such that $\|f_i\|_s = 0$ for some $i \in \{1, \dots, r\}$, then for every 1-bounded sequence $(c_N(\underline{h}))$, we have*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\underline{h} \in [N]^k} c_N(\underline{h}) \cdot \prod_{i=1}^r T^{[P_i(\underline{h})]} f_i = 0$$

in $L^2(\mu)$.

Proof. The argument is similar to the one used to prove [23, Theorem 1], where the case of polynomials with integer coefficients and $c_N(\underline{h}) := 1$ is covered, so we only sketch the points in the argument where one has to deviate slightly because of minor technical complications. The proof proceeds by

induction on a certain vector, called the weight, that is associated to each polynomial family P_1, \dots, P_r in $\mathbb{R}[t_1, \dots, t_k]$.

The inductive step is carried out by using a variant of Lemma 3.5 in the form used in equation (16) that concerns averages over $[N]^k$ (see [23, Lemma 4] for the precise statement). The argument applies verbatim in our case; the only change is that we need at various instances to replace the differences of the integer part of polynomials with the integer part of their differences; we do this with the help of Lemma 3.6, and the use of the constants $(c_N(\underline{h}))$ facilitates this task.

The base case of the induction is the case where all the polynomials are linear with respect to all variables involved. This case is covered using another induction, this time on the number r of linear functions. The inductive step is proved using [23, Lemma 4]. The only difference in our case, versus the argument used in [23, Proposition 5], appears in the proof of the estimate

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{\underline{h} \in [N]^k} \|g \cdot T^{[L(\underline{h})]} f\|_s^{2^s} \leq C_L \|f\|_{s+1}^{2^{s+1}} \tag{32}$$

for some $C_L > 0$, where $f, g \in L^\infty(\mu)$ and $L(\underline{h}) = \sum_{j=1}^k \alpha_j h_j$ for some $k \in \mathbb{N}$ and $\alpha_1, \dots, \alpha_k \in \mathbb{R}$. To obtain this bound, we first use Lemma 3.6 to show that it suffices to replace $[\sum_{j=1}^k \alpha_j h_j]$ with $\sum_{j=1}^k [\alpha_j h_j]$, and we remark that the set

$$\{([\alpha_1 h_1], \dots, [\alpha_k h_k]) : (h_1, \dots, h_k) \in \mathbb{N}^k\}$$

has bounded multiplicity and positive density (as a subset of \mathbb{N}^k). It follows that there exists $C_L > 0$ such that

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{\underline{h} \in [N]^k} \|g \cdot T^{\sum_{j=1}^k [\alpha_j h_j]} f\|_s^{2^s} \leq C_L \limsup_{N \rightarrow \infty} \mathbb{E}_{\underline{h} \in [N]^k} \|g \cdot T^{\sum_{j=1}^k h_j} f\|_s^{2^s}.$$

By [23, Lemma 8], the last expression is bounded by a constant multiple of $\|f\|_{s+1}^{2^{s+1}}$. Combining the above, we get that equation (32) holds. Finally, the base case of the induction (of the linear case) is when $r = 1$ and $P_1 = L$ is linear. To cover this case, we again use [23, Lemma 4] and reduce matters to the task of obtaining an upper bound for the expression

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{\underline{h} \in [N]^k} \left| \int \bar{f} \cdot T^{L(\underline{h})} f \, d\mu \right|.$$

By the $s = 1$ case of equation (32) (recall that $\|f\|_1 = |\int f \, d\mu|$), we get an upper bound by $C_L \|f\|_2^2$ for some $C_L > 0$. This completes the proof. □

4.5. Proof of Theorem 3.1 in the sublinear case

We are now ready to combine the ingredients of the previous subsections to complete the goal of this section, which is to prove the following result:

Proposition 4.4. *Theorem 3.1 holds in the case where all a_1, \dots, a_ℓ have fractional degree smaller than one.*

Proof. Combining Lemma 4.1 and Lemma 4.2 (for $f_{N, \underline{h}, 1} := f_1$, $N \in \mathbb{N}$, $\underline{h} \in \mathbb{N}^k$), we get that there exist $k, r \in \mathbb{N}$ and nonconstant polynomials $P_1, \dots, P_r \in \mathbb{R}[t_1, \dots, t_k]$, with pairwise nonconstant differences, such that the averages in equation (8) are bounded by an $o_N(1)$ term plus a constant multiple of

$$\mathbb{E}_{\underline{h} \in [L_N]^k} \left| \int \prod_{i=0}^r T^{[P_i(\underline{h})] + \epsilon_{i,N}} F_{i, \underline{h}} \, d\mu \right|,$$

where $P_0 := 0, F_{0,\underline{h}}, \dots, F_{r,\underline{h}} \in \{f_1, \bar{f}_1\}, \underline{h} \in \mathbb{N}^k$ and the sequences $\epsilon_{0,N}, \dots, \epsilon_{r,N}$ take values on a finite subset S of \mathbb{Z} for $N \in \mathbb{N}$. Since the limsup as $N \rightarrow \infty$ of the previous average is bounded by

$$\sum_{\epsilon_0, \dots, \epsilon_r \in S, F_0, \dots, F_r \in \{f_1, \bar{f}_1\}} \limsup_{N \rightarrow \infty} \left(\mathbb{E}_{\underline{h} \in [L_N]^k} \left| \int \prod_{i=0}^r T^{[P_i(\underline{h})] + \epsilon_i} F_i d\mu \right| \right),$$

it suffices to show that for all fixed $\epsilon_0, \dots, \epsilon_r \in \mathbb{Z}$ and $F_0, \dots, F_r \in \{f_1, \bar{f}_1\}$, we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\underline{h} \in [L_N]^k} \left| \int \prod_{i=0}^r T^{[P_i(\underline{h})] + \epsilon_i} F_i d\mu \right| = 0.$$

The last average is equal to

$$\mathbb{E}_{\underline{h} \in [L_N]^k} c_N(\underline{h}) \cdot \int \prod_{i=0}^r T^{[P_i(\underline{h})] + \epsilon_i} F_i d\mu$$

for some 1-bounded sequence $(c_N(\underline{h}))$. The result now follows from Lemma 4.3. □

5. Seminorm estimates – induction step

The goal of this section is to finish the proof of Theorem 3.1 using a PET-induction argument. The basis of the induction was covered in the previous section, and the induction step will be carried out in this section.

5.1. An example

To better illustrate our method, we first explain the details in a simple case. We take $\ell = 2$ and $a_1(t) := t^{1.5}, a_2(t) = t^{1.5} + t^{1.1}, t \in \mathbb{R}_+$. Then $\{a_1, a_2\}$ is a nice family, and our aim is to show that if $\|f_1\|_s = 0$ for some $s \in \mathbb{N}$, then

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} w_N(n) \cdot T^{[n^{1.5}]} f_1 \cdot T^{[n^{1.5} + n^{1.1}]} f_2 = 0,$$

where $w_N(n) = \Lambda'(n) \cdot c_N(n)$ for some 1-bounded sequence $(c_N(n))$.

We start by using equation (15), compose with $T^{-[n^{1.5} + n^{1.1}]}$, use Lemma 3.6 to dispose the error sequence that arises when we replace the difference of integer parts with the integer part of the difference and use the Cauchy-Schwarz inequality. We deduce that it suffices to prove convergence to zero of the averages

$$\mathbb{E}_{h_1 \in [L_N]} \left\| \mathbb{E}_{n \in [N]} w_{N,h_1}(n) \cdot T^{[(n+h_1)^{1.5} - n^{1.5} - n^{1.1}]} f_1 \cdot T^{[(n+h_1)^{1.5} + (n+h_1)^{1.1} - n^{1.5} - n^{1.1}]} f_2 \cdot T^{[-n^{1.1}]} \bar{f}_1 \right\|_{L^2(\mu)},$$

where $w_{N,h_1}(n) := (\Delta_{h_1} \Lambda')(n) \cdot c_{N,h_1}(n)$ for some 1-bounded sequence $(c_{N,h_1}(n))$. Using the mean value theorem and Lemma 3.6, we get that for the range of h_1, n we are working with, we can replace $(n + h_1)^{1.5} - n^{1.5}$ with $1.5 h_1 n^{0.5}$ and $(n + h_1)^{1.1} - n^{1.1}$ with $1.1 h_1 n^{0.1}$, which for notational simplicity we replace with $h_1 n^{0.5}$ and $h_1 n^{0.1}$, respectively. We thus arrive at the problem of proving convergence to zero of the averages

$$\mathbb{E}_{h_1 \in [L_N]} \left\| \mathbb{E}_{n \in [N]} w_{N,h_1}(n) \cdot T^{[-n^{1.1} + h_1 n^{0.5}]} f_1 \cdot T^{[h_1 n^{0.5} + h_1 n^{0.1}]} f_2 \cdot T^{[-n^{1.1}]} \bar{f}_1 \right\|_{L^2(\mu)}.$$

Performing one more time the previous operation (we compose with $T^{-[h_1n^{0.5}+h_1n^{0.1}]}$ after applying equation (15)), we arrive in a similar fashion at the following averages:

$$\mathbb{E}_{h_1, h_2 \in [L_N]} \left\| \mathbb{E}_{n \in [N]} w_{N, h_1, h_2}(n) \cdot T^{-[n^{1.1}-h_2n^{0.1}]} f_1 \cdot T^{-[n^{1.1}-h_1n^{0.5}-(h_1+h_2)n^{0.1}]} \bar{f}_1 \cdot T^{-[n^{1.1}-h_1n^{0.1}]} \bar{f}_1 \cdot T^{-[n^{1.1}-h_1n^{0.5}-h_1n^{0.1}]} f_1 \right\|_{L^2(\mu)},$$

where $w_{N, h_1, h_2}(n) := (\Delta_{h_1, h_2} \Lambda')(n) \cdot c_{N, h_1, h_2}(n)$ for some 1-bounded sequence $(c_{N, h_1, h_2}(n))$. After one more iteration of the previous operation (this time we compose with the transformation $T^{[n^{1.1}+h_1n^{0.5}+h_1n^{0.1}]}$ after applying equation (15)), we arrive at the averages

$$\mathbb{E}_{h_1, h_2, h_3 \in [L_N]} \left\| \mathbb{E}_{n \in [N]} w_{N, h_1, h_2, h_3}(n) \cdot T^{[(h_1-h_2-h_3)n^{0.1}+h_1n^{0.5}]} f_1 \cdot T^{[-(h_2+h_3)n^{0.1}]} \bar{f}_1 \cdot T^{-[h_3n^{0.1}+h_1n^{0.5}]} \bar{f}_1 \cdot T^{-[h_3n^{0.1}]} f_1 \cdot T^{[(h_1-h_2)n^{0.1}+h_1n^{0.5}]} \bar{f}_1 \cdot T^{-[h_2n^{0.1}]} f_1 \cdot T^{[h_1n^{0.5}]} f_1 \right\|_{L^2(\mu)},$$

where $w_{N, h_1, h_2, h_3}(n) := (\Delta_{h_1, h_2, h_3} \Lambda')(n) \cdot c_{N, h_1, h_2, h_3}(n)$ for some 1-bounded sequence $(c_{N, h_1, h_2, h_3}(n))$. We have now reduced to the case of fractional polynomials with 3-parameters and fractional degree smaller than 1. This case was dealt in the previous section, where we showed in Proposition 4.4 that there exists $s \in \mathbb{N}$ such that if $\|f_1\|_s = 0$, then the last averages converge to zero as $N \rightarrow \infty$.

5.2. The van der Corput operation and reduction of type

In this subsection, we define the type of a family of polynomials with real exponents and finitely many parameters and the van der Corput operation that reduces the type.

Definition. We say that two polynomials a, b with real exponents and finitely many parameters are *equivalent*, and write $a \cong b$, if the (integral) degree of $a - b$ is strictly smaller than the degree of a and b .⁹

We define the *type* of a family a_1, \dots, a_ℓ of polynomials with real exponents and finitely many parameters to be the vector that consists of the maximal degree d of the family (in the first coordinate) and the number of nonequivalent classes of degree $d, d - 1, \dots, 0$ in the other coordinates (we ignore polynomials that are identically 0).

We order the set of all possible types lexicographically, meaning $(d, k_d, \dots, k_0) > (d', k'_d, \dots, k'_0)$ if and only if in the first instance where the two vectors disagree, the coordinate of the first vector is larger than the coordinate of the second vector.

We caution the reader that $t^{2.5} \not\cong t^{2.5} + t^{2.1}$ (but $t^{2.5} \cong t^{2.5} + t^{1.1}$). Also, if $a_1(h, t) = ht^{2.5} + h^2t^{2.1}$, $a_2(h, t) = ht^{2.5}$, $a_3(h, t) = ht^{2.5} + h^2t^{2.1} + ht^{1.5}$, $a_4(h, t) = t^{0.5}$, then $a_1 \not\cong a_2$, $a_2 \not\cong a_3$, $a_1 \cong a_3$ and the family a_1, a_2, a_3, a_4 has type $(2, 2, 0, 1)$.

Recall that $L_N = [e^{\sqrt{\log N}}]$, $N \in \mathbb{N}$. We introduce a class of sequences that often occur as errors that can be eliminated using Lemma 3.6.

Definition. We say that $e : \mathbb{N}^k \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is *negligible* if

$$\lim_{N \rightarrow \infty} \max_{\underline{h} \in [L_N]^k, t \in [\sqrt{N}, N]} |e(\underline{h}, t)| = 0.$$

If $a(t)$ is a fractional polynomial, then $a(t + c)$ is also a fractional polynomial modulo negligible terms. This is the context of the next lemma, which is proved in a more general form that is better suited for our purposes.

⁹We do not choose to identify functions with the same fractional degree because if we did so, then the vdC operation that will be described shortly would not necessarily lead to families with smaller type (see the example given after the relevant definition).

Lemma 5.1. *Let $a(\underline{h}, t)$ be a polynomial with real exponents and k -parameters and degree d . Then modulo negligible terms, $a(\underline{h}, t + h_{k+1})$ is a polynomial with real exponents and $(k + 1)$ -parameters. In fact, we have*

$$a(\underline{h}, t + h_{k+1}) = \tilde{a}(\underline{h}, h_{k+1}, t) + e(\underline{h}, h_{k+1}, t), \tag{33}$$

where (below $a^{(j)}$ denotes the j th derivative of a with respect to the variable t)

$$\tilde{a}(\underline{h}, h_{k+1}, t) := \sum_{j=0}^d \frac{h_{k+1}^j}{j!} a^{(j)}(\underline{h}, t) \tag{34}$$

and $e: \mathbb{N}^{k+1} \times \mathbb{R} \rightarrow \mathbb{R}$ is negligible.

Proof. Using the Taylor expansion of $a(\underline{h}, t)$, we get that equation (33) holds with

$$e(\underline{h}, h_{k+1}, t) := \frac{h_{k+1}^{d+1}}{(d + 1)!} a^{(d+1)}(\underline{h}, \xi_{\underline{h}, h_{k+1}, t})$$

for some $\xi_{\underline{h}, h_{k+1}, t} \in [t, t + h_{k+1}]$. Since the fractional degree of a is $d + c$ for some $c \in (0, 1)$, we have

$$\max_{(\underline{h}, h_{k+1}) \in [L_N]^{k+1}, t \in [\sqrt{N}, N]} |e(\underline{h}, h_{k+1}, t)| < \frac{L_N^A}{N^{\frac{1-c}{2}}}$$

for some $A > 0$ that depends on d and the maximum degree of the coefficient polynomials of $a(\underline{h}, t)$. Since $L_N < N^\varepsilon$ for every $\varepsilon > 0$, it follows that

$$\lim_{N \rightarrow \infty} \max_{(\underline{h}, h_{k+1}) \in [L_N]^{k+1}, t \in [\sqrt{N}, N]} |e(\underline{h}, h_{k+1}, t)| = 0,$$

completing the proof. □

For example, if $a(h, t) = ht^a$ for some $a \in (2, 3)$, then modulo negligible terms (in the sense defined above), we have that $\tilde{a}(h, t + h_1)$ is equal to $ht^a + ah_1ht^{a-1} + \frac{a(a-1)}{2}h_1^2ht^{a-2}$.

Next we define an operation that we later show preserves nice families of polynomials and reduces their type.

Definition. Let $\mathcal{A} = \{a_1, \dots, a_\ell\}$ be a family of polynomials with real exponents and k -parameters and $a \in \mathcal{A}$. We define a new family of polynomials with real exponents and $(k + 1)$ -parameters $\text{vdC}(\mathcal{A}, a)$ as follows: We start with the family

$$\{\tilde{a}_i(\underline{h}, h_{k+1}, t) - a(\underline{h}, t), a_i(\underline{h}, t) - a(\underline{h}, t), i = 1, \dots, \ell\},$$

where for $i = 1, \dots, \ell$, the polynomial with real exponents and $(k + 1)$ -parameters \tilde{a}_i is as in equation (34) (so it is equal to $a_i(\underline{h}, t + h_{k+1})$ modulo negligible terms), and we remove all functions that are constant in the variable t .

Suppose for example that we start with the nice family

$$\mathcal{A} = \{t^{1.5}, t^{1.5} + t^{1.1}, t^{1.5} + t^{1.2}\}.$$

The type of this family is $(1, 3, 0)$, and the family $\text{vdC}(\mathcal{A}, t^{1.5} + t^{1.2})$ is

$$\{-t^{1.2} + 1.5ht^{0.5}, -t^{1.2} + t^{1.1} + 1.5ht^{0.5}, 1.5ht^{0.5} + 1.2ht^{0.2}, -t^{1.2}, -t^{1.2} + t^{1.1}\}$$

(note that the first and fourth functions can be identified, and the same holds for the second and the fifth), which is also nice and has smaller type, namely $(1, 2, 1)$. We remark that if we had chosen to identify functions that have the same fractional degree, then the original family would have type $(1, 1, 0)$ and the family $\text{vdC}(\mathcal{A}, t^{1.5} + t^{1.2})$ would have larger type, namely $(1, 2, 1)$.

Lemma 5.2. *Let $\mathcal{A} = \{a_1, \dots, a_\ell\}$ be a nice family of polynomials with real exponents and k -parameters such that $\text{f-deg}(a_1) > 1$. Then there exists $a \in \mathcal{A}$ such that the family $\text{vdC}(\mathcal{A}, a)$, ordered so that the first function is $\tilde{a}_1 - a$, is nice and has smaller type. Furthermore, if \mathcal{A} consists of fractional polynomials with k -parameters, then $\text{vdC}(\mathcal{A}, a)$ consists of fractional polynomials with $(k + 1)$ -parameters.*

Proof. We first remark that if \mathcal{A} consists of fractional polynomials with k -parameters and a is any fractional polynomial with k -parameters, then equation (34) implies that $\text{vdC}(\mathcal{A}, a)$ consists of fractional polynomials with $(k + 1)$ -parameters.

For $i = 1, \dots, \ell$, let \tilde{a}_i be the polynomial with real exponents and $(k + 1)$ -parameters given by equation (34). We choose $a \in \mathcal{A}$ as follows:

1. If a_1, \dots, a_ℓ do not have the same fractional degree, we let a_{i_0} be a function in the family $\{a_2, \dots, a_\ell\}$ that has minimal (positive) fractional degree and set $a = a_{i_0}$.
2. If a_1, \dots, a_ℓ have the same fractional degree, we let $i_0 \in \{1, \dots, \ell\}$ be so that $\tilde{a}_1 - a_{i_0}$ has maximal degree within the family $\tilde{a}_1 - a_1, \dots, \tilde{a}_1 - a_\ell$ and set $a = a_{i_0}$.

Claim 1. *The family $\text{vdC}(\mathcal{A}, a)$ is nice.*

By construction, all functions in $\text{vdC}(\mathcal{A}, a)$ are nonconstant (we have removed constant functions). We first show that independently of the choice of a , the difference of $\tilde{a}_1 - a$ with a function in $\text{vdC}(\mathcal{A}, a)$ is always nonconstant (in the variable t); in the process, we also show that $\text{f-deg}(\tilde{a}_1 - a) > 0$. Suppose that such a difference has the form $\tilde{a}_1 - a_i$ for some $i \in \{1, \dots, \ell\}$. It follows from Lemma 5.1 that \tilde{a}_1 contains the term $h_{k+1}a'_i(t)$, which depends nontrivially on the parameter h_{k+1} (note also that a_1, \dots, a_ℓ do not depend on this parameter). It follows from this and our assumption $\text{f-deg}(a_1) > 1$ that

$$\text{f-deg}(\tilde{a}_1 - a_i) \geq \text{f-deg}(a'_i) = \text{f-deg}(a_1) - 1 > 0, \quad i = 1, \dots, \ell.$$

It remains to cover the case where the difference of $\tilde{a}_1 - a$ with a function in $\text{vdC}(\mathcal{A}, a)$ has the form $\tilde{a}_1 - \tilde{a}_i$ for some $i \in \{2, \dots, \ell\}$. Then using Lemma 5.1 and our assumption that \mathcal{A} is nice, we get

$$\text{f-deg}(\tilde{a}_1 - \tilde{a}_i) \geq \text{f-deg}(a_1 - a_i) > 0, \quad i = 2, \dots, \ell.$$

Next we show that $\tilde{a}_1 - a$ has maximal fractional degree within the family $\text{vdC}(\mathcal{A}, a)$. Suppose first that we are in Case (i). Since $\text{f-deg}(a_{i_0}) < \text{f-deg}(a_1)$, we have that $\tilde{a}_1 - a_{i_0}$ has the same fractional degree as a_1 , which by assumption has maximal fractional degree within the family $\{a_1, \dots, a_\ell\}$. We deduce that $\tilde{a}_1 - a_{i_0}$ has maximal fractional degree within the family $\text{vdC}(\mathcal{A}, a)$. Suppose now that we are in Case (ii) and let $i \in \{1, \dots, \ell\}$. Since $a_i - a_{i_0} = (a_i - \tilde{a}_1) + (\tilde{a}_1 - a_{i_0})$ and by the choice of i_0 we have $\text{f-deg}(\tilde{a}_1 - a_{i_0}) \geq \text{f-deg}(\tilde{a}_1 - a_i)$, we deduce that

$$\text{f-deg}(\tilde{a}_1 - a_{i_0}) \geq \text{f-deg}(a_i - a_{i_0}), \quad i = 1, \dots, \ell. \tag{35}$$

Moreover, note that $\tilde{a}_i - a_{i_0} = (\tilde{a}_i - a_i) + (a_i - a_{i_0})$ and

$$\text{f-deg}(\tilde{a}_i - a_{i_0}) \geq \text{f-deg}(\tilde{a}_i - a_i) = \text{f-deg}(a_1) - 1 \geq \text{f-deg}(a_i) - 1 = \text{f-deg}(\tilde{a}_i - a_i), \tag{36}$$

where the two identities follow from Lemma 5.1, and the first estimate follows from the choice of i_0 and the second since the family \mathcal{A} is nice. We deduce from equations (35) and (36) that

$$\text{f-deg}(\tilde{a}_1 - a_{i_0}) \geq \text{f-deg}(\tilde{a}_i - a_{i_0}), \quad i = 1, \dots, \ell. \tag{37}$$

Combining equations (35) and (37), we get that $\tilde{a}_1 - a_{i_0}$ has maximal fractional degree within the family $\text{vdC}(\mathcal{A}, a)$.

Claim 2. *The family $\text{vdC}(\mathcal{A}, a)$ has smaller type.*

Using Lemma 5.1 and the definition of the degree, it is easy to verify that if for some $i \in \{1, \dots, \ell\}$, we have $a_i \not\cong a_{i_0}$, then $\text{deg}(a_i - a_{i_0}) = \text{deg}(\tilde{a}_i - a_{i_0}) = \text{deg}(a_i)$ and $a_i - a_{i_0} \cong \tilde{a}_i - a_{i_0}$, while if $a_i \cong a_{i_0}$, then $\text{deg}(a_i - a_{i_0}) < \text{deg}(a_i)$ and $\text{deg}(\tilde{a}_i - a_{i_0}) < \text{deg}(a_i)$. Using these facts, we easily get the following:

If we are in Case (i), we have that the type of \mathcal{A} has the form $(d, k_d, \dots, k_l, 0, \dots, 0)$, where $l = \text{deg}(a_{i_0})$, $k_l \geq 1$, and $d \geq 1$. Then the type of $\text{vdC}(\mathcal{A}, a)$ is $(d, k_d, \dots, k_l - 1)$ if $l = 0$, and $(d, k_d, \dots, k_l - 1, k_{l-1}, \dots, k_0)$ for some $k_0, \dots, k_{l-1} \in \mathbb{Z}_+$ if $l \geq 1$.

If we are in Case (ii), we have that the type of \mathcal{A} has the form $(d, k_d, 0, \dots, 0)$, where $d \geq 1$ and $k_d \geq 1$. Then for every $a \in \mathcal{A}$, the type of $\text{vdC}(\mathcal{A}, a)$ is $(d, k_d - 1, k_{d-1}, \dots, k_0)$ for some $k_0, \dots, k_{d-1} \in \mathbb{Z}_+$.

In both cases, the type of the family $\text{vdC}(\mathcal{A}, a)$ is smaller than the type of the family \mathcal{A} , completing the proof of Claim 2. □

5.3. Proof of Theorem 3.1

We will now use a PET-induction technique to prove Theorem 3.1. The base case of the induction was covered in the previous section, and the inductive step will be proved using equation (15) and Lemma 5.2.

Proof of Theorem 3.1. Our goal is to show that there exists $s \in \mathbb{N}$ such that if $f_{N, \underline{h}, 1} = f_1$, $\underline{h} \in [L_N]^k$, $N \in \mathbb{N}$, $\|f_1\|_s = 0$ and all other functions below are assumed to be 1-bounded, then

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\underline{h} \in [L_N]^k} \left\| \mathbb{E}_{n \in [N]} w_{N, \underline{h}}(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h}, n)]} f_{N, \underline{h}, i} \right\|_{L^2(\mu)} = 0,$$

where $w_{N, \underline{h}}(n) := (\Delta_{\underline{h}} \Lambda')(n) \cdot c_{N, \underline{h}}(n)$, $\underline{h} \in [L_N]^k$, $n \in [N]$, $N \in \mathbb{N}$ and the sequence $(c_{N, \underline{h}}(n))$ is 1-bounded.

We prove this using induction on the type of the nice family of fractional polynomials $\mathcal{A} := \{a_1, \dots, a_\ell\}$ with finitely many parameters. If $\text{f-deg}(a_1) < 1$ (then also $\text{f-deg}(a_j) < 1$ for $j = 2, \dots, \ell$), then the result follows from Proposition 4.4.

Suppose that the family $\mathcal{A} := \{a_1, \dots, a_\ell\}$ has type (d, k_d, \dots, k_0) , where $d \geq 1$, $k_d \geq 1$, $k_{d-1}, \dots, k_0 \in \mathbb{Z}_+$, and the statement holds for all families of fractional polynomials with finitely many parameters and type strictly smaller than (d, k_d, \dots, k_0) . Since $\text{deg}(a_1) \geq 1$ and a_1 is a fractional polynomial, we have that $\text{f-deg}(a_1) > 1$.

By Lemma 5.2, there exists $a \in \mathcal{A}$ such that the family $\text{vdC}(\mathcal{A}, a)$, ordered so that the first function is $\tilde{a}_1 - a$ (where \tilde{a}_1 is as in equation (34)), consists of fractional polynomials with finitely many parameters and satisfies the following:

$$\text{vdC}(\mathcal{A}, a) \text{ is nice and has type strictly smaller than } (d, k_d, \dots, k_0). \tag{38}$$

We use equation (15) for the average $\mathbb{E}_{n \in [N]}$, compose with $T^{-[a(\underline{h}, n)]}$ and then use the Cauchy-Schwarz inequality. We get that it suffices to show the following (recall that $(\Delta_{h u})(n) = u(n+h) \cdot \overline{u(n)}$):

$$\lim_{N \rightarrow \infty} \mathbb{E}_{(\underline{h}, h_{k+1}) \in [L_N]^{k+1}} \left\| \mathbb{E}_{n \in [N]} (\Delta_{h_{k+1}} w_{N, \underline{h}})(n) \cdot \prod_{i=1}^{\ell} T^{[a_i(\underline{h}, n+h_{k+1})] - [a(\underline{h}, n)]} f_{N, \underline{h}, i} \right\|_{L^2(\mu)} = 0.$$

We replace the differences of integer parts on the iterates with the integer part of their differences and also replace $a_i(\underline{h}, n + h_{k+1})$ with $\tilde{a}_i(\underline{h}, h_{k+1}, n)$, where \tilde{a}_j is associated to a_j by equation (34) of Lemma 5.1. To make these substitutions, we introduce some error sequences that take finitely many values; as usual, these sequences can be handled after we apply Lemma 3.6 (which applies without a problem since the values of n that are smaller than \sqrt{N} contribute negligibly in the average). After completing these maneuvers, we see that it suffices to show the following:

$$\lim_{N \rightarrow \infty} \mathbb{E}_{(\underline{h}, h_{k+1}) \in [L_N]^{k+1}} \left\| \mathbb{E}_{n \in [N]} w_{N, \underline{h}, h_{k+1}}(n) \cdot \prod_{i=1}^{2\ell} T^{[b_i(\underline{h}, h_{k+1}, n)] + \epsilon_{i,N}} g_{N, \underline{h}, h_{k+1}, i} \right\|_{L^2(\mu)} = 0,$$

where $\epsilon_{1,N}, \dots, \epsilon_{2\ell,N}$ take finitely many values for $N \in \mathbb{N}$,

$$w_N(\underline{h}, h_{k+1}, n) := (\Delta_{(\underline{h}, h_{k+1})} \Lambda')(n) \cdot c_{N, \underline{h}, h_{k+1}}(n)$$

for some 1-bounded sequence $(c_{N, \underline{h}, h_{k+1}}(n))$ and

$$\begin{aligned} b_i(\underline{h}, h_{k+1}, t) &:= \tilde{a}_i(\underline{h}, t + h_{k+1}) - a(\underline{h}, t), \quad i = 1, \dots, \ell, \\ b_{\ell+i}(\underline{h}, h_{k+1}, t) &:= a_i(\underline{h}, t) - a(\underline{h}, t), \quad i = 1, \dots, \ell \end{aligned}$$

and $g_{N, \underline{h}, h_{k+1}, i}$ are 1-bounded functions in $L^\infty(\mu)$ such that $g_{N, \underline{h}, h_{k+1}, 1} := f_1$ for all $(\underline{h}, h_{k+1}) \in [L_N]^{k+1}$, $N \in \mathbb{N}$. We compose with $T^{-\epsilon_{i,N}}$ inside the $L^2(\mu)$ -norm and set $h_{N, \underline{h}, h_{k+1}, i} := T^{\epsilon_{i,N} - \epsilon_{1,N}} g_{N, \underline{h}, h_{k+1}, i}$, $i = 1, \dots, 2\ell$ (then $h_{N, \underline{h}, h_{k+1}, 1} = f_1$). We get that it suffices to show that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{(\underline{h}, h_{k+1}) \in [L_N]^{k+1}} \left\| \mathbb{E}_{n \in [N]} w_{N, \underline{h}, h_{k+1}}(n) \cdot \prod_{i=1}^{2\ell} T^{[b_i(\underline{h}, h_{k+1}, n)]} h_{N, \underline{h}, h_{k+1}, i} \right\|_{L^2(\mu)} = 0. \quad (39)$$

Finally, we can remove all functions associated with iterates that do not depend on the variable n (note that by Lemma 5.2, the function b_1 is not one of them), and thus we arrive at an average with iterates given by the family $\text{vdc}(\mathcal{A}, a)$, ordered so that the first function is $\tilde{a}_1 - a$. By the choice of a , we have that equation (38) holds. Hence, the induction hypothesis applies for this family and gives that there exists $s \in \mathbb{N}$ such that if $\|f_1\|_s = 0$, then equation (39) holds. This completes the induction step and the proof. \square

Acknowledgement. The author would like to thank the two anonymous referees for their valuable comments.

Funding statement. The author was supported by the Research Grant - ELIDEK HFRI-FM17-1684.

Conflict of Interest. None.

References

- [1] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems* **7** (1987), no. 3, 337–349.
- [2] V. Bergelson, G. Kolesnik, M. Madritsch, Y. Son, R. Tichy. Uniform distribution of prime powers and sets of recurrence and van der Corput sets in \mathbb{Z}^k . *Israel J. Math.* **201** (2014), 729–760.
- [3] V. Bergelson, G. Kolesnik, Y. Son. Uniform distribution of subpolynomial functions along primes and applications. *J. Analyse Math.* **137** (2019), 135–187.
- [4] V. Bergelson, J. Moreira, F. Richter. Multiple ergodic averages along functions from a Hardy field: convergence, recurrence and combinatorial applications. Preprint 2020, [arXiv:2006.03558](https://arxiv.org/abs/2006.03558)
- [5] A. Best, A. Ferré Moragues. Polynomial ergodic averages for certain countable ring actions. To appear in *Discrete Contin. Dyn. Syst.*, [arXiv:2105.04008](https://arxiv.org/abs/2105.04008)
- [6] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial growth. *J. Analyse Math.* **112** (2010), 79–135.
- [7] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bull. Hellenic Math. Soc.* **60** (2016), 41–90.
- [8] N. Frantzikinakis. Joint ergodicity of sequences. Preprint 2021, [arXiv:2102.09967](https://arxiv.org/abs/2102.09967)

- [9] N. Frantzikinakis, B. Host, B. Kra. The polynomial multidimensional Szemerédi Theorem along shifted primes. *Isr. J. Math.* **194** (2013), 331–348.
- [10] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.* **31** (1977), 204–256.
- [11] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, Princeton, 1981.
- [12] P. Gallagher. On the distribution of primes in short intervals. *Mathematika* **23** (1976), 4–9.
- [13] B. Green, T. Tao. The primes contain arbitrarily long arithmetic progressions. *Annals Math.* **167** (2008), 481–547.
- [14] B. Green, T. Tao. Linear equations in primes. *Annals Math.* **171** (2010), 1753–1850.
- [15] H. Halberstam, H. Richert. *Sieve methods*, Academic Press, New York, 1974.
- [16] B. Host, B. Kra. Non-conventional ergodic averages and nilmanifolds. *Ann. of Math.* **161** (2005), 397–488.
- [17] B. Host, B. Kra. *Nilpotent structures in ergodic theory*. Mathematical Surveys and Monographs, vol. **236**. American Mathematical Society, Providence, RI, 2018.
- [18] H. Iwaniec, E. Kowalski. *Analytic number theory*. American Mathematical Society Colloquium Publications, vol. **53**. American Mathematical Society, Providence, RI, 2004.
- [19] D. Karageorgos, A. Koutsogiannis. Integer part independent polynomial averages and applications along primes. *Studia Mathematica* **249** (2019), 233–257.
- [20] A. Koutsogiannis. Closest integer polynomial multiple recurrence along shifted primes. *Ergodic Theory Dynam. Systems* **38** (2018), no. 2, 666–685.
- [21] E. Kowalski. Averages of Euler products, distribution of singular series and the ubiquity of Poisson distribution. *Acta Arith.* **148** (2011), 153–187.
- [22] L. Kuipers, H. Niederreiter. *Uniform distribution of sequences*. Pure and Applied Mathematics. Wiley-Interscience, New York-London-Sydney, 1974.
- [23] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Isr. J. Math.* **146** (2005), 303–316.
- [24] S. Peluse. On the polynomial Szemerédi theorem in finite fields. *Duke Math. J.* **168** (2019), 749–774.
- [25] S. Peluse, S. Prendiville. Quantitative bounds in the non-linear Roth Theorem. To appear in *Int. Math. Res. Not. (IMRN)*, [arXiv:1903.02592](https://arxiv.org/abs/1903.02592)
- [26] F. Richter. Uniform distribution in nilmanifolds along functions from a Hardy field. To appear in *J. Analyse Math.*, [arXiv:2006.02028](https://arxiv.org/abs/2006.02028)
- [27] T. Tao, T. Ziegler. The primes contain arbitrarily long polynomial progressions. *Acta Math.* **201** (2008), 213–305.
- [28] T. Tao, T. Ziegler. Narrow progressions in the primes. *Analytic Number Theory*. Springer International Publishing (2015), 357–379.
- [29] T. Wooley, T. Ziegler. Multiple recurrence and convergence along the primes. *Amer. J. of Math.* **134** (2012), 1705–1732.