JOINT ERGODICITY FOR FUNCTIONS OF POLYNOMIAL GROWTH

SEBASTIÁN DONOSO, ANDREAS KOUTSOGIANNIS AND WENBO SUN

Abstract. We provide necessary and sufficient conditions for joint ergodicity results for systems of commuting measure preserving transformations for an iterated Hardy field function of polynomial growth. Our method builds on and improves recent techniques due to Frantzikinakis and Tsinas, who dealt with multiple ergodic averages along Hardy field functions; it also enhances an approach introduced by the authors and Ferré Moragues to study polynomial iterates.

1. Introduction

A central problem in ergodic theory is the study of multiple ergodic averages of the form

$$\frac{1}{N}\sum_{n=1}^{N} T_1^{a_1(n)}f_1\cdot\ldots\cdot T_d^{a_d(n)}f_d, \tag{1}$$

where $(X,\mathcal{B},\mu,T_1,\ldots,T_d)$ is a system (that is, $(X,\mathcal{B},\mu)$ is a Borel probability space and for all $1\leq i\leq d$, $T_i\colon X\to X$ is a measurable, measure preserving transformation, i.e., $\mu(T_i^{-1}A)=\mu(A)$ for all $A\in\mathcal{B}$), for each $1\leq i\leq d$, $(a_i(n))_n$ is an appropriate integer-valued sequence, and f_i is a bounded function; for a positive integer n, T^n denotes the composition $T\circ\cdots\circ T$ of n copies of T, and $Tf(x):=f(Tx)$, $x\in X$. In particular, we are interested in the $(L^2(\mu))$ norm limiting behaviour, as $N\to\infty$, of (1) for various a_i's, and commuting T_i's (i.e., $T_iT_j=T_jT_i$). Our study deals with commuting and invertible T_i's.

Furstenberg's celebrated result ([20]), i.e., proving Szemerédi's theorem (that each dense subset of natural numbers contains arbitrarily long arithmetic progressions) by studying (1) for $T_i=T$ and $a_i(n)=in$, revolutionized the area, leading to far-reaching extensions of Szemerédi's theorem and various other profound results. For most of the latter results, the only known proofs are the ergodic theoretic ones.

For $d=1$ and $a_1(n)=n$ in (1), von Neumann's mean ergodic theorem characterizes ergodicity:[1] T is ergodic if, and only if, $\frac{1}{N}\sum_{n=1}^{N}T^nf\to\int f\,d\mu$ as $N\to\infty$. For $T=T_i$ *weakly mixing* (*w.m.* for short) (i.e., $T\times T$ is ergodic), and $a_i(n)=in$, Furstenberg showed (again in [20]) that (1) converges to $\prod_{i=1}^{d}\int f_i\,d\mu$. This result was extended in [2] by Bergelson for the case $T_1=\cdots=T_d$ being w.m. and a_i being essentially distinct integer polynomial iterates.[2] Because

2020 *Mathematics Subject Classification.* Primary: 37A05; Secondary: 37A30, 28A99, 60F99.

Key words and phrases. Polynomial functions, Hardy field functions, tempered functions, joint ergodicity.

The first author was supported by ANID/Fondecyt/1200897 and Centro de Modelamiento Matemático (CMM), FB210005, BASAL funds for centers of excellence from ANID-Chile.

[1]T is *ergodic* if $A\in\mathcal{B}, T^{-1}A=A$, implies that $\mu(A)\in\{0,1\}$.

[2]$p\in\mathbb{Q}[x]$ is an *integer polynomial* if $p(\mathbb{Z})\subseteq\mathbb{Z}$; $\{p_1,\ldots,p_d\}$ are *essentially distinct* if p_i, p_i-p_j are non-constant for all $i\neq j$.

of the aforementioned results, we call (for ergodic systems) $\prod_{i=1}^{d} \int f_i \, d\mu$ the "expected limit". So, naturally, one defines the following notion.

Definition. Let $(X, \mathcal{B}, \mu)$ be a Borel probability space, and $(S_1(n))_n, \dots, (S_d(n))_n$ be sequences of measure preserving transformations on X. We say that $(S_1(n))_n, \dots, (S_d(n))_n$ are *jointly ergodic (for μ)*, if for all functions $f_1, \dots, f_d \in L^\infty(\mu)$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} S_1(n) f_1 \cdot \ldots \cdot S_d(n) f_d = \int f_1 \, d\mu \cdot \ldots \cdot \int f_k \, d\mu, \tag{2}$$

where the convergence takes place in $L^2(\mu)$. When $d = 1$, we simply say that the sequence $(S_1(n))_n$ is *ergodic*.[3]

The first characterization of joint ergodicity is due to Berend and Bergelson [1]:

Theorem ([1])**.** *Let $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ be a system with commuting and invertible transformations. Then $(T_1^n)_n, \dots, (T_d^n)_n$ are jointly ergodic for μ if, and only if, both of the following conditions are satisfied:*

(i) *$T_i T_j^{-1}$ is ergodic for μ for all $1 \leq i, j \leq d$, $i \neq j$; and*
(ii) *$T_1 \times \cdots \times T_d$ is ergodic for $\mu^{\otimes d}$.*

This theorem, for $T_i = T^i$, where T is a w.m. transformation, implies Furstenberg's w.m. convergence result. A few years ago, Bergelson, Leibman, and Son showed (in [5]) the following result for generalized linear functions.[4]

Theorem ([5])**.** *Let $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ be a system with commuting and invertible transformations, and $\varphi_1, \dots, \varphi_d$ be generalized linear functions. Then $(T_1^{\varphi_1(n)})_n, \dots, (T_d^{\varphi_d(n)})_n$ are jointly ergodic for μ if, and only if, both of the following conditions are satisfied:*

(i) *$\left(T_i^{\varphi_i(n)} T_j^{-\varphi_j(n)}\right)_n$ is ergodic for μ for all $1 \leq i, j \leq d$, $i \neq j$; and*
(ii) *$\left(T_1^{\varphi_1(n)} \times \cdots \times T_d^{\varphi_d(n)}\right)_n$ is ergodic for $\mu^{\otimes d}$.*

This result, for $T_i = T$ and $\varphi_i(n) = [\alpha_i n]$, where T is w.m. and the α_i's are distinct real numbers, extends Furstenberg's w.m. convergence result.

Seeing the similarities of the last two results, it is reasonable to state the following problem.

Problem 1 (Joint ergodicity problem)**.** Let $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ be a system with commuting and invertible transformations. Find classes of integer-valued sequences $a_1, \dots, a_d$ so that $(T_1^{a_1(n)})_n, \dots, (T_d^{a_d(n)})_n$ are jointly ergodic for μ if, and only if, both of the following conditions are satisfied:

(i) $\left(T_i^{a_i(n)} T_j^{-a_j(n)}\right)_n$ is ergodic for μ for all $1 \leq i, j \leq d$, $i \neq j$; and
(ii) $\left(T_1^{a_1(n)} \times \cdots \times T_d^{a_d(n)}\right)_n$ is ergodic for $\mu^{\otimes d}$.

[3]Here, by saying that we have joint ergodicity, we mean that the limit in (2) exists, and it is the expected one.

[4]A *generalized linear function* $\varphi\colon \mathbb{N} \to \mathbb{Z}$ is a function of the form $\varphi(n) = [an] + e_n$, where $[\cdot]$ is the integer value, or floor, function and e_n is some special, bounded, integer-valued error term.

Answering a question due to Bergelson, we showed in [11] that the answer to Problem 1 is affirmative when $a_1, \dots, a_d$ are equal to the same integer polynomial. This result was later generalized in [10] to the case where all the a_i's are polynomials that can be grouped in a way such that polynomials in different groups have different degrees and each two polynomials in the same grouping are multiples of each other. In two recent papers [17, 18], Frantzikinakis and Kuca showed that the answer to Problem 1 is affirmative for all integer polynomials (modulo mild necessary conditions) $a_1, \dots, a_d$.

In this paper, we extend the study of Problem 1 to functions $a_1, \dots, a_d$ beyond polynomials. In literature, the multiple ergodic averages (1) with a_i's being Hardy field functions (see Section 2 for definition) of polynomial growth[5] has been studied extensively (see for instance [6, 13, 15, 19, 33, 36]). However, to the best of our knowledge, joint ergodicity results for such functions for systems with commuting transformations have not been obtained in the past. By [33, Lemma A.3], every Hardy field function h of polynomial growth can be written as

$$h(x) = s_h(x) + p_h(x) + e_h(x),$$

where s_h is a strongly non-polynomial Hardy field function,[6] p_h is a polynomial and $e_h(x) \to 0$ as $x \to \infty$. Under the additional, natural, assumption $\log x \prec s_h(x)$,[7] we have:

Theorem 1.1. *Let $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ be a system with commuting and invertible transformations, and h be a Hardy field function from $\mathcal{H}$[8] of polynomial growth, with $\log x \prec s_h(x)$. Then $(T_1^{[h(n)]})_n, \dots, (T_d^{[h(n)]})_n$ are jointly ergodic for μ if, and only if, both of the following conditions are satisfied:*

(i) *$((T_i T_j^{-1})^{[h(n)]})_n$ is ergodic for μ for all $1 \leq i, j \leq d$, $i \neq j$; and*
(ii) *$((T_1 \times \cdots \times T_d)^{[h(n)]})_n$ is ergodic for $\mu^{\otimes d}$.*

The following are some examples of functions h that we can deal with:

$$x \log x,^{9} \ \ x^e \log^2 x + x^{17}.$$

We believe that the assumption $\log \prec s_h$ can be lifted to $\log \prec h$, and that, even though we are dealing with a single sequence in Theorem 1.1, something more general holds:

Conjecture 1. *Problem 1 holds for functions $h_1, \dots, h_d$ from the Hardy field $\mathcal{H}$ of polynomial growth with $\log \prec h_i$, $1 \leq i \leq d$.*

Our method in fact yields a result more general than Theorem 1.1 (see Theorem 6.1), where h is the sum of a Hardy field function and a tempered function (see Section 6 for the definition).

[5]A function h has *polynomial growth* if it satisfies $h(x) \ll x^d$ for some $d \in \mathbb{N}$, where, for two functions $a, b : (x_0, \infty) \to \mathbb{R}$, we write $a \ll b$ if there exists a universal constant $C > 0$ so that $|a(x)| \leq C|b(x)|$ for all x.

[6]By this we mean that s_h is a Hardy field function and that, for some non-negative integer i, it satisfies $x^i \prec s_h(x) \prec x^{i+1}$, where for two functions $a, b : (x_0, \infty) \to \mathbb{R}$, we write $a \prec b$ if $|a(x)|/|b(x)| \to 0$ as $x \to \infty$.

[7]This is a natural assumption for convergence results, in the sense that for strongly non-polynomial Hardy field functions it implies equidistribution (see, e.g., [8, 13, 15]).

[8]$\mathcal{H}$ is a Hardy field that will also be defined in Section 2.

[9]This example was proven to be good for convergence very recently in [36]; what is crucial here is the fact that one can approach this function by variable polynomials. Its derivative is, modulo the constant 1, equal to $\log x$, a "bad" for convergence function as it fails to be equidistributed.

1.1. **Strategy of the paper.** In literature, the study of multiple ergodic averages (1) mainly focuses on polynomial, Hardy field and tempered functions (and combinations of them).[10] In fact, these are the only known classes of functions with the following property: if a function f of "degree k" (meaning that $n^k \ll f(n) \prec n^{k+1}$) belongs to the class, then its derivative f' not only belongs to the same class but also is of "degree $k-1$" (meaning that $n^{k-1} \ll f'(n) \prec n^k$). For this very reason, the usual approach to studying the corresponding (1), for such classes of functions, is to reduce its complexity by using variants of the van der Corput lemma combined with variations of the PET induction (see Section 3).

During the past few years there is an interest in the class of variable polynomial sequences. The study of multiple ergodic averages with such iterates is an interesting new topic on its own, with open problems (see, e.g., [16, 30]) and results which have led to variable variations of classical theorems (see [15, 22, 29, 30, 36]). Most importantly, it provides an additional tool that can be used for the study of averages with iterates coming from the suitable classes of functions mentioned above. More specifically, for iterates which are Hardy field functions (or, even more generally, "smooth enough" functions–see Section 6), we can alternatively approach them by variable polynomials of bounded degrees first, and then run the PET induction via the van der Corput's lemma, on the polynomials. This alternative approach can also treat iterates which cannot be treated by the first one, such as $x \log x$.

Our strategy is to study the joint ergodicity problem by incorporating the second approach mentioned above with the machinery created in our previous works [10, 11]. To achieve this, we first approximate the iterates in the multiple ergodic average of interest by variable polynomials. Then, we extend the concepts of PET tuples and vdC operations in [10, 11] for variable polynomials, and use them to bound the stated average by an average of ergodic averages with linear iterates (on "short intervals"). Finally, we deduce the desired result by using [10, Proposition 5.2], which is the central part of [10, 11] where concatenation theorems from [35] were crucially used. To achieve this, we need to make several adaptions to the approaches in these works. To be more precise, we need to extend the concepts of PET tuples and vdC operations to variable polynomials, i.e., of the form $[p_N(n)]$, and generalize certain seminorm estimates for multiple ergodic averages with such iterates.

As we already mentioned, [33, Lemma A.3] implies that every Hardy field function h of polynomial growth can be written as

$$h(x) = s_h(x) + p_h(x) + e_h(x),$$

where s_h is a strongly non-polynomial Hardy field function, p_h is a polynomial and $e_h(x) \to 0$ as $x \to \infty$. We will work with such functions, under the assumption that $\log x \prec s_h(x)$. In Section 2, Proposition 2.1, we show that if a function a can be written as

$$a(N + r) = p_N(r) + e_{N,r},$$

with $e_{N,r} \ll 1$, for every positive integer N and $0 \le r \le L(N)$, where L is an appropriate positive Hardy field function satisfying $1 \prec L(x) \prec x$, and $(p_N)_N$ is a variable polynomial sequence, then, to study the initial multiple averages with iterates $[a(n)]$ along $1 \le n \le N$, it suffices to study the corresponding averages with iterates $[p_N(n)]$ along $0 \le n \le L(N)$. We also prove in Proposition 2.2 a change of variables statement, which shows that if the sequence of variable

[10]A (far from complete) list here is the following: [2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 27, 29, 30, 31, 32, 36, 37].

polynomials p_N has "special" leading coefficients, then we can transform it to one with leading coefficients 1.[11] In Section 3, we extend the results on PET induction from [10, 11] to variable polynomials, which is used to reduce the complexity of multiple ergodic averages with such iterates, and eventually, via Lemma 3.5, reduce the problem to the base case, namely, the linear one. Then, in Section 4, we provide, in Proposition 4.1, a Gowers-Host-Kra-type seminorm upper bound for multiple ergodic averages for certain linear variable polynomials, and in Theorem 4.4 a bound for variable polynomials of leading coefficient 1 (using the inductive scheme of Section 3). In Section 5, we show that, for the functions h we deal with, we can combine all the ingredients proved in the previous sections to deduce our main result, Theorem 1.1, via Theorem 5.1, [11, Corollary 2.5], and [7, Theorem 1.1]. Finally, in Section 6, we explain how our method can be used to study some iterates beyond Hardy field functions.[12]

1.2. **Notation.** We denote with $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{S}^1$ the set of positive integers, non-negative integers, integers, rational numbers, real numbers, complex numbers and complex numbers of modulus 1 respectively. If X is a set, and $d \in \mathbb{N}$, X^d denotes the Cartesian product $X \times \cdots \times X$ of d copies of X. For $M, N \in \mathbb{Z}$ with $M \leq N$, let $[M, N] \coloneqq \{M, M+1, \ldots, N\}$; we also define $[N] \coloneqq \{0, \ldots, N-1\}$. We denote by e_i the i-th standard unit vector, which has 1 as its i-th coordinate and 0 elsewhere.

Let $(a(n))_n$ be a sequence of complex numbers, or a sequence of measurable functions on a probability space $(X, \mathcal{B}, \mu)$, indexed by the set of natural numbers. Throughout this article, we use the following notation for averages:

$\mathbb{E}_{n\in A} a(n) \coloneqq \frac{1}{|A|}\sum_{n\in A} a(n)$, where A is a finite subset of $\mathbb{Z}$;
$\mathbb{E}_{n\in\mathbb{Z}} a(n) \coloneqq \lim_{N\to\infty} \mathbb{E}_{n\in[-N,N]} a(n)$ if the limit exists;
$\overline{\mathbb{E}}_{n\in\mathbb{Z}} a(n) \coloneqq \overline{\lim}_{N\to\infty} \mathbb{E}_{n\in[-N,N]} a(n)$.

We also consider *iterated* averages: Let $(a(h_1, \ldots, h_s))_{h_1,\ldots,h_s\in\mathbb{Z}}$ be a multi-parameter sequence. We let

$$\mathbb{E}_{h_1,\ldots,h_s\in\mathbb{Z}} a(h_1, \ldots, h_s) \coloneqq \mathbb{E}_{h_1\in\mathbb{Z}} \ldots \mathbb{E}_{h_s\in\mathbb{Z}} a(h_1, \ldots, h_s)$$

if the limit exists, we adopt similar conventions for $\overline{\mathbb{E}}_{h_1,\ldots,h_s\in\mathbb{Z}}$ and for averages indexed in $\mathbb{N}$.

We end this section by recalling the notion of a system indexed by $(\mathbb{Z}^d, +)$, $d \in \mathbb{N}$. We say that a tuple $(X, \mathcal{B}, \mu, (T_n)_{n\in\mathbb{Z}^d})$ is a $\mathbb{Z}^d$*-measure preserving system* (or a $\mathbb{Z}^d$*-system*) if $(X, \mathcal{B}, \mu)$ is a probability space and, $T_n \colon X \to X$, $n \in \mathbb{Z}^d$, are measurable, measure preserving transformations on X such that $T_{(0,\ldots,0)} = \mathrm{id}$ and $T_n \circ T_m = T_{n+m}$ for all $n, m \in \mathbb{Z}^d$. Given d commuting and invertible transformations $T_1, \ldots, T_d$, we can naturally define a $\mathbb{Z}^d$-action as follows:

$$T_n = T_1^{n_1} \cdot \ldots \cdot T_d^{n_d}, \;\; n = (n_1, \ldots, n_d) \in \mathbb{Z}^d.$$

[11]Another indication that variable polynomial sequences with leading coefficients 1 form a good class of (variable) polynomials to deal with, is also revealed in [15], where Frantzikinakis showed that, for a single transformation, multiple ergodic averages with such iterates have the nilfactors as characteristic factors.

[12]Our approach can deal with more general iterates, namely functions a of the form

$$a(x) = h(x) + ct(x),$$

where $h \in \mathcal{H}$ is as before, $c \in \mathbb{R}$, and $t \in \mathcal{T}$, is a tempered function with $\max\{\log x, ct(x)\} \prec h(x)$, which satisfy some natural growth rate-related assumptions (see Section 6)

So, we identify the $\mathbb{Z}$-action T_i with T_{e_i} for $1 \leq i \leq d$.[13] By slightly abusing the notation, we also refer to $(X, \mathcal{B}, \mu, T_1, \ldots, T_d)$ as a $\mathbb{Z}^d$-system. Let H be a subgroup of $\mathbb{Z}^d$. We say that H is *ergodic* for a $\mathbb{Z}^d$-system $(X, \mathcal{B}, \mu, (T_n)_{n\in\mathbb{Z}^d})$ if for every $A \in \mathcal{B}$ such that $T_g A = A$ for all $g \in H$, we have that $\mu(A) \in \{0, 1\}$. In particular, we say that $(X, \mathcal{B}, \mu, (T_n)_{n\in\mathbb{Z}^d})$ is *ergodic* if $\mathbb{Z}^d$ is ergodic for the system. Finally, when it is clear, we will write $\|\cdot\|_2$ instead of $\|\cdot\|_{L^2(\mu)}$ and $\|\cdot\|_\infty$ instead of $\|\cdot\|_{L^\infty(\mu)}$.

For a set of parameters A, and a potivite real number α, we write $O_A(\alpha)$ to denote a quantity that is $\leq C_A \cdot \alpha$ for some constant $C_A > 0$ depending only on the parameters in A; if the constant C is universal, we write $O(\alpha)$ instead.

2. Reduction to variable polynomial iterates

We start with the defintion of Hardy field functions. Let B be the collection of equivalence classes of real valued functions defined on some halfline (x_0, ∞), $x_0 \geq 0$, where two functions that eventually agree are identified. These equivalence classes are called *germs* of functions. A *Hardy field* is a subfield of the ring $(B, +, \cdot)$ that is closed under differentiation.[14]

Usually, one deals with the class of logarithmico-exponential Hardy field functions, $\mathcal{LE}$, which can be handled more easily: h is a *logarithmico-exponential Hardy field function* if it is defined on some $(c, +\infty)$, $c \geq 0$, by a finite combination of symbols $+, -, \times, \div, \sqrt[n]{\cdot}, \exp, \log$ acting on the real variable x and on real constants (for more on Hardy field functions and in particular for logarithmico-exponential ones one can check [13], [15], [23]).

As in [36], we will work on the Hardy field $\mathcal{H}$ which is closed under composition and compositional inversion of functions, when defined (i.e., if $h_1, h_2 \in \mathcal{H}$ with $\lim_{x\to\infty} h_2(x) = \infty$, then $h_1 \circ h_2$, $h_2^{-1} \in \mathcal{H}$).[15]

A function h from $\mathcal{H}$ of polynomial growth has *degree* a non-negative integer $d_h \geq 0$, if $x^{d_h} \ll h(x) \prec x^{d_h+1}$ (recall from the introduction that if $x^{d_h} \prec h(x) \prec x^{d_h+1}$, then h is strongly non-polynomial).

As we mention before, our work concerns iterates involving families of variable polynomials. A sequence of real *variable polynomials* is a sequence of the form $(p_N(n))_{N,n} \subseteq \mathbb{R}$, where we assume that while the polynomials p_N might depend on N, their degrees do not.[16] The following are two examples of sequences of variable polynomials:

$$p_{N,1}(n) = \frac{n^{17}}{\sqrt{N}}, \quad p_{N,1}(n) = \Big(\frac{\sqrt{2}}{N^{e/\pi}} + \frac{N}{3}\Big)n^7 - \frac{31}{\log N}n + 1, \quad N, n \in \mathbb{N}.$$

As in [36], the main idea in our setting is that we will approximate a given function, $a \in \mathcal{H}$, by "good" variable polynomials, $(p_N)_N$, in suitable intervals (with lengths that tend to infinity). Then, as reflected in the following proposition, to study multiple ergodic averages with iterates $[a(n)]$, it suffices to study some related weighted (with some bounded error terms as weights) averages with iterates $[p_N(n)]$.

---

[13]Notice that we change our notation form T^n, $n \in \mathbb{Z}$ to T_n, $n \in \mathbb{Z}^d$ when dealing with a $\mathbb{Z}$ or, respectively, a $\mathbb{Z}^d$-action to distinguish them.

[14]We use the word *function* when we refer to elements of B (understanding that all the operations defined and statements made for elements of B are considered only for sufficiently large values of $x \in \mathbb{R}$).

[15]Notice here that $\mathcal{LE}$ does not have this property but it is contained in the Hardy field of Pfaffian functions which does ([28]).

[16]For a study on "good" variable polynomials, see [30].

Proposition 2.1. *Let $(X,\mathcal{B},\mu,T_1,\dots,T_d)$ be a system with commuting and invertible transformations. Let a be a function and $L\in\mathcal{H}$ a positive function with $1\prec L(x)\prec x$. Let $(p_N)_N$ be a sequence of functions such that for all $N\in\mathbb{N}$ and $0\leq r\leq L(N)$,*

$$a(N+r)=p_N(r)+e_{N,r},\quad \text{with } \ e_{N,r}\ll 1.$$

Assuming that

$$\limsup_{N\to\infty}\sup_{|c_n|\leq 1}\sup_{\|f_2\|_\infty,\dots,\|f_d\|_\infty\leq 1}\left\|\mathbb{E}_{0\leq n\leq L(N)}c_n\prod_{i=1}^d T_i^{[p_N(n)]}f_i\right\|_2^\kappa=0,$$

for some $\kappa\in\mathbb{N}$ and $f_1\in L^\infty(\mu)$, we have

$$\limsup_{N\to\infty}\left\|\mathbb{E}_{1\leq n\leq N}\prod_{i=1}^d T_i^{[a(n)]}f_i\right\|_2=0$$

for all $f_2,\dots,f_d\in L^\infty(\mu)$.

Proof. To show the result, by [36, Lemma 3.3],[17] it suffices to show that

$$\limsup_{R\to\infty}\mathbb{E}_{1\leq N\leq R}\left\|\mathbb{E}_{N\leq n\leq N+L(N)}\prod_{i=1}^d T_i^{[a(n)]}f_i\right\|_2^\kappa=0,$$

hence, it suffices to show

$$\limsup_{N\to\infty}\left\|\mathbb{E}_{N\leq n\leq N+L(N)}\prod_{i=1}^d T_i^{[a(n)]}f_i\right\|_2^\kappa=0.$$

Write $n=N+r$ for some $0\leq r\leq L(N)$. Since $a(N+r)=p_N(r)+e_{N,r}$, then $[a(N+r)]=[p_N(r)]+\tilde{e}_{N,r},\ \tilde{e}_{N,r}\ll 1$, hence the left hand side of the previous relation is equal to

$$\limsup_{N\to\infty}\left\|\mathbb{E}_{0\leq r\leq L(N)}\prod_{i=1}^d T_i^{[p_N(r)]+\tilde{e}_{N,r}}f_i\right\|_2^\kappa. \tag{3}$$

Since $\tilde{e}_{N,r}\ll 1$, by [36, Lemma 3.2], (3) is bounded by a constant multiple of the quantity

$$\limsup_{N\to\infty}\sup_{|c_r|\leq 1}\sup_{\|f_2\|_\infty,\dots,\|f_d\|_\infty\leq 1}\left\|\mathbb{E}_{0\leq r\leq L(N)}c_r\prod_{i=1}^d T_i^{[p_N(r)]}f_i\right\|_2^\kappa,$$

finishing the proof. □

We will demonstrate how we use the previous approach for a Hardy field function

$$h(x)=s_h(x)+p_h(x).$$

We do this with two specific examples to cover both cases, i.e., $d_{p_h}<d_{s_h}+1$, and $d_{p_h}\geq d_{s_h}+1$, where d_{p_h} is the degree of p_h and d_{s_h} is the degree of s_h (for the general case, see right after the proof of Theorem 5.1). In both cases, we will choose a positive integer K that will indicate the order of the Taylor expansion for the non-polynomial part.

[17] We remark at this point that the assumption $L\in\mathcal{H}$, where $\mathcal{H}$ is a, closed under composition and compositional inversion of functions, Hardy field, is postulated exactly so we can use this statement.

Example 1. Let $h_1(x) = \pi x + x \log x$.

Here we have $p_{h_1}(x) = \pi x$, and $s_{h_1}(x) = x \log x$, hence $d_{p_{h_1}} = 1 < 1 + d_{s_{h_1}}$. Picking $K = d_{s_{h_1}} + 1 = 2$, we have that $h_1(N + r)$ is approximated by

$$\begin{aligned} p_{N,1}(r) &= p_{h_1}(N + r) + s_{h_1}(N) + s'_{h_1}(N)r + \frac{s''_{h_1}(N)}{2} r^2 \\ &= \pi N + N \log N + (\pi + 1 + \log N)r + \frac{1}{2N} r^2, \end{aligned}$$

$0 \leq r \leq L_1(N) = N^{\frac{7}{12}}$ (we picked $L_1(x)$ as the geometric mean of $|s''_{h_1}(x)|^{-\frac{1}{2}}$ and $|s'''_{h_1}(x)|^{-\frac{1}{3}}$).

Example 2. Let $h_2(x) = \sqrt{2}x^2 + \log^2 x$.

Here we have $p_{h_2}(x) = \sqrt{2}x^2$, and $s_{h_2}(x) = \log^2 x$, hence $d_{p_{h_2}} = 2 > 1 + d_{s_{h_2}}$. Picking $K = d_{p_{h_1}} + 1 = 3$, we have that $h_2(N + r)$ is approximated by

$$\begin{aligned} p_{N,2}(r) &= p_{h_2}(N + r) + s_{h_2}(N) + s'_{h_2}(N)r + \frac{s''_{h_2}(N)}{2} r^2 + \frac{s'''_{h_2}(N)}{6} r^3 \\ &= \sqrt{2}N^2 + \log^2 N + \Big(2\sqrt{2}N + \frac{2 \log N}{N}\Big) r + \Big(\sqrt{2} + \frac{1 - \log N}{N^2}\Big) r^2 + \frac{2 \log N - 3}{3N^3} r^3, \end{aligned}$$

$0 \leq r \leq L_2(N) = N/(\log^{7/24} N)$ (here we picked $L_2(x)$ to be of the same growth rate as the geometric mean of $|s'''_{h_2}(x)|^{-\frac{1}{3}}$ and $|s^{(4)}_{h_2}(x)|^{-\frac{1}{4}}$).

As we already mentioned, and it is verified by both examples we just saw, we will deal with functions such that the corresponding variable sequence $(p_N)_N$ doesn't have leading coefficient 1. In that case, we will transform it into such. This is crucial to our study in order to use the concatenation approach from [10]. The following proposition justifies this and can be viewed as a change-of-variables procedure.

Proposition 2.2. *Let $(a_N)_N$ be a sequence of real numbers with (eventually) constant sign, $L \in \mathcal{H}$ a positive function with $1 \prec L(x) \prec x$, and $K \in \mathbb{N}$ such that*

- $\lim_{N\to\infty} L(N)|a_N|^{\frac{1}{K}} = \infty$;
- $\lim_{N\to\infty} a_N = 0$; *and*
- $L(N) \ll |a_N|^{-\frac{K+1}{K^2}}$.

If $(p_N)_N$ is a variable polynomial sequence of degree less than K, then there exist a variable polynomial sequence $(\tilde{p}_N)_N$ of degree less than K and a positive function $\tilde{L}$ with $1 \prec \tilde{L}(x) \prec x$ such that for every system $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$, $f_1 \in L^\infty(\mu)$, and $\kappa \in \mathbb{N}$, we have

$$\begin{aligned} &\limsup_{N\to\infty} \sup_{|c_n|\leq 1} \sup_{\|f_2\|_\infty,\dots,\|f_d\|_\infty\leq 1} \Big\| \mathbb{E}_{0\leq n\leq L(N)} c_n \prod_{i=1}^{d} T_i^{[a_N n^K + p_N(n)]} f_i \Big\|_2^\kappa \\ &\leq \limsup_{N\to\infty} \sup_{|c_n|\leq 1} \sup_{\|f_2\|_\infty,\dots,\|f_d\|_\infty\leq 1} \Big\| \mathbb{E}_{0\leq n\leq \tilde{L}(N)} c_n \prod_{i=1}^{d} T_i^{[n^K + \tilde{p}_N(n)]} f_i \Big\|_2^\kappa . \end{aligned} \tag{4}$$

Proof. For convenience denote $D_N := |a_N|^{\frac{1}{K}}$. We assume without loss of generality that, for large N, $a_N > 0$. We have $\lim_{N\to\infty} D_N^{-1} = \infty$. For $0 \leq n \leq L(N)$, we may write

$$n = k[D_N^{-1}] + s,$$

for some $0 \leq k \leq [\tilde{L}(N)]$, where $\tilde{L}(N) := L(N)/[D_N^{-1}]$ and $0 \leq s \leq [D_N^{-1}] - 1$.[18] Then

$$a_N n^K + p_N(n) = a_N(k[D_N^{-1}] + s)^K + p_N(k[D_N^{-1}] + s) = k^K(D_N[D_N^{-1}])^K + p_{N,s}(k)$$

for some polynomial $p_{N,s}$ of degree at most $K - 1$.

Note that, if N is large,

$$|k^K(D_N[D_N^{-1}])^K - k^K| \ll 1.$$

Indeed,

$$\begin{aligned}\left|k^K(D_N[D_N^{-1}])^K - k^K\right| &\leq Kk^K\left|D_N[D_N^{-1}] - 1\right| = Kk^K\left|D_N\left(D_N^{-1} - \{D_N^{-1}\}\right) - 1\right| \leq Kk^K D_N \\ &\leq KD_N\left(\frac{L(N)}{D_N^{-1} - 1}\right)^K \ll KD_N\left(\frac{L(N)}{D_N^{-1}}\right)^K = K\left(\frac{L(N)}{D_N^{-\frac{K+1}{K}}}\right)^K \ll 1.\end{aligned}$$

So,

$$a_N n^K + p_N(n) = k^K + p_{N,s}(k) + O(1).$$

The left hand side of (4), by using convexity, and then [36, Lemma 3.2] to deal with the bounded error terms, is bounded by a constant multiple of

$$\limsup_{N\to\infty} \mathbb{E}_{0\leq s\leq [D_N^{-1}]-1} \sup_{|c_n|\leq 1} \sup_{\|f_2\|_\infty,\ldots,\|f_d\|_\infty\leq 1} \left\|\mathbb{E}_{0\leq n\leq \tilde{L}(N)} c_n \prod_{i=1}^d T_i^{[n^K + p_{N,s}(n)]} f_i\right\|_2^\kappa.$$

Since

$$\lim_{N\to\infty} \tilde{L}(N) = \lim_{N\to\infty} L(N)D_N\frac{D_N^{-1}}{[D_N^{-1}]} = \lim_{N\to\infty} L(N)D_N = \lim_{N\to\infty} L(N)|a_N|^{\frac{1}{K}} = \infty,$$

and

$$\lim_{N\to\infty} \frac{\tilde{L}(N)}{N} = \lim_{N\to\infty} \frac{L(N)}{N} D_N\frac{D_N^{-1}}{[D_N^{-1}]} = \lim_{N\to\infty} \frac{L(N)}{N} D_N = 0,$$

by setting $\tilde{p}_N$ to be the $p_{N,s}$ which attends the maximum of

$$\sup_{|c_n|\leq 1} \sup_{\|f_2\|_\infty,\ldots,\|f_d\|_\infty\leq 1} \left\|\mathbb{E}_{0\leq n\leq \tilde{L}(N)} c_n \prod_{i=1}^d T_i^{[n^K + p_{N,s}(n)]} f_i\right\|_2^\kappa,$$

we get the result. □

In particular, for Example 1, setting $a_N = \frac{s''_{h_1}(N)}{2} = \frac{1}{2N}$, we can pick $\tilde{L}_1(N) = \frac{L_1(N)}{[a_N^{-1/2}]} = \frac{N^{\frac{7}{12}}}{[\sqrt{2N}]}$ (which grows as $\frac{N^{\frac{1}{12}}}{\sqrt{2}}$), so, for $n = k[a_N^{-1/2}] + s$, $0 \leq k \leq [\tilde{L}_1(N)]$, $0 \leq s \leq [a_N^{-1/2}] - 1$, we have

$$p_{N,1}(n) = k^2 + 2a_N[a_N^{-1/2}]k + a_N s^2 + \tilde{p}_{N,1}(k[a_N^{-1/2}] + s) + O(1),$$

where $\tilde{p}_{N,1}(r) = \pi N + N\log N + (\pi + 1 + \log N)r$ is of degree 1.

[18]This $\tilde{L}$ can actually be taken in $\mathcal{H}$ (in particular we can set $\tilde{L}(N)$ to be equal to $L(N)D_N$) by the cost of an average that goes to 0 as we lose values from a set of density 0.

Similarly, for Example 2, setting $a_N = \frac{s'''_{h_2}(N)}{6} = \frac{2\log N - 3}{3N^3}$, we can pick $\tilde{L}_2(N) = \frac{L_2(N)}{[a_N^{-1/3}]} = \frac{N}{\log^{7/24} N} \cdot \frac{1}{\left[\frac{\sqrt[3]{3}N}{\sqrt[3]{2\log N - 3}}\right]}$ (which grows as $\sqrt[3]{\frac{2}{3}} \log^{\frac{1}{24}} N$), so, for $n = k[a_N^{-1/3}] + s$, $0 \leq k \leq \tilde{L}_2(N)$, $0 \leq s \leq [a_N^{-1/3}] - 1$, we have

$$p_{N,2}(n) = k^3 + 3a_N[a_N^{-1/3}]^2 sk^2 + 3a_N[a_N^{-1/3}]s^2k + a_N s^3 + \tilde{p}_{N,2}(k[a_N^{-1/3}] + s) + O(1),$$

where $\tilde{p}_{N,2}(r) = \sqrt{2}N^2 + \log^2 N + \left(2\sqrt{2}N + \frac{2\log N}{N}\right)r + \left(\sqrt{2} + \frac{1-\log N}{N^2}\right)r^2$ is of degree 2.

3. PET induction for variable polynomials

In this section we define the van der Corput operation, which will be used, together with the van der Corput lemma (Lemma 3.1 and 3.2) and PET induction scheme, to get the required upper bounds of the expressions of interest. To achieve the latter, we also need to control the coefficients of the polynomial iterates (for which we follow [10]).

3.1. **Van der Corput lemmas and van der Corput operation.** We will use two different versions of the van der Corput lemma.

Lemma 3.1 (Lemma 4.3 of [36])**.** *Let $(u_n)_{n\in\mathbb{Z}}$ be a sequence in a Hilbert space with $\|u_n\| \leq 1$, $d \geq 1$ and $M, N \in \mathbb{N}$. Then*

$$\|\mathbb{E}_{n\in[N]}u_n\|^{2^d} \ll_d \frac{1}{M} + \left(\frac{M}{N}\right)^{2^{d-1}} + \mathbb{E}_{-M\leq m\leq M}|\mathbb{E}_{n\in[N]}\langle u_{n+m}, u_n\rangle|^{2^{d-1}}.$$

The next follows from from Chapter 21, Section 1.2, Lemma 1 of [26]:

Lemma 3.2. *Let $(u_n)_{n\in\mathbb{Z}}$ be a sequence in a Hilbert space with $\|u_n\| \leq 1$ and $M, N \in \mathbb{N}$. Then*

$$\|\mathbb{E}_{n\in[N]}u_n\|^2 \leq \frac{6M}{N} + \mathbb{E}_{x,y\in[M]}\mathbb{E}_{n\in[N]}\langle u_{n+x}, u_{n+y}\rangle.$$

The PET induction is an inductive procedure to reduce the complexity of multiple ergodic averages, which was first introduced in [2]. In this paper, we use a variation of the PET induction scheme introduced in [10], adapted to the families of variable polynomials.

We say that a sequence of polynomials $q = (q_N)_{N\in\mathbb{N}}$, $q_N : \mathbb{Z}^s \to \mathbb{R}$ is *consistent* if the degree of q_N with respect to the first variable is, for N sufficiently large, a constant. In this case this constant is defined to be the *degree* of the sequence, denoted by $\deg(q)$.

We say that a consistent sequence q is *essentially non-constant* if $\deg(q) > 0$, and that two consistent sequences q and q' are *essentially distinct* if $q - q'$ is essentially non-constant. We say that a tuple of polynomial sequences $(q_1, \ldots, q_\ell)$ is *consistent* if all of $q_i, q_i - q_j$, $i \neq j$, are consistent, and *non-degenerate* if the q_i's are essentially non-constant and essentially distinct.

Let $s \in \mathbb{N}_0$ and $\ell \in \mathbb{N}$. For $1 \leq m \leq \ell$ and $N \in \mathbb{N}$, let $q_{N,m} : \mathbb{Z}^{s+1} \to \mathbb{R}^d$ be a polynomial. Put $\mathbf{q} = (q_{N,1}, \ldots, q_{N,\ell})_N$. We say that $A = (s, \ell, \mathbf{q})$ is a *PET-tuple.*[19] The tuple $A = (s, \ell, \mathbf{q})$ is *non-degenerate* (resp. *consistent*) if $\mathbf{q}$ is non-degenerate (resp. consistent).

For each non-degenerate PET-tuple $A = (s, \ell, \mathbf{q})$ and $1 \leq t \leq \ell$, we define the *vdC-operation*, $\partial_t A$, according to the following three steps:

[19]We use s instead of $s + 1$ to highlight the number of h_i's.

Step 1: For all $1 \leq m \leq \ell$ and $N \in \mathbb{N}$, let $q'_{N,1}, \dots, q'_{N,2\ell} : \mathbb{Z}^{s+2} \to \mathbb{R}^d$ be functions defined as

$$q'_{N,m}(n; h_1, \dots, h_{s+1}) = \begin{cases} q_{N,m-\ell}(n; h_1, \dots, h_s) - q_{N,t}(n; h_1, \dots, h_s) & , \ \ell+1 \leq m \leq 2\ell \\ q_{N,m}(n + h_{s+1}; h_1, \dots, h_s) - q_{N,t}(n; h_1, \dots, h_s) & , \ 1 \leq m \leq \ell \end{cases}.$$

We use the letter n for the first variable and h_i's for the remaining ones. For convenience, we write $q'_m := (q'_{N,m})_N$ and let q'_0 denote the sequence of constant zero polynomials.

Lemma 3.3. *If* $\mathbf{q}$ *is non-degenerate, then for all* $0 \leq i, j \leq 2\ell, i \neq j$, $q'_i - q'_j$ *is consistent.*

Proof. Unpacking the definitions, it suffices to verify that the following families are consistent for all $0 \leq i, j \leq 2\ell, i \neq j$:

(i) $q_{N,i}(n; h_1, \dots, h_s) - q_{N,j}(n; h_1, \dots, h_s), N \in \mathbb{N}$;
(ii) $q_{N,i}(n + h_{s+1}; h_1, \dots, h_s) - q_{N,j}(n + h_{s+1}; h_1, \dots, h_s), N \in \mathbb{N}$;
(iii) $q_{N,i}(n + h_{s+1}; h_1, \dots, h_s) - q_{N,j}(n; h_1, \dots, h_s), N \in \mathbb{N}$.

The first case follows from the assumption that $\mathbf{q}$ is non-degenerate. The second case follows from the assumption that $\mathbf{q}$ is non-degenerate and the fact that $q_{N,i}(n + h_{s+1}; h_1, \dots, h_s) - q_{N,j}(n + h_{s+1}; h_1, \dots, h_s)$ has the same leading coefficient in the variable n as that of $q_{N,i}(n; h_1, \dots, h_s) - q_{N,j}(n; h_1, \dots, h_s)$. The third case is similar to the second one. □

Step 2: We remove from $q'_1, \dots, q'_{2\ell}$ the collections of functions q'_j which are essentially constant and the corresponding functions with those as iterates, and then put the remaining ones into groups $J_i = \{(q''_{N,i,1})_N, \dots, (q''_{N,i,t_i})_N\}$, $1 \leq i \leq r$, for some $r, t_i \in \mathbb{N}$ such that two sequences are essentially distinct if, and only if, they belong to different groups. For every $1 \leq j \leq t_i$, there exist variable polynomials $p''_{N,i,j} : \mathbb{Z}^{s+1} \to \mathbb{R}$ such that $q''_{N,i,j}(n; h_1, \dots, h_{s+1}) = q''_{N,i,1}(n; h_1, \dots, h_{s+1}) + p''_{N,i,j}(h_1, \dots, h_{s+1})$ for sufficiently large N.

Step 3: Let $q^*_{N,i} = q''_{N,i,1}$. Set $\mathbf{q}^* = (q^*_{N,1}, \dots, q^*_{N,r})_{N \in \mathbb{N}}$, and let this new PET-tuple be $\partial_t A = (s+1, r, \mathbf{q}^*)$.[20] It is clear from the construction that $\mathbf{q}^*$ and $\partial_t A$ are non-degenerate. Therefore, if A is non-degenerate, then so is $\partial_t A$.

We say that the operation $A \to \partial_t A$ is *1-inherited* if $q'_1 = q^*_1$ and we did not drop q^*_1 or group it with any other q^*_i in Step 2.

Let $A = (s, \ell, \mathbf{q})$ be a PET-tuple, where $\mathbf{q} = (q_{N,1}, \dots, q_{N,\ell})_N$ with $q_{N,i} : \mathbb{Z}^{s+1} \to \mathbb{R}^d$ being polynomials, $\kappa \in \mathbb{N}$, $(X, \mathcal{B}, \mu, (T_n)_{n \in \mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system, and $f \in L^\infty(\mu)$. For $h_1, \dots, h_s \in \mathbb{Z}$, set

$$S(A, f, \kappa, (h_1, \dots, h_s)) := \overline{\lim}_{N \to \infty} \sup_{|c_n| \leq 1} \sup_{\|g_2\|_\infty, \dots, \|g_\ell\|_\infty \leq 1} \left\| \mathbb{E}_{n \in [N]} c_n \prod_{m=1}^{\ell} T_{[q_{N,m}(n; h_1, \dots, h_s)]} g_m(x) \right\|_2^\kappa,$$

where $g_1 := f$, and

$$\begin{aligned} S(A, f, \kappa) &:= \overline{\mathbb{E}}_{h_1, \dots, h_s \in \mathbb{Z}} \overline{\lim}_{N \to \infty} \sup_{|c_n| \leq 1} \sup_{\|g_2\|_\infty, \dots, \|g_\ell\|_\infty \leq 1} \left\| \mathbb{E}_{n \in [N]} c_n \prod_{m=1}^{\ell} T_{[q_{N,m}(n; h_1, \dots, h_s)]} g_m(x) \right\|_2^\kappa \\ &= \overline{\mathbb{E}}_{h_1, \dots, h_s \in \mathbb{Z}} S(A, f, \kappa, (h_1, \dots, h_s)). \end{aligned}$$

[20]Here we abuse the notation by writing $\partial_t A$ to denote any of such operations obtained from Step 1 to 3. Strictly speaking, $\partial_t A$ is not uniquely defined as the order of grouping of $q'_{N,1}, \dots, q'_{N,2\ell}$ in Step 2 is ambiguous. However, this is done without loss of generality, since the order does not affect the value of $S(\partial_t A, \cdot)$ (see below).

Lemma 3.4. *Let $(X,\mathcal{B},\mu,(T_n)_{n\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system, $A=(s,\ell,\mathbf{q})$ be a non-degenerate PET-tuple, $f\in L^\infty(\mu)$, and $\kappa\in\mathbb{N}$. Then, for any $1\leq t\leq \ell$, $\partial_t A$ is also a non-degenerate PET-tuple. Moreover, if $A\to\partial_t A$ is 1-inherited, then*

$$S(A,f,2\kappa)\ll_{\kappa,\ell} S(\partial_t A,f,\kappa).$$

Proof. The fact that $\partial_t A$ is a non-degenerate PET-tuple was verified previously (see Step 3). We are left with proving the second conclusion. For convenience, write $\mathbf{h}:=(h_1,\dots,h_s)$ and $\mathbf{h}':=(h_1,\dots,h_{s+1})$. Suppose that $\partial_t A=(s+1,r,\mathbf{q}^*)$.

Fix $\mathbf{h}=(h_1,\dots,h_s)$. For every $N\in\mathbb{N}$, $1\leq n\leq N$, we pick $|c_{N,n}|\leq 1$, and $g_{N,m}\in L^\infty(\mu)$ with $\|g_{N,m}\|_\infty\leq 1$, $2\leq m\leq \ell$, so that

$$\Big\|\mathbb{E}_{n\in[N]}c_{N,n}\prod_{m=1}^{\ell}T_{[q_{N,m}(n;h_1,\dots,h_s)]}g_{N,m}(x)\Big\|_2^{2\kappa}$$

is $1/N$ close to

$$\sup_{|c_n|\leq 1}\sup_{\|g_2\|_\infty,\dots,\|g_\ell\|_\infty\leq 1}\Big\|\mathbb{E}_{n\in[N]}c_n\prod_{m=1}^{\ell}T_{[q_{N,m}(n;h_1,\dots,h_s)]}g_m(x)\Big\|_2^{2\kappa},$$

where $g_{N,1}:=g_1:=f$. For $M,N\in\mathbb{N}$, by Lemma 3.1, we have that

(5)
$$\begin{aligned}
&\Big\|\mathbb{E}_{n\in[N]}c_{N,n}\prod_{m=1}^{\ell}T_{[q_{N,m}(n;\mathbf{h})]}g_{N,m}\Big\|_2^{2\kappa}\\
&\ll_\kappa \mathbb{E}_{|h_{s+1}|\leq M}\Big|\mathbb{E}_{n\in[N]}\Big\langle c_{N,n}\prod_{m=1}^{\ell}T_{[q_{N,m}(n;\mathbf{h})]}g_{N,m},c_{N,n+h_{s+1}}\prod_{m=1}^{\ell}T_{[q_{N,m}(n+h_{s+1};\mathbf{h})]}g_{N,m}\Big\rangle\Big|^\kappa\\
&\qquad\qquad+\frac{1}{M}+\Big(\frac{M}{N}\Big)^\kappa\\
&=\mathbb{E}_{|h_{s+1}|\leq M}\Big|\mathbb{E}_{n\in[N]}\Big\langle c_{N,n}\prod_{m=1}^{\ell}T_{[q_{N,m}(n;\mathbf{h})]-[q_{N,t}(n;\mathbf{h})]}g_{N,m},\\
&\qquad\qquad c_{N,n+h_{s+1}}\prod_{m=1}^{\ell}T_{[q_{N,m}(n+h_{s+1};\mathbf{h})]-[q_{N,t}(n;\mathbf{h})]}g_{N,m}\Big\rangle\Big|^\kappa+\frac{1}{M}+\Big(\frac{M}{N}\Big)^\kappa\\
&=\mathbb{E}_{|h_{s+1}|\leq M}\Big|\mathbb{E}_{n\in[N]}\Big\langle c_{N,n}\prod_{m=1}^{\ell}T_{[q'_{N,m+\ell}(n;\mathbf{h}')]+\epsilon_{N,m+\ell,n,\mathbf{h}'}}g_{N,m},\\
&\qquad\qquad c_{N,n+h_{s+1}}\prod_{m=1}^{\ell}T_{[q'_{N,m}(n;\mathbf{h}')]+\epsilon_{N,m,n,\mathbf{h}'}}g_{N,m}\Big\rangle\Big|^\kappa+\frac{1}{M}+\Big(\frac{M}{N}\Big)^\kappa\\
&=\mathbb{E}_{|h_{s+1}|\leq M}\Big|\mathbb{E}_{n\in[N]}\Big\langle c_{N,n}\overline{c}_{N,n+h_{s+1}},\prod_{m=1}^{2\ell}T_{[q'_{N,m}(n;\mathbf{h}')]+\epsilon_{N,m,n,\mathbf{h}'}}g_{N,m}\Big\rangle\Big|^\kappa+\frac{1}{M}+\Big(\frac{M}{N}\Big)^\kappa,
\end{aligned}$$

where $\epsilon_{N,m,n,\mathbf{h}'}\in\{-1,0,1\}$ and $g_{m+\ell,N}:=\overline{g}_{m,N}$ for $1\leq m\leq\ell$.

Assume that $\mathbf{q}^* = (q^*_{N,1}, \dots, q^*_{N,r})_N$. Since $A \to \partial_t A$ is 1-inherited, $q^*_{N,1} = q'_{N,1}$ and we did not drop or combine $(q'_{N,1})_N$ with any other polynomial. For $1 \leq j \leq r$, let I_j be the set of $m \in \{1, \dots, 2\ell\}$ such that $(q'_{N,m})_N$ is essentially the same as $(q^*_{N,j})_N$. For $m \in I_j$, we may write $q'_{N,m} := q^*_{N,j} + \tilde{q}_{N,j,m}$ for some variable polynomial family $(\tilde{q}_{N,j,m})_N$ which does not depend on the variable n (i.e., the first variable) when N is sufficiently large. Let I_0 be the set of $m \in \{1, \dots, 2\ell\}$ such that $(q'_{N,m})_N$ is essentially constant. Then the last line of (5) is bounded by

$$\begin{aligned}(6)\quad &\mathbb{E}_{|h_{s+1}| \leq M} \Big| \mathbb{E}_{n \in [N]} \Big\langle \prod_{m \in I_0} T_{[q'_{N,m}(n;\mathbf{h}')] + \epsilon_{N,m,n,\mathbf{h}'}} \overline{g}_{N,m}, \\ &\qquad \overline{c}_{N,n} c_{N,n+h_{s+1}} \prod_{j=1}^{r} \prod_{m \in I_j} T_{[q'_{N,m}(n;\mathbf{h}')] + \epsilon_{N,m,n,\mathbf{h}'}} g_{N,m} \Big\rangle \Big|^{\kappa} + \frac{1}{M} + \Big(\frac{M}{N}\Big)^{\kappa}.\end{aligned}$$

Since for all $m \in I_0$, $q'_{N,m}(n;\mathbf{h}')$ is independent of n when N is sufficiently large, we may write $q'_{N,m}(n;\mathbf{h}') = q''_{N,m}(\mathbf{h}')$ for some polynomial $q''_{N,m}$. For any $\epsilon = (\epsilon_m)_{m \in I_0} \in \{-1, 0, 1\}^{|I_0|}$, let $A_{N,\mathbf{h}',\epsilon}$ denote the set of $n \in [N]$ such that $\epsilon_{N,m,n,\mathbf{h}'} = \epsilon_m$ for all $m \in I_0$. Then we may rewrite (6) as

$$\begin{aligned}(7)\quad &\mathbb{E}_{|h_{s+1}| \leq M} \Big| \mathbb{E}_{n \in [N]} \sum_{\epsilon \in \{-1,0,1\}^{|I_0|}} \Big\langle \mathbf{1}_{A_{N,\mathbf{h}',\epsilon}}(n) \prod_{m \in I_0} T_{[q''_{N,m}(\mathbf{h}')] + \epsilon_m} \overline{g}_{N,m}, \\ &\qquad \overline{c}_{N,n} c_{N,n+h_{s+1}} \prod_{j=1}^{r} \prod_{m \in I_j} T_{[q'_{N,m}(n;\mathbf{h}')] + \epsilon_{N,m,n,\mathbf{h}'}} g_{N,m} \Big\rangle \Big|^{\kappa} + \frac{1}{M} + \Big(\frac{M}{N}\Big)^{\kappa} \\ &\ll_\ell \mathbb{E}_{|h_{s+1}| \leq M} \sup_{\epsilon \in \{-1,0,1\}^{|I_0|}} \sup_{|c_n| \leq 1} \Big| \mathbb{E}_{n \in [N]} \Big\langle \prod_{m \in I_0} T_{[q''_{N,m}(\mathbf{h}')] + \epsilon_m} \overline{g}_{N,m}, \\ &\qquad c_n \prod_{j=1}^{r} \prod_{m \in I_j} T_{[q'_{N,m}(n;\mathbf{h}')] + \epsilon_{N,m,n,\mathbf{h}'}} g_{N,m} \Big\rangle \Big|^{\kappa} + \frac{1}{M} + \Big(\frac{M}{N}\Big)^{\kappa}.\end{aligned}$$

When N is sufficiently large, we may use the Cauchy-Schwarz inequality to bound the last line of (7) by

$$\begin{aligned}(8)\quad &\mathbb{E}_{|h_{s+1}| \leq M} \sup_{|c_n| \leq 1} \Big\| \mathbb{E}_{n \in [N]} c_n \prod_{j=1}^{r} \prod_{m \in I_j} T_{[q'_{N,m}(n;\mathbf{h}')] + \epsilon_{N,m,n,\mathbf{h}'}} g_{N,m} \Big\|_2^{\kappa} + \frac{1}{M} + \Big(\frac{M}{N}\Big)^{\kappa} \\ &= \mathbb{E}_{|h_{s+1}| \leq M} \sup_{|c_n| \leq 1} \Big\| \mathbb{E}_{n \in [N]} c_n \prod_{j=1}^{r} \prod_{m \in I_j} T_{[q^*_{N,j}(n;\mathbf{h}')] + [\tilde{q}_{N,j,m}(n;\mathbf{h}')] + \epsilon'_{N,m,n,\mathbf{h}'}} g_{N,m} \Big\|_2^{\kappa} + \frac{1}{M} + \Big(\frac{M}{N}\Big)^{\kappa},\end{aligned}$$

where $\epsilon'_{N,m,n,\mathbf{h}'} \in \{-2, 0, 2\}$. Since $\partial_t A$ is 1-inherited we have that

$$\prod_{m \in I_1} T_{[q^*_{N,m}(n;\mathbf{h}')] + [\tilde{q}_{N,j,m}(n;\mathbf{h}')] + \epsilon'_{N,m,n,\mathbf{h}'}} g_{N,m} = T_{[q^*_{N,1}(n;\mathbf{h}')] + \epsilon'_{N,1,n,\mathbf{h}'}} f.$$

Using [36, Lemma 3.2], the last line of (8) is bounded by $O_{\kappa,\ell}(1)$ times

$$\mathbb{E}_{|h_{s+1}|\leq M} \sup_{|c_n|\leq 1} \sup_{\substack{\|g^*_m\|_\infty\leq 1,\\ m\in\cup_{j=2}^r I_j}} \left\| \mathbb{E}_{n\in[N]} c_n \prod_{j=1}^{r} \prod_{m\in I_j} T_{[q^*_{N,j}(n;\mathbf{h}')]+[\tilde{q}_{N,j,m}(n;\mathbf{h}')]} g^*_m \right\|_2^\kappa + \frac{1}{M} + \left(\frac{M}{N}\right)^\kappa$$

$$\leq \mathbb{E}_{|h_{s+1}|\leq M} \sup_{|c_n|\leq 1} \sup_{\|g_2\|_\infty,\dots,\|g_r\|_\infty\leq 1} \left\| \mathbb{E}_{n\in[N]} c_n \prod_{j=1}^{r} T_{[q^*_{N,j}(n;\mathbf{h}')]} g_j \right\|_2^\kappa + \frac{1}{M} + \left(\frac{M}{N}\right)^\kappa,$$

where $g^*_1 = f$. Taking the limsup as N goes to infinity, we conclude that

$$\limsup_{N\to\infty} \sup_{|c_n|\leq 1} \sup_{\|g_2\|_\infty,\dots,\|g_\ell\|_\infty\leq 1} \left\| \mathbb{E}_{n\in[N]} c_n \prod_{m=1}^{\ell} T_{[q_{N,m}(n;\mathbf{h})]} g_m \right\|_2^{2\kappa} \leq \mathbb{E}_{|h_{s+1}|\leq M} S(\partial_t A, \kappa, \mathbf{h}') + \frac{1}{M}.$$

Letting M go to infinity we get

$$S(A, f, 2\kappa, \mathbf{h}) \ll_{\kappa,\ell} \mathbb{E}_{h_{s+1}\in\mathbb{Z}} S(\partial_t A, f, \kappa, \mathbf{h}').$$

Taking the average $\mathbb{E}_{h_1,\dots,h_s\in\mathbb{Z}}$, we obtain the desired conclusion. □

Let $A = (s, \ell, \mathbf{q} = (q_{N,1}, \dots, q_{N,\ell})_N)$ be a PET tuple. Let $\deg(A)$, the *degree* of A, be the maximum of $\deg((q_{N,j})_N), 1 \leq j \leq \ell$. We say that A is *1-standard* $\deg((q_{N,1})_N) = \deg(A)$.

Lemma 3.5. *Let A be a 1-standard and non-degenerate PET-tuple with* $\deg(A) \geq 1$. *There exist $M \in \mathbb{N}$ depending only on* $\deg(A)$, *$\ell \in \mathbb{N}$, and $i_1, \dots, i_M \in \mathbb{N}$ such that for all $1 \leq M' \leq M$, $\partial_{i_{M'-1}} \dots \partial_{i_1} A \to \partial_{i_{M'}} \dots \partial_{i_1} A$ is 1-inherited,*[21] *$\partial_{i_{M'}} \dots \partial_{i_1} A$ is 1-standard, non-degenerate, and that* $\deg(\partial_{i_M} \dots \partial_{i_1} A) = 1$.[22]

Proof. The proof is routine and almost identical to [11, Theorem 4.2]; the additional requirements "that for all $1 \leq M' \leq M$, $\partial_{i_{M'-1}} \dots \partial_{i_1} A \to \partial_{i_{M'}} \dots \partial_{i_1} A$ is 1-inherited, $\partial_{i_{M'}} \dots \partial_{i_1} A$ is a non-degenerate and 1-standard" follows directly from the proof (see also [10, Theorem 4.6], and its footnote). □

3.2. **Coefficient tracking.** While Lemma 3.5 asserts that one can always transform a PET-tuple A into a new PET-tuple $\partial_{i_{M'}} \dots \partial_{i_1} A$ of degree 1 using the vdC operations, it provides no information on the relation between the coefficients of the polynomials in $\partial_{i_{M'}} \dots \partial_{i_1} A$ and that in the original PET-tuple A. Such information will be essential in computing the upper bound of $S(A, f, \kappa)$. To overcome this difficulty, in [10, 11], we introduced a machinery to keep track of the coefficients of relevant polynomials. In this paper, we adopt an approach similar to [10] to control the coefficients. This section generalizes the results in [10, Section 5] to variable polynomials.

Let $\mathbf{p} = (p_{N,1}, \dots, p_{N,k})_N$ denotes a non-degenerate family of vectors of variable polynomials of degree at most K. Write

$$p_{N,i}(n) = \sum_{v=0}^{K} b_{N,i,v} n^v,$$

[21] $\partial_{i_K} \dots \partial_{i_1} A$ is understood as A when $K = 0$

[22] If $\deg(A) = 1$, then one can take $M = 0$, and the claim is trivial.

where $b_{N,i,v} \in \mathbb{R}^d$. For $N \in \mathbb{N}, r \in \mathbb{Q}, v \in \mathbb{N}_0$ and $0 \leq i \leq k$, we set $0 \leq i, j \leq k,\ i \neq j$, we set

$$Q_{N,r,i,v}(\mathbf{p}) := \{r(b_{N,w,v} - b_{N,i,v})\colon 0 \leq w \leq k\}.$$

For $0 \leq i, j \leq k,\ i \neq j$, let $v_{i,j}$ be the largest integer such that $b_{N,i,v_{i,j}} \neq b_{N,j,v_{i,j}}$ and set

$$G'_{N,i,j}(\mathbf{p}) := \text{span}_{\mathbb{Q}}\{b_{N,i,v_{i,j}} - b_{N,j,v_{i,j}}\}.$$

Let $\mathbf{q} = (q_{N,1}, \dots, q_{N,\ell})_N$ denote a family variable polynomials, with

$$q_{N,i}(n; h_1, \dots, h_s) = \sum_{b,a_1,\dots,a_s \in \mathbb{N}_0, b+a_1+\cdots+a_s \leq K} u_{N,i}(b, a_1, \dots, a_s) n^b h_1^{a_1} \dots h_s^{a_s}$$

for some $K \in \mathbb{N}_0$ and some $u_{N,i}(b, a_1, \dots, a_s) \in \mathbb{R}^d$. For $b, a_1, \dots, a_s \in \mathbb{N}_0$, let

$$u_N(\mathbf{q}, b; a_1, \dots, a_s) := (u_{N,1}(b, a_1, \dots, a_s), \dots, u_{N,\ell}(b, a_1, \dots, a_s))$$

and $\mathbf{u}(\mathbf{q}, b; a_1, \dots, a_s) := (u_N(\mathbf{q}, b; a_1, \dots, a_s))_N$.

Definition (Types and symbols of level data)**.** For all $v \in \mathbb{N}_0, r \in \mathbb{Q}$, and $0 \leq i \leq k$, we say that a sequence of ℓ-tuples $\mathbf{u} = (u_{N,1}, \dots, u_{N,\ell})_N$, $u_{N,i} \in \mathbb{R}$ is of *type* $\mathbf{p}(r, i, v)$ if

$$u_{N,1}, \dots, u_{N,\ell} \in Q_{N,r,i,v}(\mathbf{p}), \text{ and } u_{N,1} = r(b_{N,1,v} - b_{N,i,v})$$

when N is sufficiently large.

Let $\mathbf{u} = (u_{N,1}, \dots, u_{N,\ell})_N$ be of type $\mathbf{p}(r, i, v)$. Suppose that

$$(u_{N,1}, \dots, u_{N,\ell}) = (r(b_{N,w_1,v} - b_{N,i,v}), \dots, r(b_{N,w_\ell,v} - b_{N,i,v})),$$

for some $0 \leq w_1, \dots, w_\ell \leq k$ for all N sufficiently large. We call $\mathbf{w} := (w_1, \dots, w_\ell)$ a *symbol* of $\mathbf{u}$. (Note that we always have $w_1 = 1$.)

Definition. Let S denote the set of all $(a, a') \in \mathbb{N}_0^2$ such that a and a' are both 0 or both different than 0. Let $\mathbf{p}, \mathbf{q}$ be polynomial families of degree at least 1. We say that $\mathbf{q}$ satisfies (P1)–(P4) with respect to $\mathbf{p}$ if its level data $\mathbf{u}(\mathbf{q}, *)$ satisfy:

(P1) For all $a_1, \dots, a_s, b \in \mathbb{N}$, there exist $r \in \mathbb{Q}, 0 \leq i \leq k, v \in \mathbb{N}_0$ such that $\mathbf{u}(\mathbf{q}, b; a_1, \dots, a_s)$ is of type $\mathbf{p}(r, i, v)$. Moreover, we may choose the type and symbol for all of $\mathbf{u}(\mathbf{q}, b; a_1, \dots, a_s)$ in a way such that (P2)–(P4) hold, where:

(P2) Suppose that $\mathbf{u}(\mathbf{q}, b; a_1, \dots, a_s)$ is of type $\mathbf{p}(r, i, v)$, then $r = \binom{b+a_1+\cdots+a_s}{b,a_1,\dots,a_s}$ and $v = b + a_1 + \cdots + a_s$ (in particular, $r \neq 0$).[23]

(P3) Suppose that $\mathbf{u}(\mathbf{q}, b; a_1, \dots, a_s)$ is of type $\mathbf{p}(r, i, v)$ and $\mathbf{u}(\mathbf{q}, b'; a'_1, \dots, a'_s)$ is of type $\mathbf{p}(r', i', v')$. If $(a_1, a'_1), \dots, (a_s, a'_s) \in S$, then $i = i'$ and $\mathbf{u}(\mathbf{q}, b; a_1, \dots, a_s)$, $\mathbf{u}(\mathbf{q}, b'; a'_1, \dots, a'_s)$ share a symbol $\mathbf{w}$.

(P4) For any $\mathbf{u}(\mathbf{q}, b; a_1, \dots, a_s)$, the first coordinate w_1 of its symbol $(w_1, \dots, w_\ell)$ equals to 1.

For convenience we say that a PET-tuple $A = (s, \ell, \mathbf{q})$ satisfies (P1)–(P4) if the polynomial family $\mathbf{q}$ associated to A satisfies (P1)–(P4).

Proposition 3.6. *Let $A = (s, \ell, \mathbf{q})$ be a non-degenerate PET-tuple and $1 \leq \rho \leq \ell$. Assume that $A \to \partial_\rho A$ is 1-inherited. If A satisfies (P1)–(P4), then $\partial_\rho A$ also satisfies (P1)–(P4).*

[23]Here $\binom{b+a_1+\cdots+a_s}{b,a_1,\dots,a_s} := \frac{(b+a_1+\cdots+a_s)!}{b!a_1!\dots a_s!}$.

Proposition 3.7. *Suppose that (P1)–(P4) hold for some non-degenerate* $\mathbf{q}$ *with respect to* $\mathbf{p}$. *Then for all* $0 \leq m \leq \ell, m \neq 1$ *and* N *sufficiently large, the group*

$$H_{N,1,m}(\mathbf{q}) := \text{span}_{\mathbb{Q}}\{u_{N,1}(\mathbf{q}, b; a_1, \dots, a_s) - u_{N,m}(\mathbf{q}, b; a_1, \dots, a_s) \colon (b, a_1, \dots, a_s) \in \mathbb{N}_0^{s+1}, b \neq 0\}$$

contains at least one of the groups $G'_{N,1,j}(\mathbf{p}), 0 \leq j \leq k, j \neq 1$.

Remark 3.8. The proofs of Propositions 3.6 and 3.7 are almost identical to Propositions 5.6 and 5.7 of [10], for $L = 1$, modulo the following differences:

- Propositions 3.6 and 3.7 are for variable polynomials while Propositions 5.6 and 5.7 of [10] are about polynomials;
- the polynomials in Propositions 3.6 and 3.7 take values in $\mathbb{R}$ while the ones in Propositions 5.6 and 5.7 of [10] take values in $\mathbb{Q}$;
- the groups $H_{N,1,m}(\mathbf{q})$ and $G'_{N,1,j}(\mathbf{p})$ defined and used in Proposition 3.7 are different from the groups $H_{1,m}(\mathbf{q})$ and $G_{1,j}(\mathbf{p})$ defined and used in [10, Proposition 5.7] (we do not intersect these group with $\mathbb{Z}^d$ here).

One can easily check that the differences mentioned above do not affect the proofs in [10], and the same arguments can be used to prove Propositions 3.6 and 3.7 without difficulty. We leave the details to the interested readers.

4. Bounding multiple ergodic averages with Host-Kra seminorms

In this section we prove the Host-Kra-type bounds that we need for our main averages. More specifically, we prove Proposition 4.1 which treats the basic, linear variable polynomial case, and Theorem 4.4, which treats the case of variable polynomials of leading coefficient 1 (i.e., the ones that we are dealing with in our study).

We start with the definition of Host-Kra seminorms, which is a fundamental tool in studying problems related to multiple averages, and they were first introduced in [25] for ergodic $\mathbb{Z}$-systems. A variation of these seminorms in the context of $\mathbb{Z}^d$-systems was introduced in [24]. As in [11], we will use a slightly more general version of these characteristic factors (see also [34] for a similar approach).

For a $\mathbb{Z}^d$-system $(X, \mathcal{B}, \mu, (T_n)_{n\in\mathbb{Z}^d})$ and a subgroup H of $\mathbb{Z}^d$, $\mathcal{I}(H)$ denotes the sub-σ-algebra of $(T_h)_{h\in H}$-invariant sets, i.e., sets $A \in \mathcal{B}$ such that $T_h A = A$ for all $h \in H$. For an invariant sub-σ-algebra $\mathcal{A}$ of $\mathcal{B}$, the measure $\mu \times_{\mathcal{A}} \mu$ denotes the *relative independent product of* μ *with itself over* $\mathcal{A}$. That is, $\mu \times_{\mathcal{A}} \mu$ is the measure defined on the product space $X \times X$ as

$$\int_{X\times X} f \otimes g \; d(\mu \times_{\mathcal{A}} \mu) = \int_X \mathbb{E}(f|\mathcal{A})\mathbb{E}(g|\mathcal{A})d\mu$$

for all $f, g \in L^\infty(\mu)$.

Let $H_1, \dots, H_k$ be subgroups of $\mathbb{Z}^d$. Define

$$\mu_{H_1} = \mu \times_{\mathcal{I}(H_1)} \mu$$

and for $k > 1$,

$$\mu_{H_1,\dots,H_k} = \mu_{H_1,\dots,H_{k-1}} \times_{\mathcal{I}(H_k^{[k-1]})} \mu_{H_1,\dots,H_{k-1}},$$

where $H_k^{[k-1]}$ denotes the subgroup of $(\mathbb{Z}^d)^{2^{k-1}}$ consisting of all the elements of the form $h_k \times \cdots \times h_k$ (2^{k-1} copies of h_k) for some $h_k \in H_k$. For $f \in L^\infty(\mu)$, its *Host-Kra seminorm* $\|\!|f|\!\|_{H_1,\dots,H_k}$ is defined by

$$\|\!|f|\!\|_{H_1,\ldots,H_k}^{2^k} := \int_{X^{[k]}} \prod_{\epsilon\in\{0,1\}^k} \mathcal{C}^{|\epsilon|} f \, d\mu_{H_1,\ldots,H_k},$$

where $X^{[k]} = X \times \cdots \times X$ (2^k copies X), $|\epsilon| = \epsilon_1 + \ldots + \epsilon_k$ and $\mathcal{C}$ is the conjugation map $f \mapsto \overline{f}$.

For convenience, we adopt a flexible way to write the Host-Kra seminorms combining the aforementioned notation. For example, if $A = \{H_1, H_2\}$, then the notation $\|\!|\cdot|\!\|_{A,H_3,H_4^{\times 2},(H_i)_{i=5,6}}$ refers to $\|\!|\cdot|\!\|_{H_1,H_2,H_3,H_4,H_4,H_5,H_6}$. For $g_1,\ldots,g_t \in \mathbb{Z}^d$, we denote $\|\!|\cdot|\!\|_{T_{g_1},\ldots,T_{g_t}}$ as $\|\!|\cdot|\!\|_{H_1,\ldots,H_t}$, where each H_i is generated by g_i.

The following proposition, which has at the beginning an argument similar to that of [15, Lemma 4.7], allows us to bound weighted multiple ergodic averages with certain linear variable polynomial iterates uniformly by Host-Kra seminorms.

Proposition 4.1. *Let $(X,\mathcal{B},\mu,(T_n)_{n\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system, $\ell \in \mathbb{N}$ and $f_1,\ldots,f_\ell \in L^\infty(\mu)$ be bounded by 1. Let $k_1,\ldots,k_\ell \in \mathbb{Z}^d$ and let $(r_{N,m})_{N,m}$ be a sequence in $\mathbb{Z}^d$. If $\ell > 1$, we have that*

$$\limsup_{N\to\infty} \sup_{|c_n|\le 1} \left\| \frac{1}{N}\sum_{n=1}^{N} c_n \prod_{m=1}^{\ell} T^{k_m n + r_{N,m}} f_m \right\|_2 \ll_\ell \|\!|f_1|\!\|_{T^{k_1},T^{k_1},T^{k_1-k_2},\ldots,T^{k_1-k_\ell}}. \tag{9}$$

Furthermore, if $\ell = 1$, then the left hand side of (9) *is bounded by* $\left\|\mathbb{E}(f_1\otimes\overline{f_1}|I(T^{k_1}\times T^{k_1}))\right\|_{L^2(\mu\times\mu)}^{1/2}$.

Remark 4.2. For Proposition 4.1 to be useful in the proof of Theorem 4.4, it is crucial that the constant that appears in (9) depends only on the number of linear iterates. We highlight the fact that this would not be the case if we had, e.g., iterates of the form $[\alpha_m n]$, for vectors α_m of non-integer coordinates (the constant would then depend on $\alpha_1,\ldots,\alpha_\ell$ too).

We need the following lemma in the proof of Proposition 4.1.

Lemma 4.3. *For any sequence $f\colon \mathbb{Z}\to\mathbb{C}$ bounded by 1, if $\lim_{N\to\infty}\mathbb{E}_{n\in[-N,N]}f(n)$ exists, then*

$$\lim_{N\to\infty}\mathbb{E}_{n\in[-N,N]}\frac{2(N+1-|n|)}{N+1}f(n) = \lim_{N\to\infty}\mathbb{E}_{n\in[-N,N]}f(n).$$

Proof. We have the following relation

$$\begin{aligned}\mathbb{E}_{n\in[-N,N]}\frac{2(N+1-|n|)}{N+1}f(n) &= \frac{2}{(2N+1)(N+1)}\sum_{n=-N}^{N}\sum_{M=|n|}^{N} f(n)\\ &= \frac{2}{(2N+1)(N+1)}\sum_{M=0}^{N}\sum_{n=-M}^{M} f(n)\\ &= \frac{2}{(2N+1)(N+1)}\sum_{M=0}^{N}(2M+1)\mathbb{E}_{n\in[-M,M]}f(n).\end{aligned}$$

Noting that $\lim_{N\to\infty}\left|\frac{2}{(2N+1)(N+1)}\sum_{M=0}^{N}(2M+1)-1\right| = 0$, the claim follows. □

Proof of Proposition 4.1. If $\ell = 1$, we let $B := \left\|\mathbb{E}(f_1\otimes\overline{f_1}|I(T^{k_1}\times T^{k_1}))\right\|_{L^2(\mu\times\mu)}^{1/2}$, while if $\ell > 1$ we set $B := \|\!|f_1|\!\|_{T^{k_1},T^{k_1},T^{k_1-k_2},\ldots,T^{k_1-k_\ell}}$. For every $N \in \mathbb{N}$, and $1 \le n \le N$, pick $|c_{N,n}| \le 1$, so that the corresponding norm in the left hand side of (9) is $1/N$ close to its supremum $\sup_{|c_n|\leq 1}$. So, it suffices to show that

$$\limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}c_{N,n}\prod_{m=1}^{\ell}T^{k_m n+r_{N,m}}f_m\right\|_2 \ll_\ell B. \tag{10}$$

To this end, it suffices to show

$$\limsup_{N\to\infty}\sup_{\|f_0\|_\infty\leq 1}\frac{1}{N}\sum_{n=1}^{N}\left|\int f_0\cdot\prod_{m=1}^{\ell}T^{k_m n+r_{N,m}}f_m\,d\mu\right| \ll_\ell B^2. \tag{11}$$

Indeed, assuming (11) and using the triangle inequality, whenever $f_{N,0}\in L^\infty(\mu)$ with $\|f_{N,0}\|_\infty\leq 1$ for $N\in\mathbb{N}$, we have

$$\limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N}c_{N,n}\int f_{N,0}\cdot\prod_{m=1}^{\ell}T^{k_m n+r_{N,m}}f_m\right|\,d\mu \ll_\ell B^2. \tag{12}$$

Using (12) with the conjugate of $\frac{1}{N}\sum_{n=1}^{N}c_{N,n}\prod_{m=1}^{\ell}T^{k_m n+r_{N,m}}f_m$ in place of $f_{N,0}$, we get (10).

We now prove (11). Given $N\in\mathbb{N}$ and $\|f_0\|_\infty\leq 1$, we have

$$\begin{aligned}\left(\frac{1}{N}\sum_{n=1}^{N}\left|\int f_0\cdot\prod_{m=1}^{\ell}T^{k_m n+r_{N,m}}f_m\,d\mu\right|\right)^2 &\leq \frac{1}{N}\sum_{n=1}^{N}\left|\int f_0\cdot\prod_{m=1}^{\ell}T^{k_m n+r_{N,m}}f_m\,d\mu\right|^2\\ &= \int F_{N,0}\cdot\frac{1}{N}\sum_{n=1}^{N}S^{k_1 n}F_1\cdot\prod_{m=2}^{\ell}S^{k_m n}F_{N,m}\,d(\mu\times\mu),\end{aligned}$$

where $S=T\times T$, $F_{N,0}=T^{-r_{N,1}}f_0\otimes T^{-r_{N,1}}\overline{f_0}$, $F_1=f_1\otimes\overline{f_1}$ and $F_{N,m}=T^{r_{N,m}-r_{N,1}}f_m\otimes T^{r_{N,m}-r_{N,1}}\overline{f_m}$. Using the Cauchy-Schwarz inequality, we can bound the latter expression by

$$\left\|\frac{1}{N}\sum_{n=1}^{N}S^{k_1 n}F_1\cdot\prod_{m=2}^{\ell}S^{k_m n}F_{N,m}\right\|_{L^2(\mu\times\mu)}. \tag{13}$$

Note that this bound is uniform for all $\|f_0\|_\infty\leq 1$. Consider the case $\ell=1$. Letting $N\to\infty$ and using the von Neumann Ergodic Theorem, we get that the limit of (13) (for $\ell=1$) as $N\to\infty$ can be bounded by

$$\|\mathbb{E}(F_1|I(S^{k_1}))\|_2=\|\mathbb{E}(f_1\otimes\overline{f_1}|I(T^{k_1}\times T^{k_1}))\|_{L^2(\mu\times\mu)}.$$

Hence, we obtain the desired conclusion for $\ell=1$.

We now consider the case $\ell\geq 2$. Let $0\leq t\leq \ell-1$, and denote $\mathbf{h}':=(\mathbf{h},h_{t+1}):=(h_1,\ldots,h_{t+1})$ and $\boldsymbol{\epsilon}':=(\boldsymbol{\epsilon},\epsilon_{t+1}):=(\epsilon_1,\ldots,\epsilon_{t+1})$. For $a_1,\ldots,a_{\ell-t}\in\mathbb{Z}^d$ and $b_1,\ldots,b_t\in\mathbb{Z}^d$, consider the quantity $\tilde{S}(t,\kappa,(a_i)_{i=1}^{\ell-t},(b_i)_{i=1}^{t})$ defined as

$$\mathbb{E}_{\mathbf{h}\in[-M,M]^t}\sup_{\|F_2\|_\infty,\ldots,\|F_{\ell-t}\|_\infty\leq 1}\left\|\mathbb{E}_{n\in[N]}\Big(S^{a_1 n}\prod_{\epsilon\in\{0,1\}^t}\prod_{i=1}^{t}S^{b_i h_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1\cdot\prod_{m=2}^{\ell-t}S^{a_m n}F_m\right\|_2^{\kappa}. \tag{14}$$

Note that $\tilde{S}(0,\kappa,(k_i)_{i=1}^{\ell},\emptyset)$[24] is a bound for (13) to the power of κ.

[24]Here we adopt the natural convention that when $t=0$, $\prod_{\epsilon\in\{0,1\}^t}\Big(\prod_{i=1}^{t}S^{b_i h_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon|}$ is the identity map.

We claim that for all $0 \leq t \leq \ell-1$, and $\kappa \in \mathbb{N}$, we have

$$\tilde{S}(t, 2\kappa, (a_i)_{i=1}^{\ell-t}, (b_i)_{i=1}^{t}) \ll_{\ell,\kappa} \frac{1}{M} + \Big(\frac{M}{N}\Big)^{\kappa} + \tilde{S}(t+1, \kappa, (a_i')_{i=1}^{\ell-t-1}, (b_i')_{i=1}^{t+1}), \tag{15}$$

where $a_i' = a_i - a_{\ell-t}$ for $1 \leq i \leq \ell-t-1$, and $b_i' = b_i$ for $1 \leq i \leq t$ and $b_{t+1}' = a_1$.

Indeed, by Lemma 3.1, $\tilde{S}(t, 2\kappa, (a_i)_{i=1}^{\ell-t}, (b_i)_{i=1}^{t})$ can be bounded by $O_{\kappa,\ell}(1)$ times

$$\mathbb{E}_{\mathbf{h}\in[-M,M]^t} \sup_{\|F_2\|_\infty,\dots,\|F_{\ell-t}\|_\infty\leq 1} \Big(\mathbb{E}_{h_{s+1}\in[-M,M]}\Big|\mathbb{E}_{n\in[N]}\Big\langle\Big(S^{a_1 n}\prod_{\epsilon\in\{0,1\}^t}\prod_{i=1}^{t} S^{b_i h_i \epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1 \cdot \prod_{m=2}^{\ell-t} S^{a_m n}F_m,$$
$$\Big(S^{a_1(n+h_{t+1})}\prod_{\epsilon\in\{0,1\}^t}\prod_{i=1}^{t} S^{b_i h_i \epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1 \cdot \prod_{m=2}^{\ell-t} S^{a_m(n+h_{t+1})}F_m\Big\rangle\Big|^{\kappa}$$
$$+\frac{1}{M} + \Big(\frac{M}{N}\Big)^{\kappa}\Big).$$

Noting that

$$\Big(S^{a_1 n}\prod_{\epsilon\in\{0,1\}^t}\prod_{i=1}^{t} S^{b_i h_i \epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1 \cdot \Big(S^{a_1(n+h_{t+1})}\prod_{\epsilon\in\{0,1\}^t}\prod_{i=1}^{t} S^{b_i h_i \epsilon_i}\Big)\mathcal{C}^{|\epsilon|}\overline{F_1}$$
$$= S^{a_1 n}\prod_{\epsilon'\in\{0,1\}^{t+1}}\Big(\prod_{i=1}^{t+1} S^{b_i h_i \epsilon_i}\Big)\mathcal{C}^{|\epsilon'|}F_1,$$

and using the invariance of the measure, the previous expression equals to

$$\mathbb{E}_{\mathbf{h}\in[-M,M]^t} \sup_{\|F_2\|_\infty,\dots,\|F_{\ell-t}\|_\infty\leq 1} \Big(\frac{1}{M} + \Big(\frac{M}{N}\Big)^{\kappa}$$
$$+\mathbb{E}_{h_{t+1}\in[-M,M]}\Big|\mathbb{E}_{n\in[N]}\Big\langle S^{a_1 n}\prod_{\epsilon'\in\{0,1\}^{t+1}}\Big(\prod_{i=1}^{t+1} S^{b_i h_i \epsilon_i}\Big)\mathcal{C}^{|\epsilon'|}F_1 \cdot \prod_{m=2}^{\ell-t-1} S^{a_m n}\Big(S^{h_{t+1}}\overline{F}_m \cdot F_m\Big),$$
$$S^{a_{\ell-t}n}\Big(S^{h_{t+1}}F_{\ell-t}\cdot\overline{F}_{\ell-t}\Big)\Big\rangle\Big|^{\kappa}\Big).$$

Composing by $S^{-a_{\ell-t}n}$, we get that the previous quantity is equal to

$$\mathbb{E}_{\mathbf{h}\in[-M,M]^t} \sup_{\|F_2\|_\infty,\dots,\|F_{\ell-t}\|_\infty\leq 1} \Big(\frac{1}{M} + \Big(\frac{M}{N}\Big)^{\kappa}$$
$$+\mathbb{E}_{h_{t+1}\in[-M,M]}\Big|\mathbb{E}_{n\in[N]}\Big\langle\Big(S^{(a_1-a_{\ell-t})n}\prod_{\epsilon'\in\{0,1\}^{t+1}}\prod_{i=1}^{t+1} S^{b_i h_i \epsilon_i}\Big)\mathcal{C}^{|\epsilon'|}F_1$$
$$\cdot\prod_{m=2}^{\ell-t-1} S^{(a_m-a_{\ell-t})n}\Big(S^{h_{t+1}}\overline{F}_m\cdot F_m\Big), S^{h_{t+1}}F_{\ell-t}\cdot\overline{F}_{\ell-t}\Big\rangle\Big|^{\kappa}\Big).$$

By the Cauchy-Schwarz inequality we can bound this expression by $O_\kappa(1)$ times

$$\begin{aligned}
&\mathbb{E}_{\mathbf{h}\in[-M,M]^t} \sup_{\|F_2\|_\infty,\dots,\|F_{\ell-t}\|_\infty\leq 1} \Big(\frac{1}{M}+\Big(\frac{M}{N}\Big)^\kappa \\
&\quad + \mathbb{E}_{h_{t+1}\in[-M,M]} \Big\|\mathbb{E}_{n\in[N]}\Big(S^{(a_1-a_{\ell-t})n} \prod_{\epsilon'\in\{0,1\}^{t+1}} \prod_{i=1}^{t+1} S^{b_i h_i \epsilon_i}\Big)\mathcal{C}^{|\epsilon'|}F_1 \\
&\qquad \cdot \prod_{m=2}^{\ell-t-1} S^{(a_m-a_{\ell-t})n}\Big(S^{h_{t+1}}\overline{F}_m\cdot F_m\Big)\Big\|_2^\kappa\Big) \\
&\leq \mathbb{E}_{\mathbf{h}'\in[-M,M]^{t+1}} \sup_{\|F_2\|_\infty,\dots,\|F_{\ell-t-1}\|_\infty\leq 1} \Big\|\mathbb{E}_{n\in[N]}\Big(S^{(a_1-a_{\ell-t})n} \prod_{\epsilon'\in\{0,1\}^{t+1}} \prod_{i=1}^{t+1} S^{b_i h_i \epsilon_i}\Big)\mathcal{C}^{|\epsilon'|}F_1 \\
&\qquad \cdot \prod_{m=2}^{\ell-t-1} S^{(a_m-a_{\ell-t})n}F_m\Big\|_2^\kappa + \frac{1}{M}+\Big(\frac{M}{N}\Big)^\kappa.
\end{aligned}$$

This proves (15).

Using the inequality (15) repeatedly, starting from $\tilde{S}(0,2^\ell,(k_i)_{i=1}^\ell,\emptyset)$, and keeping track of the coefficients of a_i, we deduce that the 2^ℓ-th power of (13) is bounded by $O_\ell(1)$ times

$$\frac{1}{M}+O_\ell\Big(\frac{M}{N}\Big)+\mathbb{E}_{\mathbf{h}\in[-M,M]^{\ell-1}}\Big\|\mathbb{E}_{n\in[N]}\Big(S^{b_\ell n}\prod_{\epsilon\in\{0,1\}^{\ell-1}}\prod_{i=1}^{\ell-1}S^{b_i h_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1\Big\|_2^2,$$

where $(b_1,\dots,b_\ell)=(k_1,k_1-k_2,\dots,k_1-k_\ell)$.

Using Lemma 3.2, we get

$$\begin{aligned}
&\mathbb{E}_{\mathbf{h}\in[-M,M]^{\ell-1}}\Big\|\mathbb{E}_{n\in[N]}\Big(S^{b_\ell n}\prod_{\epsilon\in\{0,1\}^{\ell-1}}\prod_{i=1}^{\ell-1}S^{b_i h_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1\Big\|_2^2 \\
&\leq \frac{6M}{N}+\mathbb{E}_{\mathbf{h}\in[-M,M]^{\ell-1}}\mathbb{E}_{x,y\in[M]}\Big(\mathbb{E}_{n\in[N]}\Big\langle\Big(S^{b_\ell(n+x)}\prod_{\epsilon\in\{0,1\}^{\ell-1}}\prod_{i=1}^{\ell-1}S^{b_i h_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1, \\
&\qquad \Big(S^{b_\ell(n+y)}\prod_{\epsilon\in\{0,1\}^{\ell-1}}\prod_{i=1}^{\ell-1}S^{b_i h_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1\Big\rangle\Big).
\end{aligned} \tag{16}$$

Composing by $S^{-b_\ell(n+y)}$, the last line of (16) is equal to

(17)

$$\frac{6M}{N}+\mathbb{E}_{\mathbf{h}\in[-M,M]^{\ell-1}}\mathbb{E}_{x,y\in[M]}\Big(\mathbb{E}_{n\in[N]}\Big\langle\Big(S^{b_\ell(x-y)}\prod_{\epsilon\in\{0,1\}^{\ell-1}}\prod_{i=1}^{\ell-1}S^{b_ih_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1,$$
$$\prod_{\epsilon\in\{0,1\}^{\ell-1}}\Big(\prod_{i=1}^{\ell-1}S^{b_ih_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon|}F_1\Big\rangle\Big)$$
$$=\frac{6M}{N}+\mathbb{E}_{\mathbf{h}'\in[-M,M]^{\ell}}\frac{(2M+1)(M+1-|h_\ell|)}{(M+1)^2}\cdot\int_{X\times X}\prod_{\epsilon'\in\{0,1\}^{\ell}}\Big(\prod_{i=1}^{\ell}S^{b_ih_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon'|}F_1\,d(\mu\times\mu)$$
$$=\frac{6M}{N}+O\Big(\frac{1}{M}\Big)+2\mathbb{E}_{\mathbf{h}'\in[-M,M]^{\ell}}\frac{M+1-|h_\ell|}{M+1}\cdot\int_{X\times X}\prod_{\epsilon'\in\{0,1\}^{\ell}}\Big(\prod_{i=1}^{\ell}S^{b_ih_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon'|}F_1\,d(\mu\times\mu).$$

Consider the iterated average

$$\mathbb{E}_{h_1,\dots,h_\ell\in\mathbb{Z}}\int_{X\times X}\prod_{\epsilon'\in\{0,1\}^{\ell}}\Big(\prod_{i=1}^{\ell}S^{b_ih_i\epsilon_i}\Big)\mathcal{C}^{|\epsilon'|}F_1\,d(\mu\times\mu).\tag{18}$$

Inductively, using [11, Lemma 2.4 (iii)], we have that (18) equals to $\|\!|F_1|\!\|^{2^\ell}_{S^{b_1},\dots,S^{b_\ell}}$. Using Lemma 4.3 (repeatedly for (18)), the mean ergodic theorem, and the definition of Host-Kra seminorms, we have that the last line of (17) can be bounded by

$$2\|\!|F_1|\!\|^{2^\ell}_{S^{b_1},\dots,S^{b_\ell}}+\frac{6M}{N}+O\Big(\frac{1}{M}\Big).$$

By [10, Lemma 3.4],

$$\|\!|F_1|\!\|_{S^{k_1},S^{k_1-k_2},\dots,S^{k_1-k_\ell}}=\|\!|f_1\otimes\overline{f_1}|\!\|_{S^{k_1},S^{k_1-k_2},\dots,S^{k_1-k_\ell}}\le\|\!|f_1|\!\|^2_{T^{k_1},T^{k_1},T^{k_1-k_2},\dots,T^{k_1-k_\ell}}.$$

By first letting $N\to\infty$ and then $M\to\infty$, we deduce that (12) is bounded by a constant, depending only on ℓ, times $\|\!|f_1|\!\|_{T^{k_1},T^{k_1},T^{k_1-k_2},\dots,T^{k_1-k_\ell}}$, as was to be shown. □

Theorem 4.4. *Let $\ell,\kappa\in\mathbb{N}$, $A=(0,\ell,\mathbf{p})$ be a 1-standard PET-tuple with $\mathbf{p}=(p_{N,1},\dots,p_{N,\ell})_N$ of the form*

$$p_{N,j}(n)=e_jn^K+p'_{N,j}(n),\quad 1\le j\le\ell,$$

for some $K\in\mathbb{N}$ and variable polynomials $p'_{N,j}\colon\mathbb{Z}\to\mathbb{R}^\ell$ of degree less than K, when N is sufficiently large. Let $(X,\mathcal{B},\mu,(T_n)_{n\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system and $f\in L^\infty(\mu)$. Then, there exists $D=O_{\ell,K}(1)$ such that

$$\text{if }\ \|\!|f|\!\|_{\left\{T_{e_1}^{\times D},(T_{e_1-e_j})^{\times D}\right\}_{1\le j\le\ell,j\neq1}}=0,\ \text{ then we have that }\ S(A,f,\kappa)=0.$$

Moreover, in the special case where $\ell=1$,

$$\text{if }\ \mathbb{E}(f\otimes\overline{f}\mid I((T_1\times T_1)^a))=0,\ \text{ for all } a\in\mathbb{Z}\setminus\{0\},\ \text{ then }\ S(A,f,\kappa)=0.$$ [25]

[25] In particular, if $\|\!|f|\!\|_{T_1,T_1}=0$, then we have $S(A,f,\kappa)=0$ by [10, Lemma 3.4] and [11, Lemma 2.4 (iv)].

Proof. If $\ell = 1$, then denote $A' := \partial_1 \dots \partial_1 A$ with ∂_1 repeated $K-1$ times. It is not hard to compute that $A' = (K-1, 1, (q_N)_N)$, where

$$q_N(n; h_1, \dots, h_{K-1}) = e_1 K! h_1 \dots h_{K-1} n + r_N(h_1, \dots, h_{K-1})$$

for some $r_N(h_1, \dots, h_{K-1}) \in \mathbb{R}^d$ when N is sufficiently large. By Lemma 3.4, $S(A, f, 2^{K-1}) \ll_K S(A', f, 1)$. By Proposition 4.1, and the assumption that $\mathbb{E}(f \otimes \overline{f} \mid I((T_1 \times T_1)^a)) = 0$ for all $a \neq 0$, we get that

$$\limsup_{N \to \infty} \sup_{|c_n| \leq 1} \left\| \mathbb{E}_{0 \leq n \leq N} c_n T^{[K! h_1 \dots h_{K-1} n + r_N(h_1, \dots, h_{K-1})]} f \right\|_2 = 0,$$

provided that $h_1 \dots h_{K-1} \neq 0$. So $S(A', f, 1) = 0$ since the set of $(h_1, \dots, h_{K-1})$ such that $h_1 \dots h_{K-1} = 0$ is of zero density. So $S(A, f, 2^{K-1}) = 0$, which implies that $S(A, f, \kappa) = 0$.

We now consider the case $\ell > 1$. By Lemma 3.5, there exist $r \in \mathbb{N}$ depending only on K and ℓ and $i_1, \dots, i_r \in \mathbb{N}$ such that writing, $A' := \partial_{i_r} \dots \partial_{i_1} A$, we have that $\deg(A') = 1$ and A' is 1-standard and non-degenerate, and that each step $\partial_{i_{t-1}} \dots \partial_{i_1} A \to \partial_{i_t} \dots \partial_{i_1} A$ is 1-inherited. By Lemma 3.4, in order to show that $S(A, f, 2^r) = 0$, it suffices to show that $S(A', f, 1) = 0$.

Assume that $A' = (r, \ell', (q_{N,m})_N)$. Let $\mathbf{p} := (p_N(n)e_1, \dots, p_N(n)e_d)_N$ denote the initial tuple of variable polynomials in the PET-tuple A and $\mathbf{q} = (q_{N,1}, \dots, q_{N,\ell'})_N$ denote the tuple of variable polynomials in the PET-tuple A'. Since $\mathbf{p}$ satisfies (P1)–(P4) with respect to $\mathbf{p}$, by Proposition 3.6, $\mathbf{q}$ also satisfies (P1)–(P4) with respect to $\mathbf{p}$.

Since $\deg(A') = 1$, we may assume that

$$q_{N,m}(n, h_1, \dots, h_r) = c_{N,m}(h_1, \dots, h_r)n + r_{N,m}(h_1, \dots, h_r) \tag{19}$$

for some polynomials $c_{N,m} \colon \mathbb{Z}^r \to \mathbb{R}^d$ with degree, in terms of the variables $h_1, \dots, h_r$, less than K[26] and some $r_{N,m}(h_1, \dots, h_r) \in \mathbb{R}^d$ when N is sufficiently large.

Claim. For all $1 \leq m \leq \ell'$, every $c_{N,m}$ is equal to the same polynomial $c_m \colon \mathbb{Z}^r \to \mathbb{Z}^d$ when N is sufficiently large.

Write

$$q_{N,m}(n; h_1, \dots, h_s) = \sum_{b, a_1, \dots, a_s \in \mathbb{N}_0, b + a_1 + \dots + a_s \leq K} u_{N,m}(b, a_1, \dots, a_s) n^b h_1^{a_1} \dots h_s^{a_s}$$

and

$$p_{N,m}(n) = \sum_{v \in \mathbb{N}_0, v \leq K} b_{N,m,v} n^v$$

for some $u_{N,m}(b, a_1, \dots, a_s) \in \mathbb{R}^d$, $b_{N,m,v} \in \mathbb{R}^d$ for all $1 \leq m \leq \ell'$. It suffices to show that for all $a_1, \dots, a_s \in \mathbb{N}_0$ and $1 \leq m \leq \ell'$, $u_{N,m}(1, a_1, \dots, a_s)$ equals to a same vector in $\mathbb{Z}^d$ when N is large enough.

We may assume that each $\mathbf{u}(\mathbf{q}, 1; a_1, \dots, a_s)$ is associated with a type and a symbol so that (P1)–(P4) hold. Fix any $a_1, \dots, a_s \in \mathbb{N}_0$. By (P1) and (P2), we may assume that $\mathbf{u}(\mathbf{q}, 1; a_1, \dots, a_s)$ is associated with the type (r, i, v) and symbol $(1, w_2, \dots, w_{\ell'})$, where $r = \binom{1 + a_1 + \dots + a_s}{1, a_1, \dots, a_s}$ and $v = 1 + a_1 + \dots + a_s$. If all of $u_{N,m}(\mathbf{q}, 1; a_1, \dots, a_s), 1 \leq m \leq \ell'$ are $\mathbf{0}$ when N is sufficiently large, then we are done. If not, then there exists $1 \leq m \leq \ell'$ such that

[26] Here we used the obvious fact that if A is of degree at most K in terms of all the variables n, h_i, then so is $\partial_t A$.

$u_{N,m}(\mathbf{q},1;a_1,\dots,a_s)=r(b_{N,w_m,v}-b_{N,i,v})$ is not constant $\mathbf{0}$ when N is sufficiently large. Then $w_m\neq i$ and $v\leq K$.

If $v=K$, then $u_{N,m}(\mathbf{q},1;a_1,\dots,a_s)=r(b_{N,w_m,K}-b_{N,i,K})=r(e_{w_m}-e_i)\neq \mathbb{Z}^d\backslash\{\mathbf{0}\}$ when N is sufficiently large and we are done.

If $v<K$, then setting $b'=K-v+1$, we have $b'\geq 2$. By (P3), $\mathbf{u}(\mathbf{q},b';a_1,\dots,a_s)$ is of type (r',i,v'), for some v', where $r'=\binom{v'}{b',a_1,\dots,a_s}\neq 0$. By (P2), $v'=b'+a_1,\dots,a_s=K$. Since $w_m\neq i$, we conclude that $u_{N,m}(\mathbf{q},b';a_1,\dots,a_s)=r'(b_{N,w_m,K}-b_{N,i,K})=r'(e_{w_m}-e_i)\neq\mathbf{0}$ when N is sufficiently large, a contradiction to the fact that $\deg(\mathbf{q})=1$. This completes the proof of the claim.

Denote $\mathbf{h}:=(h_1,\dots,h_r)$. By the claim, we have

$$S(A',f,1)=\overline{\mathbb{E}}_{\mathbf{h}\in\mathbb{Z}^r}\limsup_{N\to\infty}\sup_{|c_n|\leq 1}\sup_{\|g_2\|_\infty,\dots,\|g_r\|_\infty\leq 1}\left\|\mathbb{E}_{0\leq n\leq N}c_n\prod_{m=1}^{\ell'}T^{c_m(\mathbf{h})n+[r_{N,m}(\mathbf{h})]}g_m\right\|_2,$$

where $g_1=f$. By Proposition 4.1,

$$\limsup_{N\to\infty}\sup_{|c_n|\leq 1}\sup_{\|g_2\|_\infty,\dots,\|g_r\|_\infty\leq 1}\left\|\mathbb{E}_{0\leq n\leq N}c_n\prod_{m=1}^{\ell'}T^{c_m(\mathbf{h})n+[r_{N,m}(\mathbf{h})]}g_m\right\|_2$$

is bounded by $C\cdot |||f|||_{T^{c_1(\mathbf{h})},T^{c_1(\mathbf{h})},T^{c_1(\mathbf{h})-c_2(\mathbf{h})},\dots,T^{c_1(\mathbf{h})-c_\ell(\mathbf{h})}}$, where C depends only on ℓ', which can be bounded in terms of ℓ and K.[27] So $S(A',f,1)$ is bounded by

$$C\cdot\overline{\mathbb{E}}_{\mathbf{h}\in\mathbb{Z}^r}|||f_1|||_{T^{c_1(\mathbf{h})},T^{c_1(\mathbf{h})},T^{c_1(\mathbf{h})-c_2(\mathbf{h})},\dots,T^{c_1(\mathbf{h})-c_\ell(\mathbf{h})}}. \tag{20}$$

Assume that

$$c_m(\mathbf{h})=\sum_{a_1,\dots,a_r\in\mathbb{N}_0,a_1+\dots+a_r\leq K}h_1^{a_1}\dots h_r^{a_r}u_m(a_1,\dots,a_r)$$

for some $u_m(a_1,\dots,a_r)\in\mathbb{Q}^d$. Let

$$H_m\coloneqq G(u_1(a_1,\dots,a_r)-u_m(a_1,\dots,a_r)\colon a_1,\dots,a_r\in\mathbb{N}_0).$$

for $0\leq m\leq\ell,m\neq 1$. By [10, Proposition 5.2], there exists $D\in\mathbb{N}$ depending only on ℓ and K such that

$$\overline{\mathbb{E}}_{\mathbf{h}\in\mathbb{Z}^r}|||f_1|||_{T^{c_1(\mathbf{h})},T^{c_1(\mathbf{h})},T^{c_1(\mathbf{h})-c_2(\mathbf{h})},\dots,T^{c_1(\mathbf{h})-c_\ell(\mathbf{h})}}=0$$

if $|||f|||_{H_m^{\times D},0\leq m\leq\ell,m\neq 1}=0$. [28]

Since $\mathbf{q}$ satisfies (P1)–(P4) with respect to $\mathbf{p}$, by Proposition 3.7, for all $0\leq m\leq r,m\neq 1$, $H_{N,1,m}(\mathbf{q})$ contains one of $G'_{N,1,j}(\mathbf{p}),0\leq j\leq d,j\neq 1$ for all N sufficiently large. In our case,

$$H_{N,1,m}(\mathbf{q})=\mathrm{span}_{\mathbb{Q}}(u_1(a_1,\dots,a_r)-u_m(a_1,\dots,a_r)\colon a_1,\dots,a_r\in\mathbb{N}_0)$$

and

$$G'_{N,1,j}(\mathbf{p})=\mathrm{span}_{\mathbb{Q}}\{e_1-e_j\}.$$

So, each $H_m:=H_{N,1,m}(\mathbf{q})\cap\mathbb{Z}^d$ contains one of $\mathrm{span}_{\mathbb{Z}}\{e_1-e_j\},0\leq j\leq d,j\neq 1$. Hence, if $|||f|||_{T_{e_1}^{\times D'},T_{e_2-e_1}^{\times D'},\dots,T_{e_d-e_1}^{\times D'}}=0$, where $D'=Dd$, then as a consequence of [11, Lemma 2.4 (v)] we have $|||f|||_{H_m^{\times D},0\leq m\leq\ell,m\neq 1}=0$, and thus the average is 0. □

[27] We remark that this is where we crucially used the fact that Proposition 4.1 depends only on the number of linear iterates.

[28] We remark at this point that it is [10, Proposition 5.2] that crucially uses concatenation results from [35].

Remark 4.5. The reason why we cannot obtain the more general assumption $\log x \prec h(x)$ instead of $\log x \prec s_h(x)$ in Theorem 1.1 (in which case we would also cover the case where h is a polynomial), is that we cannot extend Theorem 4.4 when the leading coefficients of the $p_{N,j}$'s are equal to some non-integer α. In this case, we are able to obtain an upper bound for $S(A', f, 1)$ similar to (20), but with the constant C depending on $\mathbf{h}$. This prevent us from using [10, Proposition 5.2] or any concatenation result from [35] to get a satisfactory estimate.

5. Proof of main result

We prove Theorem 1.1 in this section. Combining the estimates in the previous section, we first provide a Host-Kra seminorm upper bound for multiple ergodic averages with non-polynomial iterates.

Theorem 5.1. *Let $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ be a system with commuting and invertible transformations, a be a function, $L \in \mathcal{H}$ a positive function with $1 \prec L(x) \prec x$, and $(p_N)_N$ a sequence of functions such that for all $N \in \mathbb{N}$ and $0 \le r \le L(N)$, we have*

$$a(N+r) = p_N(r) + e_{N,r}, \quad \text{with } e_{N,r} \ll 1.$$

Assume additionally that p_N are polynomials such that, when N is sufficiently large, $\deg(p_N) = K$, for some $K \in \mathbb{N}$, and the leading coefficient of p_N equals to $a_N := a^{(K)}(N)/K!$, where we have

$$\lim_{N\to\infty} L(N)|a_N|^{\frac{1}{K}} = \infty, \quad \lim_{N\to\infty} a_N = 0, \quad \text{and } L(N) \ll |a_N|^{-\frac{K+1}{K^2}}.$$

There exists $D \in \mathbb{N}$ depending only on K and d such that if $\|\!|f_1|\!\|_{(T_1, T_1T_2^{-1}, \dots, T_1T_d^{-1})^{\times D}} = 0$, then

$$\limsup_{N\to\infty} \left\| \mathbb{E}_{1\le n\le N} T_1^{[a(n)]} f_1 \cdot \ldots \cdot T_d^{[a(n)]} f_d \right\|_2 = 0. \tag{21}$$

Moreover, when $d = 1$, (21) holds if $\|\mathbb{E}(f_1 \otimes \overline{f_1} | I((T_1 \times T_1)^a))\|_2 = 0$ for all $a \in \mathbb{Z} \backslash \{0\}$.

(A similar result holds if f_1 is replaced by any of the $f_2, \dots, f_d$.)

We briefly explain the idea of the proof of Theorem 5.1 using the Examples 1 and 2. We have already seen (in Section 2) that for these examples, the Hardy field iterates can be approximated by variable polynomials that can be transformed in such a way that their leading coefficients are equal to 1. Then, we may use Theorem 4.4 to get the desired seminorm control.

Proof of Theorem 5.1. Since $a(N+r) = p_N(r) + e_{N,r}$, $N \in \mathbb{N}$, $0 \le r \le L(N)$, it suffices by Proposition 2.1 to show

$$\limsup_{N\to\infty} \sup_{|c_n|\le 1} \sup_{\|f_2\|_\infty, \dots, \|f_d\|_\infty \le 1} \left\| \mathbb{E}_{0\le n\le L(N)} c_n \prod_{i=1}^d T_i^{[p_N(n)]} f_i \right\|_2 = 0.$$

Since $p_N(n) = a_N n^K + p'_N(n)$, for some $(p'_N)_N$ of degree less than K and $(a_N)_N, L$ satisfy the assumptions of Proposition 2.2, it suffices to show

$$\limsup_{N\to\infty} \sup_{|c_n|\le 1} \sup_{\|f_2\|_\infty, \dots, \|f_d\|_\infty \le 1} \left\| \mathbb{E}_{0\le n\le \tilde{L}(N)} c_n \prod_{i=1}^d T_i^{[n^K + \tilde{p}_N(n)]} f_i \right\|_2 = 0 \tag{22}$$

for the appropriate $(\tilde{p}_N)_N$ of degree less than K and the positive function $\tilde{L}$ with $1 \prec \tilde{L}(x) \prec x$ given by Proposition 2.2.

If (22) fails, then there exist $\epsilon > 0$ and a subsequence $(N_j)_j$ of integers such that

$$\sup_{|c_n|\le 1} \sup_{\|f_2\|_\infty,\dots,\|f_d\|_\infty\le 1} \Big\| \mathbb{E}_{0\le n\le \tilde{L}(N_j)} c_n \prod_{i=1}^{d} T_i^{[n^K+\tilde{p}_{N_j}(n)]} f_i \Big\|_2 > \epsilon \tag{23}$$

for all $j \in \mathbb{N}$. Passing to another subsequence if necessary, we may assume without loss of generality that $M_j := [\tilde{L}(N_j)]$ is strictly increasing in j.

Let $A = (0, d, \mathbf{q})$ be the 1-standard PET-tuple given by $q_{N,i}(n) = (n^K + q_N(n))e_i, 1 \le i \le d$, $N \in \mathbb{N}$, where $q_{M_j} := \tilde{p}_{N_j}$ and $q_N := 0$ otherwise. By Theorem 4.4, for $d > 1$, $S(A, f_1, 1) = 0$ if $\|f_1\|_{(T_1, T_1T_2^{-1},\dots,T_1T_d^{-1})^{\times D}} = 0$. In the $d = 1$ case, the same theorem implies $S(A, f_1, 1) = 0$ if $\big\|\mathbb{E}(f_1 \otimes \overline{f_1} | I((T_1 \times T_1)^a))\big\|_2 = 0$ for all $a \in \mathbb{Z}\setminus\{0\}$.

By the construction of A, in both cases, we have that the left hand side of (23) converges to 0, a contradiction. This finishes the proof. □

Remark 5.2. The Hardy functions of interest satisfy the conclusion of Theorem 5.1 (for some appropriate function $L \in \mathcal{H}$ and $K \in \mathbb{N}$).

Indeed, let $h(x) = s_h(x) + p_h(x) + e_h(x)$ with $\log x \prec s_h(x)$. Since we can drop the bounded error terms (see also the expression of h via variable polynomials below), it suffices to deal with the case

$$h(x) = s_h(x) + p_h(x).$$

Let d_{p_h} be the degree of p_h and d_{s_h} be the degree of s_h. If $d_{p_h} < d_{s_h} + 1$, we set $K := d_{s_h} + 1$, while if $d_{p_h} \ge d_{s_h} + 1$, we set $K := d_{p_h} + 1$. By [36, Proposition A.2][29] we have that:

$$1 \prec |s_h^{(K)}(x)|^{-\frac{1}{K}} \prec |s_h^{(K+1)}(x)|^{-\frac{1}{K+1}} \prec x,$$

and since $s_h^{(K+1)}$ is a Hardy field function, it is (eventually) monotone.

By the previous relation, we may choose $L \in \mathcal{H}$ such that

$$1 \prec |s_h^{(K)}(x)|^{-\frac{1}{K}} \prec L(x) \prec \min\left\{|s_h^{(K+1)}(x)|^{-\frac{1}{K+1}}, |s_h^{(K)}(x)|^{-\frac{K+1}{K^2}}\right\}$$

(for example take the geometric mean of the functions appearing above).[30]

Then, by the Taylor expansion, for all $N, r \in \mathbb{N}_0$, there exists $\xi_{N,r} \in [N, N + r]$, such that

$$s_h(N + r) = s_h(N) + \cdots + \frac{s_h^{(K)}(N)}{K!} r^K + \frac{s_h^{(K+1)}(\xi_{N,r})}{(K + 1)!} r^{K+1}.$$

If $0 \le r \le L(N)$, then, for N sufficiently large, using the monotonicity of $s_h^{(K+1)}$, we have that

$$\Big|\frac{s_h^{(K+1)}(\xi_{N,r}) r^{K+1}}{(K + 1)!}\Big| \le \Big|\frac{s_h^{(K+1)}(N) r^{K+1}}{(K + 1)!}\Big| \le \Big|\frac{s_h^{(K+1)}(N) L(N)^{K+1}}{(K + 1)!}\Big| \ll 1.$$

Denoting

$$p_N(r) := p_h(N + r) + s_h(N) + \cdots + \frac{s_h^{(K)}(N)}{K!} r^K,$$

[29] We can use [36, Proposition A.2] since $\log x \prec s_h(x)$.

[30] Every two Hardy field functions are comparable, hence the minimum of the right hand side is (eventually) one of the functions.

we have

$$[h(N+r)] = [p_N(r)] + e_{N,r},$$

with $e_{N,r} \ll 1$ (here we can also absorb the initial o_h term). Since the assumptions of both Proposition 2.1 and 2.2 are satisfied for h, L, p_N and $(s_h^{(K)}(N)/K!)_N, L, K$ (noticing that for all $k \geq K$ we have $h^{(k)} = s_h^{(k)}$) respectively, the function h satisfies the conclusion of Theorem 5.1.

We will now show that conditions (i) and (ii) of Theorem 1.1 are implied by the joint ergodicity of $(T_1^{[h(n)]})_n, \dots, (T_d^{[h(n)]})_n$.

Proof of the necessity of conditions (i) and (ii) in Theorem 1.1. To show (i), for any $1 \leq i, j \leq d$, $i \neq j$, setting $f_k = 1$ for $k \neq \{i, j\}$, and since strong convergence implies weak convergence, we see that

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} \int_X (T_i T_j^{-1})^{[h(n)]} f_i \cdot f_j \; d\mu = \int_X f_i \; d\mu \int_X f_j \; d\mu$$

for all $f_i, f_j \in L^\infty(\mu)$. Thus, $((T_i T_j^{-1})^{[h(n)]})_n$ is an ergodic sequence, as desired.

To prove (ii), it suffices to show that for any $f \in L^\infty(\mu^{\otimes d})$

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} (T_1 \times \dots \times T_d)^{[h(n)]} f = \int_X f \; d\mu^{\otimes d}, \tag{24}$$

where convergence takes place in $L^2(\mu^{\otimes d})$. By a standard linearity and density argument, it suffices to prove (24) for the case $f = f_1 \otimes \dots \otimes f_d$ for some $f_1, \dots, f_d \in L^\infty(\mu)$.

We claim that both sides of (24) are equal to 0 if $\|f_i\|_{T_i,T_i} = 0$ for some $1 \leq i \leq d$. Assume that $\|f_i\|_{T_i,T_i} = 0$. By [11, Lemma 2.4 (iv)], $\|f_i\|_{T_i^a,T_i^a} = 0$ for all $a \neq 0$. By the proof of [11, Lemma 5.2], this implies that $\mathbb{E}(f \otimes \overline{f} | I((S \times S)^a))\|_2 = 0$. Since h satisfies the conclusion of Theorem 5.1, we have that the left hand side of (24) is 0.

On the other hand, $\|f_i\|_{T_i,T_i} = 0$, implies that

$$\left(\int_X |\mathbb{E}(f_i | I(T_i))|^2 d\mu \right)^{1/2} = \|f_i\|_{T_i} \leq \|f_i\|_{T_i,T_i} = 0,$$

which in turn implies that $\int_X f_i \, d\mu = \int_X \mathbb{E}(f_i | I(T_i)) \, d\mu = 0$; thus the right hand side of (24) is 0.

Therefore, it suffices to prove (24) under the assumption that each f_i is measurable with respect to Z_{T_i,T_i}, the sub σ-algebra of $\mathcal{B}$ such that $\|f\|_{T_i,T_i} = 0 \Leftrightarrow \mathbb{E}(f | Z_{T_i,T_i}) = 0$ for all $f \in L^\infty(\mu)$.

Since $(T_1^{[h(n)]})_n, \dots, (T_d^{[h(n)]})_n$ are jointly ergodic, by projecting to each coordinate, we have that T_i is ergodic for μ for all $1 \leq i \leq d$. By [11, Lemma 2.7], we may approximate each f_i by finite linear combinations of eigenfunctions of T_i. So, we may assume that for each f_i we have $T_i f_i = \lambda_i f_i$, for some $\lambda_i \in \mathbb{S}^1$. If one of $f_1, \dots, f_d$ is 0 μ-a.e., then (24) holds trivially. Suppose now that none of $f_1, \dots, f_d$ is 0 a.e.. Since T_i is ergodic for each i, it follows that we may assume that $|f_i| = 1$ μ-a.e., for each i. If all the f_i's are constant, (24) holds trivially. If not, say f_{i_0} is not a constant, then $\lambda_{i_0} \neq 1$ by the ergodicity of T_{i_0}. So $\int_X f_{i_0} \; d\mu = 0$, and thus the right hand side of (24) is 0. Consequently, we have reduced matters to showing that

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} (\lambda_1 \cdots \lambda_d)^{[h(n)]} = 0.$$

This follows directly by the joint ergodicity assumption applied to the eigenfunctions $f_1, \dots, f_d$ described above, completing the proof. □

In order to show that conditions (i) and (ii) in Theorem 1.1 are sufficient for joint ergodicity, we use a criterion, first introduced by Frantzikinakis [14] and then generalized by Best and Ferré Moragues [7].

Definition ([7])**.** We say that a collection of mappings $a_1,\dots,a_k\colon \mathbb{Z}^d\to\mathbb{Z}^d$ is:

(a) *good for seminorm estimates for the system* $(X,\mathcal{B},\mu,(T_n)_{n\in\mathbb{Z}^d})$, if there exists $M\in\mathbb{N}$ such that if $f_1,\dots,f_k\in L^\infty(\mu)$ and $\|\!|f_\ell|\!\|_{(\mathbb{Z}^d)^{\times M}}=0$ for some $\ell\in\{1,\dots,k\}$, then

$$\lim_{N\to\infty}\frac{1}{N^d}\sum_{n\in[N]^d}\prod_{i=1}^{k}T_{a_i(n)}f_i=0,$$

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

(b) *good for equidistribution for the system* $(X,\mathcal{B},\mu,(T_n)_{n\in\mathbb{Z}^d})$, if for every $\alpha_1,\dots,\alpha_k\in \operatorname{Spec}\left((T_n)_{n\in\mathbb{Z}^d}\right)$, not all of them being trivial, we have

$$\lim_{N\to\infty}\frac{1}{N^d}\sum_{n\in[N]^d}\exp(\alpha_1(a_1(n))+\cdots+\alpha_k(a_k(n)))=0,$$

where $\exp(x)\coloneqq e^{2\pi i x}$ for all $x\in\mathbb{R}$, and

$$\operatorname{Spec}\left((T_n)_{n\in\mathbb{Z}^d}\right)\coloneqq\{\alpha\in\operatorname{Hom}(\mathbb{Z}^d,\mathbb{T})\colon T_nf=\exp(\alpha(n))f,\ n\in\mathbb{Z}^d,\text{ for some non-zero } f\in L^2(\mu)\}.$$

It was shown in [7, Theorem 1.1] that for an ergodic system $(X,\mathcal{B},\mu,T_1,\dots,T_d)$ (meaning that the group action generated by $T_1,\dots,T_d$ is ergodic), a collection of mappings $a_1,\dots,a_k\colon\mathbb{Z}^d\to\mathbb{Z}^d$, $(T_{a_1(n)})_n,\dots,(T_{a_k(n)})_n$ are jointly ergodic for μ[31] if, and only if, they are good for seminorm estimates and good for equidistribution for the system.

Proof of the sufficiency of conditions (i) and (ii) in Theorem 1.1. Fix any system $(X,\mathcal{B},\mu,T_1,\dots,T_d)$ that satisfies conditions (i) and (ii). We use $(X,\mathcal{B},\mu,(T_n)_{n\in\mathbb{Z}^d})$ to denote the $\mathbb{Z}^d$-system with $T_{e_i}\coloneqq T_i$ for $1\le i\le d$. Our goal is to use [7, Theorem 1.1] to show the desired joint ergodicity. To do so, we will take $h_i(n_1,\dots,n_d)\coloneqq e_ih(n_1)$ for all $(n_1,\dots,n_d)\in\mathbb{N}^d$, where e_i is the i-th canonical vector, (i.e., the h_i's depend only on the first coordinate of n). First, note that (i) and (ii) imply that $T_i, T_iT_j^{-1}$ are ergodic for all $1\le i,j\le d, i\neq j$, which also implies that our system is ergodic. By [7, Theorem 1.1], it suffices to show that for the system $(X,\mathcal{B},\mu,(T_n)_{n\in\mathbb{Z}^d})$, the mappings $h_1,\dots,h_d$ are good for seminorm estimates and good for equidistribution.

The fact that $h_1,\dots,h_d$ are good for seminorm estimates can be argued as follows: note that

$$\lim_{N\to\infty}\frac{1}{N^d}\sum_{n\in[N]^d}\prod_{i=1}^{d}T_{h_i(n)}f_i=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\prod_{i=1}^{d}T_i^{h(n)}f_i.$$

Since h satisfies the conclusion of Theorem 5.1, there exists $D\in\mathbb{N}$ depending only on d and the degree of $h(n)$ such that for all $i\in\{1,\dots,d\}$,

$$\|\!|f_i|\!\|_{(T_i,(T_iT_j^{-1})_{j\neq i})^{\times D}}=0\ \text{ implies that }\ \lim_{N\to\infty}\frac{1}{N^d}\sum_{n\in[N]^d}\prod_{i=1}^{d}T_{h_i(n)}f_i=0.$$

[31] Here we mean that, for all bounded f_i's, we have $\lim_{N\to\infty}\frac{1}{N^d}\sum_{n\in[N]^d}T_{a_1(n)}f_1\cdot\ldots\cdot T_{a_k(n)}f_k=\prod_{i=1}^k\int f_i\,d\mu$.

By [11, Corollary 2.5], since $T_i, T_iT_j^{-1}$ are ergodic for all $1 \leq i, j \leq d, i \neq j$, we get that $\|\!|f_i|\!\|_{(T_i,(T_iT_j^{-1})_{j\neq i})^{\times D}} = 0$ if, and only if, $\|\!|f_i|\!\|_{(\mathbb{Z}^d)^{\times dD}} = 0$. From this we have the good for seminorm estimates condition.

Thus, it only remains to show that the collection of $h_1, \dots, h_d$ is good for equidistribution. Suppose, for the sake of contradiction, that $h_1, \dots, h_d$ is not good for equidistribution. Then, there exist $\alpha_1, \dots, \alpha_d \in \text{Spec}\,((T_n)_{n\in\mathbb{Z}^d})$, not all of them trivial, and a subsequence $(N_j)_{j\in\mathbb{N}}$ of $\mathbb{N}$, such that

$$\lim_{j\to\infty} \frac{1}{N_j^d} \sum_{n\in[N_j]^d} \exp(\alpha_1(h_1(n)) + \cdots + \alpha_d(h_d(n))) \text{ exists and equals to } c, \tag{25}$$

for some $c \neq 0$. For $1 \leq i \leq d$, since $\alpha_i \in \text{Spec}\,((T_n)_{n\in\mathbb{Z}^d})$, there exists some nonzero $f_i \in L^2(\mu)$ such that $T_n f_i = \exp(\alpha_i(n)) f_i$ for all $n \in \mathbb{Z}^d$. Since $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ is ergodic, we have that $|f_i|$ is a non-zero constant μ-a.e.. Using (25), we have

$$\lim_{j\to\infty} \frac{1}{N_j} \sum_{n=1}^{N_j} \bigotimes_{i=1}^d T_i^{h(n)} f_i = \lim_{j\to\infty} \frac{1}{N_j^d} \sum_{n\in[N_j]^d} \bigotimes_{i=1}^d T_{h_i(n)} f_i$$
$$= \lim_{N\to\infty} \frac{1}{N_j^d} \sum_{n\in[N_j]^d} \bigotimes_{i=1}^d \exp(\alpha_i(h_i(n))) f_i = c \bigotimes_{i=1}^d f_i \not\equiv 0.$$

On the other hand, since at least one of $\alpha_1, \dots, \alpha_d$ is non-trivial, we have that $\int_{X^d} \bigotimes_{i=1}^d f_i \, d\mu^{\otimes d} = \prod_{i=1}^d \int_X f_i \, d\mu = 0$, which contradicts condition (ii). Therefore, $h_1, \dots, h_d$ are good for equidistribution. $\square$

6. An application of the method to more general iterates

In this section, we extend Theorem 1.1 to a wider class of functions.

Definition (Tempered functions)**.** Let $i \in \mathbb{N}_0$. A real-valued function t which is $(i+1)$-times continuously differentiable on (x_0, ∞) for some $x_0 \geq 0$, is called a *tempered function of degree i* (we write $d_t = i$), if the following hold:

(1) $t^{(i+1)}(x)$ tends monotonically to 0 as $x \to \infty$;
(2) $\lim_{x\to\infty} x|t^{(i+1)}(x)| = \infty$.

Tempered functions of degree 0 are called *Fejér functions.*

A big difference between Hardy field functions and tempered functions is that in the latter class, limits of ratios may not exist. In order to avoid various problematic cases, we will restrict our study to the following subclass of tempered functions (see [3], [31]):

Let $\mathcal{R} := \Big\{ g \in C^\infty(\mathbb{R}^+) : \lim_{x\to\infty} \frac{xg^{(i+1)}(x)}{g^{(i)}(x)} \in \mathbb{R} \text{ for all } i \in \mathbb{N}_0 \Big\}$;
$\mathcal{T}_i := \Big\{ g \in \mathcal{R} : \exists\, i < \alpha \leq i+1,\ \lim_{x\to\infty} \frac{xg'(x)}{g(x)} = \alpha,\ \lim_{x\to\infty} g^{(i+1)}(x) = 0 \Big\}$;
and $\mathcal{T} := \bigcup_{i=0}^\infty \mathcal{T}_i$.

It is known that every function $t \in \mathcal{T}_i$ is a tempered function of degree $d_t = i$ and satisfies the growth conditions: $x^i \log x \prec t(x) \prec x^{i+1}$ (see [3]).

We will show that our method applies to more general iterates. In particular, we will deal with functions of the form $a=h+ct$, where $c\in\mathbb{R}$, $h=s_h+p_h+e_h\in\mathcal{H}$, a Hardy field function of polynomial growth, and $t\in\mathcal{T}$, a tempered function with $\max\{\log, ct\}\prec s_h$.

As the $c=0$ case has been addressed in Theorem 5.1, we will assume without loss of generality that $c=1$ (notice that, for $c\neq 0$, the condition $\max\{\log, ct\}\prec s_h$ becomes $t\prec s_h$).

Theorem 6.1. *Let $(X,\mathcal{B},\mu,T_1,\dots,T_d)$ be a system with commuting and invertible transformations, and a be a function of the form $a=h+t$, where $h\in\mathcal{H}$, a Hardy field function of polynomial growth that satisfies $\lim_{x\to\infty}\frac{xs_h^{(d_{s_h}+1)}(x)}{s_h^{(d_{s_h})}(x)}\neq 0$ and $\lim_{x\to\infty}\frac{xs_h^{(d_{s_h}+2)}(x)}{s_h^{(d_{s_h}+1)}(x)}\neq 0$,[32] and $t\in\mathcal{T}$, a tempered function, with $t\prec s_h$. Then $(T_1^{[a(n)]})_n,\dots,(T_d^{[a(n)]})_n$ are jointly ergodic for μ if, and only if, both of the following conditions are satisfied:*

(i) *$((T_iT_j^{-1})^{[a(n)]})_n$ is ergodic for μ for all $1\leq i,j\leq d$, $i\neq j$; and*
(ii) *$((T_1\times\cdots\times T_d)^{[a(n)]})_n$ is ergodic for $\mu^{\otimes d}$.*

The following is an example of a function a that is covered by Theorem 6.1 but not covered by Theorem 1.1:

$$a(x)=x^{\pi}/\log x+x^{1/2}(2+\cos\sqrt{\log x}).$$ [33]

As with Conjecture 1, we actually expect a way more general result to hold.

Conjecture 2. *Let $(X,\mathcal{B},\mu,T_1,\dots,T_d)$ be a system and $\log\prec a_i=c_i^1h_i+c_i^2t_i$, $1\leq i\leq d$, where, for each $1\leq i\leq d$, we have $(c_i^1,c_i^2)\in\mathbb{R}^2\setminus\{(0,0)\}$, $h_i\in\mathcal{H}$ and $t_i\in\mathcal{T}$ such that h_i and t_i have different polynomial growth rates, satisfying some "standard assumptions".[34] Then $(T_1^{[a_1(n)]})_n,\dots,(T_d^{[a_d(n)]})_n$ are jointly ergodic for μ if, and only if, both of the following conditions are satisfied:*

(i) *$\left(T_i^{[a_i(n)]}T_j^{-[a_j(n)]}\right)_n$ is ergodic for μ for all $1\leq i,j\leq d$, $i\neq j$; and*
(ii) *$\left(T_1^{[a_1(n)]}\times\cdots\times T_d^{[a_d(n)]}\right)_n$ is ergodic for $\mu^{\otimes d}$.*

Remark 6.2. The reason that we don't even have Conjecture 2 for the case $a_1=\ldots=a_d$ is that, when $h\prec t$, or even when $s_h\prec t$, we cannot guarantee that the function L that we pick belongs to $\mathcal{H}$, so we cannot apply Proposition 2.1. We believe though that this is a local obstruction that should be able to be lifted. Similarly, we cannot have the case where $t\prec h$ (instead of $t\prec s_h$ that we are assuming in Theorem 6.1), since, as we mentioned in Remark 4.5, our method cannot treat integer parts of general real polynomial iterates.

To show Theorem 6.1, following the proof of Theorem 1.1, we have to show that the function a satisfies the conclusion of Theorem 5.1 (for some appropriate function L).

Dropping the $e_h(x)$ term, setting $b(x):=s_h(x)+t(x)$, it suffices to deal with functions of the form $a(x)=b(x)+p_h(x)$.

---

[32] These cannot be simultaneously 0, but if $s_{h_1}(x):=\log^2 x\prec s_{h_2}(x):=x/\log x$, then we have $d_{s_{h_1}}=d_{s_{h_2}}=0$ and $\lim_{x\to\infty}\frac{xs'_{h_1}(x)}{s_{h_1}(x)}=\lim_{x\to\infty}\frac{xs''_{h_2}(x)}{s'_{h_2}(x)}=0$.

[33] $x^{1/2}(2+\cos\sqrt{\log x})$ is not a Hardy field function (see [3]), so a is not Hardy as well.

[34] What we mean here are the usual additional assumptions we have to postulate on the growth rates of the functions (e.g., as the ones in Theorem 6.1) in order to avoid local obstructions.

If $d_{p_h} < d_{s_h} + 1$, we set $K := d_{s_h} + 1$, while when $d_{p_h} \geq d_{s_h} + 1$, we set $K := d_{p_h} + 1$. We have the following properties.

• $b^{(K)} \prec 1$.
We have

$$b^{(K)} = s_h^{(K)} + t^{(K)} \prec 1.$$

This is true since $t'(x) \ll \frac{t(x)}{x}$ by the definition of the set $\mathcal{R}$, so, iterating this, we get $t^{(k)}(x) \ll \frac{t(x)}{x^k}$ for all $k \in \mathbb{N}$; also, for all $k \in \mathbb{N}$, we have $s_h^{(k)}(x) \ll \frac{s_h(x)}{x^k}$ ([36, Proposition A.1]).

• $1/x^K \prec b^{(K)}(x)$.
We have already seen that $1/x^K \prec s_h^{(K)}(x)$ ([36, Proposition A.2]). So, under the additional assumptions on s_h, since

$$\frac{t^{(K)}(x)}{s_h^{(K)}(x)} = \left(\prod_{i=1}^{K} \frac{xt^{(i)}(x)}{t^{(i-1)}(x)} \cdot \frac{s_h^{(i-1)}(x)}{xs_h^{(i)}(x)}\right) \cdot \frac{t(x)}{s_h(x)} \to 0,$$

we have that

$$\frac{|b^{(K)}(x)|}{\frac{1}{x^K}} = \frac{|s_h^{(K)}(x)|}{\frac{1}{x^K}} \cdot \left(1 + \frac{t^{(K)}(x)}{s_h^{(K)}(x)}\right) \to \infty.$$

• $|b^{(K)}(x)|^{-\frac{1}{K}} \prec |b^{(K+1)}(x)|^{-\frac{1}{K+1}}$.
We have that

$$\lim_{x\to\infty} \frac{xb^{(K+1)}(x)}{b^{(K)}(x)} = \lim_{x\to\infty} \frac{\frac{xs_h^{(K+1)}(x)}{s_h^{(K)}(x)} + \frac{xt^{(K+1)}(x)}{t^{(K)}(x)} \cdot \frac{t^{(K)}(x)}{s_h^{(K)}(x)}}{1 + \frac{t^{(K)}(x)}{s_h^{(K)}(x)}} = \lim_{x\to\infty} \frac{xs_h^{(K+1)}(x)}{s_h^{(K)}(x)} \in \mathbb{R}.$$

So, using the fact that $1/x^K \prec b^{(K)}(x)$, we get

$$\left(b^{(K+1)}(x)\right)^K \ll \frac{\left(b^{(K)}(x)\right)^K}{x^K} \prec \left(b^{(K)}(x)\right)^{K+1}.$$

• Monotonicity of $b^{(K+1)}$.
This follows by the fact that

$$b^{(K+2)}(x) = s_h^{(K+2)}(x)\left(1 + \frac{t^{(K+2)}(x)}{s_h^{(K+2)}(x)}\right)$$

has (eventually) the same sign as $s_h^{(K+2)}$.

We are now in position to prove Theorem 6.1.

Proof of Theorem 6.1. It suffices to show that a satisfies the conclusion of Theorem 5.1. Using the first three properties we proved for the function b, we can choose, for the corresponding K, a function $L \in \mathcal{H}$ such that

$$1 \prec |b^{(K)}(x)|^{-\frac{1}{K}} \prec L(x) \prec \min\left\{|b^{(K+1)}(x)|^{-\frac{1}{K+1}}, |b^{(K)}(x)|^{-\frac{K+1}{K^2}}\right\}.$$

Indeed, we have $t^{(K)} \prec s_h^{(K)}$ and $t^{(K+1)} \prec s_h^{(K+1)}$, so $b^{(K)}$ and $b^{(K+1)}$ have the same growth rates as $s_h^{(K)}$ and $s_h^{(K+1)}$ respectively, as $b^{(K)}(x)/s_h^{(K)}(x) \to 1$ and $b^{(K+1)}(x)/s_h^{(K+1)}(x) \to 1$, hence, we can find $L \in \mathcal{H}$ (see the comments after the proof of Theorem 5.1) such that

$$1 \prec |s_h^{(K)}(x)|^{-\frac{1}{K}} \prec L(x) \prec \min\big\{|s_h^{(K+1)}(x)|^{-\frac{1}{K+1}}, |s_h^{(K)}(x)|^{-\frac{K+1}{K^2}}\big\}.$$

Then by the Taylor expansion, for all $N, r \in \mathbb{N}_0$, there exists $\xi_{N,r} \in [N, N+r]$ such that

$$b(N+r) = b(N) + \cdots + \frac{b^{(K)}(N)r^K}{K!} + \frac{b^{(K+1)}(\xi_{N,r})r^{K+1}}{(K+1)!}.$$

If $0 \le r \le L(N)$, then, for N sufficiently large, using the monotonicity of $b^{(K+1)}$, we have that

$$\Big|\frac{b^{(K+1)}(\xi_{N,r})r^{K+1}}{(K+1)!}\Big| \le \Big|\frac{b^{(K+1)}(N)r^{K+1}}{(K+1)!}\Big| \le \Big|\frac{b^{(K+1)}(N)L(N)^{K+1}}{(K+1)!}\Big| \ll 1.$$

Denoting by

$$p_N(r) := p(N+r) + b(N) + \cdots + \frac{b^{(K)}(N)r^K}{K!},$$

we have

$$[a(N+r)] = [p_N(r)] + e_{N,r},$$

with $e_{N,r} \ll 1$. Following the comments after the proof of Theorem 5.1, we have that a satisfies its conclusion, from where the result follows. $\square$

Acknowledgements. We mostly thank Andreu Ferré Moragues, with whom we started discussing this project. The second author wants to express his gratitude towards the Center for Mathematical Modeling of the University of Chile and Virginia Tech University for their hospitality. The third author also thanks the Center for Mathematical Modeling of the University of Chile for its hospitality.

References

[1] D. Berend and V. Bergelson. Jointly ergodic measure-preserving transformations, *Israel J. Math.* **49** (1984), no. 4, 307–314.

[2] V. Bergelson. Weakly mixing PET, *Ergodic Theory Dynam. Systems* **7** (1987), no. 3, 337–349.

[3] V. Bergelson and I. J. Håland Knutson. Weak mixing implies weak mixing of higher orders along tempered functions, *Ergodic Theory Dynam. Systems* **29** (2009), no. 5, 1375–1416.

[4] V. Bergelson, A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *Journal of AMS* **9** (1996), no. 3, 725–753.

[5] V. Bergelson, A. Leibman and Y. Son. Joint ergodicity along generalized linear functions, *Ergodic Theory Dynam. Systems* **36** (2016), no. 7, 2044–2075.

[6] 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.

[7] A. Best and A. Ferré Moragues. Polynomial ergodic averages for countable field actions, *Discrete and Continuous Dynamical Systems*, **42** 2022, no. 7, 3379–3413.

[8] M. D. Boshernitzan, Uniform distribution and Hardy fields, *J. Anal. Math.* **62** (1994), 225–240.

[9] Q. Chu, N. Frantzikinakis, B. Host. Ergodic averages of commuting transformations with distinct degree polynomial iterates. *Proc. of the London Math. Society.* (3), **102** (2011), 801–842.

[10] S. Donoso, A. Ferré Moragues, A. Koutsogiannis and W. Sun. Decomposition of multicorrelation sequences and joint ergodicity, Preprint 2021. arXiv: 2106.01058.

[11] S. Donoso, A. Koutsogiannis and W. Sun. Seminorms for multiple averages along polynomials and applications to joint ergodicity, *J. Anal. Math.* **146** (2022), no. 1, 1–64.

[12] S. Donoso, A. Koutsogiannis, W. Sun. Pointwise multiple averages for sublinear functions, *Ergodic Theory Dynam. Systems* **40** (2020), 1594–1618.

[13] N. Frantzikinakis. A multidimensional Szemerédi theorem for Hardy sequences of different growth, *Trans. Amer. Math. Soc.* **367** (2015), no. 8, 5653–5692.

[14] N. Frantzikinakis. Joint ergodicity of sequences, to appear in *Advances in Mathematics*. arXiv:2102.09967.

[15] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial growth, *J. d'Analyse Math.* **112** (2010), 79–135.

[16] N. Frantzikinakis. Some open problems on multiple ergodic averages, *Bulletin of the Hellenic Mathematical Society.* **60** (2016), 41-90.

[17] N. Frantzikinakis and B. Kuca. Joint ergodicity for commuting transformations and applications to polynomial sequences, Preprint 2022. arXiv:2207.12288.

[18] N. Frantzikinakis and B. Kuca. Seminorm control for ergodic averages with commuting transformations and pairwise dependent polynomial iterates, to appear in *Ergodic Theory Dynam. Systems.* arXiv:2209.11033.

[19] N. Frantzikinakis, M. Wierdl. A Hardy field extension of Szemerédi's theorem, *Adv. Math.* **222** (2009), 1–43.

[20] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204–256.

[21] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, NJ, 1981.

[22] B. Green, T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds, *Ann. of Math.* **175** (2012), 465–540.

[23] G. H. Hardy. Properties of Logarithmico-Exponential Functions. *Proc. London Math. Soc. (2)* **10** (1912), 54–90.

[24] B. Host, Ergodic seminorms for commuting transformations and applications, *Studia Math.* **195** (2009), no. 1, 31–49.

[25] B. Host, B. Kra. Nonconventional ergodic averages and nilmanifolds, *Ann. of Math.* **161** (2005), no. 1, 397–488.

[26] B. Host, B. Kra. *Nilpotent structures in ergodic theory.* Mathematical Surveys and Monographs, 236. American Mathematical Society, Providence, RI, 2018. X+427 pp.

[27] D. Karageorgos and A. Koutsogiannis. Integer part independent polynomial averages and applications along primes. *Studia Math.* **249** (2019), no. 3, 233–257.

[28] A. G. Khovanskii, *Fewnomials*, Translations of Mathematical Monographs, Volume 88, *American Mathematical Society*, 1991.

[29] Y. Kifer. Ergodic theorems for nonconventional arrays and an extension of the Szemerédi theorem, *Discrete Contin. Dyn. Syst.* **38** (2018), no. 6, 2687–2716.

[30] A Koutsogiannis. Multiple ergodic averages for variable polynomials, *Discrete Contin. Dyn. Syst.* **42** (2022), no. 9, 4637–4668.

[31] A Koutsogiannis. Multiple ergodic averages for tempered functions, *Discrete Contin. Dyn. Syst.* **41**(3) (2021), 1177–1205.

[32] A. Koutsogiannis. Integer part polynomial correlation sequences. *Ergodic Theory Dynam. Systems* **38** (2018), no. 4, 1525–1542.

[33] F. K. Richter. Uniform distribution in nilmanifolds along functions from a Hardy field, Preprint 2020. arXiv:2006.02028.

[34] W. Sun. Weak ergodic averages over dilated curves, *Ergodic Theory Dynam. Systems* **41** (2021), no. 2, 606–621.

[35] T. Tao and T. Ziegler. Concatenation theorems for anti-Gowers-uniform functions and Host-Kra characteristic factors, *Discrete Anal.* 2016, Paper No. 13, 60 pp.

[36] K. Tsinas. Joint ergodicity of Hardy field sequences, to appear in *Trans. Amer. Math. Soc.*. arXiv:2109.07941.

[37] M. Walsh. Norm convergence of nilpotent ergodic averages. *Annals of Mathematics* **175** (2012), no. 3, 1667–1688.

(Sebastián Donoso) Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, Universidad de Chile & IRL 2807 - CNRS, Beauchef 851, Santiago, Chile.

Email address: sdonoso@dim.uchile.cl

(Andreas Koutsogiannis) Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece
Email address: akoutsogiannis@math.auth.gr

(Wenbo Sun) Department of Mathematics, Virginia Tech, 225 Stanger Street, Blacksburg, VA, 24061, USA
Email address: swenbo@vt.edu