Seminorm control for ergodic averages with commuting transformations along pairwise dependent polynomials

NIKOS FRANTZIKINAKIS and BORYS KUCA

Department of Mathematics and Applied Mathematics, University of Crete, Voutes University Campus, Heraklion 71003, Greece
(e-mail: frantzikinakis@gmail.com, boryskuca@uoc.gr)

(Received 28 September 2022 and accepted in revised form 25 November 2022)

Abstract. We examine multiple ergodic averages of commuting transformations with polynomial iterates in which the polynomials may be pairwise dependent. In particular, we show that such averages are controlled by the Gowers–Host–Kra seminorms whenever the system satisfies some mild ergodicity assumptions. Combining this result with the general criteria for joint ergodicity established in our earlier work, we determine a necessary and sufficient condition under which such averages are jointly ergodic, in the sense that they converge in the mean to the product of integrals, or weakly jointly ergodic, in that they converge to the product of conditional expectations. As a corollary, we deduce a special case of a conjecture by Donoso, Koutsogiannis, and Sun in a stronger form.

Key words: joint ergodicity, ergodic averages, ergodic seminorms, characteristic factors, Host–Kra factors

2020 Mathematics Subject Classification: 37A05 (Primary); 37A30, 28D05 (Secondary)

Contents

1 Introduction 4075
1.1 Main results 4075
1.2 Extensions to other averaging schemes 4079
1.3 Outline of the article 4080
2 Ergodic background and definitions 4080
2.1 Basic notation 4080
2.2 Ergodic seminorms 4081
2.3 Dual functions and sequences 4085
2.4 Eigenfunctions and criterion for weak joint ergodicity 4086
3 Preliminary results 4087
4 Two motivating examples 4091
5 Formalism and general strategy for longer families 4096
5.1 The formalism behind the induction scheme 4097
1. Introduction

1.1. Main results. An important question in ergodic theory is to examine the limiting behavior of multiple ergodic averages of the form

\[ \frac{1}{N} \sum_{n=1}^{N} T_{1}^{p_1(n)} f_1 \cdots T_{\ell}^{p_{\ell}(n)} f_\ell. \] (1)

Here and throughout the paper, we consider a system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\) that is invertible commuting measure-preserving transformations \(T_1, \ldots, T_\ell\) acting on a Lebesgue probability space \((X, \mathcal{X}, \mu)\), polynomials \(p_1, \ldots, p_\ell \in \mathbb{Z}[n]\) that need not be distinct but are always assumed to have zero constant terms, and functions \(f_1, \ldots, f_\ell \in L^\infty(\mu)\).

The motivation for studying the limiting behavior of (1) comes from the proof of the multidimensional polynomial Szemerédi theorem [3], in which the averages from equation (1) are the central object of investigation.

It has been proved by Walsh [19] that the averages from equation (1) converge in \(L^2(\mu)\); however, little is known about the nature of the limit except in several special cases. In this paper, we examine the following question.

**Question 1.** When are the polynomials \(p_1, \ldots, p_\ell \in \mathbb{Z}[n]\)
(i) **jointly ergodic** for the system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\), in the sense that

\[ \lim_{N \to \infty} \left\| \frac{1}{N} \sum_{n=1}^{N} T_{1}^{p_1(n)} f_1 \cdots T_{\ell}^{p_{\ell}(n)} f_\ell - \int f_1 \, d\mu \cdots \int f_\ell \, d\mu \right\|_{L^2(\mu)} = 0 \] (2)

for all \(f_1, \ldots, f_\ell \in L^\infty(\mu)\)?

(ii) **weakly jointly ergodic** for the system, in the sense that

\[ \lim_{N \to \infty} \left\| \frac{1}{N} \sum_{n=1}^{N} T_{1}^{p_1(n)} f_1 \cdots T_{\ell}^{p_{\ell}(n)} f_\ell - \mathbb{E}(f_1 | \mathcal{I}(T_1)) \cdots \mathbb{E}(f_\ell | \mathcal{I}(T_\ell)) \right\|_{L^2(\mu)} = 0 \] (3)

for all \(f_1, \ldots, f_\ell \in L^\infty(\mu)\).
The first step in deriving the identities of equations (2) and (3) is usually to establish control over the $L^2(\mu)$ limit of equation (1) by one of the Gowers–Host–Kra seminorms constructed in [13], leading to the following question.

**Question 2.** When are the polynomials $p_1, \ldots, p_\ell \in \mathbb{Z}[n]$ good for seminorm control for the system $(X, X, \mu, T_1, \ldots, T_\ell)$, in the sense that there exists $s \in \mathbb{N}$ such that

$$
\lim_{N \to \infty} \left\| {1 \over N} \sum_{n=1}^N T_1^{p_1(n)} f_1 \cdots T_\ell^{p_\ell(n)} f_\ell \right\|_{L^2(\mu)} = 0 \quad (4)
$$

holds for all functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$ whenever $\|f_j\|_{s,T_j} = 0$ for some $j \in \{1, \ldots, \ell\}$?

Question 1(i) was originally posed by Bergelson and was motivated by a result of Berend and Bergelson [1] that covered the case of linear polynomials. It was investigated thoroughly by Donoso, Koutsogiannis, and Sun [7], as well as in a subsequent work of the three authors and Ferré-Moragues [6], in which they identified a set of sufficient (but not necessary) conditions under which Questions 1(i) and 2 can be answered affirmatively for general polynomials. Their conditions turned out to also be necessary when all the polynomials are equal.

In [11], we have addressed Questions 1 and 2 under the assumption that the polynomials $p_1, \ldots, p_\ell$ are pairwise independent, without any extra assumption on the system. Specifically, we have showed that for every family of pairwise independent polynomials $p_1, \ldots, p_\ell \in \mathbb{Z}[n]$, there exists $s \in \mathbb{N}$ such that the identity in equation (4) holds for all systems and all $L^\infty(\mu)$ functions under the stated seminorm assumptions. We then gave a necessary and sufficient spectral condition under which the identities in equations (2) and (3) hold.

In this paper, we drop the assumption of pairwise independence. We are thus interested in answering Questions 1 and 2 for averages from equation (1) in which some of the polynomial sequences $p_1, \ldots, p_\ell$ may be pairwise dependent or even identical. An example of this is the average

$$
{1 \over N} \sum_{n=1}^N T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot T_3^{n^2+n} f_3. \quad (5)
$$

The pairwise dependence of the polynomials $n^2, n^2, n^2 + n$ means that, contrary to the results in [11], we cannot establish the seminorm control described in Question 2 for all systems. Rather, we need to identify a special property of the system that makes the seminorm control possible. The needed property turns out to be the following.

**Definition.** (Good and very good ergodicity property) Let $\ell \in \mathbb{N}$ and $p_1, \ldots, p_\ell \in \mathbb{Z}[n]$. We say that the system $(X, X, \mu, T_1, \ldots, T_\ell)$ has the good ergodicity property for the polynomials $p_1, \ldots, p_\ell$ (sometimes we also say that the polynomials $p_1, \ldots, p_\ell$ have the good ergodicity property for the system $(X, X, \mu, T_1, \ldots, T_\ell)$, or the tuple $(T_j^{p_j(n)})_{j=1,\ldots,\ell}$ has the good ergodicity property), if whenever $p_i/c_i = p_j/c_j$ for some $i \neq j$ and non-zero $c_i, c_j \in \mathbb{Z}$ with $\gcd(c_i, c_j) = 1$, then $\mathcal{I}(T_i^{c_i} T_j^{-c_j}) = \mathcal{I}(T_i) \cap \mathcal{I}(T_j)$.
(we assume throughout that all the equalities and inclusions of $\sigma$-algebras hold up to null sets with respect to a given measure on the system), that is, a function cannot be invariant under $T_i^{c_i}T_j^{-c_j}$ except in the trivial case when it is simultaneously invariant under $T_i$ and $T_j$. If $T_i^{c_i}T_j^{-c_j}$ is ergodic for all the aforementioned indices $i, j$ and values $c_i, c_j$, then we say that the system has the very good ergodicity property for the polynomials $p_1, \ldots, p_\ell$.

Remark. As we work under the standing assumption that the polynomials have zero constant terms, the equality $p_i/c_i = p_j/c_j$ holds for some non-zero $c_i, c_j \in \mathbb{Z}$ precisely when the polynomials $p_i, p_j$ are linearly dependent.

For instance, the system $(X, X, \mu, T_1, T_2, T_3)$ has the good ergodicity property for the families $n^2, n^2, n^2 + n$ or $2n^2, 2n^2, n^2 + n$ if and only if the only functions invariant under $T_1T_2^{-1}$ are those invariant under $T_1$ and $T_2$, and it has the very good ergodicity property if $T_1T_2^{-1}$ is ergodic, that is, only constant functions are invariant under $T_1T_2^{-1}$.

We first address Question 2 for systems with the good ergodicity property.

**Theorem 1.1. (Seminorm control)** Let $D, \ell \in \mathbb{N}$ and $p_1, \ldots, p_\ell \in \mathbb{Z}[n]$ be polynomials of degrees at most $D$ with the good ergodicity property for $(X, X, \mu, T_1, \ldots, T_\ell)$. Then there exists $s \in \mathbb{N}$, depending only on $D, \ell$, such that for all $f_1, \ldots, f_\ell \in L^\infty(\mu)$, we have

$$\lim_{N \to \infty} \left\| \frac{1}{N} \sum_{n=1}^N T_1^{p_1(n)} f_1 \cdots T_\ell^{p_\ell(n)} f_\ell \right\|_{L^2(\mu)} = 0$$

whenever $\|f_j\|_{s,T_j} = 0$ for some $j \in \{1, \ldots, \ell\}$.

**Theorem 1.2. (Weak joint ergodicity)** The polynomials $p_1, \ldots, p_\ell \in \mathbb{Z}[n]$ are weakly jointly ergodic for the system $(X, X, \mu, T_1, \ldots, T_\ell)$ if and only if the following two conditions hold:

(i) the system has the good ergodicity property for the polynomials;

(ii) for all non-ergodic eigenfunctions $\chi_j \in \mathcal{E}(T_j)$, $j \in \{1, \ldots, \ell\}$, we have

$$\lim_{N \to \infty} \left\| \frac{1}{N} \sum_{n=1}^N T_1^{p_1(n)} \chi_1 \cdots T_\ell^{p_\ell(n)} \chi_\ell - \mathbb{E}(\chi_1|\mathcal{I}(T_1)) \cdots \mathbb{E}(\chi_\ell|\mathcal{I}(T_\ell)) \right\|_{L^2(\mu)} = 0.$$

(6)

Remark. Using terminology from [9, 11], condition (ii) equivalently states that the polynomials $p_1, \ldots, p_\ell$ are good for equidistribution for the system $(X, \mu, T_1, \ldots, T_\ell)$.

If additionally the transformations $T_1, \ldots, T_\ell$ are ergodic, we get the following result.
COROLLARY 1.3. (Joint ergodicity) The polynomials \( p_1, \ldots, p_\ell \in \mathbb{Z}[n] \) are jointly ergodic for the system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\) if and only if the following two conditions hold:

(i) all the transformations \( T_1, \ldots, T_\ell \) are ergodic and the system has the very good ergodicity property for the polynomials;

(ii) for eigenvalues \( \alpha_j \in \text{Spec}(T_j) \), \( j \in \{1, \ldots, \ell\} \), we have

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} e(\alpha_1 p_1(n) + \cdots + \alpha_\ell p_\ell(n)) = 0
\]

unless \( \alpha_1 = \cdots = \alpha_\ell = 0 \).

Theorem 1.1 and Corollary 1.3 extend Theorems 2.8 and 2.14 in [11] that cover the case of pairwise independent polynomials.

Theorem 1.2 and Corollary 1.3 can be put in the context of the following conjecture by Donoso, Koutsogiannis, and Sun (the version presented below is a special case of [7, Conjecture 1.5]) that was motivated by previous results of Berend and Bergelson [1]. In the statement that follows, we say that a sequence of commuting transformations \((T_n)_{n \in \mathbb{N}}\) on a probability space \((X, \mathcal{X}, \mu)\) is ergodic for \(\mu\) if

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} T_n f - \int f \, d\mu \to 0 \quad \text{in } L^2(\mu).
\]

Conjecture 1. The polynomials \( p_1, \ldots, p_\ell \in \mathbb{Z}[n] \) are jointly ergodic for the system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\) if only if the following two conditions are satisfied:

(i) for all distinct \( i, j \in \{1, \ldots, \ell\} \), the sequence \((T_i^{p_i(n)} T_j^{-p_j(n)})_{n \in \mathbb{N}}\) is ergodic for \(\mu\);

(ii) the sequence \((T_1^{p_1(n)} \times \cdots \times T_\ell^{p_\ell(n)})_{n \in \mathbb{N}}\) is ergodic for \(\mu \times \cdots \times \mu\).

Conjecture 1 thus lists conditions that have to be checked to verify the joint ergodicity of a family of polynomials for a system.

COROLLARY 1.4. Conjecture 1 holds.

In fact, our Theorem 1.2 is stronger than Corollary 1.4 in a number of ways. First, Theorem 1.2 gives a criterion for weak joint ergodicity, not just for joint ergodicity, meaning that the transformations \( T_1, \ldots, T_\ell \) need not be ergodic for us to be able to say anything meaningful. Second, our good ergodicity property lists strictly fewer conditions to check to verify joint ergodicity than the condition (i) in Conjecture 1. For instance, for the average from equation (5), the condition (i) in Conjecture 1 requires us to check the ergodicity of the three sequences \(((T_1 T_2^{-1})^{n^2})_{n \in \mathbb{N}}, (T_1^{n^2} T_3^{-(n^2+n)})_{n \in \mathbb{N}}, \) and \((T_2^{n^2} T_3^{-(n^2+n)})_{n \in \mathbb{N}}\). By contrast, the good ergodicity property of equation (5) holds if and only if \(\mathcal{I}(T_1 T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2)\), which is in any way a necessary condition for \(((T_1 T_2^{-1})^{n^2})_{n \in \mathbb{N}}\) to be ergodic.
Finally, we remark that the original version of Conjecture 1 from [7] is stated for more general tuples

\[(T_1^{p_1(n)} \cdots T_\ell^{p_\ell(n)}, \ldots, T_1^{p_1(n)} \cdots T_\ell^{p_\ell(n)})\]

(8)
a simple example of which would be the tuple

\[(T_1^{n^2} T_2^{n^2+n}, T_3^{n^2} T_4^{n^2+n}).\]

It is possible that an extension of our method would establish an analog of Theorem 1.1 for such averages. However, in addition to the fact that new problems arise, the technical complexity of some of our arguments in this paper is already formidable, and it would likely grow significantly if we wanted to tackle the more complicated averages from equation (8). We have therefore refrained from seeking an extension of Theorem 1.1 to averages of tuples as in equation (8), sticking instead to the simpler and arguably more natural averages from equation (1).

1.2. Extensions to other averaging schemes. Our arguments can be modified to cover multivariate polynomials and averages over arbitrary Følner sequences. (A sequence \((I_N)_{N \in \mathbb{N}}\) of finite subsets of \(\mathbb{Z}^D\) is called Følner, if \(\lim_{N \to \infty} |(I_N + h) \triangle I_N|/|I_N| = 0\) for every \(h \in \mathbb{Z}^D\).) While these modifications do not require any new ideas, they force us to introduce even more complicated notation and deal with straightforward but tedious technicalities. For this reason, we omit their proofs. We start with a generalization of Theorem 1.1.

**Theorem 1.5.** Let \(D, K, \ell \in \mathbb{N}\) be integers, \((I_N)_{N \in \mathbb{N}}\) be a Følner sequence on \(\mathbb{Z}^K\), and \(p_1, \ldots, p_\ell \in \mathbb{Z}[n]\) be polynomials of degree at most \(D\). Suppose that \(p_1, \ldots, p_\ell\) have the good ergodicity property for a system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\). Then there exists \(s \in \mathbb{N}\), depending only on \(D, K, \ell,\) such that for all \(1\)-bounded functions \(f_1, \ldots, f_\ell \in L^\infty(\mu),\) we have

\[
\lim_{N \to \infty} \| \frac{1}{|I_N|} \sum_{n \in I_N} T_1^{p_1(n)} f_1 \cdots T_\ell^{p_\ell(n)} f_\ell - E(f_1|\mathcal{I}(T_1)) \cdots E(f_\ell|\mathcal{I}(T_\ell)) \|_{L^2(\mu)} = 0
\]

whenever \(\|f_j\|_{s,T_j} = 0\) for some \(j \in \{1, \ldots, \ell\}\).

Theorem 1.5 and [11, Theorem 2.7] give the following generalization of Theorem 1.2.

**Theorem 1.6.** Let \(K, \ell \in \mathbb{N}\) be integers and \((I_N)_{N \in \mathbb{N}}\) be a Følner sequence on \(\mathbb{Z}^K\). The polynomials \(p_1, \ldots, p_\ell \in \mathbb{Z}[n]\) are weakly jointly ergodic for the system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\) along \((I_N)_{N \in \mathbb{N}}\), in the sense that

\[
\lim_{N \to \infty} \left\| \frac{1}{|I_N|} \sum_{n \in I_N} T_1^{p_1(n)} f_1 \cdots T_\ell^{p_\ell(n)} f_\ell - \mathbb{E}(f_1|\mathcal{I}(T_1)) \cdots \mathbb{E}(f_\ell|\mathcal{I}(T_\ell)) \right\|_{L^2(\mu)} = 0
\]

for all \(f_1, \ldots, f_\ell \in L^\infty(\mu),\) if and only if the following two conditions hold:

(i) the system has the good ergodicity property for the polynomials;
(ii) for all non-ergodic eigenfunctions $\chi_j \in \mathcal{E}(T_j)$, $j \in \{1, \ldots, \ell\}$, we have

$$\lim_{N \to \infty} \left\| \frac{1}{|I_N|} \sum_{n \in I_N} T_n^{p_1(n)} \chi_1 \cdots T_n^{p_\ell(n)} \chi_\ell - E(\chi_1|\mathcal{I}(T_1)) \cdots E(\chi_\ell|\mathcal{I}(T_\ell)) \right\|_{L^2(\mu)} = 0.$$ 

1.3. Outline of the article. We begin by recalling in §2 basic notions and results from ergodic theory, especially those related to the families of Gowers–Host–Kra and box seminorms, dual functions, as well as non-ergodic eigenfunctions. Next, we state in §3 preliminary technical lemmas that are used to prove our main results, most of which are variations of results from [6, 9, 11]. Having stated all preliminary definitions and lemmas, we discuss at length in §4 two baby cases of Theorem 1.1 that illustrate some of our techniques and point out the necessity of the good ergodicity property. We then proceed in §5 to discuss the formalism and the general strategy for handling longer families. In §6, we give more details of various maneuvers outlined in §5. These moves take the form of several highly technical propositions that play a crucial part in the inductive proof of Theorem 1.1. By presenting relevant examples, we also show various obstructions that need to be overcome to prove Theorem 1.1 in full generality. Section 7 is entirely devoted to the proof of Theorem 1.1: it contains an intricate induction scheme used for the proof and proofs of various intermediate results that together amount to Theorem 1.1. Lastly, in §8, we derive Theorem 1.2 and Corollaries 1.3 and 1.4.

Some of the techniques used in this paper were inspired by our earlier work in [11] where we dealt with pairwise independent polynomials $p_1, \ldots, p_\ell$. The lack of pairwise independence introduces serious additional complications. Consequently, we are forced to keep track of more information about the averages from equation (1) than in [11], particularly concerning the properties of the functions present therein and the coefficients of the polynomial iterates. The methods developed in this paper therefore differ in a number of places from the techniques employed in [11], and the argument from [11] is most emphatically not a special case of the argument presented in the current paper.

The need to have a better grip on the averages necessitates more extensive formalism than that in [11], making our argument rather hard to digest on a first reading. To compensate for this, we have included numerous examples that illustrate the main new obstacles and ideas in the proofs. The reader is invited to first go over these examples before delving into the details of the proofs.

2. Ergodic background and definitions

In this section, we present various notions from ergodic theory together with some basic results.

2.1. Basic notation. We start with explaining basic notation used throughout the paper.

The letters $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{N}$, and $\mathbb{N}_0$ stand for the set of complex numbers, real numbers, integers, positive integers, and non-negative integers. With $\mathbb{T}$, we denote the one-dimensional torus, and we often identify it with $\mathbb{R}/\mathbb{Z}$ or with $[0, 1)$. We let $[N] := \{1, \ldots, N\}$ for any $N \in \mathbb{N}$. With $\mathbb{Z}[n]$, we denote the collection of polynomials with integer coefficients.
For an element $t \in \mathbb{R}$, we let $e(t) := e^{2\pi i t}$.

If $a : \mathbb{N}^s \to \mathbb{C}$ is a bounded sequence for some $s \in \mathbb{N}$ and $A$ is a non-empty finite subset of $\mathbb{N}^s$, we let

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

We commonly use the letter $\ell$ to denote the number of transformations in our system or the number of functions in an average while the letter $s$ usually stands for the degree of ergodic seminorms. We normally write tuples of length $\ell$ in bold, e.g. $b \in \mathbb{Z}^\ell$, and we underline tuples of length $s$ (or $s + 1$, or $s - 1$) that are typically used for averaging, e.g. $h \in \mathbb{Z}^s$. For a vector $b = (b_1, \ldots, b_\ell) \in \mathbb{Z}^\ell$ and a system $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$, we let

$$
T^b := T_1^{b_1} \cdots T_\ell^{b_\ell},
$$

and we denote the $\sigma$-algebra of $T^b$ invariant functions by $\mathcal{I}(T^b)$. For $j \in [\ell]$, we set $e_j$ to be the unit vector in $\mathbb{Z}^\ell$ in the $j$th direction, and we let $e_0 = 0$, so that $T^{e_j} = T_j$ for $j \in [\ell]$ and $T^{e_0}$ is the identity transformation.

We often write $\epsilon \in \{0, 1\}^s$ for a vector of 0s and 1s of length $s$. For $\epsilon \in \{0, 1\}^s$ and $h, h' \in \mathbb{Z}^s$, we set:

- $\epsilon \cdot h := \epsilon_1 h_1 + \cdots + \epsilon_s h_s$;
- $|h| := |h_1| + \cdots + |h_s|$;
- $h^\epsilon := (h_1^\epsilon, \ldots, h_s^\epsilon)$, where $h_j^0 := h_j$ and $h_j^1 := h_j'$ for $j = 1, \ldots, s$.

We let $C_\mathbb{Z} := \mathbb{C}$ be the complex conjugate of $z \in \mathbb{C}$.

For a tuple $\eta \in \mathbb{N}_0^\ell$ and $I \subset [\ell]$, we define the restriction $\eta|_I := (\eta_i)_{i \in I}$.

2.2. Ergodic seminorms. We review some basic facts about two families of ergodic seminorms: the Gowers–Host–Kra seminorms and the box seminorms.

2.2.1. Gowers–Host–Kra seminorms. Given a system $(X, \mathcal{X}, \mu, T)$, we will use the family of ergodic seminorms $\| \cdot \|_{s, T}$, also known as Gowers–Host–Kra seminorms, which were originally introduced in [13] for ergodic systems. A detailed exposition of their basic properties can be found in [14, Ch. 8]. These seminorms are inductively defined for $f \in L^\infty(\mu)$ as follows (for convenience, we also define $\| \cdot \|_0$, which is not a seminorm):

$$
\| f \|_{0, T} := \int f \, d\mu,
$$

and for $s \in \mathbb{N}_0$, we let

$$
\| f \|_{s+1, T} := \lim_{H \to \infty} \mathbb{E}_{H \in [H]} \| \Delta_T h f \|_{s, T}^{2^s},
$$

where

$$
\Delta_T h f := f \cdot T^h \overline{f}, \quad h \in \mathbb{Z},
$$

is the multiplicative derivative of $f$ with respect to $T$. The limit can be shown to exist by successive applications of the mean ergodic theorem, and for $f \in L^\infty(\mu)$ and $s \in \mathbb{N}_0$, we have $\| f \|_{s, T} \leq \| f \|_{s+1, T}$ (see [13] or [14, Ch. 8]). It follows immediately from the
definition that
\[ \|f\|_{1,T} = \|\mathbb{E}(f|\mathcal{I}(T))\|_{L^2(\mu)}, \]
where \(\mathcal{I}(T) := \{ f \in L^2(\mu) : Tf = f \}\). We also have
\[ \|f\|_{s,T}^{2s'} = \lim_{H_1 \to \infty} \cdots \lim_{H_s \to \infty} \mathbb{E}_{h_1 \in [H_1]} \cdots \mathbb{E}_{h_s \in [H_s]} \int \Delta_{s,T;\hbar} f \, d\mu, \tag{9} \]
where for \(\hbar = (h_1, \ldots, h_s) \in \mathbb{Z}^s\), we let
\[ \Delta_{s,T;\hbar} f := \Delta_{T;h_1} \cdots \Delta_{T;h_s} f = \prod_{\epsilon \in \{0,1\}^s} C^{[\epsilon]} T^{\pm \hbar} \]
be the multiplicative derivative of \(f\) of degree \(s\) with respect to \(T\).

It can be shown that we can take any \(s' \leq s\) of the iterative limits to be simultaneous limits (that is average over \([H]\)^{s'} and let \(H \to \infty\)) without changing the value of the limit in equation (9). This was originally proved in [13] using the main structural result of [13]; a more ‘elementary’ proof can be deduced from [4, Lemma 1.12] once the convergence of the uniform Cesàro averages is known (and yet another proof can be found in [12, Lemma 1]).

For \(s' := s\), this gives the identity
\[ \|f\|_{s,T}^{2s} = \lim_{H \to \infty} \mathbb{E}_{h \in [H]} \int \Delta_{s,T;\hbar} f \, d\mu. \tag{10} \]
Moreover, for \(1 \leq s' \leq s\), we have
\[ \|f\|_{s,T}^{2s'} = \lim_{H \to \infty} \mathbb{E}_{h \in [H]^{s-s'}} \|\Delta_{s-s',T;\hbar} f\|_{s-s'}^{2s'} . \tag{11} \]

It has been established in [13] for ergodic systems and in [14, Ch. 8, Theorem 14] for general systems that the seminorms are intimately connected with a certain family of factors of the system. Specifically, for every \(s \in \mathbb{N}\), there exists a factor \(Z_s(T) \subseteq \mathcal{X}\), known as the Host–Kra factor of degree \(s\), with the property that
\[ \|f\|_{s,T} = 0 \quad \text{if and only if } f \text{ is orthogonal to } Z_{s-1}(T). \]
Equivalently, \(\|\cdot\|_{s,T}\) defines a norm on the space \(L^2(Z_{s-1}(T))\) (for a proof, see [14, Theorem 15, Ch. 9]).

2.2.2. Box seminorms. More generally, we use analogs of equation (10) defined with regards to several commuting transformations. These seminorms originally appeared in the work of Host [12]; their finitary versions are often called box seminorms, and we sometimes employ this terminology. Let \((\mathcal{X}, \mathcal{A}, \mu, T_1, \ldots, T_\ell)\) be a system. For each \(f \in L^\infty(\mu)\), \(h \in \mathbb{Z}\), and \(b \in \mathbb{Z}^\ell\), we define
\[ \Delta_{b;h} f := f \cdot T^{bh} \]
and for \(\hbar \in \mathbb{Z}^s\) and \(b_1, \ldots, b_s \in \mathbb{Z}^\ell\), we let
\[ \Delta_{b_1,\ldots,b_s;\hbar} f := \Delta_{b_1;h_1} \cdots \Delta_{b_s;h_s} f = \prod_{\epsilon \in \{0,1\}^s} C^{[\epsilon]} T^{b_1 \epsilon_1 h_1 + \cdots + b_s \epsilon_s h_s} f. \]
We let
\[ \| f \|_{b} := \int f \, d\mu \]
and
\[
\| f \|_{b_1, \ldots, b_t}^{2^{t+1}} := \lim_{H \to \infty} \lim_{h_r \to \infty} \lim_{h_1 \in [H]} \cdots \lim_{h_{t-1} \in [H]} \Delta_{b_{t+1}; h} f \|_{b_1, \ldots, b_t}^{2^t}. \tag{12}
\]
In particular, if \( b_1 = \cdots = b_t = b \), then \( \Delta_{b_1, \ldots, b_t} = \Delta_{s,Tb; b} \) and \( \| \cdot \|_{s, Tb} \). We remark that these seminorms were defined in a slightly different way in [12] and the above identities were established in [12, §2.3].

Iterating equation (12), we get the identity
\[
\| f \|_{b_1, \ldots, b_t}^{2^t} = \lim_{H_1 \to \infty} \cdots \lim_{H_{t-1} \to \infty} \lim_{h_{t-1} \in [H_{t-1}]} \cdots \lim_{h_1 \in [H_1]} \int \Delta_{b_1; h_1} \cdots \Delta_{b_t; h_t} f \, d\mu, \tag{13}
\]
which extends equation (9). In a complete analogy with the remarks made for the Gowers–Host–Kra seminorms, we have the following: using [12, Lemma 1] which implies the convergence of the uniform Cesàro averages over \( h \in \mathbb{Z}^s \) of \( \Delta_{b_1, \ldots, b_t; h} f \, d\mu \) and [4, Lemma 1.12], we get that we can take any \( s' \leq s \) of the iterative limits to be simultaneous limits (that is, average over \( [H]^{s'} \) and let \( H \to \infty \)) without changing the value of the limit in equation (13). Taking \( s' = s \) gives the identity
\[
\| f \|_{b_1, \ldots, b_t}^{2^t} = \lim_{H \to \infty} \lim_{h \in [H]} \int \Delta_{b_1, \ldots, b_t; h} f \, d\mu. \tag{14}
\]
More generally, for any \( 1 \leq s' \leq s \) and \( f \in L^\infty(\mu) \), we get the identity
\[
\| f \|_{b_1, \ldots, b_t}^{2^t} = \lim_{H \to \infty} \lim_{h \in [H]} \int \Delta_{b_{s+1}, \ldots, b_t; h} f \|_{b_1, \ldots, b_t}^{2^t}, \tag{15}
\]
which generalizes equation (11).

As a consequence of the identity in equation (14), we observe that for any permutation \( \sigma : [s] \to [s] \), we have
\[
\| f \|_{b_1, \ldots, b_t} = \| f \|_{b_{\sigma(1)}, \ldots, b_{\sigma(t)}},
\]
that is, the order of taking the vectors \( b_1, \ldots, b_t \) does not matter.

As an example of a box seminorm that is not a Gowers–Host–Kra seminorm, consider \( s = 2 \) and the vectors \( e_1 = (1, 0), e_2 = (0, 1) \), in which case
\[
\| f \|_{e_1, e_2}^4 = \lim_{H \to \infty} \lim_{h_1, h_2 \in [H]} \int f \cdot T_{1}^{h_1} T_{2}^{h_2} f \cdot T_{1}^{h_1} T_{2}^{h_2} f \, d\mu.
\]
More generally, for \( s = 2 \) and \( a = (a_1, a_2), b = (b_1, b_2) \), we have
\[
\| f \|_{a, b}^4 = \lim_{H \to \infty} \lim_{h_1, h_2 \in [H]} \int f \cdot T_{1}^{a_1 h_1} T_{2}^{a_2 h_2} T_{1}^{b_1 h_2} T_{2}^{b_2 h_1} f \cdot T_{1}^{a_1 h_1 + b_1 h_2} T_{2}^{a_2 h_2 + b_2 h_1} f \, d\mu.
\]
If the vector \( a \) repeats \( s \) times, we abbreviate it as \( a^{\times s} \), e.g.
\[
\| f \|_{a^{\times 2}, b^{\times 3}, c} = \| f \|_{a, a, b, b, b, c}.
\]
Box seminorms satisfy the following Gowers–Cauchy–Schwarz inequality [12, Proposition 2]:

$$\limsup_{H \to \infty} \mathbb{E}_{h \in [H]^r} \left| \prod_{t \in [0,1]^r} \mathcal{C}^{(c)}(T^{b_1 e_1 + \cdots + b_r e_r}) f \mu \right| \leq \prod_{s \in [0,1]^r} \|f\|_{b_1, \ldots, b_s}.$$  \hspace{1cm} (16)

(One can replace the limsup with a limit since it is known to exist.)

We frequently bound one seminorm in terms of another. An inductive application of equation (12), or alternatively a simple application of the Gowers–Cauchy–Schwarz inequality in equation (16), yields the following monotonicity property:

$$\|f\|_{b_1, \ldots, b_s} \leq \|f\|_{b_1, \ldots, b_s, b_{s+1}},$$ \hspace{1cm} (17)

a special case of which is the aforementioned bound $\|f\|_{s,T} \leq \|f\|_{s+1, T}$ for any $f \in L^\infty(\mu)$ and system $(X, \mathcal{X}, \mu, T)$.

In many of our arguments, we have to deal simultaneously both with a collection of transformations and their powers. The relevant box seminorms are compared in the following lemma.

**Lemma 2.1.** [11, Lemma 3.1] Let $\ell, s \in \mathbb{N}$, $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system, $f \in L^\infty(\mu)$ be a function, $b_1, \ldots, b_s \in \mathbb{Z}^\ell$ be vectors, and $r_1, \ldots, r_s \in \mathbb{Z}$ be non-zero. Then

$$\|f\|_{b_1, \ldots, b_s} \leq \|f\|_{r_1 b_1, \ldots, r_s b_s},$$

and if $s \geq 2$, we additionally get the bound

$$\|f\|_{r_1 b_1, \ldots, r_s b_s} \leq (r_1 \cdot \cdots \cdot r_s)^{1/2s} \|f\|_{b_1, \ldots, b_s}.$$

The following lemma allows us to compare box seminorms depending on the invariant $\sigma$-algebras of the transformations involved.

**Lemma 2.2.** Let $\ell, s \in \mathbb{N}$, $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system of commuting transformations, and $b_1, \ldots, b_s, c_1, \ldots, c_s \in \mathbb{Z}^\ell$ be vectors with the property that $\mathcal{I}(T^{b_i}) \subseteq \mathcal{I}(T^{c_i})$ for each $i \in [s]$. Then $\|f\|_{b_1, \ldots, b_s} \leq \|f\|_{c_1, \ldots, c_s}$ for each $f \in L^\infty(\mu)$.

**Proof.** We prove this by induction on $s$. For $s = 1$, we simply have

$$\|f\|_{b_1} = \mathbb{E}(f(\mathcal{I}(T^{b_1}))) \|L^{2, (\mu)} \leq \mathbb{E}(f(\mathcal{I}(T^{c_1}))) \|L^{2, (\mu)} = \|f\|_{c_1},$$

where we use the fact that $\|\mathbb{E}(f|A)\|_{L^2(\mu)} \leq \|\mathbb{E}(f|B)\|_{L^2(\mu)}$ whenever $A \subseteq B$. For $s > 1$, we use the induction formula for seminorms and the result for $s = 1$ to deduce that

$$\|f\|_{b_1, \ldots, b_s} \leq \lim_{H \to \infty} \mathbb{E}_{h \in [H]^{\ell-1}} \|\Delta_{b_1, \ldots, b_{s-1}, h} f\|_{b_s}^{2s-1} \leq \|f\|_{b_1, \ldots, b_{s-1}, e_s} = \|f\|_{c_1, \ldots, c_s}.$$
2.3. Dual functions and sequences. Let $s \in \mathbb{N}$ and $\{0, 1\}^s = \{0, 1\}^s \setminus \{0\}$. For a system $(X, \mathcal{X}, \mu, T)$ and $f \in L^\infty(\mu)$, we define

$$D_{s,T}(f) := \lim_{M \to \infty} \mathbb{E}_{m \in [M]^s} \prod_{\xi \in [0,1]^s} \mathcal{C}^{|\xi|}T^{\varepsilon \cdot m}f$$

(the limit exists in $L^2(\mu)$ by [13]). We call $D_{s,T}(f)$ the dual function of $f$ of level $s$ with respect to $T$. The name comes because of the identity

$$\|f\|_{s,T}^{2^s} = \int f \cdot D_{s,T}(f) \, d\mu,$$

a consequence of which is that the span of dual functions of degree $s$ is dense in $L^1(\mathcal{Z}_{s-1}(T))$.

Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system. Using the identities in equations (15) and (18), we get

$$\|f\|_{s',T_1,\ldots,T_\ell}^{2^{s'+1}} = \lim_{H \to \infty} \mathbb{E}_{h \in [H]^\ell} \int \Delta_{b_1,\ldots,b_{s'+1}}f \cdot D_{s',T_1,\ldots,T_\ell}(\Delta_{b_1,\ldots,b_{s'+1}}f) \, d\mu,$$

the special case of which is

$$\|f\|_{s,T_1,\ldots,T_\ell}^{2^{s+1}} = \lim_{H \to \infty} \mathbb{E}_{h \in [H]^\ell} \int \Delta_{b_1,\ldots,b_{s+1}}f \cdot \mathbb{E}(\Delta_{b_1,\ldots,b_{s+1}}f) \cdot D_{T_1,\ldots,T_\ell}(\mathbb{E}(\Delta_{b_1,\ldots,b_{s+1}}f)) \, d\mu.$$
for some (finite) linear combinations of dual functions. Applying the triangle inequality and the pigeonhole principle, we deduce that there exists $k$ for which

$$\limsup_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell - 1]} T_j^{p_j(n)} f_j \cdot \mathcal{D}(p_\ell(n)) \right\|_{L^2(\mu)} > 0,$$

where $\mathcal{D}(n)(x) := T_n^p \mathcal{D}_s \mathcal{S}_{mk} g_k(x)$ for $n \in \mathbb{N}$ and $x \in X$. This way, we essentially replace the term $T^{p_\ell(n)} f_\ell$ in the original average by the more structured piece $\mathcal{D}(p_\ell(n))$.

### 2.4. Eigenfunctions and criterion for weak joint ergodicity

Following [10], we define the notion of eigenfunctions that appears in the statement of Theorem 1.2.

**Definition.** Let $(X, \mathcal{X}, \mu, T)$ be a system, $\chi \in L^\infty(\mu)$, and $\lambda \in L^\infty(\mu)$ be a $T$-invariant function. We say that $\chi \in L^\infty(\mu)$ is a non-ergodic eigenfunction with eigenvalue $\lambda$ if

(i) $|\chi(x)|$ has value 0 or 1 for $\mu$-almost every $x \in X$ and $\lambda(x) = 0$ whenever $\chi(x) = 0$;
(ii) $T \chi = \lambda \chi$, $\mu$-almost everywhere (a.e.).

We denote the set of non-ergodic eigenfunctions with respect to $T$ by $E(T)$. For ergodic systems, a non-ergodic eigenfunction is either the zero function or a classical unit modulus eigenfunction. For general systems, each function $\chi \in E(T)$ satisfies $\chi(Tx) = e^{\phi(x)} \chi(x)$ for some $T$-invariant set $E \subseteq X$ and measurable $T$-invariant function $\phi : X \to \mathbb{T}$.

**Definition.** (Weak joint ergodicity) We say that a collection of sequences $a_1, \ldots, a_\ell : \mathbb{N} \to \mathbb{Z}$ is weakly jointly ergodic for the system $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$, if

$$\lim_{N \to \infty} \left\| \frac{1}{N} \sum_{n=1}^N T_1^{a_1(n)} f_1 \cdots T_\ell^{a_\ell(n)} f_\ell - \mathbb{E}(f_1|\mathcal{I}(T_1)) \cdots \mathbb{E}(f_\ell|\mathcal{I}(T_\ell)) \right\|_{L^2(\mu)} = 0$$

for all $f_1, \ldots, f_\ell \in L^\infty(\mu)$.

The notion of non-ergodic eigenfunction is important for us because of the following criterion for weak joint ergodicity from [11].

**Theorem 2.4.** (Criterion for weak joint ergodicity [11, Theorem 2.5]) The collection of sequences $a_1, \ldots, a_\ell : \mathbb{N} \to \mathbb{Z}$ is weakly jointly ergodic for the system $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$, if and only if the following two properties hold:

(i) there exists $s \in \mathbb{N}$ such that for every $m \in [\ell]$, we have

$$\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} T_1^{a_1(n)} f_1 \cdots T_\ell^{a_\ell(n)} f_\ell \right\|_{L^2(\mu)} = 0$$

for all $f_1, \ldots, f_\ell \in L^\infty(\mu)$ with $f_j \in E(T_j)$ for $j \in [m + 1, \ldots, \ell]$, whenever $\|f_m\|_{s,T_m} = 0$;

(ii) for all non-ergodic eigenfunctions $\chi_j \in E(T_j)$, $j \in [\ell]$, we have

$$\lim_{N \to \infty} \mathbb{E}_{n \in [N]} T_1^{a_1(n)} \chi_1 \cdots T_\ell^{a_\ell(n)} \chi_\ell = \mathbb{E}(\chi_1|\mathcal{I}(T_1)) \cdots \mathbb{E}(\chi_\ell|\mathcal{I}(T_\ell))$$

(19)
in $L^2(\mu)$. 

N. Frantzikinakis and B. Kuca
When $T_1, \ldots, T_\ell$ are ergodic, the condition in equation (19) can be restated as follows:

$$
\lim_{N \to \infty} \mathbb{E}_{n \in [N]} e(\alpha_1 a_1(n) + \cdots + \alpha_\ell a_\ell(n)) = 0
$$

for all $\alpha_j \in \text{Spec}(T_j)$, $j \in [\ell]$, not all zero. Here,

$$
\text{Spec}(T) := \{ \alpha \in \mathbb{T} : T \chi = e(\alpha) \chi \text{ for some } \chi \in \mathcal{E}(T) \}.
$$

We will apply Theorem 2.4 in the proof of Theorem 1.2. The first condition will be satisfied thanks to the stronger result proved in Theorem 1.1.

3. Preliminary results

In this section, we gather auxiliary results needed in the proof of Theorem 1.1. We start with the following simple lemma from [11, Lemma 5.2], which allows us to pass from averages of sequences $a_h-a_{h'}$ to averages of sequences $a_h$.

**Lemma 3.1.** Let $(a_h)_{h \in \mathbb{N}}$ be a sequence of non-negative real numbers. Then

$$
\mathbb{E}_{h, h' \in [H]} a_h - a_{h'} \leq \mathbb{E}_{h \in [H]} a_h
$$

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

Subsequently, we state a result that allows us to replace a function $f_m$ in the original average by a more structured averaged term $\tilde{f}_m$ that encodes the information about the original average. This idea originates in the finitary works on the polynomial Szemerédi theorem by Peluse and Prendiville [15–17], and it has been successfully applied in the ergodic theoretic setting in [5, 9, 11]. The version below differs from earlier formulations because we additionally show that if the to-be-replaced function $f_m$ is measurable with respect to some sub-$\sigma$-algebra $\mathcal{A}$, then the same can be assumed about the function that replaces it. In our applications, $\mathcal{A}$ will always be either the full $\sigma$-algebra $\mathcal{X}$ or the invariant sub-$\sigma$-algebra of some measure-preserving transformation.

**Lemma 3.2.** (Introducing averaged functions) Let $a_1, \ldots, a_\ell : \mathbb{N} \to \mathbb{Z}$ be sequences, $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system, and 1-bounded functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$ be such that

$$
\limsup_{N \to \infty} \mathbb{E}_{n \in [N]} e(\alpha_1 a_1(n) \cdot \cdots \cdot \alpha_\ell a_\ell(n)) \geq \delta
$$

for some $\delta > 0$. Let $m \in [\ell]$. Suppose moreover that $f_m$ is $\mathcal{A}$-measurable for some sub-$\sigma$-algebra $\mathcal{A} \subseteq \mathcal{X}$. Then there exist $N_k \to \infty$ and $g_k \in L^\infty(\mu)$, with $\|g_k\|_{L^\infty(\mu)} \leq 1$, $k \in \mathbb{N}$, such that for

$$
\tilde{f}_m := \lim_{k \to \infty} \mathbb{E}_{n \in [N_k]} T_{-a_m(n)} T_{j} f_j,
$$

where the limit is a weak limit, we have

$$
\limsup_{k \to \infty} \mathbb{E}_{n \in [N_k]} T_{-a_m(n)} e(\tilde{f}_m) \cdot \prod_{j \in [\ell], j \neq m} T_{-a_m(n)} T_{j} f_j \geq \delta^4.
$$
Proof. Let \( \{\tilde{N}_k\}_{k \in \mathbb{N}} \) be an increasing sequence of integers for which
\[
\left\| \mathbb{E}_{n \in [\tilde{N}_k]} T_1^{a_1(n)} f_1 \cdots T_\ell^{a_\ell(n)} f_\ell \right\|_{L^2(\mu)} \geq \delta. \tag{20}
\]
We set
\[
\tilde{g}_N := \mathbb{E}_{n \in [N]} T_1^{a_1(n)} f_1 \cdots T_\ell^{a_\ell(n)} f_\ell
\]
for every \( N \in \mathbb{N} \). The weak compactness of \( L^2(\mu) \) implies that there exists a subsequence \( (N_k)_{k \in \mathbb{N}} \) of \( (\tilde{N}_k)_{k \in \mathbb{N}} \) for which the sequence
\[
F_k := \mathbb{E}_{n \in [N_k]} T_{m}^{-a_m(n)} g_k \cdot \prod_{j \in [\ell], j \neq m} T_{m}^{-a_m(n)} T_{j}^{a_j(n)} f_j,
\]
where \( g_k := \tilde{g}_{N_k}, k \in \mathbb{N} \), converges weakly to a 1-bounded function \( \tilde{f}_m \).

We observe from equation (20) that
\[
\delta^2 \leq \int g_k \cdot \mathbb{E}_{n \in [N_k]} T_1^{a_1(n)} f_1 \cdots T_\ell^{a_\ell(n)} f_\ell \, d\mu
= \int f_m \cdot \mathbb{E}_{n \in [N_k]} T_{m}^{-a_m(n)} g_k \cdot \prod_{j \in [\ell], j \neq m} T_{m}^{-a_m(n)} T_{j}^{a_j(n)} f_j \, d\mu.
\]
Taking \( k \to \infty \), using the \( \mathcal{A} \)-measurability of the 1-bounded function \( f_m \), and applying the Cauchy–Schwarz inequality, we get
\[
\delta^2 \leq \int f_m \cdot \tilde{f}_m \, d\mu = \int f_m \cdot \mathbb{E}(\tilde{f}_m|\mathcal{A}) \, d\mu \leq \| \mathbb{E}(\tilde{f}_m|\mathcal{A}) \|_{L^2(\mu)}.
\]
Hence,
\[
\delta^4 \leq \| \mathbb{E}(\tilde{f}_m|\mathcal{A}) \|_{L^2(\mu)}^2 = \int \mathbb{E}(\tilde{f}_m|\mathcal{A}) \cdot \tilde{f}_m \, d\mu
= \lim_{k \to \infty} \int \mathbb{E}(\tilde{f}_m|\mathcal{A}) \cdot \mathbb{E}_{n \in [N_k]} T_{m}^{-a_m(n)} g_k \cdot \prod_{j \in [\ell], j \neq m} T_{m}^{-a_m(n)} T_{j}^{a_j(n)} f_j \, d\mu
= \lim_{k \to \infty} \int \tilde{g}_k \cdot \mathbb{E}_{n \in [N_k]} T_{m}^{a_m(n)} \mathbb{E}(\tilde{f}_m|\mathcal{A}) \cdot \prod_{j \in [\ell], j \neq m} T_{j}^{a_j(n)} f_j \, d\mu.
\]
An application of the Cauchy–Schwarz inequality gives the result. \( \square \)

We now present two different versions of the dual-difference interchange result that we use in our smoothing argument. While the second version in the proposition below has already been used in [11], the first one is novel since the extra information that it provides has not been required in earlier arguments.

**Proposition 3.3. (Dual-difference interchange)** Let \((X, X', \mu, T_1, \ldots, T_\ell)\) be a system, \( s, s' \in \mathbb{N}, b_1, \ldots, b_{s+1}, c \in \mathbb{Z}^\ell \) be vectors, \((f_{n,k})_{n,k \in \mathbb{N}} \subseteq L^\infty(\mu)\) be 1-bounded, and \( f \in L^\infty(\mu) \) be defined by
\[
f := \lim_{k \to \infty} \mathbb{E}_{n \in [N_k]} f_{n,k}.
\]
for some $N_k \to \infty$, where the average is assumed to converge weakly.

(i) If
\[
\|\mathbb{E}(f | \mathcal{I}(T^c))\|_{b_1, \ldots, b_{s+1}} > 0,
\]
then there exist 1-bounded functions $u_{h, h'}$, invariant under both $T^{b_{s+1}}$ and $T^c$, for which the inequality
\[
\liminf_{H \to \infty} \frac{1}{E_{h, h' \in \{H\}^s}} \limsup_{k \to \infty} \frac{1}{n \in \{N_k\}} \int \Delta_{b_1, \ldots, b_s; h - h'} f_{n,k} \cdot u_{h, h'} \, d\mu > 0
\]
holds.

(ii) If
\[
\|f\|_{b_1, \ldots, b_{s+1}} > 0,
\]
then
\[
\liminf_{H \to \infty} \frac{1}{E_{h, h' \in \{H\}^s}} \limsup_{k \to \infty} \frac{1}{n \in \{N_k\}} \int \Delta_{b_1, \ldots, b_s; h - h'} f_{n,k} \cdot \prod_{j=1}^{2^s} C_{h, h'} \, d\mu > 0
\]
for some 1-bounded dual functions $C_{h, h'}$ of $T$ of level $s'$.

For the proof of Proposition 3.3, we need the following version of the Gowers–Cauchy–Schwarz inequality from [11].

**Lemma 3.4.** (Twisted Gowers–Cauchy–Schwarz inequality) Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system, $s \in \mathbb{N}$, $b_1, \ldots, b_s \in \mathbb{Z}^\ell$ be vectors, and the functions $(f_\varepsilon)_{\varepsilon \in \{0,1\}^s}$, $(u_h)_{h \in \mathbb{N}^s} \subseteq L^\infty(\mu)$ be 1-bounded. Then for every $H \in \mathbb{N}$, we have
\[
\left| \mathbb{E}_{h \in \{H\}^s} \left[ \prod_{\varepsilon \in \{0,1\}^s} T^{b_1 \varepsilon_1 + \cdots + b_s \varepsilon_s} f_\varepsilon \cdot u_h \, d\mu \right]^{2^s} \right| \leq \mathbb{E}_{h, h' \in \{H\}^s} \left[ \prod_{\varepsilon \in \{0,1\}^s} C_{h, h'}^{\varepsilon} \right] d\mu.
\]

We also record two simple observations.

**Lemma 3.5.** Let $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ be a system, $a, b \in \mathbb{Z}^\ell$ be vectors, and $f \in L^\infty(\mu)$ be a function. Then the following two properties hold.

(i) If $f$ is invariant under $T^a T^{-b}$, then $\|f\|_{s, T^a} = \|f\|_{s, T^b}$ for any $s \in \mathbb{N}$.

(ii) If $f$ is invariant under $T^a$, then so is the function $\mathbb{E}(f | \mathcal{I}(T^b))$.

**Proof.** For part (i), we notice that
\[
\|f\|_{s, T^a} = \lim_{H \to \infty} \mathbb{E}_{h \in \{H\}^s} \left[ \prod_{\varepsilon \in \{0,1\}^s} C_{h, h'}^{\varepsilon} \right] \mathbb{E}(f | \mathcal{I}(T^b))
\]

In some cases, the average is assumed to converge weakly.
For part (ii), we use the fact that $T^a, T^b$ commute and the $T^a$-invariance of $f$ to observe that $T^a T^{hb} f = T^{hb} T^a f = T^{hb} f$. From this and the mean ergodic theorem, it follows that

$$T^a \mathbb{E}(f | \mathcal{I}(T^b)) = \lim_{H \to \infty} \mathbb{E}_{h \in [H]} T^a T^{hb} f = \lim_{H \to \infty} \mathbb{E}_{h \in [H]} T^{hb} f = \mathbb{E}(f | \mathcal{I}(T^b)),$$

and so $\mathbb{E}(f | \mathcal{I}(T^b))$ is invariant under $T^a$.

**Proof of Proposition 3.3.** Part (ii) follows from [11, Proposition 5.7], and so we only prove (i). Letting $u_h := \mathbb{E}(\Delta_{b_1,\ldots,b_s} \mathbb{E}(f | \mathcal{I}(T^s)) | \mathcal{I}(T^{b,s+1}))$, we deduce from equation (21) that

$$\lim_{H \to \infty} \mathbb{E}_{h \in [H]} \int \Delta_{b_1,\ldots,b_s} \mathbb{E}(f | \mathcal{I}(T^s)) \cdot u_h \, d\mu > 0.$$

The $T^c$-invariance of $\mathbb{E}(f | \mathcal{I}(T^c))$ implies the $T^c$-invariance of $\Delta_{b_1,\ldots,b_s} \mathbb{E}(f | \mathcal{I}(T^s))$, so the functions $u_h$ are invariant under $T^c$ by Lemma 3.5 (their $T^{b,s+1}$-invariance is trivial). Using the $T^c$-invariance of $u_h$ and the properties of conditional expectations, we deduce that

$$\lim_{H \to \infty} \mathbb{E}_{h \in [H]} \int \Delta_{b_1,\ldots,b_s} \mathbb{E}(f | \mathcal{I}(T^c)) \cdot u_h \, d\mu$$

$$= \lim_{H \to \infty} \mathbb{E}_{h \in [H]} \int \prod_{e \in \{0,1\}^s \setminus \{1\}} C_1 [T^{b_1 e_1 h_1 + \cdots + b_s e_s h_s} \mathbb{E}(f | \mathcal{I}(T^c))]$$

$$T^{b_1 h_1 + \cdots + b_s h_s} f \cdot \mathbb{E}(u_h | \mathcal{I}(T^c)) \, d\mu.$$

For $e \in \{0,1\}^s \setminus \{1\}$, we let $f_e := C_1 \mathbb{E}(f | \mathcal{I}(T^c))$. We deduce from the previous identity and the fact $f = \lim_{k \to \infty} \mathbb{E}_{n \in [N_k]} f_{n,k}$, where convergence is in the weak sense, that

$$\lim_{H \to \infty} \lim_{k \to \infty} \mathbb{E}_{n \in [N_k]} \mathbb{E}_{h \in [H]} \int \prod_{e \in \{0,1\}^s \setminus \{1\}} T^{b_1 e_1 h_1 + \cdots + b_s e_s h_s} f_e$$

$$\times T^{b_1 h_1 + \cdots + b_s h_s} f_{n,k} \cdot u_h \, d\mu > 0.$$

For fixed $k, n, H \in \mathbb{N}$, we apply Lemma 3.4 with $f_{\perp} := f_{n,k}$, obtaining

$$\lim \inf_{H \to \infty} \lim \sup_{k \to \infty} \mathbb{E}_{n \in [N_k]} \mathbb{E}_{h \in [H]} \int \Delta_{b_1,\ldots,b_s} \mathbb{E}(f_{\perp} | \mathcal{I}(T^c)) \cdot u_h \, d\mu > 0,$$

where

$$u_{h,h'} := T^{-(b_1 h'_1 + \cdots + b_s h'_s)} \left( \prod_{e \in \{0,1\}^s} C_1 [u_h e] \right).$$

The functions $u_{h,h'}$ are both $T^{b,s+1}$ and $T^c$ invariant given that each $(u_h)_{h \in \mathbb{N}}$ is and the transformations $T_1, \ldots, T_s$ commute. The result follows from the fact that the limsup of a sum is at most the sum of the limsup.

The proposition below enables a transition between qualitative and soft quantitative results. Its proof uses rather abstract functional analytic arguments and the mean
convergence result of Walsh [19]. If we instead use the mean convergence result of Zorin-Kranich [20] we can also get a variant that deals with averages over an arbitrary Følner sequence on \( \mathbb{Z}^D \).

**Proposition 3.6.** (Soft quantitative control [11, Proposition A.1]) Let \( m, \ell, s \in \mathbb{N} \) with \( m \in [\ell] \), \( p_1, \ldots, p_\ell \in \mathbb{Z}[n] \) be polynomials, and \( (X, \mathcal{X}, \mu, T_1, \ldots, T_\ell) \) be a system. Let \( \mathcal{Y}_1, \ldots, \mathcal{Y}_\ell \subseteq \mathcal{X} \) be sub-\( \sigma \)-algebras. Suppose that for all \( f_j \in L^\infty(\mathcal{Y}_j, \mu) \), \( j \in [\ell] \), the seminorm \( \|f_m\|_s \) controls the average

\[
\mathbb{E}_{n \in [N]}^\mathbb{E} \prod_{\ell} T_{\ell}^{p_\ell(n)} f_1 \cdots T_{\ell}^{p_\ell(n)} f_\ell \tag{22}
\]

in that equation (22) converges to 0 in \( L^2(\mu) \) whenever \( \|f_m\|_s = 0 \). Then for every \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that if \( f_j \in L^\infty(\mathcal{Y}_j, \mu) \), \( j \in [\ell] \) are \( 1 \)-bounded and \( \|f_m\|_s \leq \delta \), then

\[
\lim_{N \to \infty} \mathbb{E}_{n \in [N]}^\mathbb{E} \prod_{\ell} T_{\ell}^{p_\ell(n)} f_1 \cdots T_{\ell}^{p_\ell(n)} f_\ell \|_{L^2(\mu)} \leq \varepsilon.
\]

Finally, we need the following polynomial ergodic theorem (PET) result that gives box seminorm control for averages with extra terms involving dual functions. It extends [6, Theorem 2.5] that did not involve dual functions. We remark that these arguments are proved by combining a complicated variant of Bergelson’s original PET technique [2] with concatenation results of Tao and Ziegler [18].

**Proposition 3.7.** (Box seminorm control [11, Proposition B.1]) Let \( d, \ell, D, L \in \mathbb{N}, \eta \in [\ell]^\ell \), and \( p_1, \ldots, p_\ell, q_1, \ldots, q_L \in \mathbb{Z}[n] \) be polynomials of degrees at most \( D \). Let \( p_0 = 0 \) and \( p_j(n) = \sum_{i=0}^D a_{ji} n^i \) for \( j \in [\ell] \). Suppose that \( \text{deg} p_\ell = D \) and \( d_{ij} := \text{deg}(p_\ell e_{\eta_i} - p_j e_{\eta_j}) > 0 \) for every \( j = 0, \ldots, \ell - 1 \). Then there exist \( s \in \mathbb{N} \), depending only on \( d, D, \ell, \eta, \) and non-zero vectors

\[
b_1, \ldots, b_s = \{a_{\ell d_{ij}} e_{\eta_i} - a_{\ell d_{ij}} e_{\eta_j} : j = 0, \ldots, \ell - 1\},
\]

with the following property: for every system \( (X, \mathcal{X}, \mu, T_1, \ldots, T_\ell) \), functions \( f_1, \ldots, f_\ell \in L^\infty(\mu) \), and sequences of functions \( D_1, \ldots, D_L \in \mathcal{D}_d \), we have

\[
\lim_{N \to \infty} \mathbb{E}_{n \in [N]} \prod_{\ell} T^{p_\ell(n)}_{\eta_j} f_j \prod_{j=1}^L D_j(q_j(n)) \|_{L^2(\mu)} = 0
\]

whenever \( \|f_\ell\|_{b_1, \ldots, b_s} = 0 \).

Due to the monotonicity property of box seminorms in equation (17), we may (and will) always assume that \( s \geq 2 \) in Proposition 3.7 (this is necessary to apply Lemma 2.1 in some of our arguments in §7).

4. **Two motivating examples**
In this section, we prove Theorem 1.1 for the family \( n^2, n^2, n^2 + n \) and sketch the changes needed to handle the family \( n^2, n^2, 2n^2 + n \). These two cases illustrate some (but not all) key ideas needed in the proof of Theorem 1.1 in a simple setting. Additional complications
arise for more general families and the ideas needed to overcome them will be illustrated with examples given in subsequent sections.

Example 1. (Seminorm control for a monic family of length 3) Our goal is to prove the following result.

**PROPOSITION 4.1.** (Seminorm control for a monic family of length 3) There exists $s \in \mathbb{N}$ such that for every system $(X, \mathcal{X}, \mu, T_1, T_2, T_3)$ satisfying $\mathcal{I}(T_1 T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2)$ and all functions $f_1, f_2, f_3 \in L^\infty(\mu)$, the average

$$\frac{\mathbb{E}}{n \in [N]} T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot T_3^{n^2+n} f_3$$

converges to 0 in $L^2(\mu)$ whenever $\|f_3\|_{s,T_3} = 0$.

We subsequently show in Corollary 4.5 that the same conclusion holds if we assume $\|f_i\|_{s,T_i} = 0$ for $i = 1, 2$, instead.

By Proposition 3.7, there exist vectors $b_1, \ldots, b_{s+1} \in \{e_3, e_3 - e_1, e_3 - e_2\}$ such that

$$\lim_{N \to \infty} \frac{\mathbb{E}}{n \in [N]} T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot T_3^{n^2+n} f_3 = 0$$

whenever $\|f_3\|_{b_1,\ldots,b_{s+1}} = 0$. The goal is to inductively replace all the vectors $b_1, \ldots, b_{s+1}$ in the seminorm different from $e_3$ by $e_3^{s,s'}$ for some $s' \in \mathbb{N}$, which is achieved in the following proposition.

**PROPOSITION 4.2.** (Box seminorm smoothing) Let $b_1, \ldots, b_{s+1} \in \{e_3, e_3 - e_1, e_3 - e_2\}$ and $(X, \mathcal{X}, \mu, T_1, T_2, T_3)$ be a system satisfying $\mathcal{I}(T_1 T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2)$. Suppose that equation (25) holds for all functions $f_1, f_2, f_3 \in L^\infty(\mu)$ whenever $\|f_3\|_{b_1,\ldots,b_{s+1}} = 0$. Then there exists $s' \in \mathbb{N}$, depending only on $s$, such that equation (25) holds whenever $\|f_3\|_{b_1,\ldots,b_{s},e_3^{s,s'}} = 0$.

Proposition 4.1 follows from Proposition 3.7 and an iterative application of Proposition 4.2.

Passing from a control by $\|f_3\|_{b_1,\ldots,b_{s+1}}$ in Proposition 4.2 to a control by $\|f_3\|_{b_1,\ldots,b_{s},e_3^{s,s'}}$ follows a two-step ping-pong strategy similar to the one used in [11].

Using the control by $\|f_3\|_{b_1,\ldots,b_{s},e_3^{s_1}}$ for some $i = 1, 2$ and $s_1 \in \mathbb{N}$, and then we use this auxiliary control to go back and control the average from equation (42) by $\|f_3\|_{b_1,\ldots,b_{s},e_3^{s,s'}}$ for some $s' \in \mathbb{N}$. We call the two steps outlined above ping and pong.

The assumption $\mathcal{I}(T_1 T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2)$ is crucial for the following special case that will be invoked in the proof of the general case of Proposition 4.1.

**PROPOSITION 4.3.** There exists $s \in \mathbb{N}$ such that for every system $(X, \mathcal{X}, \mu, T_1, T_2)$ satisfying $\mathcal{I}(T_1 T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2)$ and all functions $f_1, f_2, f_3 \in L^\infty(\mu)$, the average

$$\frac{\mathbb{E}}{n \in [N]} T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot T_2^{n^2+n} f_3$$

converges to 0 in $L^2(\mu)$ whenever one of $\|f_1\|_{s,T_1}, \|f_2\|_{s,T_2}, \|f_3\|_{s,T_2}$ is 0.
Proof. We prove that $\|f_3\|_{s,T_2} = 0$ implies
\[
\lim_{N \to \infty} \mathbb{E}_{n \in [N]} T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot T_3^{n^2+n} f_3 = 0 \quad (27)
\]
for some $s \in \mathbb{N}$; the other cases follow similarly. By Proposition 3.7, the equality in equation (27) holds under the assumption that $\|f_3\|_{\mathcal{F}_{t+1} \times (s_2-e_1) \times s_2} = 0$ for some $s_1, s_2 \in \mathbb{N}_0$ (which are absolute in that they do not depend on the system or the functions). The assumption $I(T_1 T_2^{-1}) = I(T_1) \cap I(T_2)$ implies that $I(T_1 T_2^{-1}) \subseteq I(T_2)$. Together with Lemma 2.2, this gives
\[
\|f_3\|_{\mathcal{F}_{t+1} \times (s_2-e_1) \times s_2} \leq \|f_3\|_{\mathcal{F}(s_1+2) \times s_2} = \|f_3\|_{s_1+2,T_2},
\]
and so equation (27) holds whenever $\|f_3\|_{s,T_2} = 0$ for $s = s_1 + s_2$.

Proposition 4.3 is invoked in the ping step of the proof of Proposition 4.2; in the pong step, we invoke the result below.

**PROPOSITION 4.4.** Let $d, L \in \mathbb{N}$. Then there exists $s \in \mathbb{N}$ depending only on $d$ and $L$ such that for all systems $(X, \mathcal{X}, \mu, T_1, T_2, T_3)$, functions $f_1, f_3 \in L^\infty(\mu)$, and sequences $\mathcal{D}_1, \ldots, \mathcal{D}_L \in \mathcal{D}_d$, the average
\[
\mathbb{E}_{n \in [N]} T_1^{n^2} f_1 \cdot \prod_{j=1}^L \mathcal{D}_j(n^2) \cdot T_3^{n^2+n} f_3 \quad (28)
\]
converges to 0 in $L^2(\mu)$ whenever $\|f_3\|_{s,T_3} = 0$.

Proposition 4.4 follows from [11, Proposition 8.4] since the two non-dual terms in equation (28) involve the pairwise independent polynomials $n^2$ and $n^2 + n$.

Having stated all the needed auxiliary results, we are finally in the position to prove Proposition 4.2.

**Proof of Proposition 4.2.** Suppose that (25) fails. Then $\|f_3\|_{b_1,\ldots,b_{s+1}} > 0$. If $b_{s+1} = e_3$, then $\|f_3\|_{b_1,\ldots,b_{s+1}} = 0$, so we assume without loss of generality that $b_{s+1} = e_3 - e_2$, the case $b_{s+1} = e_3 - e_1$ being identical.

**Step 1 (ping):** Obtaining auxiliary control by a seminorm of $f_2$.

By Proposition 3.2 (applied with $A = \mathcal{X}$), we replace $f_3$ by $f_3$ so that
\[
\lim_{N \to \infty} \mathbb{E}_{n \in [N]} T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot T_3^{n^2+n} f_3 \|_{L^2(\mu)} > 0.
\]
We set $f_{j,b_1,b'} := \Delta_{b_1,\ldots,b_1,b,b'} f_j$ for $j \in [3]$ and apply Proposition 3.3(i) (with $c = 0$) to conclude that
\[
\liminf_{H \to \infty} \mathbb{E}_{b_1,b' \in [H]} \lim_{N \to \infty} \mathbb{E}_{n \in [N]} T_1^{n^2} f_{1,b_1,b'} \cdot T_2^{n^2} f_{2,b_1,b'} \cdot T_3^{n^2+n} u_{b_1,b'} \|_{L^2(\mu)} > 0
\]
for some 1-bounded functions $u_{b_1,b'}$ invariant under $T_3 T_2^{-1}$. The invariance property implies that
\[
T_3^{n^2+n} u_{b_1,b'} = T_2^{n^2+n} u_{b_1,b'},
\]
and hence
\[
\liminf_{H \to \infty} \lim_{N \to \infty} \mathbb{E}_{h, h' \in [H]} \mathbb{E}_{n \in [N]} T_1 n^2 f_1 h + h' T_2 n^2 f_2 h + h' T_2 n^2 + n u_{h, h'} L_2(\mu) > 0.
\]
Consequently, there exist a set \( B \subseteq \mathbb{N} \) of positive lower density and \( \varepsilon > 0 \) such that
\[
\lim_{N \to \infty} \mathbb{E}_{n \in [N]} T_1 n^2 f_1 h + h' T_2 n^2 f_2 h + h' T_2 n^2 + n u_{h, h'} L_2(\mu) > \varepsilon
\]
for all \((h, h') \in B\). Each of the averages in equation (29) takes the form of equation (26); we therefore apply Propositions 4.3 and 3.6 to obtain \( s_1 \in \mathbb{N} \) (independent of the system or the functions) and \( \delta > 0 \) (which depends only on \( \varepsilon \) and the system but not the functions) such that
\[
\| f_2 h + h' \|_{s_1, T_2} > \delta
\]
for all \((h, h') \in B\). Hence,
\[
\liminf_{H \to \infty} \mathbb{E}_{h, h' \in [H]} \mathbb{E}_{n \in [N]} \| \Delta_{b_1 \ldots b_s} f_2 \|_{s_1, T_2} > 0.
\]
Together with Lemma 3.1, the inductive formula for seminorms in equation (15), and Hölder inequality, the inequality in equation (30) implies that
\[
\| f_2 \|_{b_1 \ldots b_s, e_s^x} > 0.
\]
We deduce from this that the seminorm \( \| f_2 \|_{b_1 \ldots b_s, e_s^x} \) controls the average (24).

This seminorm control is not particularly useful as an independent result since the vectors \( b_1, \ldots, b_s \) may not involve the transformation \( T_2 \) in any way. However, it is of great importance as an intermediate result applied in the next step of our argument.

**Step 2 (pong):** Obtaining control by a seminorm of \( f_3 \).

Using our assumption that equation (25) fails, we now replace \( f_2 \) by \( \tilde{f}_2 \) and deduce from Proposition 3.2 (applied again with \( A = X \)) that
\[
\lim_{N \to \infty} \mathbb{E}_{n \in [N]} T_1 n^2 f_1 T_2 n^2 \tilde{f}_2 T_2 n^2 + n f_3 L_2(\mu) > 0.
\]
From Proposition 3.3 (with \( c = 0 \)), it follows that
\[
\liminf_{H \to \infty} \mathbb{E}_{h, h' \in [H]} \lim_{N \to \infty} \mathbb{E}_{n \in [N]} T_1 n^2 f_1 h + h' \cdot \mathcal{D}_{h, h'} (n^2) T_3 n^2 + n f_3 h + h' L_2(\mu) > 0,
\]
where \( \mathcal{D}_{h, h'} \) is a product of \( 2^s \) elements of \( \mathcal{D}_{s_1} \). Once again, there exists a set \( B' \subseteq \mathbb{N} \) of positive lower density and \( \varepsilon > 0 \) such that
\[
\mathbb{E}_{n \in [N]} T_1 n^2 f_1 h + h' \cdot \mathcal{D}_{h, h'} (n^2) T_3 n^2 + n f_3 h + h' L_2(\mu) > \varepsilon
\]
for \((h, h') \in B'\). Each average indexed in equation (31) is of the form
\[
\mathbb{E}_{n \in [N]} T_1 n^2 g_1 \cdot \prod_{j=1}^{2^s} \mathcal{D}_{j} (n^2) T_3 n^2 + n g_3
\]
for \( D_1, \ldots, D_{2^r} \in \mathcal{D}_s \). By Proposition 4.4, the averages from equation (32) are controlled by \( \| g_3 \|_{s', T_3} \) for some \( s' \in \mathbb{N} \) depending only on \( s \), and so Proposition 3.6 gives \( \delta > 0 \) such that

\[
\| f_{3, h, h'} \|_{s', T_3} > \delta
\]

for all \( (h, h') \in B' \). (Specifically, we invoke Proposition 3.6 for averages of the form \( \mathbb{E}_{n \in \mathbb{N}} T_1^{n^2} g_1 \cdot \prod_{j=1}^{2^r} T_2^{n^2} g_j' \cdot T_3^{n^2+n} g_3 \), where \( g_j' \) is \( \mathcal{Z}_s(T_2) \)-measurable for each \( j \in [2^r] \). One can show that an average like this is qualitatively controlled by \( \| g_3 \|_{s', T_3} \) by approximating functions \( g_1', \ldots, g_{2^r}' \) by linear combinations of dual functions using Proposition 2.3 and then applying Proposition 3.7.) Hence,

\[
\liminf_{H \to \infty} \mathbb{E}_{h, h' \in [H]^s} \| \Delta_{b_1, \ldots, b_s, e_{3^s'}} f_3 \|_{s', T_3} > 0.
\]

Together with Lemma 3.1, the Hölder inequality, and the inductive formula in equation (15) for the seminorms, this implies that \( \| f_3 \|_{b_1, \ldots, b_s, e_{3^s'}} > 0 \), which is what we claim.

Finally, we show how we can use Proposition 4.1 to obtain control of the average from equation (24) by seminorms of other terms.

**Corollary 4.5.** There exists \( s \in \mathbb{N} \) such that for every system \((X, X', \mu, T_1, T_2, T_3)\) satisfying \( \mathcal{I}(T_1 T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2) \) and all functions \( f_1, f_2, f_3 \in L^\infty(\mu) \), the average from equation (24) converges to 0 in \( L^2(\mu) \) whenever one of \( \| f_1 \|_{s, T_1}, \| f_2 \|_{s, T_2}, \| f_3 \|_{s, T_3} \) is 0.

**Proof.** The statement that for some absolute \( s \in \mathbb{N} \), the identity \( \| f_3 \|_{s, T_3} = 0 \) implies the vanishing of the \( L^2(\mu) \) limit of equation (24) follows from Propositions 3.7 and 4.1, so the content of Corollary 4.5 is to show control by other terms. Suppose that

\[
\lim_{N \to \infty} \mathbb{E}_{n \in \mathbb{N}} T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot T_3^{n^2+n} f_3 \neq 0.
\]

We then apply Proposition 2.3 and the pigeonhole principle to find a 1-bounded dual function \( D_{s, T_3} g \) such that

\[
\lim_{N \to \infty} \mathbb{E}_{n \in \mathbb{N}} T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot \mathcal{D}(n^2 + n),
\]

where \( \mathcal{D}(n) := T_3^n D_{s, T_3} g \). By Proposition 3.7, we have \( \| f_2 \|_{e_{3^s'}, (e_2, e_1) = e_{3^s'}} > 0 \) for some absolute \( s' \in \mathbb{N} \). The ergodicity assumption of \( T_1 T_2^{-1} \) and Lemma 2.2 imply that \( \| f_2 \|_{e_{3^s'}} > 0 \), and an analogous argument gives \( \| f_1 \|_{e_{2^s'}} > 0 \).

The argument in Example 1 is relatively clean because the leading coefficients of the polynomials are all 1. When this is not the case, minor modifications are required as explained in the next example.

**Example 2.** (Seminorm control for a non-monic family of length 3) Consider the average

\[
\mathbb{E}_{n \in \mathbb{N}} T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot T_3^{2n^2+n} f_3. \tag{33}
\]
By Proposition 3.7, this average is controlled by \( \| f_j \|_{b_1, \ldots, b_s+1} \) for some \( s \in \mathbb{N} \) and \( b_1, \ldots, b_{s+1} \in \{2e_3, 2e_3 - e_2\} \), and we want to pass toward the control by \( \| f_j \|_{b_1, \ldots, b_s, e_2} \). Suppose that the \( L^2(\mu) \) limit of equation (33) does not vanish, and suppose moreover that \( b_{s+1} = 2e_3 - e_2 \). Arguing as in the proof of Proposition 4.2, we arrive at the inequality

\[
\lim_{H \to \infty} \mathbb{E} \limsup_{N \to \infty} \left\| \mathbb{E}_{n[N]} \left( T_1^{n_1} f_1, b_1, b_1' \cdot T_2^{n_2} f_2, b_2, b_2' \cdot T_3^{n_3} f_3, b_3, b_3' \right) \right\|_{L^2(\mu)} > 0,
\]

where \( f_j, b_j, b_j' \) for \( j \in [2] \) and the functions \( u_{b, b'} \) are \( T_2 T_3^{-2} \) invariant. We can no longer apply the invariance property in the same way as before since the polynomial \( 2n^2 + n \) is not divisible by 2. Instead, we first split \( \mathbb{N} \) into the odd and even part and then apply the triangle inequality to deduce that

\[
\lim_{N \to \infty} \mathbb{E} \limsup_{r \in \{0, 1\}} \liminf_{H \to \infty} \mathbb{E}_{b,h \in [H]} \left\| T_1^{(2n+r)2} f_1, b, h \cdot T_2^{(2n+r)2} f_2, b, h \cdot T_3^{(2n+r)2} f_3, b, h \cdot \right\|_{L^2(\mu)} > 0.
\]

Only then can we apply the \( T_2 T_3^{-2} \) invariance of \( u_{b, b'} \) to obtain the identity

\[
T_3^{(2n+r)^2 + (2n+r)} u_{b, b'} = T_2^{4n^2 + (4r+1)n} T_3^{2r^2 + r} u_{b, b'}.
\]

It follows that for some \( r_0 \in \{0, 1\} \), we have

\[
\limsup_{H \to \infty} \mathbb{E}_{b, h \in [H]} \left\| T_1^{4n^2 + 4r_0n} f_1, b, h' \cdot T_2^{4n^2 + 4r_0n} f_2, b, h' \cdot T_3^{4n^2 + (4r_0+1)n} u_{b, h'} \right\|_{L^2(\mu)} > 0.
\]

upon setting \( f_j, b_j, b_j' := T_1^{r_0} f_j, b_1, b_1' \) for \( j \in [2] \) and \( u_{b, h'} := T_3^{2r_0 + r_0} u_{b, h'} \). The rest of the argument proceeds analogously except that we invoke an analog of Proposition 4.3 for the tuple \( (T_1^{4n^2 + 4r_0n}, T_2^{4n^2 + 4r_0n}, T_3^{4n^2 + (4r_0+1)n}) \). The important part about this new tuple is that the first two polynomials are again pairwise dependent while the last one is pairwise independent with any of the first two, and that the new tuple retains the good ergodicity property.

5. Formalism and general strategy for longer families

We move on towards deriving Theorem 1.1 for longer families. To prove it for averages

\[
\mathbb{E}_{n[N]} \prod_{j \in [\ell]} T_j^{p_j(n)} f_j,
\]

we need to analyze more complicated averages of the form

\[
\mathbb{E}_{n[N]} \prod_{j \in [\ell]} T_n^{p_j(n)} f_j \prod_{j \in [L]} D_j(q_j(n))
\]

that appear at the intermediate steps of the proof of Theorem 1.1, much like averages from equations (26) and (28) show up at the intermediate steps of the proof of Proposition 4.1.
a special case of Theorem 1.1 for the family \( n^2, n^2, n^2 + n \). In equations (34) and (35), \( p_1, \ldots, p_\ell, \rho_1, \ldots, \rho_\ell, q_1, \ldots, q_L \) are (not necessarily distinct) polynomials with integer coefficients and zero constant terms, \((X, X, \mu, T_1, \ldots, T_\ell)\) is a system, \( f_1, \ldots, f_\ell \in L^\infty(\mu) \) are 1-bounded functions, and \( D_1, \ldots, D_L \in \mathcal{D} \) are sequences of functions. Since \( D_j(q_j(n)) \) has the form \( T_{\pi_j}^{\rho_j(n)} g_j \) for some \( \pi_j \in [\ell] \) and \( g_j \in L^\infty(\mu) \), the averages from equation (35) converge in \( L^2(\mu) \) by [19]. The same comment applies for all limits involving dual sequences that appear in the rest of the paper.

The purpose of this section is to introduce a formalism that helps us to meaningfully discuss averages from equation (35). This will be done in §5.1. Subsequently, we give in §5.2 an overview of the strategy used to prove Theorem 1.1. The details of various moves discussed in §5.2 will be presented in §6.

While discussing various examples in this and the next sections, we often say informally that for \( j \in [\ell] \), the average from equation (35) is controlled by a \( T_{\eta_j} \)-seminorm of \( f_j \) (or that we have seminorm control of equation (35) by a \( T_{\eta_j} \)-seminorm of \( f_j \)) if for all \( d, L \in \mathbb{N} \), there exists \( s \in \mathbb{N} \) such that if the \( T_{\eta_j} \)-seminorm of \( f_j \) vanishes, then the \( L^2(\mu) \) limit of equation (35) is 0 for all sequences \( D_1, \ldots, D_L \in \mathcal{D}_d \) and all functions \( f_1, \ldots, f_\ell \in L^\infty(\mu) \) satisfying some explicitly stated invariance properties. We also say informally that we have seminorm control over the average from equation (35) if we have seminorm control by a \( T_{\eta_j} \)-seminorm of \( f_j \) for every \( j \in [\ell] \).

5.1. The formalism behind the induction scheme. We start by introducing a handy formalism used for the induction scheme in the proof of Theorem 1.1. We often associate the average from equation (35) with the tuple \( (T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]} \). This tuple does not contain any information about the polynomials \( q_1, \ldots, q_L \), but this is not necessary. These terms play no role in our inductive argument and can be easily disposed of using Proposition 3.7. The only thing they do influence is the degree \( s \) of the seminorm with which we end up controlling the average from equation (35).

**Definition.** (Indexing data) For an average from equation (35) or the associated tuple \( (T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]} \), we let \( \ell \) be its length, \( d := \max_{j \in [\ell]} \deg \rho_j \) be its degree, and \( \eta \) be its indexing tuple. For \( j \in [\ell] \), we set \( d_j := \deg \rho_j \). We let \( K_1 \) be the maximum number of pairwise independent polynomials within the family \( \rho_1, \ldots, \rho_\ell \) (we set \( K_1 := 1 \) if \( \ell = 1 \) or every two polynomials are pairwise dependent). We partition \( [\ell] = \bigcup_{t \in [K_1]} J_t \), where \( j_1, j_2 \) belong to the same \( J_t \) if and only if \( \rho_{j_1}, \rho_{j_2} \) are linearly dependent. Thus, \( J_1, \ldots, J_{K_1} \) partitions \([\ell]\) into index sets corresponding to families of pairwise dependent polynomials. Furthermore, we define

\[ \mathcal{L} := \{ j \in [\ell] : \deg \rho_j = d \} \]

to be the set of indices corresponding to polynomials of maximum degree, and we rearrange \( J_1, \ldots, J_{K_1} \) so that \( \mathcal{L} = \bigcup_{t \in [K_2]} J_t \) for some \( K_2 \leq K_1 \). We also let \( K_3 := |\mathcal{L}| \) be the number of maximum degree polynomials, and we notice that \( K_2 \leq K_3 \leq \ell \).

Sometimes, we denote \( \mathcal{L} = \mathcal{L}(\rho_1, \ldots, \rho_\ell), J_t = J_t(\rho_1, \ldots, \rho_\ell), \) and \( K_i = K_i(\rho_1, \ldots, \rho_\ell) \) to emphasize the dependence on a specific family of polynomials.
Example 3. Consider the tuple

$$(T_1^n, T_2^n, T_3^{n^2+n}, T_4^{2n^2+2n}, T_5^{n^2+2n}, T_6^n, T_2^{n^2+3n}).$$

(36)

It has length 7, degree 2, indexing tuple $\eta = (1, 2, 3, 4, 5, 6, 2)$, $K_1 = 5$ (corresponding to five pairwise independent polynomials $n^2, n^2 + n, n^2 + 2n, n^2 + 3n$), $K_2 = 4$ (corresponding to four quadratic pairwise independent polynomials $n^2, n^2 + n, n^2 + 2n, n^2 + 3n$), $K_3 = 6$ (corresponding to six quadratic polynomials appearing in equation (36)), $\mathcal{L} = \{1, 2, 3, 4, 5, 7\}$, and partition $\mathcal{J}_1 = \{1, 2\}, \mathcal{J}_2 = \{3, 4\}, \mathcal{J}_3 = \{5\}, \mathcal{J}_4 = \{7\}, \mathcal{J}_5 = \{6\}$.

The rationale behind introducing the partition $\mathcal{J}_1, \ldots, \mathcal{J}_{K_1}$ and the indexing tuple $\eta$ is as follows. As part of our assumptions, we know that $T_i^{\beta_i} T_j^{-\beta_j}$ is ergodic whenever the polynomials $\rho_i, \rho_j$ are dependent, $b_i, b_j$ are their leading coefficients, and $\beta_i := b_i / \gcd(b_i, b_j), \beta_j := b_j / \gcd(b_i, b_j)$. Introducing the partition $\mathcal{J}_1, \ldots, \mathcal{J}_{K_1}$ allows us to keep track of pairs $(i, j)$ for which we have these ergodicity properties. The reason for introducing the indexing tuple $\eta$ is that in our induction procedure, we gradually replace a transformation $T_{\eta_n}$ in the tuple $(T_{\eta_n}^{\rho_j(n)})_{j \in [\ell]}$ by a different transformation $T_{\eta_n}$, and so $\eta$ keeps track of these substitutions.

**Definition.** (Good ergodicity property along $\eta$) Let $\eta \in [\ell]^\ell$ be an indexing tuple and $b_1, \ldots, b_\ell$ be the leading coefficients of the polynomials $\rho_1, \ldots, \rho_\ell$. We say that the system $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ has the good ergodicity property along $\eta$ for the polynomials $\rho_1, \ldots, \rho_\ell$ if for every distinct $\eta_{j_1}, \eta_{j_2}$ belonging to the same $\mathcal{J}_t$ with $t \in [K_1]$, we have

$$\mathcal{I}(T_{\eta_{j_1}}^{\beta_{j_1}} T_{\eta_{j_2}}^{-\beta_{j_2}}) = \mathcal{I}(T_{\eta_{j_1}}) \cap \mathcal{I}(T_{\eta_{j_2}}),$$

(37)

where $\beta_j := b_j / \gcd(b_{j_1}, b_{j_2})$ for $j = j_1, j_2$. In other words, the only functions invariant under $T_{\eta_{j_1}}^{\beta_{j_1}} T_{\eta_{j_2}}^{-\beta_{j_2}}$ are those simultaneously invariant under $T_{\eta_{j_1}}$ and $T_{\eta_{j_2}}$. In particular, having the good ergodicity property corresponds to having the good ergodicity property for the polynomials $\rho_1, \ldots, \rho_\ell$ along the identity indexing tuple $\eta_0 := (1, \ldots, \ell)$. We similarly say that the tuple $(T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}$ has the good ergodicity property if $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ has the good ergodicity property along $\eta$ for $\rho_1, \ldots, \rho_\ell$.

What this definition captures is that every time we encounter in our tuple two transformations $T_{\eta_{j_1}}, T_{\eta_{j_2}}$ whose indices $\eta_{j_1}, \eta_{j_2}$ lie in the same set $\mathcal{J}_t$, the identity in equation (37) is satisfied. For instance, the tuple in equation (36) has the good ergodicity property along $\eta$ precisely when

$$\mathcal{I}(T_1 T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2) \quad \text{and} \quad \mathcal{I}(T_3 T_4^{-2}) = \mathcal{I}(T_3) \cap \mathcal{I}(T_4).$$

The first identity corresponds to comparing the pairs $(j_1, j_2) = (1, 2), (1, 7)$ corresponding to the occurrences of transformations $T_1$ and $T_2$ from the cell $\mathcal{J}_1$ ($T_1$ occurs at the index 1 and $T_2$ occurs at the indices 2 and 7). The second identity comes from comparing the pairs $(j_1, j_2) = (3, 4), (4, 5)$ corresponding to the transformations $T_3, T_4$ from the cell $\mathcal{J}_2$ ($T_3$ occurs at the indices 3 and 5 whereas $T_4$ occurs at the index 4).
The guiding principle behind our arguments is that we derive seminorm control of the average from equation (35) by inductively applying seminorm control of an average that is ‘simpler’ than the original average in an appropriate sense. For instance, in Proposition 4.2 and Corollary 4.5, we obtained seminorm control for the tuple \((T_n^2, T_n^2, T_n^{2+n})\) from Example 1 by invoking seminorm control for the following tuples:

- \((T_n^2, T_n^2, T_n^{2+n})\) in the ping step of the smoothing argument in Proposition 4.2;
- \((T_n^2, \ast, T_n^{2+n})\) in the pong step of the smoothing argument in Proposition 4.2 (the asterisk is introduced purely for convenience; it denotes the term replaced by a product of dual functions);
- \((T_n^2, T_n^2, \ast)\) in Corollary 4.5.

The relative complexity of a tuple or an average is captured by the following notion.

**Definition.** (Type) The type of \((T_{\rho_j(n)}^{\ell(n)})_{j \in [\ell]}\) is the tuple \(w = (w_1, \ldots, w_{K_2})\), where

\[w_t := |\{j \in \mathcal{L} : \eta_j \in \mathcal{J}_t\}| = |\{j \in [\ell] : \deg \rho_j = d, \eta_j \in \mathcal{J}_t\}|\]

counts the number of times the transformations \((T_j)_{j \in \mathcal{J}_t}\) appear in \((T_{\rho_j(n)}^{\ell(n)})_{j \in [\ell]}\) with a polynomial iterate of maximum degree. (We note here that the type of \((T_{\rho_j(n)}^{\ell(n)})_{j \in [\ell]}\) depends not just on \(n\) and the polynomials \(\rho_1, \ldots, \rho_\ell\), but also on the ordering of the sets \(\mathcal{J}_1, \ldots, \mathcal{J}_K\). We do not record this dependence explicitly, instead fixing some ordering of \(\mathcal{J}_1, \ldots, \mathcal{J}_K\) a priori.) We note that \(|w| = w_1 + \cdots + w_{K_2} = K_3\). We say that the type \(w\) is basic if it has the form \(w = (K_3, 0, \ldots, 0)\).

For instance, the tuple in equation (36) has type \((3, 3, 0, 0)\): this is because \(T_1, T_2\) corresponding to \(\mathcal{J}_1\) occur thrice, as do the transformations \(T_3, T_4\) corresponding to \(\mathcal{J}_2\), while the transformations \(T_5, T_7\) corresponding to \(\mathcal{J}_3\) and \(\mathcal{J}_4\) do not occur at all. We do not care about the occurrence of \(T_6\) since it has a linear iterate.

It is instructive to see what happens when the polynomials \(\rho_1, \ldots, \rho_\ell\) are pairwise independent. In that case, \(\mathcal{J}_t = \{j_t\}\) for every \(t \in [\ell]\) and \(w_t\) counts the number of times the transformation \(T_{j_t}\) appears among \((T_{\rho_j(n)})_{j \in \mathcal{L}}\), or equivalently the number of times that \(T_{j_t}\) attains a polynomial iterate of maximal degree. So for pairwise independence polynomials, this notion of type recovers the concept of type from [11, §8.2] (up to permuting \(\mathcal{J}_1, \ldots, \mathcal{J}_K\)).

For the set of tuples in \(\mathbb{N}_0^{K_2}\) of length \(K_3\), we say \(w' < w\) if there exists \(\kappa \in [K_2 - 1]\) such that for all \(t \in [\kappa]\), we have \(w'_t = w_t\), and either \(w'_{\kappa+1} = 0 < w_{\kappa+1}\) or \(w'_{\kappa+1} > w_{\kappa+1} > 0\). For instance, we have the following chain of type inequalities:

\[(4, 0, 0) < (3, 0, 1) < (3, 1, 0) < (2, 0, 2) < (2, 2, 0) < (2, 1, 1) < (1, 0, 3) < (1, 2, 1) < (1, 1, 2).\]

The first, third, fifth, sixth, and eighth inequalities follow from the condition \(w'_{\kappa+1} > w_{\kappa+1} > 0\) while the second, fourth, and seventh inequalities are consequences of the condition \(w'_{\kappa+1} = 0 < w_{\kappa+1}\). This is a rather atypical ordering, but it turns out to determine well which of the tuples \((T_{\rho_j(n)}^{\ell(n)})_{j \in [\ell]}\), \((T_{\rho_j(n)}^{\ell(n)})_{j \in [\ell]}\) is ‘simpler’ than the other.
The motivation for this particular choice of ordering is that in the arguments to come, we will be passing from a tuple \((T_{n_j}^{(\rho_j)}(n))_{j \in [\ell]}\) of type \(w\) to another tuple \((T_{n_j}^{(\rho_j)}(n))_{j \in [\ell]}\) of type \(w' < w\) in two ways. In one of them, the type \(w'\) will meet the condition \(w'_{k+1} > w_{k+1} > 0\) while in the other one, it will satisfy the condition \(w'_{k+1} = 0 < w_{k+1}\). Arguing this way, we arrive in finitely many steps at a tuple of a basic type \(w = (K_3, 0, \ldots, 0)\), which constitutes the base case of our induction. This transition will be explained in greater detail at the end of §5.2 and illustrated in Example 10.

**Lemma 5.1.** For fixed \(K_2, K_3 \in \mathbb{N}\), let \(A := \{w \in \mathbb{N}_0^{K_2} : w_1 + \cdots + w_{K_2} = K_3\}\). Then \(<\) defines a strict partial order on \(A\).

**Proof.** It is clear that \(<\) is asymmetric and irreflexive, so it remains to show that it is transitive. Suppose that \(w'' < w', w' < w\), and let \(\kappa_1, \kappa_2 \in [K_2 - 1]\) be indices such that \(w''_t = w'_t\) for all \(t \in [\kappa_2]\) but not for \(t = \kappa_2 + 1\) and \(w'_t = w_t\) for all \(t \in [\kappa_1]\) but not for \(t = \kappa_1 + 1\). If \(\kappa_2 < \kappa_1\), then either \(w''_{\kappa_2+1} = 0 < w'_{\kappa_2+1} = w_{\kappa_2+1}\) or \(w''_{\kappa_2+1} > w'_{\kappa_2+1} = w_{\kappa_2+1} > 0\), and so \(w'' < w\). Otherwise \(\kappa_2 \geq \kappa_1\), in which case either \(w''_{\kappa_2+1} = w'_{\kappa_2+1} = 0 < w_{\kappa_2+1}\) or \(w''_{\kappa_2+1} \geq w'_{\kappa_2+1} > w_{\kappa_2+1} > 0\), implying \(w'' < w\) again. \(\square\)

## 5.2. The general strategy

In this section, we outline how to obtain a seminorm control of a given tuple using seminorm control for tuples of lower type or shorter length.

**Definition.** (Controllable and uncontrollable tuples) Let \(t_w := \max\{t : w_t > 0\}\) be the last non-zero index of \(w\). We call a tuple \((T_{n_j}^{(\rho_j)}(n))_{j \in [\ell]}\) of a non-basic type \(w\) (or the corresponding average) **controllable**, if there exists an index \(m \in [\ell]\) such that

- \(\eta_m \in J_{t_w}\);
- for every other \(i \in [\ell]\) with \(\eta_i = \eta_m\), we have \(\rho_i \neq \rho_m\).

If \(m\) satisfies the aforementioned assumption, we say that it satisfies the **controllability condition**; in this case, Proposition 3.7 guarantees that the average from equation (35) is controlled by \(\| f_m \|_{b_1, \ldots, b_s}\) for non-zero vectors \(b_1, \ldots, b_s\). If no such index \(m\) exists, we call the tuple \((T_{n_j}^{(\rho_j)}(n))_{j \in [\ell]}\) **uncontrollable**.

The previous notions of controllability are supposed to capture whether Proposition 3.7 is applicable to the relevant tuples in our setting.

**Example 4.** (Controllable versus uncontrollable tuples) Consider the following two tuples:

\[
(T_{n_1}^2, T_{n_2}^2, T_{n_3}^2, T_{n_4}^2, T_{n_5}^2, T_{n_1}^{n_2+n}, T_{n_5}^{n_2+2n}, T_{n_5}^{n_2+2n}),
\]

(38)

\[
(T_{n_1}^2, T_{n_2}^2, T_{n_3}^2, T_{n_4}^2, T_{n_5}^2, T_{n_1}^{n_2+n}, T_{n_4}^{n_2+n}, T_{n_5}^{n_2+2n}).
\]

(39)

Defining the partitions \(\mathcal{I}_1 = \{1, 2, 3, 4\}, \mathcal{I}_2 = \{5, 6\}, \mathcal{I}_3 = \{7, 8\}\) corresponding to the independent polynomials \(n^2, n^2 + n, n^2 + 2n\) respectively, the first tuple has type \((5, 3, 0)\) while the second one has type \((6, 2, 0)\), and for both tuples, we have \(t_w = 2\). The first one is controllable because for the index \(m = 5\), the only values \(i \neq m\) such that \(\eta_i = 5\) are \(i = 7, 8\) corresponding to the polynomial \(n^2 + 2n\), which is distinct from \(n^2 + n\).
By contrast, the tuple in equation (39) does not possess an index satisfying the controllability condition: the only indices \( m \in [8] \) with \( \eta_m \in \mathcal{S}_2 \) are \( m = 7, 8 \), and we have both \( \eta_7 = \eta_8 \) and \( \rho_7(n) = n^2 + 2n = \rho_8(n) \). Hence, this tuple is uncontrollable.

Our strategy for proving seminorm control will work rather differently for controllable and uncontrollable tuples. Suppose first that the average from equation (35) with tuple \((T_{\eta_j} \rho_j(n))_{j \in [\ell]}\) of a non-basic type is controllable, and that an index \( m \) satisfies the controllability condition. Then Proposition 3.7 guarantees that the average from equation (35) is controlled by \( \| f_m \|_{b_1, \ldots, b_{s+1}} \) for some vectors \( b_1, \ldots, b_{s+1} \) from equation (23), and the controllability implies that these vectors are indeed non-zero. We want to show that this average is also controlled by \( \| f_m \|_{b_1, \ldots, b_{s'}, e_{\eta_m}^{s'}} \) for some \( s' \in \mathbb{N} \) via a seminorm smoothing argument that generalizes Proposition 4.2. We then iterate this result \( s \) more times to get control by a \( T_{\eta_m} \)-seminorm of \( f_m \). The seminorm smoothing argument follows a ping-pong strategy much like in the proof of Proposition 4.2. We first show that control by \( \| f_m \|_{b_1, \ldots, b_{s+1}} \) implies control by \( \| f_i \|_{b_1, \ldots, b_{s'}, e_{\eta_i}^{s+1}} \) for some \( s_1 \in \mathbb{N} \) and \( i \neq m \), and then we use this auxiliary result to obtain control by \( \| f_m \|_{b_1, \ldots, b_{s'}, e_{\eta_m}^{s'}} \).

The main idea behind the ping step of the seminorm smoothing argument is to show that a seminorm control of the average from equation (35) can be deduced from a seminorm control of a family of averages of the form

\[
\lim_{N \to \infty} \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{\eta_j}^{\rho_j'(n)} f_j' \prod_{j \in [L']} D_j'(q_j'(n)) \tag{40}
\]

for some polynomials \( \rho_1', \ldots, \rho_\ell', q_1', \ldots, q_{\ell'}', \in \mathbb{Z}[n] \), 1-bounded functions \( f_1', \ldots, f_{\ell}' \in L^\infty(\mu) \), and sequences of functions \( D_1', \ldots, D_{\ell'}' \in \mathcal{D} \). Moreover, the indexing tuple \( \eta' = [\ell] \) is obtained from \( \eta \) by changing \( \eta_m \) into \( \eta_i \) for some \( i \neq m \), that is, the passage from equation (35) to equation (40) goes by replacing \( T_{\eta_m} \) at index \( m \) with \( T_{\eta_i} \). Importantly, the new average from equation (40) satisfies several key properties:

(i) it has a lower type than the original average, so that we can argue by induction;
(ii) the new average retains the good ergodicity property of the original average;
(iii) the functions \( f_j' \) in the new average satisfy some invariance properties;
(iv) as long as the aforementioned invariance properties are satisfied, the new average from equation (40) is controlled by the seminorm \( \| f_j' \|_{s, T_{\eta_j}'} \) for each \( j \in [\ell] \) and some \( s \in \mathbb{N} \).

Proposition 6.1 explains the exact way in which we pass from averages from equation (35) to those from equation (40) so that the property (i) is satisfied, and Proposition 6.4 establishes the property (ii). Proposition 6.5 then ensures that the functions \( f_j' \) in equation (40) satisfy needed invariance properties.

We note though that the new average from equation (40) need not be controllable. For instance, if we take the average corresponding to the tuple from equation (38), then in the ping step, we replace \( T_{\eta_7}^{n^2+n} \) by the same iterate of one of \( T_1, T_2, T_3, T_4 \). The new average is then uncontrollable, as is the tuple from equation (39), corresponding to replacing \( T_{\eta_7}^{n^2+n} \).
by $T_1^{n^2+n}$. Hence, controllability may not be preserved while performing the procedure outlined above.

In the pong step of the smoothing argument for equation (35), we deal with averages of the form

$$\mathbb{E}_{n \in [N]} \prod_{j \in [\ell], j \neq i} T_{\rho_j(n)} f_j'' \cdot \prod_{j \in [L']} D_j''(q_j''(n)).$$

(41)

Crucially, each function $f_j''$ is invariant under some composition of $T_{\eta_j}$ and $T_j$. This allows us to replace (some iterate of) $T_{\eta_j}$ in equation (41) by (some iterate of) $T_j$, a procedure that we call flipping, and show that an average from equation (41) essentially equals an average of the form

$$\mathbb{E}_{n \in [N]} \prod_{j \in [\ell], j \neq i} T_j^{\rho_j''(n)} f_j''' \cdot \prod_{j \in [L']} D_j''(q_j'''(n)).$$

(42)

of length $\ell - 1$. The details of how flipping is performed are presented in Proposition 6.6.

An inductive application of a suitable modification of Theorem 1.1 then gives a control of equation (42) by a $T_j$-seminorm of $f_j'''$ for each $j \neq i$, and the invariance property of $f_j''$ translates it into a control of equation (41) by a $T_{\eta_j}$-seminorm of $f_j''$ for each $j \neq i$. A straightforward argument analogous to one at the end of the proof of Proposition 4.2 gives a control of equation (35) by a $T_{\eta_m}$-seminorm of $f_m$.

If the average from equation (35) is uncontrollable, then we proceed rather differently. The previous strategy breaks right at the start since there is no index $m$ satisfying the controllability condition. Consequently, whichever index $m$ with $\eta_m \in \mathcal{I}_{tw}$ we take, we cannot employ Proposition 3.7 to bound the seminorm by $|||f_m|||_{b_1, \ldots, b_{s+1}}$ for non-zero vectors $b_1, \ldots, b_{s+1}$. What we use instead is the inductive assumption that the functions $f_j$ at indices $j$ with $\eta_j \in \mathcal{I}_{tw}$ are invariant under a composition of (some power of) $T_{\eta_j}$ and (some power of) $T_j^{-1}$. Using this invariance property, we perform flipping once more to replace the original average from equation (35) by a new average

$$\lim_{N \to \infty} \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{\eta_j'''}^{\rho_j'''}(n) f_j''' \cdot \prod_{j \in [L']} D_j'''(q_j'''(n)),$$

(43)

where $\eta''' = j$ whenever $\eta_j \in \mathcal{I}_{tw}$; the details are provided in Corollary 6.7. This new average has the good ergodicity property and is controllable. Importantly, it has a lower type, which is established in Proposition 6.8. We can then obtain seminorm control of equation (35) by inductively invoking the seminorm control of equation (43). The seminorm control of equation (43) is proved in turn by the smoothing argument for controllable averages described above.

Thus, whether the tuple $(T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}$ of a non-basic type is controllable or not, the idea is to control it by a Gowers–Host–Kra seminorm by invoking seminorm control for tuples of lower type or smaller length that naturally appear when examining $(T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}$. If the tuple $(T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}$ of type $w$ is controllable, we will invoke seminorm control for tuples of type $w'$ satisfying $w'_t = w_t$ for $t \in [\kappa]$ and $w'_{\kappa+1} > w_{\kappa+1} > 0$ for some
κ ∈ [K2 − 1] in the ping step of the smoothing argument. Specifically, the new type \( w' \) is obtained from \( w \) by the type operation defined in equation (45). In the pong step of the smoothing argument, we will use seminorm control for tuples of length \( ℓ − 1 \). If the tuple \( (T^{ρ_j(n)}_{η_j})_{j∈[ℓ]} \) is uncontrollable, we will invoke seminorm control for tuples of type \( w' \) satisfying \( w'_t = w_t \) for \( t ∈ [κ] \) and \( w'_{κ+1} = 0 < w_{κ+1} \) for some \( κ ∈ [K2 − 1] \); the details are given in Proposition 6.8(v). The way in which we apply seminorm control for tuples of lower type motivates the choice of our somewhat weird ordering on types.

Reducing to tuples of lower type this way and noting that the tuples of length \( ℓ \) can have at most \((ℓ + 1)^ℓ\) distinct types, we arrive after finitely many steps at tuples of basic type \( w = (K_3, 0, \ldots, 0) \), that is, those in which all the transformations come from the same class \( I_1 \). Tuples of basic type will serve as the basis for our induction procedure. For instance, the tuple \( (T^{n^2}_{n_1}, T^{n_2}_{n_2}, T^{n_3+n}_{n_2}) \) from Example 1 has basic type \((3, 0)\) because it only involves the transformations \( T_1, T_2 \) whose indices belong to the set \( I_1 = \{1, 2\} \) (corresponding to the polynomial \( n^2 \)); however, the type \((2, 1)\) of \( (T^{n_1}_{n_1}, T^{n_2}_{n_2}, T^{n_3+n}_{n_3}) \) is not basic because this tuple involves both the transformations \( T_1, T_2 \) and the transformation \( T_3 \) with an index from the set \( I_2 = \{3\} \) (corresponding to the polynomial \( n^2 + n \)).

6. Further maneuvers and obstructions for longer families
Having presented the general strategy for proving Theorem 1.1 for longer families, we move on to discuss in detail the specific maneuvers outlined in §5.2. In this section, we state and prove various partial results that give substance to the moves discussed in §5.2. We also discuss a number of obstructions that appear in the process and have to be overcome before we can give a complete proof of Theorem 1.1. All of the above are illustrated with examples that will hopefully make the abstract statements in this and the next section more comprehensible to the reader. We then move on in §7 to prove Theorem 1.1.

The plan for this section is as follows. In §6.1, we discuss how to obtain a tuple of lower type in the ping step of the seminorm smoothing argument for controllable tuples. Section 6.2 exhibits the necessity of assuming that the functions appearing in the averages from equation (40) have some invariance properties. In particular, we show by examples how these properties are essentially used to tackle tuples of basic type and to perform the pong step of the seminorm smoothing argument for controllable tuples. We also give details of the flipping procedure that relies on these invariance properties. Subsequently, we discuss in §6.3 how flipping can be used to reduce an uncontrollable tuple to a controllable tuple of a lower type. Finally, we combine the details of the aforementioned moves in §6.4 and show how we can reach a tuple of a basic type in a finite number of steps.

6.1. Reducing controllable tuples to tuples of lower type in the ping step. As explained in §5.2, in the ping part of the smoothing argument, we will replace the original tuple \( (T^{ρ_j(n)}_{η_j})_{j∈[ℓ]} \) by a new tuple \( (T^{ρ'_j(n)}_{η'_j})_{j∈[ℓ]} \). The new indexing tuple \( η' \) will be defined via the operation

\[
(τ_m η)_{j} := \begin{cases} 
η_j, & j \neq m, \\
η_i, & j = m,
\end{cases}
\] (44)
for some distinct values $m, i \in \mathcal{I}$. This indexing tuple corresponds to replacing the term $T_{\eta_{m}^{(n)}}$ in $(T_{\eta_{j}^{(n)}})_{j \in [\ell]}$ by $T_{\eta_{i}^{(n)}}$, and all the other terms $T_{\eta_{j}^{(n)}}$ by $T_{\eta_{j}^{(n)}}$. The new tuple $(T_{\eta_{j}^{(n)}})_{j \in [\ell]}$ has to be chosen carefully: it must preserve the good ergodicity property and allow for seminorm control. Lastly, it must have a lower type. For this reason, we let $\rho_{m}$ be an index $j$ (n)

\begin{equation}
(\sigma_{1,2} w)_{t} := \begin{cases}
  w_{t}, & t \neq t_{1}, t_{2}, \\
  w_{1} - 1, & t = t_{1}, \\
  w_{t_{2}} + 1, & t = t_{2}.
\end{cases}
\end{equation}

For instance, $\sigma_{32}(3, 2, 2) = (3, 3, 1)$. As a consequence of our ordering on types, we have $\sigma_{1,2} w < w$ whenever $t_{2} < t_{1}$ (the assumption $w_{t_{2}} > 0$ is crucial here), so in particular $(3, 3, 1) < (3, 2, 2)$.

Proposition 6.1, which we are about to state now, specifies how these tuples of lower type are picked, what form they take, and what properties they enjoy. It will be used in our smoothing argument in Proposition 7.5 in that the tuple of lower type for which we invoke the induction hypothesis in the ping step is constructed in Proposition 6.1.

**Proposition 6.1. (Type reduction for controllable tuples)** Let $\ell \in \mathbb{N}$, $\eta \in [\ell]^{\ell}$ be an indexing tuple and $\rho_{1}, \ldots, \rho_{\ell} \in \mathbb{Z}[n]$ be polynomials with leading coefficients $b_{1}, \ldots, b_{\ell}$. Let also $(X, \mathcal{X}, \mu, T_{1}, \ldots, T_{\ell})$ be a system and $(T_{\eta_{j}^{(n)}})_{j \in [\ell]}$ be a tuple of a non-basic type $w$ whose last non-zero index is $t_{w}$. Suppose that $(T_{\eta_{j}^{(n)}})_{j \in [\ell]}$ is controllable. Then there exists $\lambda \in \mathbb{N}$ such that for every $r \in \{0, \ldots, \lambda - 1\}$, there exist an index $t_{w}' \in \text{Supp}(w)$ distinct from $t_{w}$, an index $i \in [\ell]$ with $\eta_{i} \in \mathcal{I}_{t_{w}'}$, and a tuple $(T_{\eta_{j}^{(n)}})_{j \in [\ell]}$ satisfying the following properties.

(i) The type $w'$ of the tuple $(T_{\eta_{j}^{(n)}})_{j \in [\ell]}$ satisfies $w' = \sigma_{t_{w} t_{w}'} w < w$.

(ii) The indexing tuple $\eta'$ is given by $\eta' := \tau_{m} \eta$ for some $m \in [\ell]$ satisfying the controllability condition (recall that $\tau_{m}$ is defined in equation (44)).

(iii) The polynomials $\rho_{1}, \ldots, \rho_{\ell}$ have integer coefficients and zero constant terms, and they take the form

$$
\rho_{j}^{(n)} := \begin{cases}
  \rho_{j}(\lambda n + r) - \rho_{j}(r), & j \neq m, \\
  b_{i} \cdot (\rho_{j}(\lambda n + r) - \rho_{j}(r)), & j = m.
\end{cases}
$$

We remark that when the leading coefficients of $\rho_{1}, \ldots, \rho_{\ell}$ are 1, then $\rho_{j} = \rho_{j}$ for every $j \in [\ell]$, so the property (iii) becomes trivial.

**Proof.** Let $t_{w}$ be the last non-zero index of the type $w$ of $(T_{\eta_{j}^{(n)}})_{j \in [\ell]}$. By the controllability of the tuple, there exists $m \in [\ell]$ with $\eta_{m} \in \mathcal{I}_{t_{w}}$ such that $\eta_{m'} = \eta_{m}$ implies that $\rho_{m'}$ and $\rho_{m}$ are distinct. Let $t_{w}' \in \text{Supp}(w)$ be an index different from $t_{w}$ (it exists since the
Definition. (Descendants) Let \( w \) be a non-basic index with \( \eta_i \in \mathcal{I}_{\ell} \). We define \( \eta' := \tau_{mi} \eta \), meaning that we replace \( T_{mi} \) by \( T_{\eta_i} \) and keep the other transformations the same.

We let \( \lambda \in \mathbb{N} \) be the smallest number for which \( (\lambda/b_m) \rho_m \in \mathbb{Z}[n] \) (equivalently, \( \lambda \) is the smallest number such that \( b_m \) divides the coefficients of the polynomials \( \rho_m(\lambda n + r) - \rho_m(r) \) for \( r \in \mathbb{Z} \)). We also fix an arbitrary \( r \in \{0, \ldots, \lambda - 1\} \). We then define the new polynomials \( \rho'_1, \ldots, \rho'_\ell \) by the formula

\[
\rho'_j(n) := \begin{cases} 
\rho_j(\lambda n + r) - \rho_j(r), & j \neq m, \\
b_i & j = m.
\end{cases}
\]

The new polynomials are in \( \mathbb{Z}[n] \) by the choice of \( \lambda \), and it is not hard to check that they have zero constant terms. Lastly, the new tuple \( (T_{\eta_j}^{\rho'_j(n)})_{j \in [\ell]} \) has the type \( w' = a_{\tau_w t_w} w \), which is strictly smaller than \( w \) by the assumption that \( t'_w < t_w \) (which follows from \( t_w' \neq t_w \) and the assumption that \( t_w \) is the last non-zero index).

Example 5. (Examples of type reduction) We show how Proposition 6.1 has been implicitly applied to the two tuples from §4.

(i) The tuple \( (T_1^{n^2}, T_2^{n^2}, T_3^{n^2+n}) \) discussed at length in §4 has type \((2, 1)\) corresponding to the partition \( \mathcal{J}_1 = \{1, 2\}, \mathcal{J}_2 = \{3\} \), and its good ergodicity property means that \( \mathcal{I}(T_1T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2) \), that is, the only \( T_1T_2^{-1} \)-invariant functions are those invariant simultaneously under \( T_1 \) and \( T_2 \). In the proof of Proposition 4.2, we applied the type reduction once (with the operation \( \tau_{32} \)) to obtain the new tuple \( (T_1^{n^2}, T_2^{n^2}, T_3^{n^2+n}) \) with indexing tuple \((1, 2, 2)\) and basic type \((3, 0)\).

(ii) The tuple \( (T_1^{n^2}, T_2^{n^2}, T_3^{2n^2+n}) \) presented at the end of §4 also has type \((2, 1)\) corresponding to the partition \( \mathcal{J}_1 = \{1, 2\}, \mathcal{J}_2 = \{3\} \), and its good ergodicity property also states that \( \mathcal{I}(T_1T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2) \). In the ping step of the smoothing argument, we obtained (upon taking \( r_0 = 1 \)) the new tuple \( (T_1^{4n^2+4n}, T_2^{4n^2+4}, T_2^{4n^2+5n}) \) by performing the operation \( \tau_{32} \). This new tuple also has the indexing tuple \((1, 2, 2)\) and basic type \((3, 0)\), and its ergodicity property is the same as for the original tuple.

Definition. (Descendants) Let \( p_1, \ldots, p_\ell \in \mathbb{Z}[n] \) be polynomials with leading coefficients \( a_1, \ldots, a_\ell \) and \( \eta \in [\ell] \) be an indexing tuple. We say that the polynomials \( \rho_1, \ldots, \rho_\ell \) are descendants of \( p_1, \ldots, p_\ell \) along \( \eta \), if there exists \( \lambda \in \mathbb{N} \) and \( r \in \{0, \ldots, \lambda - 1\} \) such that \( \rho_j(n) = a_\eta_j / a_\eta_j (p_j(\lambda n + r) - p_j(r)) \). If this is the case, we also say the tuple \( (T_{\eta_j}^{\rho'_j(n)})_{j \in [\ell]} \) is a descendant of the tuple \( (T_{j}^{\rho_j(n)})_{j \in [\ell]} \).

Descendancy enjoys the following transitivity property.

Lemma 6.2. (Descendancy is transitive) Suppose that the polynomials \( \rho_1, \ldots, \rho_\ell \in \mathbb{Z}[n] \) are descendants of \( p_1, \ldots, p_\ell \) along \( \eta \), and let \( \rho'_j(n) := a_{\eta'_j} / a_{\eta'_j} (\rho_j(\lambda' n + r') - \rho_j(r')) \) for all \( j \in [\ell] \), where \( \eta' \in [\ell] \), \( \lambda' \in \mathbb{N} \), \( r' \in \{0, \ldots, \lambda - 1\} \) and \( a_1, \ldots, a_\ell \) are the leading coefficients of \( p_1, \ldots, p_\ell \). Then \( \rho'_1, \ldots, \rho'_\ell \) are descendants of \( p_1, \ldots, p_\ell \) along \( \eta' \).
Proof. Let \( \lambda \in \mathbb{N} \) and \( r \in \{0, \ldots, \lambda - 1\} \) be such that \( \rho_j(n) = a_{\eta j} / a_j(p_j(\lambda n + r) - p_j(r)) \). Then a direct computation gives that

\[
\rho'_j(n) = \frac{a'_{\eta j}}{a_j}(p_j(\lambda' n + \lambda r' + r) - p_j(\lambda r' + r)),
\]
giving the claim.

In particular, we get the following corollary of interest to us that follows from a straightforward combination of Proposition 6.1 and Lemma 6.2.

**Corollary 6.3.** (Type reduction preserves descendancy) Let \( \ell \in \mathbb{N} \), \( \eta \in [\ell]^{\ell} \) be an indexing tuple, \( p_1, \ldots, p_\ell, \rho_1, \ldots, \rho_\ell \in \mathbb{Z}[n] \) be polynomials, and \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\) be a system. Suppose that \((T_{\eta j}^{\rho_j(n)})_{j \in [\ell]}\) is a tuple of a non-basic type \( \omega \) that is a descendant of \((T_j^{p_j(n)})_{j \in [\ell]}\). Then the tuple \((T_{\eta j}^{\rho'_j(n)})_{j \in [\ell]}\) constructed from \((T_{\eta j}^{\rho_j(n)})_{j \in [\ell]}\) in Proposition 6.1 is also a descendant of \((T_j^{p_j(n)})_{j \in [\ell]}\).

**Proof.** Suppose that \( \rho_1, \ldots, \rho_\ell \) are descendants of \( p_1, \ldots, p_\ell \) along \( \eta \) by assumption. Letting \( a_j, b_j \) be the leading coefficients of \( p_j, \rho_j \), respectively, and \( \rho'_j \) be as defined in Proposition 6.1, we have \( b_j = a_{\eta j}(\lambda')^d_j \) for some \( \lambda \in \mathbb{Z} \), where \( d_j := \deg p_j = \deg \rho_j = \deg \rho'_j \). Thus,

\[
\rho'_j(n) = \rho_j(\lambda n + r) - \rho_j(r) = \frac{a'_{\eta j}}{a_j}(\rho_j(\lambda n + r) - \rho_j(r))
\]

for \( j \neq m \) and

\[
\rho'_j(n) = \frac{b_i}{b_m}(\rho_j(\lambda n + r) - \rho_j(r)) = \frac{a_{\eta j}}{a_{\eta m}}(\rho_j(\lambda n + r) - \rho_j(r))
\]

for \( j = m \), where we use \( \eta'_m = \eta_i \) and \( \eta'_j = \eta_j \) for \( j \neq m \). Hence, the polynomials \( \rho'_1, \ldots, \rho'_\ell \) satisfy the condition of Lemma 6.2, implying the claim.

Descendant tuples enjoy the following important properties.

**Proposition 6.4.** (Properties of descendants) Let \( \ell \in \mathbb{N} \), \( \eta \in [\ell]^{\ell} \) be an indexing tuple, \( p_1, \ldots, p_\ell, \rho_1, \ldots, \rho_\ell \in \mathbb{Z}[n] \) be polynomials, and \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\) be a system. Suppose that \((T_j^{p_j(n)})_{j \in [\ell]}\) has the good ergodicity property and \((T_{\eta j}^{\rho_j(n)})_{j \in [\ell]}\) is its descendant. Then the following hold.

(i) We have

\[
\Sigma(\rho_1, \ldots, \rho_\ell) = \Sigma(p_1, \ldots, p_\ell),
\]

\[
K_i(\rho_1, \ldots, \rho_\ell) = K_i(p_1, \ldots, p_\ell), \quad i \in [3],
\]

\[
I_t(\rho_1, \ldots, \rho_\ell) = I_t(p_1, \ldots, p_\ell), \quad t \in [K_1].
\]

(ii) The tuple \((T_{\eta j}^{\rho'_j(n)})_{j \in [\ell]}\) has the good ergodicity property.
Property (i) ensures that when passing to descendants, we do not need to redefine the partition $\mathcal{J}_1, \ldots, \mathcal{J}_{K_1}$. Property (ii) is crucial because it shows that descendants retain the essential ergodicity properties of the original tuple.

**Proof.** Part (i) follows from the fact that for every $j \in [\ell]$, the polynomials $p_j$ and $\rho_j$ have the same degree, and that $p_j$, $p_j'$ are linearly dependent if and only if $\rho_j$, $\rho_j'$ are. We therefore move on to proving part (ii). Let $b_j$ be the leading coefficient of $\rho_j$, $a_j$ be the leading coefficient of $p_j$, and $d_j := \deg p_j = \deg \rho_j$ for every $j \in [\ell]$. To check that the tuple $(T^{\rho_j(n)} \eta_j(j))_{j \in [\ell]}$ has the good ergodicity property, we need to show that if $\eta_{j_1}, \eta_{j_2}$ are distinct elements of the same set $\mathcal{J}_j$, then
\[
\mathcal{I}(T^{\rho_{j_1}} \eta_{j_1} - \rho_{j_2}) = \mathcal{I}(T^{\eta_{j_1}}) \cap \mathcal{I}(T^{\eta_{j_2}}),
\]
where
\[
\beta_j := b_j / \gcd(b_{j_1}, b_{j_2})
\]
for $j = j_1, j_2$. By construction, $b_j = a_{j_1} \lambda^{d_j}$ for some $\lambda \in \mathbb{N}$, and so
\[
\beta_j = a_{j_1} / \gcd(a_{j_1}, a_{j_2}) =: \alpha_{j_1}
\]
for $j = j_1, j_2$. The assumption $\eta_{j_1}, \eta_{j_2} \in \mathcal{J}_t$ for some fixed $t$ implies that $p_{\eta_{j_1}}, p_{\eta_{j_2}}$ are linearly dependent, and additionally $p_{\eta_{j_1}} / \alpha_{\eta_{j_1}} = p_{\eta_{j_2}} / \alpha_{\eta_{j_2}}$. Since $\alpha_{\eta_{j_1}}, \alpha_{\eta_{j_2}}$ are coprime, the good ergodicity property of $(T^{\rho_j(n)} \eta_j(j))_{j \in [\ell]}$ implies that
\[
\mathcal{I}(T^{\alpha_{\eta_{j_1}}} \eta_{j_1} - \alpha_{\eta_{j_2}}) = \mathcal{I}(T^{\eta_{j_1}}) \cap \mathcal{I}(T^{\eta_{j_2}}).
\]
The equality in equation (46) follows from this and the identification in equation (47).

**Example 6.** (Type reduction for non-monic polynomials) We present one more example to show how Proposition 6.1 is applied iteratively for more complicated tuples, and how properties listed in Proposition 6.4 are retained when passing to lower-type descendant tuples. Consider the tuple
\[
(T_1^{n^2}, T_2^{3n^2}, T_3^{2n^2}, T_4^{2n^2+n}, T_5^{n^2+n}, T_6^{n^2+n}, T_7^n),
\]
and assume that it has the good ergodicity property, that is,
\[
\mathcal{I}(T_1 T_2^{-3}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2), \quad \mathcal{I}(T_1 T_3^{-2}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_3),
\]
\[
\mathcal{I}(T_2^{-3} T_3^{-2}) = \mathcal{I}(T_2) \cap \mathcal{I}(T_3), \quad \mathcal{I}(T_5 T_6^{-1}) = \mathcal{I}(T_5) \cap \mathcal{I}(T_6).
\]
This tuple has length 7, degree 2, $K_1 = 5$, $K_2 = 4$, $K_3 = 6$, and $\mathcal{L} = \{1, 2, 3, 4, 5, 6\}$. If we define the partition $\mathcal{J}_1 = \{1, 2, 3\}$, $\mathcal{J}_2 = \{5, 6\}$, $\mathcal{J}_3 = \{4\}$, $\mathcal{J}_4 = \{7\}$, then the tuple has type $w_0 = (3, 2, 1)$; we recall that the term $T_7^n$ plays no part in the type consideration since the polynomial $\rho_{07}(n) = n$ has a lower degree. (Perhaps a more natural way to define the partition would be to have $\mathcal{J}_1 = \{1, 2, 3\}$, $\mathcal{J}_2 = \{4\}$, $\mathcal{J}_3 = \{5, 6\}$, $\mathcal{J}_4 = \{7\}$. However, then the tuple would have type $(3, 1, 2)$, and reducing to the tuple of basic type by iteratively applying Proposition 6.1 would take more steps. This shows that choosing the partition strategically can save on the number of iterations of Proposition 6.1.
needed to reach a tuple of basic type.) The tuple in equation (48) has the basic indexing tuple \( \eta_0 = (1, 2, 3, 4, 5, 6, 7) \). The tuple is controllable, and 4 satisfies the controllability condition, so in the first step, we replace \( T_4 \) (this corresponds to us wanting to first get a \( T_4 \)-seminorm control over the tuple in equation (48)). We are then provided with an index \( i_0 \in I_1 \cup I_2 \) (say, \( i_0 = 1 \)), and we get the new indexing tuple

\[
\eta_1 := \tau_{m_0} \eta_0 = \tau_{41} \eta_0 = (1, 2, 3, 1, 5, 6, 7).
\]

The leading coefficient 2 of \( 2n^2 + n \) does not divide the linear coefficient, and the smallest \( \lambda_0 \in \mathbb{N} \) such that 2 divides the coefficients of \( \lambda_0(2n^2 + n) \) is \( \lambda_0 = 2 \). In performing the ping step of the seminorm smoothing argument for the tuple in equation (48), we will want to apply the \( T_4^{-2} \) invariance of some function \( u \) to replace \( T_4^{2n^2+n} u \) by \( T_4^{q(n)} u' \) for some \( q \in \mathbb{Z}[n] \) and a function \( u' \) related in some way to \( u \). We cannot do this directly since \( \frac{1}{2}(2n^2 + n) \notin \mathbb{Z}[n] \), but we can do this ‘piecewise’ by splitting \( \mathbb{N} \) into arithmetic progressions \( (2N + r)_{r=0,1} \), and considering the two cases separately (see the sketch of the seminorm smoothing argument for \( n^2, n^2, 2n^2 + n \) at the end of §4 to see how this was done for that family). We therefore replace the original polynomials \( \rho_0, \ldots, \rho_7 \) by new polynomials

\[
\rho_{1j}(n) := \begin{cases} 
\rho_{0j}(2n + r_0) - \rho_{0j}(r_0), & j \neq 4, \\
\frac{1}{2}(\rho_{04}(2n + r_0) - \rho_{04}(r_0)), & j = 4,
\end{cases}
\]

for some \( r_0 \in \{0, 1\} \) (the choice of \( r_0 \) is not ours). Assuming that \( r_0 = 1 \), we obtain the new tuple

\[
(T_4^{4n^2+4n}, T_2^{12n^2+12n}, T_3^{8n^2+8n}, T_4^{4n^2+5n}, T_5^{4n^2+6n}, T_6^{4n^2+6n}, T_7^{2n}).
\]

(49)

The type of the new tuple is \( w_1 = \sigma_{31} w_0 = (4, 2, 0) \) since we now have four transformations with indices coming from \( I_1 \) and two transformations coming from \( I_2 \). This type is lower than the original type \( w_0 \), and so we have successfully obtained a tuple of lower type. The new tuple is controllable, with \( m = 5, 6 \) both satisfying the controllability condition.

Although we replaced the polynomials \( \rho_0, \ldots, \rho_7 \) by new ones, we note that for any \( j_1, j_2 \in [7] \), the polynomials \( \rho_{1j_1}, \rho_{1j_2} \) are pairwise dependent if and only if \( \rho_{0j_1}, \rho_{0j_2} \) are, and not only that: if they are pairwise dependent, then \( \rho_{1j_1}/c_1 = \rho_{1j_2}/c_2 \) if and only if \( \rho_{0j_1}/c_1 = \rho_{0j_2}/c_2 \) for any non-zero integers \( c_1, c_2 \). Moreover, if \( \eta_{1j_1} = \eta_{1j_2} \), then the leading coefficients of \( \rho_{1j_1} \) and \( \rho_{1j_2} \) are identical. These two observations ensure that the ergodicity conditions on \( T_1T_2^{-3}, T_1T_3^{-2}, T_2^2T_3^{-2}, T_5T_6^{-1} \), which constitute the assumption that the original tuple in equation (48) has the good ergodicity property, carry on to the new tuple in equation (49), implying that it also enjoys the good ergodicity property. This exemplifies the claim from Proposition 6.4 that descendants of tuples with the good ergodicity property inherit the property.

The type \( w_1 \) is not basic, and so we continue the procedure. This time, we pick some \( m_1 \in \mathcal{I}_2 \), say \( m_1 = 5 \) (it satisfies the controllability condition, as does 6, the other possible choice). We are then handed an index \( i_1 \in I_1 \) (say, \( i_1 = 3 \)), so that

\[
\eta_2 := \tau_{m_1} \eta_1 = \tau_{53} \eta_1 = (1, 2, 3, 1, 3, 6, 7).
\]
The leading coefficient 4 of $4n^2 + 6n$ does not divide the linear term, and so we replace the polynomials $\rho_1, \ldots, \rho_{17}$ by new polynomials of the form

$$\rho_{2j}(n) := \begin{cases} 
\rho_{1j}(2n + r_1) - \rho_{1j}(r_1), & j \neq 5, \\
\frac{8}{2}(\rho_{15}(2n + r_1) - \rho_{15}(r_1)), & j = 5,
\end{cases}$$

for some $r_1 \in \{0, 1\}$ (we pass from $n$ to $2n + r_1$ because 2 is the smallest natural number $\lambda_1$ such that the leading coefficient 4 of $\rho_{15}$ divides the coefficients of $\lambda_1 \rho_{15}$). Hence, the new tuple takes the form (upon assuming $r_1 = 0$)

$$(T_1^{16n^2 + 8n}, T_2^{48n^2 + 24n}, T_3^{32n^2 + 16n}, T_1^{16n^2 + 10n}, T_3^{32n^2 + 24n}, T_6^{16n^2 + 12n}, T_7^{4n}). \quad (50)$$

The tuple in equation (50) still has the good ergodicity property; this is once again a consequence of two facts:

- for any $j_1, j_2 \in [7]$ and non-zero $c_1, c_2 \in \mathbb{Z}$, we have $\rho_{2j_1}/c_1 = \rho_{2j_2}/c_2$ if and only if $\rho_{0j_1}/c_1 = \rho_{0j_2}/c_2$;

- for any $j_1, j_2 \in [7]$, if $\eta_{2j_1} = \eta_{2j_2}$, then $\rho_{2j_1}, \rho_{2j_2}$ have identical leading coefficients.

We remark though that to ensure the ergodicity property of equation (50), we no longer need the original assumption $\mathcal{I}(T_5 T_6^{-1}) = \mathcal{I}(T_5) \cap \mathcal{I}(T_6)$ because the transformation $T_5$ is not present.

The new tuple in equation (50) has type $w_2 = \sigma_21 w_1 = (5, 1, 0)$, which is still not basic, and so we continue the procedure one more time. The only index left in $\mathcal{J}_1$ is 6, and it satisfies the controllability assumption, so we replace $T_6$ this time. We are given an index $i_2 \in \mathcal{J}_1$ (say, $i_2 = 1$), so that

$$\eta_3 := \tau_{m_2} \eta_1 = (1, 2, 3, 1, 3, 1, 7).$$

Since the leading coefficient 16 of $\rho_{21}$ does not divide the coefficients of the polynomial $\rho_{26}(n) = 16n^2 + 12n$, and the smallest $\lambda_2 \in \mathbb{N}$ for which 16 divides the coefficients of $\lambda_2 \rho_{26}(n) = \lambda_2(16n^2 + 12n)$ is $\lambda_2 = 4$, we define the new polynomials to be

$$\rho_{3j}(n) := \begin{cases} 
\rho_{2j}(4n + r_2) - \rho_{2j}(r_2), & j \neq 6, \\
16(\rho_{26}(4n + r_2) - \rho_{26}(r_2)), & j = 6,
\end{cases}$$

for some $r_2 \in \{0, 1, 2, 3\}$. Assuming, say, $r_2 = 3$, we get the new tuple

$$(T_1^{256n^2 + 416n}, T_2^{768n^2 + 1248n}, T_3^{512n^2 + 832n}, T_1^{256n^2 + 424n}, T_3^{512n^2 + 864n}, T_1^{256n^2 + 432n}, T_7^{16n}).$$

This tuple has the basic type $w_2 = \sigma_21 w_2 = (6, 0, 0)$, and so the procedure halts. A similar argument as before shows also that it enjoys the good ergodicity property.

Lastly, we observe that $\eta_3 |_{\mathcal{J}_1} = \eta_3 |_{\{1, 2, 3\}}$ is constant, that is, while performing the type reduction procedure, we did not replace the transformations at indices from $\mathcal{J}_1$. This is a special case of property (vi) from Proposition 6.8, which will play an important role in the proof of Proposition 7.2, a seminorm control argument for tuples of basic type.

6.2. The role of invariance properties. Proposition 6.1 ensures that the lower type tuples to which we pass in the ping step of the smoothing argument have the good ergodicity properties.
property. However, this is not enough. For more complicated tuples, we also need to assume that the functions appearing in the associated average have some invariance properties, otherwise the induction breaks. The example that we present now displays the necessity of this extra information. We sketch how—reducing the original tuple to tuples of shorter length or lower type—we eventually arrive at averages for which we cannot obtain seminorm control unless the functions appearing in the averages satisfy certain invariance properties. We emphasize that our goal in this example is not to give a complete proof of seminorm control, but rather to point out the necessity of the invariance assumptions. Therefore, we assume without proof when convenient that we have seminorm control over certain tuples of lower type or shorter length.

Example 7. (The necessity of invariance properties) Consider the average

$$\mathbb{E}_{n \in [N]} T_1^n f_1 \cdot T_2^n f_2 \cdot T_3^{n+2} f_3 \cdot T_4^{n+4} f_4.$$  \hfill (51)

It has length 4, degree 2, and type $w = (2, 2)$, corresponding to the partition $I_1 = \{1, 2\}$, $I_2 = \{3, 4\}$. Suppose that equation (51) has the good ergodicity property, that is,

$$I(T_1 T_2^{-1}) = I(T_1) \cap I(T_2) \quad \text{and} \quad I(T_3 T_4^{-1}) = I(T_3) \cap I(T_4).$$

We illustrate the steps that need to be taken to show that this average is controlled by $\|f_4\|_{s,T_4}$ for some $s \in \mathbb{N}$.

By Proposition 3.7, the average from equation (51) is controlled by the seminorm $\|f_4\|_{b_1,\ldots,b_{s+1}}$ for some vectors

$$b_1, \ldots, b_{s+1} \in \{e_4, e_4 - e_3, e_4 - e_2, e_4 - e_1\}.$$  

We want to replace the vector $b_{s+1}$ by (multiple copies of) $e_4$. Iterating this procedure gives a control of equation (51) by a $T_d$-seminorm of $f_4$.

If $b_{s+1} = e_4$, there is nothing to check. If $b_{s+1} = e_4 - e_3$, then this is the consequence of the good ergodicity property of the average and Lemma 2.2. So the only cases to check are when $b_{s+1}$ equals $e_4 - e_2$ or $e_4 - e_1$. Without loss of generality, we assume that $b_{s+1} = e_4 - e_2$.

Suppose that the limit of equation (51) is non-zero. Arguing as in the proof of Proposition 4.1, we deduce that

$$\lim_{H \to \infty} \mathbb{E}_{h,h' \in [H]} \lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} T_1^n f_{1,h,h'} \cdot T_2^n f_{2,h,h'} \cdot T_3^{n+2} f_{3,h,h'} \cdot T_4^{n+4} u_{h,h'} \right\|_{L^2(\mu)} > 0$$

for some $T_3 T_2^{-1}$-invariant functions $u_{h,h'}$ as well as functions $f_{j,h,h'} := \Delta_{b_1,\ldots,b_j \cdot h,h' - h} f_j$ for $j \in [4]$. The invariance property of the functions $u_{h,h'}$ implies that

$$\lim_{H \to \infty} \mathbb{E}_{h,h' \in [H]} \lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} T_1^n f_{1,h,h'} \cdot T_2^n f_{2,h,h'} \cdot T_3^{n+2} f_{3,h,h'} \cdot T_4^{n+4} u_{h,h'} \right\|_{L^2(\mu)} > 0.$$  \hfill (52)
Each of the averages inside the liminf above is of the form

\[
\mathbb{E}_{n \in [N]} T_1^{n^2} g_1 \cdot T_2^{n^2} g_2 \cdot T_3^{n^2 + n} g_3 \cdot T_2^{n^2 + n} g_4
\]  

for 1-bounded functions \( g_1, g_2, g_3, g_4 \in L^\infty(\mu) \) of which \( g_4 \) is \( T_4 T_2^{-1} \)-invariant. The averages from equation (53) are controllable, with 3 satisfying the controllability condition, and they have type \((3,1)\), which is lower than the type \((2,2)\) of the original average from equation (51). Assuming inductively that we have the seminorm control of averages from equation (53) by a \( T_3 \)-seminorm of \( f_3 \) (while we only use this particular control, our inductive assumption will guarantee that we control averages from equation (53) by a relevant seminorm of other functions, too), we can deduce from equation (52) (like in the proof of Proposition 4.2) that

\[
\| f_3 \|_{b_1, \ldots, b_s, e_3^{s_1}} > 0
\]

for some \( s_1 \in \mathbb{N} \), and similarly for other terms. This completes the ping step. For the pong step, this auxiliary control and Proposition 3.3 imply that

\[
\liminf_{H \to \infty} \mathbb{E}_{h, h' \in [H]'} \lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} T_1^{n^2} f_{1, h, h'} \cdot T_2^{n^2} f_{2, h, h'} \times \prod_{j=1}^{2'} D_j(n^2 + n) \cdot T_4^{n^2 + n} f_{4, h, h'} \right\|_{L^2(\mu)} > 0
\]

for some \( D_j \in D_{s_1} \). Each average in equation (54) takes the form

\[
\mathbb{E}_{n \in [N]} T_1^{n^2} g_1 \cdot T_2^{n^2} g_2 \cdot \prod_{j=1}^{2'} D_j(n^2 + n) \cdot T_4^{n^2 + n} g_4.
\]

Assuming inductively that averages of the form in equation (55) are controlled by a \( T_4 \)-seminorm of the last term, we get the desired claim \( \| f_4 \|_{b_1, \ldots, b_s, e_4^{s'}} > 0 \) for some \( s' \in \mathbb{N} \) using an argument similar to one in the proof of Proposition 4.2.

We have showed how a seminorm control of the original average from equation (51) by a \( T_4 \)-seminorm of \( f_4 \) follows from the seminorm control of the averages from equations (53) and (55). We have not proved, however, that these auxiliary averages are indeed controlled by Gowers–Host–Kra seminorms, assuming instead that this follows by induction. It turns out that obtaining a seminorm control of the averages from equation (53) involves an interesting twist in that the argument makes essential use of the assumption that the function \( g_4 \) is \( T_4 T_2^{-1} \)-invariant. We sketch the steps taken in the seminorm smoothing argument for this average under the extra invariance assumption to show where this invariance property comes up and why it is necessary.

In proving the seminorm control of equation (53), we first prove that the average is controlled by a \( T_3 \)-seminorm of \( g_3 \) since \( T_3 \) is the only transformation with index in \( I_2 \). Arguing as above (using Proposition 3.7 for equation (53), assuming that the \( L^2(\mu) \) limit
of equation (53) is non-zero and mimicking the proof of Proposition 4.2), we deduce that
\[ \lim_{H \to \infty} \lim_{n \to \infty} \mathbb{E} T_n^2 g_1 \cdot T_2^2 g_2 \cdot T_3^2 g_3 \cdot T_4^2 g_4 \geq 0 \]
for 1-bounded functions \( u_{h, h'} \) that are all invariant either under \( T_3 T_1^{-1} \) or under \( T_3 T_2^{-1} \). Then we use the relevant invariance property to replace each \( T_3^2 g_3 \) by \( T_1^2 g_1 \) or \( T_2^2 g_2 \). Hence, in the ping step of the seminorm smoothing argument for equation (53), we need to invoke seminorm control of averages of the form
\[ \mathbb{E} T_1^2 g_1 \cdot T_2^2 g_2 \cdot T_3^2 g_3 \cdot T_4^2 g_4, \]
and
\[ \mathbb{E} T_1^2 g_1' \cdot T_2^2 g_2' \cdot T_3^2 g_3' \cdot T_4^2 g_4', \]
where \( g_4' \) is \( T_4 T_2^{-1} \)-invariant while \( g_3' \) is invariant under \( T_3 T_1^{-1} \) and \( T_3 T_2^{-1} \) respectively. Both of them have basic type.

We show that for arbitrary \( g_1', g_2', g_3', g_4' \), without the aforementioned invariance assumptions, we would not be able to control the average from equation (56) by Gowers–Host–Kra seminorms; specifically, we could not control it by a \( T_2 \)-seminorm of \( g_4' \). Conversely, this is achievable if \( g_3', g_4' \) are invariant under \( T_3 T_2^{-1} \), \( T_4 T_2^{-1} \), respectively. Assuming for simplicity that \( g_1' = g_2 := 1 \), we have that the second average equals

\[ \mathbb{E} T_1^2 g_1 \cdot T_2^2 g_2 \cdot T_3^2 g_3 \cdot T_4^2 g_4 = \mathbb{E} T_1^2 g_1' \cdot T_2^2 g_2' \cdot T_3^2 g_3' \cdot T_4^2 g_4', \]

and so without any additional assumptions, the average from equation (56) is in general not controlled by a \( T_2 \)-seminorm of \( g_3' \) or a \( T_2 \)-seminorm of \( g_4' \). However, the invariance assumptions on \( g_3', g_4' \) give us

\[ \mathbb{E} T_2^2 g_3' \cdot g_4' = \mathbb{E} T_3^2 g_3' \cdot T_4^2 g_4', \]

and by Proposition 3.7, these averages can be controlled by \( \| g_4' \|_{e_4' \cdot (e_4 - e_3)} \cdot s \) for some \( s \in \mathbb{N} \). Then the assumption \( \mathcal{I}(T_4 T_3^{-1}) \subseteq \mathcal{I}(T_4) \) and Lemma 2.2 give \( \| g_4' \|_{e_4' \cdot (e_4 - e_3)} \cdot s \leq \| g_4' \|_{e_3} \), and so a \( T_4 \)-seminorm of \( g_4' \) does control the average from equation (57). Using the invariance property once again, this time along with Lemma 3.5, we get \( \| g_4' \|_{e_3} \cdot T_4 = \| g_4' \|_{e_3} \cdot T_2 \), so a \( T_2 \)-seminorm of \( g_4' \) controls equation (57) and hence equation (56).

To get control over equation (56) by a \( T_2 \)-seminorm of \( g_4' \) without any simplifying assumptions on \( g_1', g_2' \), we have to run a more complicated argument. Combining Proposition 3.7, the ergodic condition on \( T_1 T_2^{-1} \) and Lemma 2.2, we first obtain control of equation (56) by a \( T_1 \)-seminorm of \( g_1' \) and a \( T_2 \)-seminorm of \( g_2' \). Assuming that the \( L^2(\mu) \) limit of the average from equation (56) is positive, we use this newly established control, decompose \( g_1' \) using Proposition 2.3, and apply the pigeonhole principle to show that the average

\[ \mathbb{E} T_1^2 g_1' \cdot T_2^2 g_2' \cdot T_3^2 g_3' \cdot T_4^2 g_4' \]
has a non-vanishing limit. The invariance properties of $g'_3$ and $g'_4$ imply that the average from equation (58) equals
\[
\mathbb{E}_{n \in [N]} \ D(n^2) \cdot T_2^{n^2} g'_2 \cdot T_3^{n^2+n} g'_3 \cdot T_4^{n^2+n} g'_4.
\]
for which a seminorm control by a $T_4$-seminorm of $g'_4$ follows by inductively invoking seminorm control for averages of length 3. The invariance property of $g'_4$ implies once again that $\|g'_4\|_{s', T_4} = \|g'_4\|_{s', T_2}$ for any $s' \in \mathbb{N}$. It follows that a $T_2$-seminorm of $g'_4$ controls equation (59), and hence also equations (58) and (56).

The invariance properties also come up in the pong step of the smoothing argument for equation (53). In this part, we encounter averages of the form
\[
\mathbb{E}_{n \in [N]} \prod_{j=1}^{L} D_j(n^2) \cdot T_2^{n^2} g''_2 \cdot T_3^{n^2+n} g''_3 \cdot T_4^{n^2+n} g''_4.
\]
Moreover, the function $g''_4$ is $T_4 T_2^{-1}$-invariant because it is essentially a multiplicative derivative of $g'_4$. By a similar reason as before, such averages could not be controlled by Gowers–Host–Kra seminorms for arbitrary $g''_2$, $g''_3$, $g''_4$ without the invariance assumption. However, thanks to the invariance assumption, the average from equation (60) equals
\[
\lim_{N \to \infty} \mathbb{E}_{n \in [N]} \prod_{j=1}^{L} D_j(n^2) \cdot T_2^{n^2} g''_2 \cdot T_3^{n^2+n} g''_3 \cdot T_4^{n^2+n} g''_4,
\]
which is controlled by $\|g''_4\|_{s', T_4}$ for some $s' \in \mathbb{N}$. (We can assume this inductively, or we can prove that a $T_4$-seminorm of $g''_4$ controls equation (61) in essentially the same way as we argued in Proposition 4.2.) Using the invariance property of $g''_4$ again together with Lemma 3.5, we deduce that $\|g''_4\|_{s', T_4} = \|g''_4\|_{s', T_2}$, and so a $T_2$-seminorm of $g''_4$ does control equation (60). An argument similar to the one at the end of the proof of Proposition 4.2 implies that a $T_2$-seminorm of $g_4$ controls equation (53).

The example above shows that it is crucial to keep track of the invariance properties of the functions appearing in our averages; these invariance properties turn out to be indispensable for applying Proposition 3.7 to averages of basic type, while obtaining seminorm control in the pong step of the argument, or—as we see later on—for handling uncontrollable averages. Using the invariance property to replace an average like equations (58) and (60) for which we cannot have seminorm control by an average like equations (59) and (61), respectively, which is controlled by Gowers–Host–Kra seminorms, exemplifies the flipping technique that will be presented in detail in Proposition 6.6.

Recalling how we have performed the ping and pong steps in Examples 1, 2, and 7, we observe that in the ping step, we pass from the average
\[
\mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{\eta_j(n)} f_j \cdot \prod_{j \in [L]} D_j(q_j(n))
\]
(62)
Moreover, the functions $u_{h'}$ are invariant under $T_{\eta i}^\rho_m$ where $\rho_m$ are the leading coefficients of $\rho_m$ and $\rho_i$, but they also retain whatever invariance property $f_m$ has. Thus, by passing from equation (62) to equation (63), we do not lose any invariance properties of the original functions, but rather gain new ones.

Similarly, in the pong step, we pass to averages

$$\mathbb{E}_{n \in [N]} \prod_{j \in [\ell], \ j \neq i} T_{\eta_j}^{\rho_j(n)} f_{j,h',h'} \cdot T_{\eta_i}^{\rho_i(n)} u_{h'} \cdot \prod_{j \in [L']} D_j'(q_j'(n)),$$

where $f_{j,h',h'} := \Delta_{b_1, \ldots, b_s} f_j$ for some vectors $b_1, \ldots, b_s \in \mathbb{Z}^\ell$. In particular, the functions $f_{j,h',h'}$ are invariant under whatever transformations the functions $f_j$ are invariant. Moreover, the functions $u_{h'}$ are invariant under $T_{\eta m}^{a_m} T_{\eta i}^{-a_i}$, where $a_m$ and $a_i$ are the leading coefficients of $\rho_m$ and $\rho_i$.

In our arguments, we will ensure that the functions $f_j$ have the good invariance property with respect to the original polynomials $p_1, \ldots, p_{\ell}$, that is, there exists $\gamma \in \mathbb{N}$ such that for every $j \in [\ell]$, the function $f_j$ is invariant under $(T_{\eta j}^{a_j} T_j^{-a_j})^\gamma$. Let $I$ be a (possibly infinite) indexing set. We say that a collection $(f_1, \ldots, f_{\ell}) \in I$ has the good invariance property along $\eta$ with respect to polynomials $p_1, \ldots, p_{\ell}$ if there exists $\gamma \in \mathbb{N}$ such that $(f_1, \ldots, f_{\ell}) \in I$ has the $\gamma$-invariance property along $\eta$ with respect to $p_1, \ldots, p_{\ell}$ for every $i \in I$.

If $\eta = (1, \ldots, \ell)$ is the identity tuple, there is nothing to check and any collection of functions has the 1-invariance property with respect to any polynomial family. The property only becomes non-trivial when $\eta$ is not the identity tuple.

In our arguments, we will ensure that the functions $f_j$ in the average

$$\mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{\eta j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L']} D_j(q_j(n)),$$

obtained by a sequence of reductions from the original average $\mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_j^{p_j(n)}$, have the good invariance property with respect to the original polynomials $p_1, \ldots, p_{\ell}$, that is, there exists $\gamma \in \mathbb{N}$ such that for every $j \in [\ell]$, the function $f_j$ is invariant under $(T_{\eta j}^{a_j} T_j^{-a_j})^\gamma$, where $a_j$ is the leading coefficient of the polynomial $p_j$. The need to keep
track of the invariance property with respect to the original polynomials is explained in Example 8 below. Before we state this example, however, we prove that the invariance properties get preserved when passing from the tuple of functions \((f_1, \ldots, f_\ell)\) to the tuples in equations (64) and (65).

**Proposition 6.5.** (Propagation of invariance properties) Let \(\ell, s \in \mathbb{N}, \eta \in \{\ell\}^\ell\) be an indexing tuple, \(\eta' := \tau_m^\ell \eta\) for distinct \(m, \ell \in \{\ell\}^\ell\) be another indexing tuple, \(p_1, \ldots, p_\ell \in \mathbb{Z}[n]\) be polynomials with leading coefficients \(a_1, \ldots, a_\ell, (X, X, \mu, T_1, \ldots, T_\ell)\) be a system, and \(b_1, \ldots, b_s \in \mathbb{Z}^\ell\) be vectors. Suppose that for some \(\gamma \in \mathbb{N}\), the functions \((f_1, \ldots, f_\ell)\) have the \(\gamma\)-invariance property along \(\eta\) with respect to the polynomials \(p_1, \ldots, p_\ell\). Consider the functions \((f_{1, \eta}^j, \ldots, f_{\ell, \eta}^j)_{h \in \mathbb{Z}^s}\), where \(f_{j, \eta}^j := \Delta_{b_1, \ldots, b_s, h, f_j}^{a_\ell}(T, \eta)\) for \(j \neq m\), and \(f_{m, \eta}^j\) is a function invariant under both \(S_1 := (T_{\eta_1}^\ell T_{\eta_m}^\ell T_{\eta_m}^\ell T_{\eta_m}^\ell)^{\gamma_1}\) and \(S_2 := (T_{\eta_1}^\ell T_{\eta_m}^\ell T_{\eta_m}^\ell T_{\eta_m}^\ell)^{\gamma_2}\) for some \(\gamma_1, \gamma_2 \in \mathbb{N}\) independent of \(h\). Then \((f_{1, \eta}^j, \ldots, f_{\ell, \eta}^j)_{h \in \mathbb{Z}^s}\) has the good invariance property along \(\eta'\) with respect to \(p_1, \ldots, p_\ell\).

**Proof of Proposition 6.5.** For \(j \neq m\), the functions \(f_j\) are invariant under \((T_{\eta_1}^\ell T_{\eta_2}^\ell)^{\gamma_2}\) for some non-zero \(\gamma \in \mathbb{Z}\) independent of \(h \in \mathbb{Z}^s\), and so are their translations

\[C_{\ell\eta}^\ell T_{e_1, h_1, \ldots, e_s, h_s} f_j.\]

The identity \(\eta' = \eta, j, \) which holds for \(j \neq m\), and the fact that \(f_{j, \eta}^j\) is a product of \((T_{\eta_1}^\ell T_{\eta_2}^\ell)^{\gamma_2}\)-invariant functions, imply that \(f_{j, \eta}^j\) is itself invariant under \((T_{\eta_1}^\ell T_{\eta_2}^\ell)^{\gamma_2}\).

For \(j = m\), the functions \(f_{j, \eta}^j\) are invariant under

\[S_1^\gamma_1 S_2^\gamma_2 = (T_{\eta_1}^\ell T_{\eta_2}^\ell T_{\eta_1}^\ell T_{\eta_2}^\ell)^{\gamma_1 \gamma_2} = (T_{\eta_1}^\ell T_{\eta_2}^\ell)^{\gamma_1 \gamma_2} = (T_{\eta_1}^\ell T_{\eta_2}^\ell)^{\gamma_1 \gamma_2}\]

by noting \(\eta''_m = \eta_i\) and combining the two invariance properties that these functions enjoy. Letting \(\gamma' := \text{lcm}(\gamma, \gamma_1 \gamma_2)\), it follows that for every \(h \in \mathbb{Z}^s\), the collection \((f_{1, \eta}^j, \ldots, f_{\ell, \eta}^j)\) has the \(\gamma'\)-invariance property along \(\eta'\) with respect to \(p_1, \ldots, p_\ell\). \(\square\)

To get desirable seminorm control over the intermediate tuples encountered in Proposition 6.1, it is not sufficient to keep track of the most immediate invariance properties. This is illustrated by the example below.

**Example 8.** (The necessity of composed invariance properties) Consider the average

\[\mathbb{E}_{n \in [N]} T_1^{n^2} f_1 \cdot T_2^{n^2} f_2 \cdot T_3^{n^2+n} f_3 \cdot T_4^{n^2+n} f_4 \cdot T_5^{n^2+n} f_5 \cdot T_6^{n^2+n} f_6. \quad (66)\]

It has length 6, degree 2, and type \((2, 2, 2)\) corresponding to the partition

\[\mathcal{J}_1 = \{1, 2\}, \quad \mathcal{J}_2 = \{3, 4\}, \quad \mathcal{J}_3 = \{5, 6\}.\]

We assume that it has the good ergodicity property, that is,

\[\mathcal{I}(T_1 T_2^{-1}) = \mathcal{I}(T_1) \cap \mathcal{I}(T_2), \quad \mathcal{I}(T_3 T_4^{-1}) = \mathcal{I}(T_3) \cap \mathcal{I}(T_4), \quad \mathcal{I}(T_5 T_6^{-1}) = \mathcal{I}(T_5) \cap \mathcal{I}(T_6).\]

Suppose we want to perform the seminorm smoothing argument to obtain a control of the associated average by the \(T_6\)-seminorm of \(f_6\). We iteratively pass to averages of lower type
as in Proposition 6.1 (all of which turn out to be controllable), and we show that to get seminorm control for the average of basic type at which we arrive, we need to keep track of not just the latest invariance properties that the functions in the intermediate averages enjoy, but of all the invariance properties that the functions in earlier intermediate averages enjoyed.

**Step 1:** Reducing to an average of basic type.

In the ping part of the seminorm smoothing argument for equation (66), we replace $T_6$ in the original average from equation (66) by some $T_i$ with $i \in \{4\} = \mathcal{I}_1 \cup \mathcal{I}_2$, arriving at, say, averages

$$\mathbb{E}_{n \in [N]} T_1^{n^2} f_{11} \cdot T_2^{n^2} f_{12} \cdot T_3^{n^2+n} f_{13} \cdot T_4^{n^2+n} f_{14} \cdot T_5^{n^2+2n} f_{15} \cdot T_4^{n^2+2n} f_{16}. \quad (67)$$

The new tuple in equation (67) has type $(2, 3, 1)$ and is controllable, with the index 5 satisfying the controllability condition. The functions inside take the form

$$f_{1j} := \begin{cases} \Delta b_{11, \ldots, b_{1j}; h-h'} f_j, & j \neq 6, \\ u_{1, h-h'}, & j = 6, \end{cases}$$

where the functions $u_{1, h-h'}$ are $T_6 T_4^{-1}$-invariant.

To obtain seminorm control of the average from equation (67), we need to perform the seminorm smoothing argument for this tuple. We aim to control it first by a $T_5$-seminorm of $f_{15}$ since 5 satisfies the controllability condition and is the only index left in $\mathcal{I}_3$. As guided by Proposition 6.1, in the ping step of the smoothing argument, we replace $T_5$ in equation (67) by some $T_i$ with $i \in \{4\}$. When $i = 1$, for instance, we end up with averages

$$\mathbb{E}_{n \in [N]} T_1^{n^2} f_{21} \cdot T_2^{n^2} f_{22} \cdot T_3^{n^2+n} f_{23} \cdot T_4^{n^2+n} f_{24} \cdot T_5^{n^2+2n} f_{25} \cdot T_4^{n^2+2n} f_{26}. \quad (68)$$

of type $(3, 3, 0)$. The functions in equation (68) take the form

$$f_{2j} := \begin{cases} \Delta b_{21, \ldots, b_{2j}; h-h'} f_j, & j \neq 5, \\ u_{2, h-h'}, & j = 5, \end{cases}$$

where $u_{2, h-h'}$ are $T_5 T_4^{-1}$-invariant. We note by Proposition 6.5 that the functions $f_{26}$ retain the $T_6 T_4^{-1}$-invariance of $f_{16}$.

The indices 3, 4, 6 in the average from equation (68) all satisfy the controllability condition, so if we want to get seminorm control of this average, we should first control it by a relevant seminorm of one of the functions $f_{23}, f_{24}, f_{26}$. Suppose that we choose to obtain a seminorm control of the tuple in equation (68) with respect to $f_{26}$ first. Then we would replace $T_4$ at index 6 by $T_i$ for any $i \in \{2\} = \mathcal{I}_1$ in the ping step, getting, say, averages

$$\mathbb{E}_{n \in [N]} T_1^{n^2} f_{31} \cdot T_2^{n^2} f_{32} \cdot T_3^{n^2+n} f_{33} \cdot T_4^{n^2+n} f_{34} \cdot T_5^{n^2+2n} f_{35} \cdot T_1^{n^2+2n} f_{36}. \quad (69)$$
of type \((4, 2, 0)\). The functions in equation (69) take the form

\[
f_{3j} := \begin{cases} 
\Delta b_{31, \ldots, b_{33} h h' f_{2j}}, & j \neq 6, \\
u_{3, h h'}, & j = 6.
\end{cases}
\]

The function \(f_{35}\), being a multiplicative derivative of a \(T_3 T_1^{-1}\)-invariant function, is itself invariant under \(T_5 T_1^{-1}\). The function \(f_{36}\) is invariant not only under \(T_4 T_1^{-1}\), but also under \(T_6 T_4^{-1}\) thanks to Proposition 6.5. It is crucial that \(f_{36}\) retains the \(T_6 T_4^{-1}\)-invariance of \(f_{26}\), and we shall return to this point shortly.

The average from equation (69) is controllable, and so to arrive at an average of a basic type, we need to perform this procedure two more times. Both the indices 3 and 4 satisfy the controllability condition, so we want to get seminorm control in terms of one of \(f_{33}, f_{34}\)—say we choose \(f_{34}\). To obtain control of the tuple in equation (69) by a \(T_4\)-seminorm of \(f_{34}\), we replace \(T_4\) by \(T_i\) for any \(i \in [2] = \mathcal{I}_1\) in the ping step, arriving at, say, averages

\[
\mathbb{E}_{n \in \mathbb{N}} T_{i_1}^{n^2} f_{41} \cdot T_{i_2}^{n^2} f_{42} \cdot T_{i_3}^{n^2+n} f_{43} \cdot T_{i_4}^{n^2+n} f_{44} \cdot T_{i_5}^{n^2+2n} f_{45} \cdot T_{i_6}^{n^2+2n} f_{46}
\]

of type \((5, 1, 0)\) if \(i = 2\). The functions in equation (70) take the form

\[
f_{4j} := \begin{cases} 
\Delta b_{41, \ldots, b_{44} h h' f_{3j}}, & j \neq 4, \\
u_{4, h h'}, & j = 4.
\end{cases}
\]

The function \(f_{45}\) retains the \(T_5 T_1^{-1}\)-invariance of \(f_{35}\) while \(f_{46}\), like \(f_{36}\), is invariant under \(T_4 T_1^{-1}\) and \(T_6 T_4^{-1}\). Moreover, the function \(f_{44}\) is \(T_4 T_1^{-1}\)-invariant.

Finally, the only index in the average from equation (70) satisfying the controllability condition is 3, so if we want to obtain control of the tuple in equation (70), we first want to get this in terms of a \(T_3\)-seminorm of \(f_{43}\). Applying Proposition 6.1, we end up replacing \(T_3\) by, say, \(T_2\), getting averages

\[
\mathbb{E}_{n \in \mathbb{N}} T_{i_1}^{n^2} f_{51} \cdot T_{i_2}^{n^2} f_{52} \cdot T_{i_3}^{n^2+n} f_{53} \cdot T_{i_4}^{n^2+n} f_{54} \cdot T_{i_5}^{n^2+2n} f_{55} \cdot T_{i_6}^{n^2+2n} f_{56}.
\]

The functions in equation (71) take the form

\[
f_{5j} := \begin{cases} 
\Delta b_{51, \ldots, b_{55} h h' f_{4j}}, & j \neq 3, \\
u_{5, h h'}, & j = 3.
\end{cases}
\]

in particular, the functions \(f_{54}, f_{55}, f_{56}\) retain the invariance properties of \(f_{44}, f_{45}, f_{46}\) and \(f_{33}\) is \(T_3 T_2^{-1}\)-invariant.

**Step 2**: Handling an average of basic type.

The average from equation (71) has basic type \((6, 0, 0)\), and so we want to control it by appropriate seminorms using Proposition 3.7. We show that without the assumption that \(f_{56}\) is invariant under both \(T_4 T_1^{-1}\) and \(T_6 T_4^{-1}\), we cannot control this average by a \(T_1\)-seminorm of \(f_{36}\), and conversely that this goal can be achieved with both of these assumptions.
We first note that Proposition 3.7, the ergodicity condition $I(T_1T_2^{-1}) = I(T_1) \cap I(T_2)$, and Lemma 2.2 allow us to control it by a $T_1$-seminorm of $f_{51}$ and by a $T_2$-seminorm of $f_{52}$. (At the same time, Proposition 3.7, our ergodicity assumptions, and Lemma 2.2 alone cannot be used to control equation (71) by $T_2$-seminorms of $f_{53}$ or $f_{54}$ or by $T_1$-seminorms of $f_{55}$ or $f_{56}$. Without additional information about the invariance properties of the functions, Proposition 3.7, our ergodicity assumptions, and Lemma 2.2 could only give control of equation (71) by $T_1$-seminorms of $f_{51}$, which is insufficient for our purposes.) Suppose that the $L^2(\mu)$ limit of equation (71) is positive. Decomposing $f_{51}$ using Proposition 2.3 and then applying the pigeonhole principle, we deduce the existence of $D \in \mathcal{D}$ such that

$$\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} D(n^2) \cdot T_2^{n^2} f_{52} \cdot T_2^{n^2+n} f_{53} \cdot T_2^{n^2+2n} f_{54} \cdot T_1^{n^2+2n} f_{55} \cdot T_1^{n^2+2n} f_{56} \right\|_{L^2(\mu)} > 0.$$  \hspace{1cm} (72)

We now want to obtain a seminorm control of the average in equation (72) by inductively invoking seminorm control for some average of length 5. To this end, we attempt to proceed like in Example 7. That is, we use the invariance of $f_{53}$, $f_{54}$, $f_{55}$, $f_{56}$ under $T_3 T_2^{-1}$, $T_4 T_2^{-1}$, $T_5 T_1^{-1}$, $T_4 T_1^{-1}$, respectively, to conclude that the average in equation (72) equals

$$\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} D(n^2) \cdot T_2^{n^2} f_{52} \cdot T_3^{n^2+n} f_{53} \cdot T_4^{n^2+n} f_{54} \cdot T_5^{n^2+2n} f_{55} \cdot T_4^{n^2+2n} f_{56} \right\|_{L^2(\mu)} = 0.$$  \hspace{1cm} (73)

However, without any extra information, we could not control the average from equation (73) using a $T_2$-seminorm of $f_{55}$ or $T_1$-seminorm of $f_{56}$. Suppose for simplicity that $D$ is a constant sequence and $f_{52} = f_{53} = f_{54} = 1$. Then equation (73) reduces to

$$\lim_{N \to \infty} \mathbb{E}_{n \in [N]} T_5^{n^2+2n} f_{55} \cdot T_4^{n^2+2n} f_{56}.$$  \hspace{1cm} (74)

We know nothing about the composition $T_3 T_4^{-1}$, and so without additional input, we cannot control equation (74) by a Gowers–Host–Kra seminorm.

This is the moment when we have to use the additional $T_6 T_4^{-1}$-invariance of $f_{56}$. Since $T_4 f_{56} = T_6 f_{56}$, we can replace $T_4$ in equation (74) by $T_6$. Then Proposition 3.7 gives us control over equation (74) by $\|f_{56}\|_{\mathcal{E}_s, (\mathcal{E}_5 - \mathcal{E}_6) \times s}$ for some $s \in \mathbb{N}$. The ergodicity condition on $T_6 T_5^{-1}$ and the $T_6 T_1^{-1}$-invariance of $f_{56}$ then give $\|f_{56}\|_{\mathcal{E}_s, (\mathcal{E}_5 - \mathcal{E}_6) \times s} \leq \|f_{56}\|_{2s, T_6} = \|f_{56}\|_{2s, T_1}$, and so this latter seminorm controls equation (72), and hence also equation (71).

If we want to control equation (72) by a $T_1$-seminorm of $f_{56}$ without the simplifying assumptions on $D$ and $f_{52}$, $f_{53}$, $f_{54}$, we proceed similarly. (There is no special reason why we would want to control equation (72) by a seminorm of $f_{56}$ instead of other functions. We just aim to illustrate that using the extra $T_6 T_4^{-1}$-invariance of $f_{56}$, this can be done.) Using the $T_6 T_4^{-1}$-invariance of $f_{56}$, we rewrite equation (73) as

$$\lim_{N \to \infty} \mathbb{E}_{n \in [N]} D(n^2) \cdot T_2^{n^2} f_{52} \cdot T_3^{n^2+n} f_{53} \cdot T_4^{n^2+n} f_{54} \cdot T_5^{n^2+2n} f_{55} \cdot T_6^{n^2+2n} f_{56}.$$  \hspace{1cm} (75)
Then we inductively apply the fact that we have seminorm control for averages of length 5 of the form in equation (75) to control this average, and hence also equation (71), by a \( T_6 \)-seminorm of \( f_{56} \). Subsequently, the \( T_6 T_4^{-1} \)-invariance of \( f_{56} \) (resulting from its \( T_6 T_4^{-1} \)- and \( T_4 T_1^{-1} \)-invariance) and Lemma 3.5 give control of equation (71) by a \( T_1 \)-seminorm of \( f_{56} \).

We note that the argument above would not work if we only used the ‘new’ property of \( f_{56} \) of being invariant under \( T_4 T_1^{-1} \) rather than the ‘combined’ property of being invariant under \( T_6 T_1^{-1} \). The important point that this example shows is that it is not enough to keep track of the new invariance properties that we obtain at each stage of the ping argument and forget the old ones. Rather, we need to keep track of the invariance property with respect to a composition of the original and the most recent transformations, which in our example are \( T_6 \) and \( T_1 \).

Examples 7 and 8 reveal that to obtain seminorm control of a controllable average of basic type and in the pong step of the seminorm smoothing argument, we need to substitute the transformations \( T_j \) for \( T_{nj} \) to arrive at an average for which we have seminorm control. The following proposition allows us to do just that.

**Proposition 6.6.** (Flipping) Let \( \gamma, \ell, L \in \mathbb{N} \), \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\) be a system, \( \eta \in [\ell]^\ell \) be an indexing tuple, \( A \subset [\ell] \) be a subset of indices, and \( p_1, \ldots, p_\ell, \rho_1, \ldots, \rho_\ell, q_1, \ldots, q_L \in \mathbb{Z}[n] \) be polynomials. Suppose that:

(i) the tuple \((T^{(\rho_j(n))}_{\eta_j})_{j \in [\ell]}\) is a descendant of \((T^{(p_j(n))}_{j})_{j \in [\ell]}\);

(ii) \( f_1, \ldots, f_\ell \in L^\infty(\mu) \) are 1-bounded functions having the \( \gamma \)-invariance property along \( \eta \) with respect to \( p_1, \ldots, p_\ell \).

Then there exist 1-bounded functions \( f'_1, \ldots, f'_\ell, \) polynomials \( \rho'_1, \ldots, \rho'_\ell, q'_1, \ldots, q'_L \in \mathbb{Z}[n], \) and an indexing tuple \( \eta' \) with the following properties:

(i) the tuple \( \eta' \) takes the form

\[
\eta'_j = \begin{cases} j, & j \in A, \\
\eta_j, & j \notin A; \end{cases}
\]

(ii) \((T^{(\rho'_j(n))}_{\eta'_j})_{j \in [\ell]}\) is a descendant of \((T^{(p'_j(n))}_{j})_{j \in [\ell]}\);

(iii) \( f'_1, \ldots, f'_\ell \in L^\infty(\mu) \) are 1-bounded and have the \( \gamma \)-invariance property along \( \eta' \) with respect to \( p_1, \ldots, p_\ell; \) moreover, for every \( s \geq 2 \) and \( j \in [\ell], \) they satisfy the bound \( \| f'_j \|_{L^s, T_{\eta'_j}} \leq C \| f_j \|_{L^s, T_{\eta_j}} \) for some \( C > 0 \) depending only on \( \gamma \) and the leading coefficients of \( p_1, \ldots, p_\ell; \) lastly, \( f'_j = 1 \) whenever \( f_j = 1; \)

(iv) we have the inequality

\[
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T^{(\rho_j(n))}_{\eta_j} f_j \cdot \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} \leq \lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T^{(\rho'_j(n))}_{\eta'_j} f'_j \cdot \prod_{j \in [L]} D_j(q'_j(n)) \right\|_{L^2(\mu)}.
\]
We note that if the leading coefficients of \( p_1, \ldots, p_\ell \) are all 1, and the good invariance property takes the form of \( f_j \) being invariant under \( T_{\eta_j}T_j^{-1} \) for all \( j \in [\ell] \), then Proposition 6.6 is straightforward, and in fact for every \( N \in \mathbb{N} \), we have

\[
\mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{\eta_j}^{\rho_j(n)} f_j \prod_{j \in [L]} D_j(q_j(n)) = \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{\eta_j}^{\rho_j(n)} f_j \prod_{j \in [L]} D_j(q_j(n))
\]

and \( \|f_j\|_{s,T_{\eta_j}} = \|f_j\|_{s,T_j} \) for all \( j \in [\ell], s \in \mathbb{N} \). The need for the more complicated statement of Proposition 6.6 comes from tedious but uninspiring technicalities that appear when the polynomials \( p_1, \ldots, p_\ell \) have leading coefficients distinct from 1.

We have used the flipping technique twice in Example 7: in equation (57) and when passing from equation (60) to equation (61) to obtain a seminorm control on the former using the seminorm control on the latter. We also used it in Example 8 to get seminorm control of equation (71). We will also use it shortly to handle uncontrollable tuples.

**Proof of Proposition 6.6.** For each \( j \in [\ell] \), let \( a_j \) be the leading coefficient of \( p_j \) and \( \gamma_j \in \mathbb{N} \) be the smallest natural number such that \( f_j \) is invariant under \( (T_{\eta_j}^{a_j} T_j^{-a_j})^{\gamma_j} \) (in particular, \( \gamma_j = 1 \) if \( \eta_j = j \)). Let \( \gamma \in \mathbb{N} \) be the smallest natural number such that \( a_{\gamma_j} \gamma_j \) divides the coefficients of \( \gamma \rho_j \) for every \( j \in A \). We then define

\[
f_j' := T_{\eta_j}^{\rho_j(r)} f_j, \quad q_j'(n) := q_j(\gamma n + r), \quad \rho_j'(n) := \frac{a_{\gamma_j}'}{a_{\gamma_j}} (\rho_j(\gamma n + r) - \rho_j(r))
\]

for some \( r \in \{0, \ldots, \lambda - 1\} \) to be chosen later, and we observe that \( \rho_j' \in \mathbb{Z}[n] \) for every \( j \in [\ell] \). By definition of \( \eta_j' \) and the \( \gamma \)-invariance of \( f_1, \ldots, f_\ell \) along \( \eta_j \), the functions \( f_1', \ldots, f_\ell' \) are \( \gamma \)-invariant along \( \eta_j' \). Moreover, if \( f_j = 1 \), then so is \( f_j' \). Lastly, they satisfy the bound \( \|f_j'\|_{s,T_j} \ll a_{\gamma_j} \|f_j\|_{s,T_j} \) for every \( s \geq 2 \) and \( j \in [\ell] \); this is trivial for \( j \notin A \), and if \( j \in A \), then

\[
\|f_j'\|_{s,T_j} = \|f_j'\|_{s,T_j} = \|f_j\|_{s,T_j} \leq \|f_j\|_{s,T_j} = \|f_j\|_{s,T_j} = \|f_j\|_{s,T_j} \ll a_{\gamma_j} \|f_j\|_{s,T_j},
\]

where we use the fact that \( f_j \) is a composition of \( f_j' \) with respect to a measure-preserving transformation, the invariance property of \( f_j \), Lemma 3.5, and both directions of Lemma 2.1.

We move on to prove the inequality in equation (76). For \( j \notin A \), where \( \eta_j' = \eta_j \), we simply have \( T_{\eta_j}^{\gamma_j(\gamma n + r)} f_j = T_{\eta_j}'^{\rho_j'(n)} f_j' \). For \( j \in A \), the invariance property of \( f_j \) gives us the identity

\[
T_{\eta_j}^{\gamma_j(\gamma n + r)} f_j = (T_{\eta_j}^{a_{\gamma_j} \gamma_j})^{(\rho_j(\gamma n + r) - \rho_j(r))/a_{\gamma_j} \gamma_j} T_{\eta_j}^{\gamma_j(\gamma n + r)} f_j = T_{\eta_j}'^{\rho_j'(n)} f_j'.
\]

Splitting \( \mathbb{N} \) into \( (\gamma \cdot \mathbb{N} + r)_{r \in \{0, \ldots, \gamma - 1\}} \), we deduce from the pigeonhole principle that there exists \( r \in \{0, \ldots, \gamma - 1\} \) for which equation (76) holds.

From the construction of the polynomials \( \rho_1', \ldots, \rho_\ell' \), the assumption that \( (T_{\eta_j}^{\gamma_j(\gamma n + r)})_{j \in [\ell]} \) is a descendant of \( (T_{\eta_j}^{\gamma_j(\gamma n + r)})_{j \in [\ell]} \) and Lemma 6.2, it follows that \( (T_{\eta_j}'^{\rho_j'(n)})_{j \in [\ell]} \) is also a descendant of \( (T_{\eta_j}'^{\rho_j'(n)})_{j \in [\ell]} \).
6.3. Handling uncontrollable tuples. We have explained in the previous sections that if a tuple \((T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}\) is controllable, then we control it by a Gowers–Host–Kra seminorm using a seminorm smoothing argument. If the tuple is uncontrollable, however, we use the following variant of the flipping technique from Proposition 6.6 to bound the \(L^2(\mu)\) norm of the associated average by an \(L^2(\mu)\) norm of a controllable average.

**Corollary 6.7. (Flipping uncontrollable tuples)** Let \(\gamma, \ell, L \in \mathbb{N}, (X, \mathcal{X}, \mu, T_1, \ldots, T_{\ell})\) be a system, \(\eta \in [\ell]^\ell\) be an indexing tuple, and \(p_1, \ldots, p_{\ell}, \rho_1, \ldots, \rho_{\ell}, q_1, \ldots, q_L \in \mathbb{Z}[n]\) be polynomials. Suppose that

1. the tuple \((T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}\) is uncontrollable of type \(w\) with the last non-zero index \(t_w\), and it is a descendant of \((T_{j}^{\rho_j(n)})_{j \in [\ell]}\);
2. \(f_1, \ldots, f_{\ell} \in L^\infty(\mu)\) are 1-bounded functions having the \(\gamma\)-invariance property along \(\eta\) with respect to \(p_1, \ldots, p_{\ell}\).

Then there exist 1-bounded functions \(f'_1, \ldots, f'_{\ell}\), polynomials \(\rho'_1, \ldots, \rho'_{\ell}, q'_1, \ldots, q'_{\ell} \in \mathbb{Z}[n]\), and an indexing tuple \(\eta'\) with the following properties:

1. the tuple \(\eta'\) takes the form
   \[
   \eta'_j = \begin{cases} 
   j & \eta_j \in \mathcal{I}_{t_w}, \\
   \eta_j & \eta_j \notin \mathcal{I}_{t_w};
   \end{cases}
   \]
2. \((T_{\eta'_j}^{\rho'_j(n)})_{j \in [\ell]}\) is a descendant of \((T_{j}^{\rho_j(n)})_{j \in [\ell]}\);
3. \(f'_1, \ldots, f'_{\ell} \in L^\infty(\mu)\) are 1-bounded and have the \(\gamma\)-invariance property along \(\eta'\) with respect to \(p_1, \ldots, p_{\ell}\); moreover, for every \(s \geq 2\) and \(j \in [\ell]\), they satisfy the bound \(\|f'_j\|_{S,T_{\eta'_j}} \leq C\|f_j\|_{S,T_{\eta_j}}\) for some \(C > 0\) depending only on \(\gamma\) and the leading coefficients of \(p_1, \ldots, p_{\ell}\); lastly, \(f'_j = 1\) whenever \(f_j = 1\);
4. we have the inequality
   \[
   \lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{\eta_j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} \leq \lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{\eta'_j}^{\rho'_j(n)} f'_j \cdot \prod_{j \in [L]} D_j(q'_j(n)) \right\|_{L^2(\mu)}.
   \]

Corollary 6.7 follows from Proposition 6.6 by taking \(A = \{j \in [\ell] : \eta_j \in \mathcal{I}_{t_w}\}\).

We emphasize that Corollary 6.7 by itself does not guarantee that the tuple \((T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}\) is of lower type than the tuple \((T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}\). However, this will be the case when we apply it to all the tuples \((T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}\) that appear in our inductive procedure. The crucial ingredient in achieving this type reduction will be the property (iv) in Proposition 6.8 enjoyed by all the tuples showing up in our arguments.

**Example 9.** Consider the tuple

\[
(T_1^{n^2}, T_2^n, T_3^n, T_4^{n^2}, T_5^{n^2+n}, T_6^{n^2}, T_7^{n^2+2n}, T_8^{n^2+2n})
\]
from Example 4. It is a descendant of the tuple

\[(T_1^2, T_2^2, T_3^2, T_4^2, T_5^2 + n, T_6^2 + n, T_7^2 + 2n, T_8^2 + 2n), \tag{78}\]

obtained by four applications of Proposition 6.1, in which we substitute \(T_8\) for \(T_7\) at the index 8, \(T_5\) for \(T_7\) at the index 7, \(T_1\) for \(T_5\) at the index 5, and \(T_1\) for \(T_6\) at the index 6. Corollary 6.7 gives that if \(f_1, \ldots, f_8 \in L^\infty(\mu)\) are functions such that \(f_5, f_6, f_7, f_8\) are invariant under \(T_5 T_1^{-1}, T_6 T_1^{-1}, T_7 T_5^{-1}, T_8 T_5^{-1}\), respectively, then we have

\[
\lim_{N \to \infty} \| \mathbb{E}_{n \in [N]} T_1^n f_1 \cdot T_2^n f_2 \cdot T_3^n f_3 \cdot T_4^n f_4 \cdot T_5^n + n f_5 \cdot T_6^n + n f_6 \cdot T_7^n + 2n f_7 \cdot T_8^n + 2n f_8 \|_{L^2(\mu)} \leq \lim_{N \to \infty} \| \mathbb{E}_{n \in [N]} T_1^n f_1 \cdot T_2^n f_2 \cdot T_3^n f_3 \cdot T_4^n f_4 \cdot T_5^n + n f_5 \cdot T_6^n + n f_6 \|
\]

\[
\times T_7^n + 2n f_7 \cdot T_8^n + 2n f_8 \|_{L^2(\mu)}
\]

(in fact, a closer look guarantees that we get an equality and not just for the \(L^2(\mu)\) limits, but for each finite average). We moreover obtain the equality of seminorms \(\|f_7\|_{s, T_7} = \|f_7\|_{s, T_5}, \|f_8\|_{s, T_8} = \|f_8\|_{s, T_5}\), and the new tuple

\[(T_1^2, T_2^2, T_3^2, T_4^2, T_5^2 + n, T_6^2 + n, T_7^2 + 2n, T_8^2 + 2n) \tag{79}\]

is a descendant of equation (78). Importantly, the new tuple in equation (79) has type \((6, 0, 2)\), which is lower than the type \((6, 2, 0)\) of equation (77). Applying Corollary 6.7 to equation (77), we have thus successfully replaced it by a tuple of lower type. Lastly, the new tuple in equation (79) is controllable as the indices 7, 8 satisfy the controllability condition.

6.4. Recapitulation. We conclude this section with Proposition 6.8, which forms an inductive framework for the proof of Theorem 1.1 in the next section. Combining the content of Proposition 6.1 and Corollary 6.7, it shows that—starting with a tuple \((T_j^{p_j(n)}(n))_{n \in [\ell]}\) satisfying the good ergodicity property—we reach a tuple of basic type in a finite number of steps. For tuples of basic type, seminorm control will follow from arguments made in Proposition 7.2, and then we will use this fact and induction to go back and get seminorm control for the original tuple \((T_j^{p_j(n)}(n))_{n \in [\ell]}\).

**Proposition 6.8.** (Inductive framework) Let \(\ell \in \mathbb{N}, p_1, \ldots, p_\ell \in \mathbb{Z}[n]\) be polynomials, and \((X, X', \mu, T_1, \ldots, T_\ell)\) be a system. Suppose that the tuple \((T_j^{p_j(n)}(n))_{n \in [\ell]}\) has the good ergodicity property. Then there exists \(r \in \mathbb{N}\) (which depends only on \(p_1, \ldots, p_\ell\) but can be bounded purely in terms of \(\ell\)) and a sequence of tuples

\[
(T_{n_{0j}}^{p_{0j}(n)}(n))_{n \in [\ell]} \to (T_{n_{1j}}^{p_{1j}(n)}(n))_{n \in [\ell]} \to (T_{n_{2j}}^{p_{2j}(n)}(n))_{n \in [\ell]} \to \ldots \to (T_{n_{rj}}^{p_{rj}(n)}(n))_{n \in [\ell]}
\]

with types \(w_0, \ldots, w_r\) such that \((T_{n_{kj}}^{p_{kj}(n)}(n))_{n \in [\ell]} := (T_j^{p_j(n)}(n))_{n \in [\ell]}\) and for \(k \in \{0, \ldots, r - 1\}\), the following properties hold.

(i) If the tuple \((T_{n_{kj}}^{p_{kj}(n)}(n))_{n \in [\ell]}\) is controllable, then the tuple \((T_{n_{(k+1)j}}^{p_{(k+1)j}(n)}(n))_{n \in [\ell]}\) is chosen using Proposition 6.1.
(ii) If the tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) is uncontrollable, then the tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) is chosen using Corollary 6.7.

(iii) The tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) is a descendant of \((T_j^{\rho_j(n)})_{j \in [\ell]}\). In particular, it has the good ergodicity property.

(iv) For the indexing tuple \(\eta_{k+1} = (\eta_{(k+1)1}, \ldots, \eta_{(k+1)\ell})\), if \(j \in \mathcal{I}\), then either \(\eta_{(k+1)j} = j\) or \(\eta_{(k+1)j} \in \mathcal{I}_{t'}\) for some \(t' < t\).

(v) We have \(w_{k+1} < w_k\), that is, the tuple \((T_{\eta_{k+1}j}^{\rho_j(n)})_{j \in [\ell]}\) has a lower type than the tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\).

(vi) The tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) has basic type and the restriction \(\eta_r | \mathcal{I}_1\) is the identity sequence.

For the rest of the paper, we call a tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) a proper descendant of the tuple \((T_j^{\rho_j(n)})_{j \in [\ell]}\) if it appears in one of the sequences of tuples constructed from \((T_j^{\rho_j(n)})_{j \in [\ell]}\) using Proposition 6.8.

**Proof.** Let \(k \in \{0, \ldots, r − 1\}\). Suppose that the tuples \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\), \ldots, \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) are already constructed and satisfy the properties (i)–(v). If the tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) has basic type, we halt. Otherwise, we choose the tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) using Proposition 6.1 if \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) is controllable and using Corollary 6.7 otherwise, so that the properties (i) and (ii) are satisfied. (We remark that by the property (iv) applied to the tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\), we have \(\eta_{kj} = j\) whenever \(j \in \mathcal{I}_1\), and hence the type \(w_k = (w_{k1}, \ldots, w_{k\ell})\) of \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) satisfies \(w_{k1} > 0\). By the assumption that \(w_k\) is not basic, we also have \(w_{kt} > 0\) for some \(t > 1\); therefore, we have at least two non-zero indices in \(w_k\) and we can act as in Proposition 6.1.) We note from Corollaries 6.3 and 6.7 that the new tuple is a descendant of \((T_j^{\rho_j(n)})_{j \in [\ell]}\) and by Proposition 6.4, it has the good ergodicity property and hence the property (iii) holds as well. The property (iv) holds by induction and the way the tuple \((T_{\eta_{k+1}j}^{\rho_j(n)})_{j \in [\ell]}\) is constructed using Proposition 6.1 or Corollary 6.7.

For the property (v), we first note that if \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) is controllable, then the property (v) holds for \((T_{\eta_{k+1}j}^{\rho_j(n)})_{j \in [\ell]}\) by Proposition 6.1. If \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) is uncontrollable, then we get from Corollary 6.7 that \(w_{(k+1)t} = 0 < w_{kt}\), where \(t_k := t_{w_k}\) is the last non-zero index of \(w_k\), and we deduce from property (iv) that \(w_{(k+1)t} = w_{kt}\) for \(1 \leq t < t_k\). This point is important, so we explain it in words. What happens is that when we apply Corollary 6.7, the index \(w_{(k+1)t}\) goes down to 0 (since all the transformations with indices from \(\mathcal{I}_{t_k}\) get flipped), but the new transformations appearing in their place have indices from \(\mathcal{I}_{t_k+1}, \ldots, \mathcal{I}_{K_2}\), as given by the property (iv). Hence, \(w_{k+1} < w_k\) in this case as well.

It follows from the property (v) and the fact that there are at most \((K_3 + 1)^{K_2}\) possible types for tuples in \(\ell \leq K_3\) with sum of coordinates \(K_3\) that the sequence eventually terminates.

And since, by our construction, it can only terminate on the basic type \((K_3, 0, \ldots, 0)\), there exists \(r < (K_3 + 1)^{K_2} \leq (\ell + 1)^{\ell}\) such that the tuple \((T_{\eta_{kj}}^{\rho_j(n)})_{j \in [\ell]}\) has basic type.
Theorem 1.1 follows by induction from the result below upon setting $d = L = 0$ and letting $η$ be the identity tuple.

7. The proof of Theorem 1.1

7.1. Induction scheme. In all statements in this section, we work in the setting of Proposition 6.8, that is, all the lower type tuples at which we arrive from some original average are those constructed in Proposition 6.8 and have the properties listed there.

Theorem 1.1 follows by induction from the result below upon setting $d = L = 0$ and letting $η$ be the identity tuple.

**Example 10.** (Iterative reduction to tuples of lower type) For the tuple in equation (78) from Example 9, Proposition 6.8 would give, among other options, the following sequence of tuples:

$$(T^n_1, T^n_2, T^n_3, T^n_4, T^n_5, T^n_6, T^n_7, T^n_8) \rightarrow (T^n_1, T^n_2, T^n_3, T^n_4, T^n_5, T^n_6, T^n_7, T^n_8) \rightarrow (T^n_1, T^n_2, T^n_3, T^n_4, T^n_5, T^n_6, T^n_7, T^n_8) \rightarrow (T^n_1, T^n_2, T^n_3, T^n_4, T^n_5, T^n_6, T^n_7, T^n_8) \rightarrow (T^n_1, T^n_2, T^n_3, T^n_4, T^n_5, T^n_6, T^n_7, T^n_8) \rightarrow (T^n_1, T^n_2, T^n_3, T^n_4, T^n_5, T^n_6, T^n_7, T^n_8) \rightarrow (T^n_1, T^n_2, T^n_3, T^n_4, T^n_5, T^n_6, T^n_7, T^n_8).$$

Their types, starting from the top tuple, are

$$(4, 2, 2) > (4, 3, 1) > (4, 4, 0) > (5, 3, 0) > (6, 2, 0) > (6, 0, 2) > (7, 0, 1) > (8, 0, 0),$$

which shows that at each step, the new tuple has a lower type than its predecessor. The sixth tuple, counting from the top, has been obtained via Corollary 6.7 (since the fifth tuple is uncontrollable, as explained in Examples 4 and 9), while all the other tuples have been obtained using Proposition 6.1. Each subsequent tuple is a descendant of the original tuple, and so each of them has the good ergodicity property thanks to Proposition 6.4. Lastly, the final tuple has basic type, and moreover for its indexing tuple $η$, the restriction $η|_{[3]} = η|_{[4]}$ is an identity because no substitution has taken place at the first four indices.
\[ f_1, \ldots, f_\ell \in L^\infty(\mu) \] with the good invariance property along \( \eta \) with respect to \( p_1, \ldots, p_\ell \) and all sequences of functions \( D_1, \ldots, D_L \in D_d \), we have

\[
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{\eta_j}^{p_j(n)} f_j \cdot \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} = 0 \tag{80}
\]

whenever \( \| f_j \|_{s, T_{\eta_j}} = 0 \) for some \( j \in [\ell] \).

A word of explanation is necessary for the statement of Proposition 7.1. We need both polynomials \( p_1, \ldots, p_\ell \) and \( \rho_1, \ldots, \rho_\ell \). The reason is that for our induction to work, we need the functions \( f_1, \ldots, f_\ell \) to have the good invariance property with respect to the original family \( p_1, \ldots, p_\ell \) rather than the descendant family \( \rho_1, \ldots, \rho_\ell \). This is necessary for a number of reasons: to prove seminorm control for averages of basic types in the proof of Proposition 7.2; to apply Proposition 6.6 in the pong step of Proposition 7.5; to derive Proposition 7.1 from Proposition 7.4 for controllable tuples; and to invoke Corollary 6.7 for uncontrollable tuples in Proposition 7.1. The necessity of keeping track of the invariance property with regards to the original polynomial family has also been explained in Step 2 of Example 8.

Technically, the value \( s \) in Proposition 7.1 depends also on \( \eta \), but since the number of possible tuples \( \eta \) is bounded in terms of \( \ell \), this dependence can be removed.

We first prove Proposition 7.1 for averages from equation (35) of basic type. This will serve as the base for induction for Propositions 7.1, 7.4, and 7.5.

**Proposition 7.2. (Seminorm control of basic types)** Let \( d, D, \ell, L \in \mathbb{N}, \eta \in [\ell]^\ell \) and \( p_1, \ldots, p_\ell, q_1, \ldots, q_L \in \mathbb{Z}[n] \) be polynomials of degrees at most \( D \). Suppose that the polynomials \( p_1, \ldots, p_\ell \) have the good ergodicity property for the system \((X, X, \mu, T_1, \ldots, T_\ell)\). Let \((T_{\eta_j}^{p_j(n)})_{j \in [\ell]}\) be a proper descendant of the tuple \((T_{j}^{p_j(n)})_{j \in [\ell]}\), and suppose that the type \( w \) of \((T_{\eta_j}^{p_j(n)})_{j \in [\ell]}\) is basic. Then there exists \( s \in \mathbb{N} \), depending only on \( d, D, \ell, L \), such that for all 1-bounded functions \( f_1, \ldots, f_\ell \in L^\infty(\mu) \) with the good invariance property along \( \eta \) with respect to \( p_1, \ldots, p_\ell \) and all sequences of functions \( D_1, \ldots, D_L \in D_d \), we have equation (80) whenever \( \| f_j \|_{s, T_{\eta_j}} = 0 \) for some \( j \in [\ell] \).

A special case of Proposition 7.2 has been sketched in Step 2 of Example 8, and we invite the reader to compare the abstract proof presented below with the argument in Step 2 of Example 8.

**Proof.** We induct on the length \( \ell \) of the average. If \( \ell = 1 \), the statement holds by Proposition 3.7. We therefore assume that \( \ell > 1 \), and we will prove Proposition 7.2 for fixed \( \ell > 1 \) by invoking Proposition 7.1 for an average of length \( \ell - 1 \). More specifically, we will show first that there exists an index \( m \) satisfying the controllability condition, and that we can control the average by a \( T_{\eta_m} \)-seminorm of \( f_m \). Then we will replace \( f_m \) by a dual function using Proposition 2.3 and the pigeonhole principle, flip the other transformations \( T_{\eta_j} \) into \( T_j \) using Proposition 6.6, and invoke Proposition 7.1 for averages of length \( \ell - 1 \) to obtain seminorm control in terms of other functions.
Take any \( m \in \mathcal{I}_1 \); we assume for simplicity that \( m = \ell \). Proposition 6.8(vi) implies that \( \eta_\ell = \ell \), and moreover that \( \ell \) satisfies the controllability condition. This fact and Proposition 3.7 imply that the identity in equation (80) holds whenever \( \| f_\ell \|_{b_1, \ldots, b_s} = 0 \) for some
\[
b_1, \ldots, b_s \in \{ b_\ell e_\ell - b_i e_{\eta_i} : i \in \Sigma \cup \{0 \} \setminus \{\ell\},
\]
where \( s \in \mathbb{N} \) depends only on \( d, D, \ell, L \), the numbers \( b_\ell, b_i \) are the coefficients of \( \rho_\ell, \rho_i \) of degree \( d_\ell := \deg(\rho_\ell e_\ell - \rho_i e_{\eta_i}) \) whenever \( i \neq 0 \), and for \( i = 0 \), we simply set \( b_0 e_{\eta_0} = 0 \). Using the monotonicity property of equation (17), we can assume without loss of generality that \( s \geq 2 \). Since the tuple \( (T_{\eta_j}^{\rho_j(n)})_{j \in \ell} \) has a basic type, it follows that the indices \( \eta_i \) in equation (81) come from the set \( \mathcal{I}_1 \). We have to show that each transformation \( T^{b_1}, \ldots, T^{b_s} \) is either a non-zero iterate of \( T_\ell \) or its invariant functions are invariant under a bounded power of \( T_\ell \). For \( k \in \{s\} \), let \( b_k := b_\ell e_\ell - b_i e_{\eta_i} \). If \( i = 0 \), then \( T^{b_k} \) is indeed a non-zero iterate of \( T_\ell \). If \( i \in \mathcal{I}_1 \), then \( \eta_i = i \neq \ell \) by the property from Proposition 6.8(vi) that \( \eta|\mathcal{I}_1 \) is the identity tuple. Then \( T^{b_k} = T^{b_\ell}_k T^{-b_i}_i = (T^{b_\ell}_k T^{-b_i}_i)^{\gcd(b_\ell, b_i)} \) for coprime integers \( b_\ell, b_i \), and the good ergodicity property of \( \rho_1, \ldots, \rho_\ell \) along \( \eta \) (Propositions 6.8(iii)) implies that \( \mathcal{I}(T^{b_\ell}_k T^{-b_i}_i) \subseteq \mathcal{I}(T_\ell) \). If \( i \notin \mathcal{I}_1 \cup \{0\} \), then we split into the cases \( \eta_i \neq \ell \) and \( \eta_i = \ell \). In the former case, we once again use the good ergodicity property of \( \rho_1, \ldots, \rho_\ell \) along \( \eta \) to conclude that \( T^{b_k} = (T^{b_\ell}_k T^{-b_i}_i)^{\gcd(b_\ell, b_i)} \) and \( \mathcal{I}(T^{b_\ell}_k T^{-b_i}_i) \subseteq \mathcal{I}(T_\ell) \). In the latter case, the pairwise independence of \( \rho_i \) and \( \rho_\ell \) implies that \( b_k \) is non-zero, and hence \( T^{b_k} \) is a non-zero iterate of \( T_\ell \). Rearranging \( b_1, \ldots, b_s \), it follows that
\[
\| f_\ell \|_{b_1, \ldots, b_s} = \| f_\ell \|_{c_1 e_\ell, \ldots, c_s e_\ell, b_{s+1}, \ldots, b_s}
\]
for some \( 0 \leq s' \leq s \), non-zero integers \( c_1, \ldots, c_{s'} \), and transformations \( T^{b_{s'+1}}, \ldots, T^{b_s} \), with the property that for every \( j \in \{s' + 1, \ldots, s\} \), there exists \( b_j' \in \mathbb{Z}_\ell \) and non-zero \( c_j \in \mathbb{Z} \) such that \( b_j = c_j b_j' \) and \( \mathcal{I}(T^{b_j'}_j) \subseteq \mathcal{I}(T^{e_\ell}) \). By Lemmas 2.1 and 2.2, we have
\[
\| f_\ell \|_{b_1, \ldots, b_s} = \| f_\ell \|_{c_1 e_\ell, \ldots, c_j e_\ell, c_{s'+1} e_\ell, b_{s'+1}, \ldots, c_s e_\ell} \ll_{c_1, \ldots, c_s} \| f_\ell \|_{s, T_\ell}
\]
implying that equation (80) holds whenever \( \| f_\ell \|_{s, T_\ell} = 0 \).

We now use the seminorm control at \( \ell \) with Proposition 2.3 and the pigeonhole principle to deduce that if equation (80) fails, then it also fails when \( T_{\eta}^{\rho_{\ell}(n)} f_\ell \) is replaced by a sequence \( D_{L+1}(\rho_{\ell}(n)) \) with \( D_{L+1} \in \mathcal{D}_\Sigma \). Letting \( q_{L+1} := \rho_\ell \) for simplicity, it is enough to show that
\[
\lim_{N \to \infty} \left\| \sum_{\ell \in \mathcal{N}} \mathbb{E} \prod_{j \in [\ell-1]} T_{\eta_j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L+1]} D_j(q_j(n)) \right\|_{L^2(\mu)} > 0
\]
implies \( \| f_j \|_{s', T_{\eta_j}} > 0 \) for every \( j \in [\ell-1] \) and some \( s' \in \mathbb{N} \) depending only on \( d, D, \ell, L \). By Proposition 6.6, there exist 1-bounded functions \( f'_1, \ldots, f'_{\ell-1} \) and polynomials \( \rho'_1, \ldots, \rho'_{\ell-1} \) with the good ergodicity property for the system such that
\[
\lim_{N \to \infty} \left\| \sum_{\ell \in \mathcal{N}} \mathbb{E} \prod_{j \in [\ell-1]} T_j^{\rho_j(n)} f'_j \cdot \prod_{j \in [L+1]} D_j(g_j(n)) \right\|_{L^2(\mu)} > 0
\]
and \( \|f_j\|_{s',T_j} \leq C \|f_j\|_{s',T_{n_j}} \) for all \( j \in [\ell - 1] \), \( s' \geq 1 \), and some constant \( C > 0 \) depending only on \( \gamma \) and the leading coefficients of \( p_1, \ldots, p_\ell \). Invoking inductively the case \( \ell - 1 \) of Proposition 7.1, we deduce that \( \|f_j\|_{s',T_j} > 0 \) for some \( s' \in \mathbb{N} \) depending only on \( d, D, \ell, L \), and hence \( \|f_j\|_{s',T_{n_j}} > 0 \). This proves the claim. \( \square \)

We also need the following quantitative version of Proposition 7.1.

**Proposition 7.3.** (Soft quantitative estimates) Let \( d, D, \gamma, \ell, L \in \mathbb{N} \), \( \eta \in [\ell]^{\ell} \), and \( p_1, \ldots, p_\ell \), \( q_1, \ldots, q_L \in \mathbb{Z}[n] \) be polynomials of degrees at most \( D \). Suppose that the polynomials \( p_1, \ldots, p_\ell \) have the good ergodic property for the system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\). Let \((T_{n_j}^{\rho_j(n)})_{j \in [\ell]} \) be a proper descendant of the tuple \((T_j^{\rho_j(n)})_{j \in [\ell]} \). Then there exists \( s \in \mathbb{N} \), depending only on \( d, D, \ell, L \), with the following property: for any \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that for all \( 1\)-bounded functions \( f_1, \ldots, f_\ell \in L^\infty(\mu) \) that are \( \gamma \)-invariant along \( \eta \) with respect to \( p_1, \ldots, p_\ell \) and all sequences of functions \( D_1, \ldots, D_L \in \mathcal{D}_d \), we have

\[
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{n_j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} < \varepsilon
\]

whenever \( \|f_j\|_{s, T_{n_j}} < \delta \) for some \( j \in [\ell] \).

**Proof.** We prove Proposition 7.3 for fixed \( d, D, \gamma, \ell, L, \eta \) by assuming Proposition 7.1 for the same parameters.

Let \( s \in \mathbb{N} \) be as in the statement of Proposition 7.1 (so in particular it only depends on \( d, D, \ell, L \)) and \( b_1, \ldots, b_\ell \) be the leading coefficients of \( \rho_1, \ldots, \rho_\ell \). Fix \( \pi \in [\ell]^L \). We first prove the following qualitative claim: for all \( f_1, \ldots, f_\ell, g_1, \ldots, g_L \in L^\infty(\mu) \), where \( f_j \) is \((T_{n_j}^{b_j} T_j^{-b_j})^\gamma\)-invariant for each \( j \in [\ell] \) and \( g_j \) is \( \mathcal{Z}_d(T_{\pi_j}) \)-measurable for each \( j \in [L] \), we have

\[
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{n_j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L]} T_{\pi_j}^{q_j(n)} g_j \right\|_{L^2(\mu)} = 0 \quad (82)
\]

whenever \( \|f_j\|_{s, T_{n_j}} = 0 \) for some \( j \in [\ell] \).

Fix \( f_1, \ldots, f_\ell, g_1, \ldots, g_L \) and suppose that equation (82) fails. Using Proposition 2.3 and the pigeonhole principle, we deduce that there exist dual functions \( g_1', \ldots, g_L' \) of \( T_{\pi_1}, \ldots, T_{\pi_L} \) of level \( d \) such that

\[
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{n_j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L]} T_{\pi_j}^{q_j(n)} g_j' \right\|_{L^2(\mu)} > 0.
\]

Setting \( D_j(n) := T_{\pi_j}^{n} g_j' \) and using Proposition 7.1, we deduce that \( \|f_j\|_{s, T_{n_j}} > 0 \) for all \( j \in [\ell] \), and so the claim follows.

We combine the claim above with Proposition 3.6 for

\[
\mathcal{Y}_j := \begin{cases} \mathcal{I}(T_{n_j}^{b_j} T_j^{-b_j})^\gamma, & 1 \leq j \leq \ell, \\ \mathcal{Z}_d(T_{\pi_j}), & \ell + 1 \leq j \leq L, \end{cases}
\]
deducing the following: for every \( \varepsilon > 0 \), there exists \( \delta_\pi > 0 \) such that for all 1-bounded functions \( f_1, \ldots, f_\ell, g_1, \ldots, g_L \in L^\infty(\mu) \) with \( f_j \in L^\infty(I((T_{\eta_j}^{bn_j} T_j^{n_j}(-)^{\eta_j})) \) for \( j \in [\ell] \) and \( g_j \in L^\infty(Z_d(T_{\pi_j})) \) for \( j \in [L] \), we have

\[
\lim_{N \to \infty} \left\| \sum_{n \in [N]} \prod_{j \in [\ell]} T_{\eta_j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L]} T_{\pi_j}^{\eta_j(n)} g_j \right\|_{L^2(\mu)} < \varepsilon
\]

whenever \( \| f_j \|_{s,T_{\eta_j}^{\rho_j(n)}} < \delta_\pi \) for some \( j \in [\ell] \).

Proposition 7.3 follows by taking \( \delta := \min(\delta_\pi : \pi \in [\ell]^L) \) and recalling that dual functions of \( T_j \) of order \( d \) are \( Z_d(T_j) \)-measurable, and hence for every \( j \in [L] \), the sequence \( D_j(q_j(n)) \) has the form \( D_j(q_j(n)) = T_{\pi_j}^{\eta_j(n)} g_j \) for some \( \pi_j \in [\ell] \) and a 1-bounded \( Z_d(T_j) \)-measurable function \( g_j \in L^\infty(\mu) \).

For controllable tuples, Proposition 7.1 will be deduced from the following result.

**Proposition 7.4.** (Iterated box seminorm smoothing) Let \( d, D, \ell, L \in \mathbb{N} \), \( \eta \in [\ell]^\ell \), and \( p_1, \ldots, p_\ell, q_1, \ldots, q_L \in \mathbb{Z}[n] \) be polynomials of degrees at most \( D \). Suppose that the polynomials \( p_1, \ldots, p_\ell \) have the good ergodicity property for the system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\). Let \((T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}\) be a proper descendant of the tuple \((T_j^{\rho_j(n)})_{j \in [\ell]}\).

Suppose that \((T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}\) of a non-basic type is controllable, and let \( m \) be an index satisfying the controllability condition. Then there exist \( s, s' \in \mathbb{N} \), depending only on \( d, D, \ell, L \), such that for all functions \( f_1, \ldots, f_\ell \in L^\infty(\mu) \) with the good invariance property along \( \eta \) with respect to \( p_1, \ldots, p_\ell \) and all sequences of functions \( D_1, \ldots, D_L \in \mathcal{D}_d \), we obtain equation (80) whenever \( \| f_m \|_{s,T_{im}^{\rho_j(n)}} = 0 \).

Proposition 7.4 is a consequence of Proposition 3.7, followed by an iterated application of the smoothing result given below.

**Proposition 7.5.** (Box seminorm smoothing) Let \( d, D, \ell, L \in \mathbb{N} \), \( \eta \in [\ell]^\ell \) and \( p_1, \ldots, p_\ell, q_1, \ldots, q_L \in \mathbb{Z}[n] \) be polynomials of degrees at most \( D \). Suppose that the polynomials \( p_1, \ldots, p_\ell \) have the good ergodicity property for the system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\). Let \((T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}\) be a proper descendant of the tuple \((T_j^{\rho_j(n)})_{j \in [\ell]}\).

Suppose that \((T_{\eta_j}^{\rho_j(n)})_{j \in [\ell]}\) of a non-basic type is controllable, and let \( m \) be an index satisfying the controllability condition. Then for all \( s \geq 2 \) and vectors \( b_1, \ldots, b_{s+1} \) satisfying equation (23), there exists \( s' \in \mathbb{N} \), depending only on \( d, D, \ell, L, s \), with the following property: for all functions \( f_1, \ldots, f_\ell \in L^\infty(\mu) \) with the good invariance property along \( \eta \) with respect to \( p_1, \ldots, p_\ell \) and all sequences of functions \( D_1, \ldots, D_L \in \mathcal{D}_d \), if \( \| f_m \|_{b_1, \ldots, b_{s+1}} = 0 \) implies equation (80), then equation (80) also holds under the assumption that \( \| f_m \|_{b_1, \ldots, b_s, e_m^{s'}} = 0 \).

We explain now the induction scheme whereby we prove Propositions 7.1–7.5. Roughly speaking, the proofs proceed by the induction on the length \( \ell \) of an average and—for fixed \( \ell \)—by induction on type, where the base case are averages of basic types. More precisely, the induction scheme goes as follows.
(i) For tuples \((T_{ρ_j}^{(n)})_{j∈[ℓ]}\) of length \(ℓ\) of type \(w\), Proposition 7.3 follows from Proposition 7.1 as proved before.

(ii) Tuples \((T_{ρ_j}^{(n)})_{j∈[ℓ]}\) of length \(ℓ = 1\) have necessarily basic type, and Propositions 7.1 and 7.2 follow easily from Proposition 3.7.

(iii) For tuples \((T_{ρ_j}^{(n)})_{j∈[ℓ]}\) of length \(ℓ > 1\) and basic type, Proposition 7.1 is a consequence of Proposition 7.2.

(iv) For tuples \((T_{ρ_j}^{(n)})_{j∈[ℓ]}\) of length \(ℓ > 1\) and non-basic type \(w\), we prove Proposition 7.5 only under the assumption of controllability. This proof goes by inductively invoking Proposition 7.3 in two cases: for tuples of length \(ℓ\) and type \(w' < w\), and for tuples of length \(ℓ − 1\). An iterative application of Proposition 7.5 then yields Proposition 7.4 for tuples of length \(ℓ\) and type \(w\).

(v) For controllable tuples \((T_{ρ_j}^{(n)})_{j∈[ℓ]}\) of length \(ℓ > 1\) and non-basic type \(w\), we prove Proposition 7.1 by invoking Proposition 7.4 for tuples of length \(ℓ\) and type \(w' < w\).

(vi) Lastly, for uncontrollable tuples \((T_{ρ_j}^{(n)})_{j∈[ℓ]}\) of length \(ℓ > 1\) and non-basic type \(w\), we prove Proposition 7.1 by inductively invoking Proposition 7.1 for tuples of length \(ℓ\) and type \(w' < w\).

The way in which step (vi) is carried out has been illustrated in Example 9, in which seminorm control for the uncontrollable average from equation (77) is deduced from seminorm control for the lower-type controllable average from equation (79). The example below summarizes how steps (iv) and (v) proceed for a controllable average.

**Example 11.** (Inductive steps for a controllable average) Consider the tuple

\[
(T_{n}^{2}, T_{n}^{2}, T_{n}^{2+n}, T_{2}^{n^{2+n}})
\]  

(83)

of type \((3, 1)\), which is a descendant of the tuple

\[
(T_{n}^{2}, T_{n}^{2}, T_{n}^{2+n}, T_{4}^{n^{2+n}})
\]  

(84)

from Example 7 of type \((2, 2)\). While proving Proposition 7.5 for equation (83), we invoke in the ping step Proposition 7.3 for tuples

\[
(T_{n}^{2}, T_{2}^{n^{2+n}}, T_{1}^{n^{2+n}}, T_{2}^{n^{2+n}}) \quad \text{and} \quad (T_{n}^{2}, T_{n}^{2}, T_{2}^{n^{2+n}}, T_{2}^{n^{2+n}}).
\]

They are proper descendants of the original tuple in equation (84), have basic type \((4, 0)\), and Proposition 7.1 follows for them from Proposition 7.2. In the pong step of the proof of Proposition 7.5 for equation (83), we inductively invoke Proposition 7.3 for the following tuples of length 3, obtained by replacing the first and second term respectively by dual functions:

\[
(*, T_{2}^{n^{2}}, T_{3}^{n^{2+n}}, T_{2}^{n^{2+n}}) \quad \text{and} \quad (T_{1}^{n^{2}}, *, T_{3}^{n^{2+n}}, T_{2}^{n^{2+n}}).
\]

Finally, once we prove the \(T_{3}\)-seminorm control of the third term in equation (83) using Proposition 7.4, an iterated version of Proposition 7.5, we derive seminorm control of other terms in equation (83) as follows. Replacing the third term in equation (83) by a
dual function using the newly established $T_3$-control, Proposition 2.3, and the pigeonhole principle, we get a tuple
\[ (T_1^{n^2_2}, T_2^{n^2_2}, *, T_2^{n^2+n}) , \]
and then we apply Proposition 6.6 to flip the $T_2$ in the last term into $T_4$, obtaining the tuple
\[ (T_1^{n^2_2}, T_2^{n^2_2}, *, T_4^{3n^2+3n}) . \]
This tuple has length 3, and it is good for seminorm control by Proposition 7.1 applied inductively to tuples of length 3. Going back, this gives us seminorm control of the other terms in equation (83).

7.2. Proof of Proposition 7.5. We prove Proposition 7.5 for the tuple $(T_{ρ_j}^{ρ_j(n)})_{j∈[ℓ]}$ of non-basic type $w$ with the last non-zero index $t$, which is a proper descendant of $(T_{ρ_j}^{ρ_j(n)})_{j∈[ℓ]}$, by assuming that Proposition 7.1 holds for tuples of length $ℓ - 1$ as well as length $ℓ$ and type $w'.<w$. For simplicity of notation, we assume that $m = ℓ$ satisfies the controllability condition.

By Proposition 3.7, the vector $b_{s+1}$ is non-zero and takes the form $b_{s+1} = b_ℓ e_{η_ℓ} - b_i e_{η_i}$ for some $i ∈ \{0, \ldots, ℓ - 1\}$, where $b_ℓ, b_i$ are the coefficients of $p_ℓ, p_i$ of degree $d_ℓ := \deg(p_ℓ e_{η_ℓ} - p_i e_{η_i})$. If $i = 0$, then $b_ℓ ≠ 0$ since $b_{s+1}$ is non-zero. If $η_i = η_ℓ$, then the controllability condition implies that $ρ_i, ρ_ℓ$ are independent, and so $T^{b_{s+1}}$ is a non-zero iterate of $T_{η_ℓ}$. In both these cases, the result follows from the bound
\[ ||b_ℓ||b_{1, \ldots, b_s, c}e_{η_ℓ} ≤ ||b_ℓ||b_{1, \ldots, b_s, e_{η_ℓ}} \]
for any $c ≠ 0$, which is a consequence of Lemma 2.1.

For the case $η_i ≠ η_ℓ$ and $η_i ∈ J$, recall first that $η_ℓ ∈ J_ℓ$ by assumption, and the good ergodicity property of $p_1, \ldots, p_ℓ$ implies that $ρ_1, \ldots, ρ_ℓ$ have the good ergodicity property along $η$. Hence, we have $b_ℓ = β_ℓ \gcd(b_ℓ, b_i)$, $b_i = β_i \gcd(b_ℓ, b_i)$ for coprime integers $β_ℓ, β_i ∈ \mathbb{Z}$ such that $I(T^{β_ℓ e_{η_ℓ}} - β_i e_{η_i}) ⊆ I(T_ℓ)$. By Lemmas 2.1 and 2.2, we have
\[ ||b_ℓ||b_{1, \ldots, b_s, β_i e_{η_ℓ} - β_i e_{η_i} ≤ ||b_ℓ||b_{1, \ldots, b_s, e_{η_ℓ}} . \]

The last remaining case to consider, and the most difficult one, is when $η_i ∈ J_ℓ$ with $t' ≠ t$ and $b_ℓ, b_i ≠ 0$. The proof of Proposition 7.5 in this case follows the same two-step strategy that was explained in Example 1, but we also have to take into account additional complications explained in Examples 7 and 8. We first obtain the control of equation (35) by $||f_i||b_{1, \ldots, b_s, e_{η_1}}$ for some $s_1 ∈ \mathbb{N}$ depending only on $d, D, ℓ, L, s$. This is accomplished by using the control by $||f_i||b_{1, \ldots, b_{s+1}}$, given by assumption, for an appropriately defined function $\tilde{f}_i$ in place of $f_i$. Subsequently, we repeat the procedure by applying the newly established control by $||f_i||b_{1, \ldots, b_s, e_{η_1}}$ for a function $\tilde{f}_i$ in place of $f_i$. This gives us the claimed result.
Step 1 (ping): Obtaining control by a seminorm of $f_i$.

Suppose that

$$\lim_{N \to \infty} \left\| \mathbb{E} \prod_{j \in \ell} T_{\eta_j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} > 0. \quad (85)$$

The good invariance property of $f_1, \ldots, f_\ell$ implies that $f_\ell$ is invariant under $(T_{\eta_\ell}^{a_\ell}, T_\ell^{-a_\ell})^\gamma = T^c$ for some non-zero $\gamma > 0$ and $c := \gamma(a_{\eta_\ell}e_{\eta_\ell} - a_\ell e_\ell)$, where $a_{\eta_\ell}, a_\ell$ are the leading coefficients of $p_{\eta_\ell}$ and $p_\ell$. Combining this with Lemma 3.2, we deduce that

$$\lim_{N \to \infty} \left\| \mathbb{E} \prod_{j \in \ell-1} T_{\eta_j}^{\rho_j(n)} f_j \cdot T_{\eta_\ell}^{\rho_\ell(n)} \mathbb{E}(f_\ell | \mathcal{I}(T^c)) \cdot \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} > 0$$

for some function

$$f_\ell := \lim_{k \to \infty} \mathbb{E} T_{\eta_\ell}^{-\rho_\ell(n)} g_k \cdot \prod_{j \in \ell-1} T_{\eta_j}^{-\rho_j(n)} T_{\eta_j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L]} T_{\eta_\ell}^{-\rho_\ell(n)} D_j(q_j(n)),$$

where the limit is a weak limit. Then our assumption gives

$$|||\mathbb{E}(f_\ell | \mathcal{I}(T^c))|||_{b_1, \ldots, b_{s+1}} > 0.$$ 

By Proposition 3.3, we get

$$\lim \inf_{H \to \infty} \mathbb{E} \prod_{h, h' \in [H]^\ell} T_{\eta_\ell}^{-\rho_\ell(n)} g_k \cdot \prod_{j \in \ell-1} T_{\eta_j}^{-\rho_j(n)} T_{\eta_j}^{\rho_j(n)} f_j \cdot \prod_{j \in [L]} T_{\eta_\ell}^{-\rho_\ell(n)} D_j(q_j(n)) \right\|_{L^2(\mu)} > 0,$$

where $u_{h, h'}$ are 1-bounded and invariant under both $T^{b_{s+1}}$ and $T^c$, and

$$\mathcal{D}_j^{h, h'}(n) := \Delta_{b_1, \ldots, b_s, h, h'} D_j(n)$$

is a product of $2^s$ elements of $\mathcal{D}_d$. As a consequence of the $T^{b_{s+1}}$-invariance of $u_{h, h'}$, we have

$$T_{\eta_\ell}^{b_{s+1}} u_{h, h'} = T_{\eta_\ell}^{b_{s+1}} u_{h, h'}, \quad (86)$$

where we use the identity $T_{\eta_\ell}^{b_{s+1}} = T_{\eta_\ell}^{b_{s+1}} T^{b_{s+1}}$. Let $\lambda \in \mathbb{N}$ be the smallest natural number such that $b_\ell$ divides the coefficients of $\lambda \rho_\ell$. By the triangle inequality,

$$\lim \inf_{H \to \infty} \mathbb{E} \prod_{h, h' \in [H]^\ell} \prod_{r \in [0, \ldots, \lambda-1]} T_{\eta_\ell}^{(\lambda r + r)} (\Delta_{b_1, \ldots, b_s, h, h'} f_j) \cdot T_{\eta_\ell}^{\rho_\ell(\lambda r + r)} u_{h, h'} \times \prod_{j \in [L]} \mathcal{D}_j^{h, h'}(q_j(\lambda r + r)) \right\|_{L^2(\mu)} > 0.$$
Using equation (86) and the pigeonhole principle, and setting \( \eta' = \tau_{\ell i} \eta \), we get that for some \( r_0 \in \{0, \ldots, \lambda - 1\} \), we have

\[
\limsup_{H \to \infty} \frac{\mathbb{E}}{h, h' \in [H]} \left| \lim_{N \to \infty} \mathbb{E} \left[ n \in [N] \prod_{j \in [\ell]} T_{\eta_j}^{\rho'_j(n)} \cdot \prod_{j \in [L]} D'_{j, h, h'}(q'_j(n)) \right] \right|_{L^2(\mu)} > 0, \tag{87}
\]

where

\[
\rho'_j(n) := \begin{cases} 
\rho_j(\lambda n + r_0) - \rho_j(r_0), & j \in [\ell - 1], \\
\frac{b_j}{b_\ell} (\rho_\ell(\lambda n + r_0) - \rho_\ell(r_0)), & j = \ell,
\end{cases}
\]

and \( q'_j(n) := q_j(\lambda n + r_0) \). We note that the polynomials \( \rho'_1, \ldots, \rho'_\ell \) are as in Proposition 6.1.

It follows from equation (87) that there exists a set \( B \subseteq \mathbb{N}^{2s} \) of positive upper density and \( \varepsilon > 0 \) such that

\[
\lim_{N \to \infty} \left| \mathbb{E} \left[ n \in [N] \prod_{j \in [\ell]} T_{\eta_j}^{\rho'_j(n)} \cdot \prod_{j \in [L]} D'_{j, h, h'}(q'_j(n)) \right] \right|_{L^2(\mu)} \geq \varepsilon \tag{88}
\]

for every \( (h, h') \in B \).

By assumption, \( t \) is the last non-zero index of \( w \), implying that \( t' < t \). Hence, the new tuple \( (T_{\eta_j}^{\rho'_j(n)})_{j \in [\ell]} \) has type \( w' = \sigma_{tt'} w < w \). Furthermore, by Proposition 6.5, the functions \( f_{j, h, h'} \) have the good invariance property along \( \eta' \) with respect to \( p_1, \ldots, p_\ell \). We therefore inductively apply Proposition 7.3 for tuples of length \( \ell \) and type \( w' \) to each average from equation (88). This allows us to conclude that there exist \( s_1 \in \mathbb{N} \), depending only on \( d, D, \ell, L, s \), and \( \delta > 0 \) such that

\[
\| \Delta_{b_1, \ldots, b_s, h - h'} f_i \|_{s_1, T_{\eta_j}} \geq \delta
\]

for \( (h, h') \in B \). Hence,

\[
\limsup_{H \to \infty} \frac{\mathbb{E}}{h, h' \in [H]^s} \| \Delta_{b_1, \ldots, b_s, h - h'} f_i \|_{s_1, T_{\eta_j}} > 0. \tag{89}
\]

Together with Lemma 3.1, the inductive formula for seminorms in equation (15), and Hölder inequality, the inequality in equation (89) implies that

\[
\| f_i \|_{b_1, \ldots, b_s, e_{\eta_j}^{s_1}} > 0,
\]

and so the seminorm \( \| f_i \|_{b_1, \ldots, b_s, e_{\eta_j}^{s_1}} \) controls the average from equation (35).

**Step 2 (pong):** Obtaining control by a seminorm of \( f_\ell \).

To get the claim that \( \| f_\ell \|_{b_1, \ldots, b_s, e_{\eta_j}^{s_1}} \) controls the average for some \( s' \in \mathbb{N} \) depending only on \( d, D, \ell, L, s \), we repeat the procedure once more with \( f_i \) in place of \( f_\ell \). From equation (85), it follows that
Proposition 6.5 implies that the functions $g_j$ give polynomials where the limit is a weak limit. Then the previous result gives

$$
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell], j \neq i} T_{\eta_j}^{\rho_j(n)} f_j \cdot T_{\eta_i}^{\rho_i(n)} \tilde{f}_i \cdot \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} > 0
$$

for some function

$$
\tilde{f}_i := \lim_{k \to \infty} \mathbb{E}_{n \in [N_k]} T_{\eta_i}^{-\rho_i(n)} g_k \cdot \prod_{j \in [\ell], j \neq i} T_{\eta_j}^{-\rho_j(n)} T_{\eta_i}^{\rho_i(n)} f_j \cdot \prod_{j \in [L]} T_{\eta_j}^{-\rho_j(n)} D_j(q_j(n))
$$

where the limit is a weak limit. Then the previous result gives

$$
\|\tilde{f}_i\|_{b_1, \ldots, b_1, \eta_{\xi_1}} > 0.
$$

By Proposition 3.3, we get

$$
\liminf_{H \to \infty} \mathbb{E}_{L \in [H]} \lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell], j \neq i} T_{\eta_j}^{\rho_j(n)} (\Delta_{b_1, \ldots, b_1, h-h'} f_j) \right\|_{L^2(\mu)} > 0
$$

where

$$
\mathcal{D}_{j, h, h'}(n) := \begin{cases} 
\Delta_{b_1, \ldots, b_1, h-h'} D_j(n), & j \in [L],
\prod_{\xi \in \{0,1\}^s} C(\xi) \mathcal{D}_{s_1, T_{\eta_i}} (\Delta_{b_1, \ldots, b_1, h-h'} \tilde{f}_i), & j = L + 1.
\end{cases}
$$

Thus, the sequence of functions $\mathcal{D}_{j, h, h'}$ is a product of $2^s$ elements of $\mathcal{D}_j$ if $j \in [L]$, and it is a product of $2^s$ elements of $\mathcal{D}_{s_1}$ for $j = L + 1$. Consequently, there exists $\varepsilon > 0$ and a set $B' \subset \mathbb{N}^{2s}$ of positive lower density such that for every $(h, h') \in B'$, we have

$$
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell], j \neq i} T_{\eta_j}^{\rho_j(n)} (\Delta_{b_1, \ldots, b_1, h-h'} f_j) \cdot \prod_{j \in [L+1]} \mathcal{D}_{j, h, h'}(q_j(n)) \right\|_{L^2(\mu)} > \varepsilon.
$$

Proposition 6.5 implies that the functions $(g_{1, h, h'}, \ldots, g_{\ell, h, h'})_{(h, h') \in \mathbb{N}^{2s}}$ given by

$$
g_{j, h, h'} := \begin{cases} 
\Delta_{b_1, \ldots, b_1, h-h'} f_j, & j \neq i,
1, & j = i.
\end{cases}
$$

have the good invariance property along $\eta$ with respect to $p_1, \ldots, p_\ell$. Proposition 6.6 then gives polynomials $\rho_1', \ldots, \rho_\ell', q_1', \ldots, q_\ell' \in \mathbb{Z}[n]$, 1-bounded functions $g_{1, h, h'}', \ldots, g_{\ell, h, h'}'$ with $g_{i, h, h'}' := 1$ such that

$$
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell], j \neq i} T_{\eta_j}^{\rho_j(n)} g_{j, h, h'}' \cdot \prod_{j \in [L+1]} \mathcal{D}_{j, h, h'}(q_j'(n)) \right\|_{L^2(\mu)} > \varepsilon. \quad (90)
$$
It also gives a constant $C > 0$ depending only on $\gamma$ and the leading coefficients of $p_1, \ldots, p_\ell$ such that
\[
\|g_{j, h', h''}^{\prime} \|_{s', T_j} \leq C \| \Delta_{b_1, \ldots, b_\ell} h' f_j \|_{s', T_{n_\ell}}
\]
for all $j \in [\ell] \setminus \{i\}$, $(h', h'') \in B'$, and $s' \geq 2$. Observing that the averages from equation (90) have length $\ell - 1$ and $\rho_1', \ldots, \rho_{\ell - 1}', \rho_{\ell + 1}', \ldots, \rho_\ell'$ have the good ergodicity property for the system $(X, \mathcal{X}, \mu, T_1, \ldots, T_{\ell - 1}, T_{\ell + 1}, \ldots, T_{\ell})$ (another consequence of Propositions 6.6 and 6.4), we conclude from equation (91) and Proposition 7.3 that there exist $s' \in \mathbb{N}$ (depending only on $d, D, \ell, L, s$) and $\delta > 0$ satisfying
\[
\| \Delta_{b_1, \ldots, b_\ell} h' f_j \|_{s', T_{n_\ell}} > \delta
\]
for every $(h', h'') \in B'$. Consequently, we deduce that
\[
\lim_{H \to \infty} \inf_{h', h'' \in [H]'} \mathbb{E} \| \Delta_{b_1, \ldots, b_\ell} h' f_j \|_{s', T_{n_\ell}} > 0.
\]

It then follows from Lemma 3.1 and the Hölder inequality that $\| f_{\ell} \|_{b_1, \ldots, b_{\ell}, s_{n_\ell} < s'} > 0$, as claimed.

7.3. Proof of Proposition 7.1. We induct on the length $\ell$ of the average, and for each fixed $\ell$, we further induct on type. In the base case $\ell = 1$, Proposition 7.1 follows directly from Proposition 3.7. We assume therefore that the average has length $\ell > 1$ and type $w$, and the statement holds for averages of length $\ell - 1$ as well as length $\ell$ and type $w' < w$. If the type $w$ is basic, then Proposition 7.1 follows from Proposition 7.2, so we assume that $w$ is not basic. We argue differently on whether the average is controllable or not.

Case 1: Controllable averages.

If the average is controllable, then Proposition 7.4 gives $s \in \mathbb{N}$ depending only on $d, D, \ell, L$ as well as $m \in [\ell]$ such that equation (80) holds whenever $\| f_m \|_{s, T_{n_m}} = 0$. Suppose now that
\[
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell]} T_{n_j}^{\rho_j(n)} f_j \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} > 0.
\]
Applying the fact that $\| f_m \|_{s, T_{n_m}}$ controls this average, Proposition 2.3, and the pigeonhole principle, we replace $f_m$ by a dual function of level $s$, so that we have
\[
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell], \ j \neq m} T_{n_j}^{\rho_j(n)} f_j \cdot D(\rho_j(n)) \cdot \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} > 0
\]
for some $D \in \mathcal{D}_s$. By Proposition 6.6, there exist 1-bounded functions $(f_j')_{j \in [\ell]}$ with $f_j' = 1$ and polynomials $\rho_1', \ldots, \rho_{\ell}'$, $q_1', \ldots, q_{\ell}' \in \mathbb{Z}[n]$ such that
\[
\lim_{N \to \infty} \left\| \mathbb{E}_{n \in [N]} \prod_{j \in [\ell], \ j \neq m} T_{n_j}^{\rho_j(n)} f_j' \cdot D(\rho_j(n)) \cdot \prod_{j \in [L]} D_j(q_j(n)) \right\|_{L^2(\mu)} > 0.
\]
Proofs of joint ergodicity results

8. Proof of joint ergodicity results

In this section, we derive Theorem 1.2 and Corollaries 1.3 and 1.4. We start with two observations that connect the notions of joint ergodicity and weak joint ergodicity. Their proofs are straightforward and hence we skip them.

**Lemma 8.1.** Let \((X, \mathcal{X}, \mu, T)\) be a system. Suppose that there exists a sequence \(a : \mathbb{N} \to \mathbb{Z}\) such that \((T^{a(n)})_{n \in \mathbb{N}}\) is ergodic for \(\mu\). Then \(T\) is ergodic.

**Lemma 8.2.** Let \(a_1, \ldots, a_\ell : \mathbb{N} \to \mathbb{Z}\) be sequences and \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\) be a system. The sequences are jointly ergodic for the system if and only if they are weakly jointly ergodic and the transformations \(T_1, \ldots, T_\ell\) are ergodic.

We continue with the proof of Theorem 1.2.

**Proof of Theorem 1.2.** Suppose first that the polynomials \(p_1, \ldots, p_\ell\) have the good ergodicity property for the system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\). By Theorem 1.1, this implies that \(p_1, \ldots, p_\ell\) are good for the seminorm control for the system \((X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)\). This result, property (ii), and Theorem 2.4 imply that \(p_1, \ldots, p_\ell\) are weakly jointly ergodic for the system.
Conversely, suppose that the polynomials are weakly jointly ergodic for the system. The condition (ii) follows by taking $f_1, \ldots, f_\ell$ to be non-ergodic eigenfunctions of respective transformations. To prove condition (i), suppose that $p_i/c_i = p_j/c_j$ for some $i \neq j$ and coprime integers $c_i, c_j$, and there exists a function $f$ invariant under $T_i^{c_i} T_j^{-c_j}$ that is not simultaneously invariant under $T_i$ and $T_j$. The invariance property of $f$ gives $T_i^{c_i} f = T_j^{c_j} f$, and the same holds for $\overline{T}$. The coprimeness of $c_i, c_j$ implies that the polynomial $p_i/c_i = p_j/c_j$ has integer coefficients, and so we have

$$
\mathbb{E} \left[ T_i^{p_i(n)} f \cdot T_j^{p_j(n)} \overline{f} \right] = \mathbb{E} \left[ T_i^{p_i(n)} |f|^2 \right] = \mathbb{E} \left[ T_j^{p_j(n)} |f|^2 \right],
$$

which by the weak joint ergodicity of $p_1, \ldots, p_\ell$ converges to $\mathbb{E}(|f|^2 |I(T_i)) = \mathbb{E}(|f|^2 |I(T_j))$ in $L^2(\mu)$. However, the weak joint ergodicity of $p_1, \ldots, p_\ell$ implies that

$$
\lim_{N \to \infty} \mathbb{E} \left[ T_i^{p_i(n)} f \cdot T_j^{p_j(n)} \overline{f} \right] = \mathbb{E}(f |I(T_i)) \cdot \mathbb{E}(f |I(T_j))
$$

in $L^2(\mu)$. Hence, $\mathbb{E}(|f|^2 |I(T_i)) = \mathbb{E}(f |I(T_i)) \cdot \mathbb{E}(f |I(T_j))$. The properties of the conditional expectation and the Cauchy–Schwarz inequality imply that

$$
\int \|\mathbb{E}(f |I(T_i)) \cdot \mathbb{E}(f |I(T_j))\| d\mu \leq \|\mathbb{E}(f |I(T_i))\|_{L^2(\mu)} \cdot \|\mathbb{E}(f |I(T_j))\|_{L^2(\mu)}
$$

$$
\leq \|f\|^2_{L^2(\mu)} = \int |f|^2 d\mu = \int \mathbb{E}(|f|^2 |I(T_i)) d\mu.
$$

The two inequalities above become an equality precisely when $f = \mathbb{E}(f |I(T_i)) = \mathbb{E}(f |I(T_j))$ holds $\mu$-a.e., that is, when $f$ is simultaneously invariant under $T_i$ and $T_j$, and so either this is the case, contradicting the assumptions on $f$, or $\mathbb{E}(|f|^2 |I(T_i)) \neq \mathbb{E}(f |I(T_i)) \cdot \mathbb{E}(f |I(T_j))$, contradicting the weak joint ergodicity of $p_1, \ldots, p_\ell$.

We now derive Corollary 1.3 from Theorem 1.2.

**Proof of Corollary 1.3.** By Lemma 8.2, the polynomials $p_1, \ldots, p_\ell$ are jointly ergodic for the system $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ if and only if they are weakly jointly ergodic for this system and the transformations $T_1, \ldots, T_\ell$ are ergodic. Theorem 1.2 in turn implies that this is equivalent to the system having the good ergodicity property, the transformations $T_1, \ldots, T_\ell$ being ergodic, and equation (6) holding for all eigenfunctions. Since the transformations are ergodic, all the eigenfunctions $\chi_j$ of $T_j$ satisfy $T_j \chi_j = \lambda_j \chi_j$ for a constant $\lambda_j$, and so the condition of equation (6) reduces to equation (7) upon taking $\lambda_j = e(\alpha_j)$ and realizing that $\int \chi_j d\mu = 0$ unless $\alpha_j = 0$. Lastly, the good ergodicity property and the ergodicity of the transformations $T_1, \ldots, T_\ell$ jointly imply the very good ergodicity property.

Finally, we prove Corollary 1.4.

**Proof of Corollary 1.4.** The forward direction follows from [7, Proposition 5.3], and so it is enough to deduce the reverse direction. Our goal is to show that the conditions (i) and (ii) in the statement of Conjecture 1 imply the conditions (i) and (ii) in the statement of Corollary 1.3. The condition (ii) in Conjecture 1, that is, the ergodicity of...
(T^{p_i(n)}_1, \ldots, T^{p_\ell(n)}_\ell)_{n \in \mathbb{N}}$, implies the condition (ii) in Corollary 1.3 by taking eigenfunctions. By Lemma 8.1, it also implies the ergodicity of $T_1, \ldots, T_\ell$ because each sequence $(T^{p_j(n)}_j)_{n \in \mathbb{N}}$ is ergodic.

To establish the very good ergodicity property of $p_1, \ldots, p_\ell$, suppose that $p_i/c_i = p_j/c_j$ for some relatively prime $c_i, c_j \in \mathbb{Z}$, and $f$ is a non-constant function invariant under $T^{c_i}_i T^{-c_j}_j$. Letting $q := p_i/c_i = p_j/c_j$ and noting that it has integer coefficients due to the coprimeness of $c_i, c_j$, we observe that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} f(T^{p_i(n)}_i T^{-p_j(n)}_j) = \lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} f(T^{c_i}_i T^{-c_j}_j)q(n) = f$$

contradicting the ergodicity of $(T^{p_i(n)}_i T^{-p_j(n)}_j)_{n \in \mathbb{N}}$.

**Acknowledgements.** We thank the anonymous referee for many useful suggestions. The authors were supported by the Hellenic Foundation for Research and Innovation, Project No. 1684.

**References**

[1] D. Berend and V. Bergelson. Jointly ergodic measure preserving transformations. *Israel J. Math.* 49(4) (1984), 307–314.

[2] V. Bergelson. Weakly mixing PET. *Ergod. Th. & Dynam. Sys.* 7(3) (1987), 337–349.

[3] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden’s and Szemerédi’s theorems. *J. Amer. Math. Soc.* 9 (1996), 725–753.

[4] V. Bergelson and A. Leibman. Cubic averages and large intersections. Recent trends in ergodic theory and dynamical systems. *Contemp. Math.* 631 (2015), 5–19.

[5] A. Best and A. Ferré-Moragues. Polynomial ergodic averages for certain countable ring actions. *Discrete Contin. Dyn. Syst.* 42(7) (2022), 3379–3413.

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

[7] S. Donoso, A. Koutsogiannis and W. Sun. Seminorms for multiple averages along polynomials and applications to joint ergodicity. *J. Anal. Math.* 146 (2022), 1–64.

[8] N. Frantzikinakis. A multidimensional Szemerédi theorem for Hardy sequences of polynomial growth. *Trans. Amer. Math. Soc.* 367 (2015), 5653–5692.

[9] N Frantzikinakis. Joint ergodicity of sequences. *Adv. Math.* to appear. Preprint, 2022, arXiv:2102.09967.

[10] N. Frantzikinakis and B. Host. Weighted multiple ergodic averages and correlation sequences. *Ergod. Th. & Dynam. Sys.* 38(1) (2018), 81–142.

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

[12] B. Host. Ergodic seminorms for commuting transformations and applications. *Studia Math.* 195(1) (2009), 31–49.

[13] B. Host and B. Kra. Non-conventional ergodic averages and nilmanifolds. *Ann. of Math.* (2) 161 (2005), 397–488.

[14] B. Host and B. Kra. *Nilpotent Structures in Ergodic Theory (Mathematical Surveys and Monographs, 236).* American Mathematical Society, Providence, RI, 2018.

[15] S. Peluse. On the polynomial Szemerédi theorem in finite fields. *Duke Math. J.* 168 (2019), 749–774.

[16] S. Peluse. Bounds for sets with no polynomial progressions. *Forum Math. Pi* 8 (2020), e16.

[17] S. Peluse and S. Prendiville. Quantitative bounds in the non-linear Roth theorem. Preprint, 2022, arXiv:1903.02592.

[18] T. Tao and T. Ziegler. Concatenation theorems for anti-Gowers-uniform functions and Host–Kra characteristic factors. *Discrete Analysis* 13 (2016), 60.

[19] M. Walsh. Norm convergence of nilpotent ergodic averages. *Ann. of Math.* (2) 175(3) (2012), 1667–1688.

[20] P. Zorin-Kranich. Norm convergence of multiple ergodic averages on amenable groups. *J. Anal. Math.* 130 (2016), 219–241.