

# The Density Finite Sums Theorem

By BRYNA KRA and JOEL MOREIRA and FLORIAN K. RICHTER  
and DONALD ROBERTSON

September 12, 2025

## Abstract

For any set  $A$  of natural numbers with positive upper Banach density and any  $k \geq 1$ , we show the existence of an infinite set  $B \subset \mathbb{N}$  and a shift  $t \geq 0$  such that  $A - t$  contains all sums of  $m$  distinct elements from  $B$  for all  $m \in \{1, \dots, k\}$ . This can be viewed as a density analog of Hindman's finite sums theorem. Our proof reveals the natural relationships among infinite sumsets, the dynamics underpinning arithmetic progressions, and homogeneous spaces of nilpotent Lie groups.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Infinite sumsets . . . . .	1
1.2	Outline of the proof of Theorem 1.2 . . . . .	3
1.3	Acknowledgments . . . . .	4
<b>2</b>	<b>Translation to measure preserving systems</b>	<b>4</b>
<b>3</b>	<b>A variation on recurrence in <math>X^{k+1}</math></b>	<b>7</b>
<b>4</b>	<b>Constructing a progressive measure</b>	<b>9</b>
4.1	Two motivating examples . . . . .	9
4.2	Pronilfactors . . . . .	10
4.3	Constructing the measure $\sigma$ . . . . .	11
4.4	Outlining the proof of Theorem 3.3 . . . . .	13
<b>5</b>	<b>Proof of Theorem 4.8</b>	<b>13</b>
<b>6</b>	<b>Proof of Theorem 4.9</b>	<b>15</b>
6.1	Ingredients in the proof of Theorem 4.9 . . . . .	15
6.2	An extension of Furstenberg's multiple recurrence theorem . . . . .	16
6.3	Uniformity norms . . . . .	18
6.4	Proof of Theorem 6.2 . . . . .	22

## 1. Introduction

### 1.1. Infinite sumsets

In 1927, van der Waerden [32] proved a classic result in Ramsey theory: no matter how one partitions the natural numbers  $\mathbb{N} = \{1, 2, 3, \dots\}$  into finitely many pieces, some piece

contains arbitrarily long arithmetic progressions. Settling a long standing conjecture of Erdős and Turán, Szemerédi [30] proved in 1975 a density version of van der Waerden’s theorem. Namely, he showed that any set  $A \subset \mathbb{N}$  whose *upper Banach density*, meaning

$$d^*(A) := \limsup_{N \rightarrow \infty} \left\{ \frac{|A \cap \{M + 1, \dots, M + N\}|}{N} : M \in \mathbb{N} \right\},$$

is positive contains arbitrarily long arithmetic progressions.

Around the same time, in a major breakthrough, Hindman [13] settled a conjecture of Graham and Rothschild showing that no matter how one partitions the natural numbers into finitely many pieces, some piece always contains an IP-set (also called a finite sums set)

$$\{b(i_1) + \dots + b(i_m) : m \in \mathbb{N}, i_1 < \dots < i_m \in \mathbb{N}\} \quad (1.1)$$

for some strictly increasing sequence  $b: \mathbb{N} \rightarrow \mathbb{N}$ . Erdős [5] then tried to find the underlying behavior connecting these two results, writing “I have tried to formulate a conjecture which would be in the same relation to Hindman’s theorem as Szemerédi’s theorem is to van der Waerden’s.” We believe that our main theorem comes as close as is possible to giving a density version.

**Theorem 1.1.** *If  $A \subset \mathbb{N}$  has positive upper Banach density then for every  $k \in \mathbb{N}$  there is a strictly increasing sequence  $b: \mathbb{N} \rightarrow \mathbb{N}$  and an integer  $t \geq 0$  such that*

$$\{b(i_1) + \dots + b(i_m) : 1 \leq m \leq k, i_1 < \dots < i_m \in \mathbb{N}\} \quad (1.2)$$

*is a subset of  $A - t$ .*

The shift is in general necessary: for example, the odd numbers do not contain a configuration of the form (1.2) when  $k \geq 2$ . Furthermore, an example of Straus (published by others, see for example [14]) exhibits a set with density arbitrarily close to 1 that contains no shift of an IP-set, explaining the need to curtail  $m$ .

The  $k = 2$  case of Theorem 1.1 was proved in our previous work [24]. The methods used there do not generalize to the setting of Theorem 1.1, which was conjectured in [24, Conjecture 1.5] and [25, Conjecture 2.4]. Apart from the initial translation of Theorem 1.1 into a dynamical statement (Theorem 1.2), our proof of Theorem 1.1 is new even in the case  $k = 2$  and gives, in particular, a simpler proof of the main result in [24].

Theorem 1.1 follows a long history of results on infinite sumset configurations in large subsets of the integers, initiated by various conjectures of Erdős from the late 1970s and early 1980s (see [8, p. 85], [5, p. 305], [6, pp. 57–58], and [7, p. 105]). Hindman [14, 15] gave various examples and provided reformulations and refinements of some of these conjectures. Early results include the existence [29] of sumsets  $B + C$  with  $B \subset \mathbb{N}$  infinite and  $C \subset \mathbb{N}$  finite but arbitrarily large, and the existence [1] of restricted sumset configurations  $\{b + c : b \in B, c \in C, b < c\}$  where  $B$  and  $C$  are infinite. In [4], it was proved that any set of density strictly greater than  $1/2$  contains  $B + C$  with both  $B$  and  $C$  infinite. In [28], the weakest of Erdős’s conjectures was resolved. A more streamlined proof [17] simplified many of the technical arguments in [28]. The generalization to  $k$ -fold sumsets  $B_1 + \dots + B_k$  with  $B_1, \dots, B_k \subset \mathbb{N}$  infinite was proved by the authors in [23]. Recent related results include work on analogs of Erdős’s conjectures in more general groups [3], work on unrestricted sumsets [22, 21], and

work on sumsets in the primes [31]. We refer to our survey [25] for further references and variations.

Our proof of Theorem 1.1 relies on ergodic theory: the first step is to reduce the combinatorial problem to a dynamical statement, following Furstenberg’s framework for recasting combinatorial questions in terms of properties of orbits in a measure preserving system. Our main dynamical theorem, from which we derive Theorem 1.1 in Section 2, states the following.

**Theorem 1.2.** *Let  $T$  be a homeomorphism of a compact metric space  $X$  and let  $\mu$  be a  $T$ -invariant Borel probability measure on  $X$ . For every point  $a \in X$  with*

$$\mu(\overline{\{T^n a : n \in \mathbb{N}\}}) = 1,$$

*for every  $k \in \mathbb{N}$ , and for every open set  $E \subset X$  with  $\mu(E) > 0$ , there exist an integer  $t \geq 0$ , points  $x_1, \dots, x_k \in X$ , and a strictly increasing sequence  $c: \mathbb{N} \rightarrow \mathbb{N}$  such that*

$$\lim_{n \rightarrow \infty} T^{c(n)} a = x_1, \quad \lim_{n \rightarrow \infty} T^{c(n)} x_1 = x_2, \quad \dots, \quad \lim_{n \rightarrow \infty} T^{c(n)} x_{k-1} = x_k, \quad (1.3)$$

*all hold, and  $x_i \in T^{-t} E$  for all  $i = 1, \dots, k$ .*

The proof that Theorem 1.2 implies Theorem 1.1 is contained in Section 2 and uses by now well-understood techniques. The proof of Theorem 1.2 then occupies the rest of the paper.

## 1.2. Outline of the proof of Theorem 1.2

We describe the strategy used to prove Theorem 1.2, referring as needed to terminology and notation introduced later in the paper.

We start by introducing a new class of measures on  $X^{k+1}$  defined by a recurrence property which we call *progressive* (Definition 3.1). We show that if a measure on  $X^{k+1}$  is progressive, then any open set  $U \subset X^{k+1}$  with positive measure contains a point  $(a, x_1, \dots, x_k)$  satisfying (1.3) – we call such points *Erdős progressions* (Definition 2.1). Thus every progressive measure is supported on the closure of the set of Erdős progressions.

Our goal, then, is to associate to each system  $(X, \mu, T)$  a measure  $\sigma$  on  $X^{k+1}$  that is progressive. In the extreme case when the system is a rotation on a compact abelian group, there is a concise algebraic description of the set of Erdős progressions, which allows for a simple and explicit construction of a progressive measure (see Example 4.1). More generally, if  $X$  is a nilmanifold and  $(X, \mu, T)$  is a nilsystem (see Definition 4.3), one can also give an explicit description of a progressive measure as the Haar measure on a suitable sub-nilmanifold. In the opposite case of weakly mixing systems, we can easily show that the product measure is progressive due to the randomness inherent in such a system (see Example 4.2).

Motivated by the insights of these two contrasting examples, we construct in Section 4 for any system  $(X, T)$  a measure  $\sigma$  on  $X^{k+1}$  (see Definition 4.5) consisting of a structured component arising from the maximal pronilfactor of  $(X, T)$  and a random component that is relatively independent over the maximal pronilfactor. The proof of our main result then

reduces to two tasks:

- (i) Showing that for some  $t \in \mathbb{N}$ , we have  $\sigma(\{a\} \times T^{-t}E \times \cdots \times T^{-t}E) > 0$ ;
- (ii) Proving that  $\sigma$  is a progressive measure.

Since (i) and (ii) together imply that  $\{a\} \times T^{-t}E \times \cdots \times T^{-t}E$  contains an Erdős progression, namely, a point  $(a, x_1, \dots, x_k)$  satisfying (1.3), the conclusion of Theorem 1.2 follows.

The proof of (i) is carried out in Section 5 and uses Furstenberg’s multiple recurrence theorem, that the maximal pronilfactor is the characteristic factor for multiple ergodic averages, that products of pronilsystems are pronilsystems, and that orbit closures in pronilsystems are uniquely ergodic. Our proof of (ii), which occupies Section 6, includes a novel extension (Theorem 6.3) of Furstenberg’s multiple recurrence theorem, several advanced properties of uniformity norms (established in Section 6.3), and again basic facts about pronilsystems such as being closed under products and that every orbit closure in a pronilsystem is uniquely ergodic.

### 1.3. Acknowledgments

The first author was partially supported by the National Science Foundation grant DMS-2348315. The second author was supported by the EPSRC Frontier Research Guarantee grant EP/Y014030/1. The third author was supported by the Swiss National Science Foundation grant TMSGI2-211214. We thank Ethan Acklesberg and Ioannis Kousek for numerous helpful comments on an earlier version of the article and Borys Kuca for providing useful references, and we thank the International Centre for Mathematical Science and Northwestern University for hosting and funding activities at which the authors collaborated. Finally we thank the referee for helpful suggestions that improved the article.

## 2. Translation to measure preserving systems

In this section, we show that Theorem 1.1 follows from Theorem 1.2.

By a *topological system* we mean a pair  $(X, T)$  where  $X$  is a compact metric space and  $T: X \rightarrow X$  is a homeomorphism. Given a compact metric space  $Y$ , let  $\mathcal{M}(Y)$  denote the set of all Borel probability measures on  $Y$ . By a *measure preserving system* we mean a tuple  $(X, \mu, T)$  where  $(X, T)$  is a topological system and  $\mu \in \mathcal{M}(X)$  is invariant under  $T$ . All spaces are implicitly endowed with their Borel  $\sigma$ -algebras and so we do not include the  $\sigma$ -algebras in our notation. A measure preserving system is said to be *ergodic* if any  $T$ -invariant Borel subset of  $X$  has either measure 0 or measure 1.

Building on Furstenberg’s dynamical approach to combinatorial problems [12], in our earlier work [24] we established a connection between sumsets in  $\mathbb{N}$  of the form (1.2) for  $k = 2$  and “Erdős progressions” in topological systems. We extend that definition to handle sumsets of the form (1.2).

**Definition 2.1** (cf. [24, Definition 2.1]). Given a topological system  $(X, T)$  and  $k \in \mathbb{N}$ , a point

$$(x_0, x_1, \dots, x_{k-1}, x_k) \in X^{k+1}$$

is a  $(k + 1)$ -term Erdős progression if there exists a strictly increasing sequence  $c : \mathbb{N} \rightarrow \mathbb{N}$  such that

$$T^{c(n)}x_0 \rightarrow x_1, \quad T^{c(n)}x_1 \rightarrow x_2, \quad \dots \quad T^{c(n)}x_{k-1} \rightarrow x_k \quad (2.1)$$

as  $n \rightarrow \infty$ .

The next lemma demonstrates the connection between  $(k + 1)$ -term Erdős progressions and sumsets.

**Lemma 2.2.** *Fix  $k \in \mathbb{N}$ , a topological system  $(X, T)$ , and open sets  $U_1, \dots, U_k \subset X$ . If there exists an Erdős progression  $(x_0, \dots, x_k) \in X^{k+1}$  with  $x_j \in U_j$  for  $j = 1, \dots, k$ , then there exists a strictly increasing sequence  $b : \mathbb{N} \rightarrow \mathbb{N}$  such that*

$$\{b(i_1) + \dots + b(i_m) : i_1 < \dots < i_m \in \mathbb{N}\} \subset \{n \in \mathbb{N} : T^n x_0 \in U_m\}$$

for  $m = 1, \dots, k$ .

*Proof.* The special case  $k = 2$  was established in [24, Theorem 2.2], and the proof of the general case is essentially the same, with only slightly more complicated notation.

Let  $c : \mathbb{N} \rightarrow \mathbb{N}$  be a strictly increasing sequence such that (2.1) holds as  $n \rightarrow \infty$ . Set  $U_{j,0} = U_j$  for all  $1 \leq j \leq k$ .

Each  $U_{j,0}$  is a neighborhood of  $x_j$ , and so there exists  $n_1 \in \mathbb{N}$  such that  $b(1) := c(n_1)$  satisfies  $T^{b(1)}x_{j-1} \in U_{j,0}$  for  $j = 1, \dots, k$ . Set  $U_{j,1} := U_{j,0} \cap T^{-b(1)}U_{j+1,0}$  for each  $j = 1, \dots, k - 1$  and set  $U_{k,1} := U_k$ .

Note that  $U_{j,1}$  is an open set that contains  $x_j$  for  $j = 1, \dots, k$  and so there exists  $n_2 > n_1$  such that  $b(2) := c(n_2)$  satisfies  $T^{b(2)}x_{j-1} \in U_{j,1}$  for  $j = 1, \dots, k$ . We then set  $U_{j,2} := U_{j,1} \cap T^{-b(2)}U_{j+1,1}$  for each  $j = 1, \dots, k - 1$  and set  $U_{k,2} := U_k$ .

We continue inductively, noting that for each  $i \in \mathbb{N}$  and  $j = 1, \dots, k$  the open set  $U_{j,i}$  contains  $x_j$ , so that we can find some  $n_i > n_{i-1}$  such that  $b(i) := c(n_i)$  satisfies  $T^{b(i)}x_{j-1} \in U_{j,i-1}$  for  $j = 1, \dots, k$ . We then let  $U_{j,i} := U_{j,i-1} \cap T^{-b(i)}U_{j+1,i-1}$  for each  $j = 1, \dots, k - 1$  and set  $U_{k,i} := U_k$ .

Let  $m \leq k$  and fix  $i_1 < \dots < i_m$  in  $\mathbb{N}$ . We need to verify that  $T^{b(i_1)+\dots+b(i_m)}x_0 \in U_m$ . We have

$$T^{b(i_m)}x_0 \in U_{1,i_m-1} \subset U_{1,i_m-1}$$

because  $i_m - 1 \geq i_{m-1}$ . Noting that, from the construction,  $U_{1,i_m-1} \subset T^{-b(i_{m-1})}U_{2,i_{m-1}-1}$ , and since  $i_{m-1} - 1 \geq i_{m-2}$ , we have

$$T^{b(i_{m-1})+b(i_m)}x_0 \in T^{b(i_{m-1})}U_{1,i_m-1} \subset U_{2,i_{m-1}-1} \subset U_{2,i_{m-2}}$$

Proceeding in this manner, we deduce that

$$T^{b(i_{m-h})+\dots+b(i_{m-1})+b(i_m)}x_0 \in U_{h+1,i_{m-h-1}}$$

for  $h = 0, \dots, m - 1$ . As  $U_{m,0} = U_m$ , the desired result is the case  $h = m - 1$ .  $\square$

To connect Lemma 2.2 to Theorem 1.1, we use a version of Furstenberg's correspondence principle.

**Theorem 2.3** (Correspondence principle, cf. [11, 12]). *Given a set  $A \subset \mathbb{N}$  with positive upper Banach density, there exists a measure preserving system  $(X, \mu, T)$ , a point  $a \in X$  with  $\mu(\overline{\{T^n a : n \in \mathbb{N}\}}) = 1$ , and an open set  $E \subset X$  such that  $A = \{n \in \mathbb{N} : T^n a \in E\}$  and  $\mu(E) > 0$ .*

Throughout we work with invertible measure preserving transformations, but for our combinatorial conclusions, we need the measure to be supported on the forward orbit closure of  $a$ . This explains the slight differences in the formulation of the correspondence principle from those used to prove density regularity of finite configurations, such as arithmetic progressions.

Using Theorem 2.3 and Lemma 2.2, we can quickly derive Theorem 1.1 from Theorem 1.2.

*Proof that Theorem 1.2 implies Theorem 1.1.* Suppose  $A \subset \mathbb{N}$  has positive upper Banach density. Invoking Theorem 2.3, we find a system  $(X, \mu, T)$ , a point  $a \in X$ , and an open set  $E \subset X$  with  $\mu(E) > 0$  and such that  $A = \{n \in \mathbb{N} : T^n a \in E\}$  and  $\mu(\overline{\{T^n a : n \in \mathbb{N}\}}) = 1$ . Fix  $k \in \mathbb{N}$ . Theorem 1.2 gives an integer  $t \geq 0$ , points  $x_1, \dots, x_k \in X$  with  $T^t x_i \in E$  for each  $i = 1, \dots, k$ , and a strictly increasing sequence  $c: \mathbb{N} \rightarrow \mathbb{N}$  such that (1.3) holds, which is equivalent to the assertion that  $(a, x_1, \dots, x_k)$  forms an Erdős progression. Setting  $U_j = T^{-t}E$  for all  $1 \leq j \leq k$ , the desired conclusion then follows from Lemma 2.2.  $\square$

We are left with establishing Theorem 1.2. We end this section with a standard argument reducing Theorem 1.2 to the statement in Theorem 2.4. We start by recalling some definitions. A *Følner sequence* in  $\mathbb{N}$  is a sequence  $\Phi = (\Phi_N)_{N \in \mathbb{N}}$  of finite nonempty subsets of  $\mathbb{N}$  satisfying

$$\lim_{N \rightarrow \infty} \frac{|(\Phi_N + 1) \cap \Phi_N|}{|\Phi_N|} = 1.$$

Note that, throughout, although we work with homeomorphisms we only use Følner sequences in  $\mathbb{N}$ . We restrict to such Følner sequences because we must be able to conclude that the sequence  $(b(n))_{n \in \mathbb{N}}$  that we construct is a sequence in  $\mathbb{N}$  and not just in  $\mathbb{Z}$ . This necessitates our assumption in Theorems 1.2 and 2.3 that the forward orbit closure of the point  $a$  have full measure. Recall that whenever  $A \subset \mathbb{N}$  has positive upper Banach density, there is a Følner sequence  $\Phi = (\Phi_N)_{N \in \mathbb{N}}$  such that the limit

$$\lim_{N \rightarrow \infty} \frac{|A \cap \Phi_N|}{|\Phi_N|}$$

exists and is positive.

Given a measure preserving system  $(X, \mu, T)$ , a point  $a \in X$  is *generic* for  $\mu$  along a Følner sequence  $\Phi = (\Phi_N)_{N \in \mathbb{N}}$ , written  $a \in \text{gen}(\mu, \Phi)$ , if

$$\mu = \lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \delta_{T^n a},$$

where  $\delta_x$  is the Dirac measure at the point  $x \in X$  and the limit is taken in the weak\* topology.

**Theorem 2.4.** *Let  $(X, \mu, T)$  be an ergodic measure preserving system. For every  $k \in \mathbb{N}$ , every Følner sequence  $\Phi$ , every  $a \in \text{gen}(\mu, \Phi)$ , and every open set  $E \subset X$  with  $\mu(E) > 0$ , there exists an integer  $t \geq 0$  and an Erdős progression of the form  $(a, x_1, \dots, x_k)$  with  $x_i \in T^{-t}E$  for  $i = 1, \dots, k$ .*

*Proof that Theorem 2.4 implies Theorem 1.2.* Suppose  $X, T, a, \mu, E$  are as in Theorem 1.2. Using ergodic decomposition of the measure  $\mu$ , we can find a  $T$ -invariant ergodic measure  $\tilde{\mu}$  such that  $\tilde{\mu}(E) \geq \mu(E)$  and  $\tilde{\mu}(\overline{\{T^n a : n \in \mathbb{N}\}}) = 1$ . It follows from [12, Proposition 3.9] that there exists a Følner sequence  $\Phi$  in  $\mathbb{N}$  such that  $a \in \text{gen}(\tilde{\mu}, \Phi)$ .

By Theorem 2.4, for each  $k \in \mathbb{N}$  there exist an Erdős progression  $(a, x_1, \dots, x_k)$  and an integer  $t \geq 0$  such that  $x_i \in T^{-t}E$  for  $i = 1, \dots, k$ . Therefore, there exists a strictly increasing sequence  $c: \mathbb{N} \rightarrow \mathbb{N}$  such that (1.3) holds.  $\square$

The proof of Theorem 2.4 occupies the rest of the paper.

### 3. A variation on recurrence in $X^{k+1}$

Starting with this section, we depart significantly from earlier approaches [28, 17, 23, 24] to similar problems.

Fix a topological system  $(X, T)$ . If  $(a, x_1, \dots, x_k)$  is an Erdős progression and  $U_j$  is a neighborhood of  $x_j$  for  $1 \leq j \leq k$ , then there is  $n \in \mathbb{N}$  such that  $a \in T^{-n}U_1$  and  $x_j \in U_j \cap T^{-n}U_{j+1}$  for  $1 \leq j \leq k-1$ . In  $X^{k+1}$ , we can write this as

$$(a, x_1, \dots, x_{k-1}, x_k) \in (X \times U_1 \times \dots \times U_{k-1} \times U_k) \cap T_\Delta^{-n}(U_1 \times U_2 \times \dots \times U_k \times X)$$

where

$$T_\Delta = \underbrace{T \times T \times \dots \times T}_{k+1 \text{ times}}$$

denotes the product transformation. This leads us to consider when the intersection

$$(X \times U_1 \times \dots \times U_{k-1} \times U_k) \cap T_\Delta^{-n}(U_1 \times U_2 \times \dots \times U_k \times X)$$

is nonempty, motivating our next definition.

**Definition 3.1.** Let  $(X, T)$  be a topological system and  $k \in \mathbb{N}$ . We say that a probability measure  $\tau \in \mathcal{M}(X^{k+1})$  is *progressive* if for all open sets  $U_1, \dots, U_k \subset X$  with

$$\tau(X \times U_1 \times \dots \times U_k) > 0$$

there exist infinitely many  $n \in \mathbb{N}$  such that

$$\tau((X \times U_1 \times \dots \times U_{k-1} \times U_k) \cap T_\Delta^{-n}(U_1 \times U_2 \times \dots \times U_k \times X)) > 0. \quad (3.1)$$

The notion of a measure being *progressive* depends on the value of  $k$  but we omit this from our terminology.

Though we view this as a form of recurrence, it differs from standard notions in several ways. For example, the measure  $\tau$  is not assumed to be  $T_\Delta$  invariant, and (3.1) is not of the form  $\mu(B \cap S^{-n}B)$ . While usually there is no distinction between requiring a single return

to the set and infinitely many returns, in our setting of distinct sets  $U_i$ , one can construct choices of the sets showing that the two notions are different.

The next proposition shows that a progressive probability measure produces Erdős progressions.

**Proposition 3.2.** *Fix  $k \in \mathbb{N}$  and let  $(X, T)$  be a topological system. Let  $a \in X$ , let  $U_1, \dots, U_k \subset X$  be open sets, and let  $\tau \in \mathcal{M}(X^{k+1})$  be a progressive measure satisfying  $\tau(\{a\} \times X^k) = 1$ . If  $\tau(X \times U_1 \times \dots \times U_k) > 0$ , then there is an Erdős progression  $(a, x_1, \dots, x_k)$  with  $x_j \in U_j$  for all  $1 \leq j \leq k$ .*

*Proof.* Let  $V = U_1 \times \dots \times U_k$ . Using the assumption that  $\tau$  is progressive, we can find  $c(1) \in \mathbb{N}$  such that

$$\tau((X \times V) \cap T_{\Delta}^{-c(1)}(V \times X)) > 0.$$

Since  $V$  is open, it follows that the intersection above contains some point  $(v_0, v_1, \dots, v_k)$  in the support of  $\tau$ . From the assumption that the closed set  $\{a\} \times X^k$  has full measure, it follows that  $v_0 = a$ . Letting  $V_1 \subset X^k$  be an open ball centered at  $(v_1, \dots, v_k)$  with a sufficiently small radius, we have all the following properties.

- $\tau(\{a\} \times V_1) > 0$
- $\text{diam}(V_1) \leq \frac{1}{2} \text{diam}(V)$
- $\{a\} \times \overline{V_1} \subset (\{a\} \times V) \cap T_{\Delta}^{-c(1)}(V \times X)$

We proceed inductively, setting  $V_0 = V$  and constructing, for each  $n \in \mathbb{N}$ , an open set  $V_n \subset X^k$  and a recurrence time  $c(n) \in \mathbb{N}$  satisfying the following properties:

- (i)  $\overline{V_n} \subset V_{n-1}$  and  $\text{diam}(V_n) \leq \frac{1}{2} \text{diam}(V_{n-1})$ ;
- (ii)  $\tau(\{a\} \times V_n) > 0$ ;
- (iii)  $\{a\} \times \overline{V_n} \subset T_{\Delta}^{-c(n)}(V_{n-1} \times X)$ ;
- (iv)  $c(n+1) > c(n)$ .

The case  $n = 1$  is already established. The inductive step is the same: using (ii) and the hypotheses on  $\tau$  we can find  $c(n+1) > c(n)$  for which

$$(\{a\} \times V_n) \cap T_{\Delta}^{-c(n+1)}(V_n \times X)$$

has positive  $\tau$  measure. Since  $V_n$  is open, it follows that there exists a point in this intersection that belongs to the support of  $\tau$ . By intersecting a small ball around that point with  $\{a\} \times X^k$ , we can find an open set  $V_{n+1} \subset X^k$  satisfying properties (i)–(iii).

We now produce the desired Erdős progression. Using property (i), the intersection

$$\bigcap_{n=1}^{\infty} \{a\} \times \overline{V_n}$$

is a singleton. Calling this point  $(a, x_1, \dots, x_k)$ , then property (iii) gives that

$$T_{\Delta}^{c(n)}(a, x_1, \dots, x_{k-1}) \in V_{n-1}$$

for all  $n \in \mathbb{N}$ . It thus follows that

$$\lim_{n \rightarrow \infty} T_{\Delta}^{c(n)}(a, x_1, \dots, x_{k-1}) = (x_1, x_2, \dots, x_k),$$

as desired. □

We remark that the Erdős progression found in the proof of Proposition 3.2 is in the support of  $\tau$ . Therefore it follows that the support of  $\tau$  contains a dense set of Erdős progressions. In our earlier work [24] we constructed a measure which was not only progressive, but had the property that almost every  $x \in X^3$  is an Erdős progression. Due to the reliance on Proposition 3.2 in the current work, we do not know if the measure we produce has this stronger property. It would be interesting to know whether the conclusion of Proposition 3.2 can be strengthened to match earlier work.

Using Proposition 3.2, we can now derive Theorem 2.4 from the following result.

**Theorem 3.3.** *Let  $(X, \mu, T)$  be an ergodic measure preserving system. For every  $k \in \mathbb{N}$ , every Følner sequence  $\Phi$ , every  $a \in \text{gen}(\mu, \Phi)$ , and every open set  $E \subset X$  with  $\mu(E) > 0$ , there exists an integer  $t \geq 0$  and a progressive measure  $\sigma \in \mathcal{M}(X^{k+1})$  such that*

$$\sigma(X \times T^{-t}E \times \cdots \times T^{-t}E) > 0 \tag{3.2}$$

and  $\sigma(\{a\} \times X^k) = 1$ .

*Proof that Theorem 3.3 implies Theorem 2.4.* We apply Proposition 3.2 with  $U_j = T^{-t}E$  for all  $1 \leq j \leq k$  and  $\tau = \sigma$ . The hypotheses of Theorem 3.3 allow us to conclude from Proposition 3.2 that there is an Erdős progression  $(a, x_1, \dots, x_k)$  with  $x_j \in U_j$  for all  $1 \leq j \leq k$ . □

## 4. Constructing a progressive measure

### 4.1. Two motivating examples

To motivate our approach to constructing a progressive measure, we consider two special cases that represent complementary behaviors: irrational rotations and weakly mixing systems.

**Example 4.1.** Let  $X = \mathbb{T} = \mathbb{R}/\mathbb{Z}$  be the unit circle and let  $\alpha \in \mathbb{R}$  be irrational. The map  $T: X \rightarrow X$  given by  $Tx = x + \alpha \pmod{1}$  is ergodic with respect to the Haar measure on  $\mathbb{T}$ . In this case,  $(a, x_1, \dots, x_k)$  is an Erdős progression if and only if it is an arithmetic progression in  $\mathbb{T}$ , meaning that

$$(a, x_1, \dots, x_k) = (a, a + \beta, \dots, a + k\beta)$$

for some  $\beta \in \mathbb{T}$ . In particular, the set of all Erdős progressions starting at  $a$  is a subtorus of  $\mathbb{T}^{k+1}$ . Writing  $\sigma$  for the Haar measure on this subtorus, we can describe  $\sigma$  dynamically via the formula

$$\sigma = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \delta_a \times \delta_{T^n a} \times \cdots \times \delta_{T^{kn} a}$$

with the limit taken in the weak\* topology. Open sets  $U_1, \dots, U_k \subset \mathbb{T}$  satisfy

$$\sigma(\mathbb{T} \times U_1 \times \cdots \times U_k) > 0$$

precisely when there exists  $\beta \in \mathbb{T}$  with  $a + i\beta \in U_i$  for  $i = 1, \dots, k$ . In this case, choosing an integer  $n$  such that  $n\alpha \approx \beta$ , we have that  $a + i\beta \in U_i \cap T^{-n}U_{i+1}$  for all  $i = 1, \dots, k-1$ . This in turn implies that

$$\sigma\left(\left(\mathbb{T} \times U_1 \times \cdots \times U_k\right) \cap T_{\Delta}^{-n}\left(U_1 \times \cdots \times U_k \times \mathbb{T}\right)\right) > 0$$

showing that  $\sigma$  is progressive.

In view of this example, a natural candidate for  $\sigma$  in the general setting is

$$\sigma = \lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \delta_a \times \delta_{T^n a} \times \cdots \times \delta_{T^{kn} a} \quad (4.1)$$

where  $\Phi$  is a Følner sequence along which the limit exists. where one would pass to a subsequence of  $(\Phi_N)_{N \in \mathbb{N}}$  if necessary to guarantee the existence of the limit. Unfortunately, even in the complementary case of weakly mixing systems, we are unable to prove that the measure defined by (4.1) is progressive. It would be interesting to know whether it always is. This makes it necessary for us to take a different approach for weakly mixing systems, illustrated by the following example.

**Example 4.2.** If  $(X, \mu, T)$  is weakly mixing, then we can take  $\sigma$  to be the product measure  $\sigma = \delta_a \times \mu \times \cdots \times \mu \in \mathcal{M}(X^{k+1})$ . To verify that  $\sigma$  is progressive, fix  $U_1, \dots, U_k \subset X$  with  $\mu(U_i) > 0$  for each  $i$ . Weak mixing implies each of the sets  $\{n \in \mathbb{N} : \mu(U_i \cap T^{-n}U_{i+1}) > 0\}$  has full density with respect to every Følner sequence. Setting  $U_0 = U_{k+1} := X$  for convenience, we conclude that

$$\begin{aligned} \{n \in \mathbb{N} : \sigma((U_0 \times \cdots \times U_k) \cap T_{\Delta}^{-n}(U_1 \times \cdots \times U_{k+1})) > 0\} \\ \supset \bigcap_{i=0}^k \{n \in \mathbb{N} : \mu(U_i \cap T^{-n}U_{i+1}) > 0\} \end{aligned}$$

has full density with respect to every Følner sequence. This shows that  $\sigma$  is progressive and hence satisfies the conclusion of Theorem 3.3. In fact, one can prove that for  $\mu \times \cdots \times \mu$ -almost every  $(x_1, \dots, x_k) \in X^k$  the point  $(a, x_1, \dots, x_k)$  is an Erdős progression; when  $\mu$  is weakly mixing with respect to  $T$ , it follows that  $\mu \times \cdots \times \mu$  is ergodic with respect to  $T_{\Delta}$  and hence almost every point is both generic for and in the support of this measure.

To define  $\sigma$  in general, we construct a measure that combines structured behavior, like that of Example 4.1, with mixing behavior, like that of Example 4.2. To make this precise, we use the notion of pronilfactors.

## 4.2. Pronilfactors

Let  $(X, \mu, T)$  and  $(Y, \nu, S)$  be measure preserving systems. We say that  $(Y, \nu, S)$  is a *factor* of  $(X, \mu, T)$  if there exists an almost surely defined and measurable map  $\pi: X \rightarrow Y$  called the *factor map* such that  $\pi \circ T = S \circ \pi$  holds  $\mu$ -almost everywhere and  $\pi\mu = \nu$ . Of particular importance for us are the factors which are nilsystems.

**Definition 4.3** (Nilsystems, pronilsystems). Let  $G$  be an  $s$ -step nilpotent Lie group and let

$\Gamma \subset G$  be a discrete and cocompact subgroup. The compact manifold  $X = G/\Gamma$  is an *s-step nilmanifold*.

The group  $G$  acts on  $X$  by left translation, and there is a unique Borel probability measure  $\mu$  on  $X$  that is invariant under this action, the *Haar measure* on  $X$ . Letting  $T: X \rightarrow X$  denote left translation by a fixed element of  $G$ , the resulting measure preserving system  $(X, \mu, T)$  is called an *s-step nilsystem* (as usual endowed with the Borel  $\sigma$ -algebra).

An inverse limit of *s-step nilsystems* is called an *s-step pronilsystem*.

We note that we can view the inverse limit in the category of measure preserving systems or in the category of topological dynamical systems, as both describe the same system.

A topological system  $(X, T)$  is *uniquely ergodic* if there is a unique  $T$ -invariant measure. A pronilsystem is uniquely ergodic if and only if every point is generic for  $\mu$  along every Følner sequence. It is a classical result that for all *s-step pronilsystems*  $(X, \mu, T)$  the properties (i) minimal; (ii) transitive; (iii) ergodic; (iv) uniquely ergodic; are equivalent. (See [20, Theorem 11, Chapter 11] for a summary and discussion.) Moreover, every orbit closure in a pronilsystem supports a unique invariant measure (see [27, 26]).

For each  $s \in \mathbb{N}$ , every ergodic system  $(X, \mu, T)$  has a maximal factor that is isomorphic to an *s-step pronilsystem* (see [20, Chapter 16]); this system is called the *s-step pronilfactor* of  $(X, \mu, T)$  and is denoted  $(Z_s, m_s, T)$ . By a standard abuse of notation we use  $T$  to denote the transformation not only in the system but also in the pronilfactor.

### 4.3. Constructing the measure $\sigma$

We are now ready to construct the probability measure  $\sigma$  on  $X^{k+1}$  that we later show to be progressive. This is carried out in two steps: first we define a progressive measure on  $Z_{k-1}^{k+1}$  where  $(Z_{k-1}, m_{k-1}, T)$  is the  $(k-1)$ -step pronilfactor of the given system  $(X, \mu, T)$ , and then we lift this to a measure on  $X^{k+1}$ .

For the first step, when  $(X, \mu, T)$  is a pronilsystem, the measure  $\sigma$  described in (4.1) is progressive; this follows from Theorem 4.9 below but a direct proof is more involved than the special case in Example 4.1 of a rotation.

For the second step, we need to overcome the obstacle that the definition of  $\sigma$  in (4.1) depends on the single point  $a \in \text{gen}(\mu, \Phi)$ , but the map  $\pi_s: X \rightarrow Z_s$  to the *s-step pronilfactor* is, a priori, only a measurable map and only defined almost everywhere. To remedy this issue, we restrict attention to systems where the map  $\pi$  is continuous.

**Definition 4.4.** We say that  $(X, \mu, T)$  has *topological pronilfactors* if for every  $s \in \mathbb{N}$ , there is a continuous factor map  $\pi_s: X \rightarrow Z_s$ , where  $(Z_s, m_s, T)$  denotes the *s-step pronilfactor* of  $(X, \mu, T)$ .

Fortunately, it is possible to reduce Theorem 3.3 to the case when  $(X, \mu, T)$  has topological pronilfactors. This maneuver has been used in related situations, such as [19], and was first adapted to this setting by Host [17]. See [23, Section 5] for a more complete discussion. The reduction of Theorem 3.3 to the case of topological pronilfactors is proven in the next section

If  $(Y, \nu, S)$  is a factor of  $(X, \mu, T)$  with factor map  $\pi: X \rightarrow Y$ , then the *conditional expectation* of a function of  $f \in L^2(\mu)$  onto  $Y$ , denoted  $\mathbb{E}(f | Y)$ , is the unique function in

$L^2(\nu)$  such that

$$\int_{\pi^{-1}(B)} f \, d\mu = \int_B \mathbb{E}(f \mid Y) \, d\nu$$

for all Borel measurable sets  $B \subset Y$ .

Using this, we can define the measure  $\sigma$ .

**Definition 4.5.** Let  $(X, \mu, T)$  be an ergodic system with topological pronilfactors and fix  $k \in \mathbb{N}$ . Let  $(Z_{k-1}, m_{k-1}, T)$  denote the  $(k-1)$ -step pronilfactor of  $(X, \mu, T)$  and let  $\pi_{k-1}: X \rightarrow Z_{k-1}$  denote the corresponding continuous factor map. Given a Følner sequence  $(\Phi_N)_{N \in \mathbb{N}}$  and  $a \in \text{gen}(\mu, \Phi)$  define the measure  $\xi \in \mathcal{M}(Z_{k-1}^k)$  by setting

$$\xi = \lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \delta_{T^n(\pi_{k-1}a)} \otimes \delta_{T^{2n}(\pi_{k-1}a)} \otimes \cdots \otimes \delta_{T^{kn}(\pi_{k-1}a)}. \quad (4.2)$$

We then define the measure  $\sigma \in \mathcal{M}(X^{k+1})$  via the formula

$$\int_{X^{k+1}} f_0 \otimes f_1 \otimes \cdots \otimes f_k \, d\sigma = f_0(a) \int_{Z_{k-1}^k} \mathbb{E}(f_1 \mid Z_{k-1}) \otimes \cdots \otimes \mathbb{E}(f_k \mid Z_{k-1}) \, d\xi \quad (4.3)$$

for all continuous functions  $f_0, f_1, \dots, f_k: X \rightarrow \mathbb{C}$ .

The limit in (4.2) exists and is in fact independent of  $\Phi$ , since  $(Z_{k-1}^k, T \times T^2 \times \cdots \times T^k)$  is a topological pronilsystem and hence every point has a uniquely ergodic orbit closure. It follows from this that  $\xi$  is the unique measure on  $\mathcal{M}(Z_{k-1}^k)$  such that  $(\pi_{k-1}(a), \dots, \pi_{k-1}(a))$  is a generic point (along  $\Phi$ ) with respect to the transformation  $T \times T^2 \times \cdots \times T^k$ .

We stress that the measures  $\xi$  and  $\sigma$  depend on the choice of the point  $a$ , as well as the system  $(X, \mu, T)$  and the positive integer  $k$ . For ease of reading, we omit these dependencies from the notation. Note also that while  $\xi$  is defined on a  $k$ -fold product space indexed by  $\{1, \dots, k\}$ , the measure  $\sigma$  is defined on a  $(k+1)$ -fold product space indexed by  $\{0, 1, \dots, k\}$  and distinguishes the zeroth coordinate.

For a measure  $\tau \in \mathcal{M}(X^k)$  and each  $i = 1, \dots, k$ , we let  $\tau_i$  denote the  $i$ -th marginal of  $\tau$ , meaning the projection of  $\tau$  onto the  $i$ -th coordinate. Given two measures  $\tau, \tau'$  defined on the same space, we say that  $\tau \leq \tau'$  if  $\tau(A) \leq \tau'(A)$  for all measurable sets  $A$ .

**Lemma 4.6.** Fix an ergodic system  $(X, \mu, T)$  with topological pronilfactors, an integer  $k \in \mathbb{N}$ , a Følner sequence  $(\Phi_N)_{N \in \mathbb{N}}$ , and a point  $a \in \text{gen}(\mu, \Phi)$ . For every  $1 \leq i \leq k$ , the  $i$ -th marginal  $\sigma_i$  of the measure  $\sigma$  in Definition 4.5 satisfies  $\sigma_i \leq i\mu$ .

*Proof.* Note that  $\xi_i$ , the  $i$ -th marginal of the measure  $\xi$  from (4.2), is  $T^i$ -invariant. Since the system  $(Z_{k-1}, T)$  is uniquely ergodic, we have that  $\xi_i \leq im_{Z_{k-1}}$  and hence

$$\int_{Z_{k-1}} g \, d\xi_i \leq i \int_{Z_{k-1}} g \, dm_{Z_{k-1}} \quad (4.4)$$

whenever  $g: Z_{k-1} \rightarrow [0, 1]$  is Borel measurable. Therefore, for any Borel set  $A \subset X$ , we have

$$\sigma_i(A) = \int_{Z_{k-1}} \mathbb{E}(\mathbf{1}_A \mid Z_{k-1}) \, d\xi_i \leq i \int_{Z_{k-1}} \mathbb{E}(\mathbf{1}_A \mid Z_{k-1}) \, dm_{Z_{k-1}} = i\mu(A). \quad \square$$

We also record the following which follows directly from the definition.

**Lemma 4.7.** *The measure  $\sigma$  defined in Definition 4.5 is invariant under  $\text{Id} \times T \times T^2 \times \cdots \times T^k$ .*

#### 4.4. Outlining the proof of Theorem 3.3

To prove Theorem 3.3, we need to verify that  $\sigma$  has the required properties, which are captured by the following two theorems.

**Theorem 4.8.** *Fix an ergodic system  $(X, \mu, T)$  with topological pronilfactors. Let  $k \in \mathbb{N}$ , let  $(\Phi_N)_{N \in \mathbb{N}}$  be a Følner sequence, let  $a \in \text{gen}(\mu, \Phi)$ , and let  $\sigma$  be the measure described in Definition 4.5. Then for every open set  $E \subset X$  with  $\mu(E) > 0$ , there is an integer  $t \geq 0$  such that  $\sigma(X \times T^{-t}E \times \cdots \times T^{-t}E) > 0$ .*

**Theorem 4.9.** *Fix an ergodic system  $(X, \mu, T)$  with topological pronilfactors. Let  $k \in \mathbb{N}$ , let  $(\Phi_N)_{N \in \mathbb{N}}$  be a Følner sequence, and let  $a \in \text{gen}(\mu, \Phi)$ . Then the measure  $\sigma$  described in Definition 4.5 is progressive.*

These ingredients, together with the ability to pass to an extension with topological pronilfactors, combine to give a quick proof of Theorem 3.3.

*Proof that Theorems 4.8 and 4.9 together imply Theorem 3.3.* Fix an ergodic system  $(X, \mu, T)$ , a Følner sequence  $(\Phi_N)_{N \in \mathbb{N}}$ , a point  $a \in \text{gen}(\mu, \Phi)$ , and an open set  $E \subset X$  with  $\mu(E) > 0$ . By [23, Proposition 5.7 and Lemma 5.8], there exist an ergodic system  $(\tilde{X}, \tilde{\mu}, \tilde{T})$  with topological pronilfactors and a continuous factor map  $\pi: \tilde{X} \rightarrow X$  such that  $(X, \mu, T)$  is a factor of  $(\tilde{X}, \tilde{\mu}, \tilde{T})$ . Moreover, there is a Følner sequence  $(\tilde{\Phi}_N)_{N \in \mathbb{N}}$  and  $\tilde{a} \in \text{gen}(\tilde{\mu}, \tilde{\Phi})$  such that  $\pi(\tilde{a}) = a$ . Let  $\tilde{\sigma}$  denote the measure on  $(\tilde{X})^{k+1}$  defined by (4.3) and let  $\tilde{E} := \pi^{-1}(E)$ . Applying Theorems 4.8 and 4.9 to  $\tilde{\sigma}$ , it follows that  $\tilde{\sigma}$  is a progressive measure, that  $\tilde{\sigma}(\{\tilde{a}\} \times X^k) = 1$ , and that there exists an integer  $t \geq 0$  satisfying  $\tilde{\sigma}(\tilde{X} \times \tilde{T}^{-t}\tilde{E} \times \cdots \times \tilde{T}^{-t}\tilde{E}) > 0$ .

Letting  $\sigma$  to be the push-forward of  $\tilde{\sigma}$  under the map  $\pi \times \pi \times \cdots \times \pi$ , it follows that  $\sigma$  satisfies the conclusion of Theorem 3.3.  $\square$

## 5. Proof of Theorem 4.8

We are left with proving Theorems 4.8 and 4.9. The proof of the former is given in this section, whereas the proof of the latter is in Section 6. In both proofs we relate certain dynamical averages of  $\sigma$  to the Furstenberg joining of  $(X, \mu, T)$ .

**Definition 5.1** (The Furstenberg joining). Given a measure preserving system  $(X, \mu, T)$  and  $k \in \mathbb{N}$ , let  $\lambda$  be the Borel probability measure on  $X^{k+1}$  uniquely determined by

$$\int_{X^{k+1}} f_0 \otimes f_1 \otimes \cdots \otimes f_k \, d\lambda = \lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \int_X f_0 \cdot T^n f_1 \cdots T^{kn} f_k \, d\mu \quad (5.1)$$

for all  $f_0, f_1, \dots, f_k \in L^\infty(\mu)$ . The measure  $\lambda$  determined this way is called the *Furstenberg joining* of  $(X, \mu, T)$ .

This joining is well defined, as it is shown in [18] that the limit exists and is independent of the choice of Følner sequence  $\Phi = (\Phi_N)_{N \in \mathbb{N}}$ .

As we show, Theorem 4.8 follows quite quickly from the following theorem and Furstenberg's multiple recurrence theorem.

**Theorem 5.2.** *Fix an ergodic measure preserving system  $(X, \mu, T)$  with topological pronilfactors,  $k \in \mathbb{N}$ , a Følner sequence  $(\Phi_N)_{N \in \mathbb{N}}$  and  $a \in \text{gen}(\mu, \Phi)$ . Let  $\sigma$  be the measure described in Definition 4.5. For any  $g_1, \dots, g_k \in C(X)$  we have*

$$\lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \int_{X^{k+1}} T_\Delta^n(\mathbf{1} \otimes g_1 \otimes \dots \otimes g_k) d\sigma = \int_{X^{k+1}} \mathbf{1} \otimes g_1 \otimes \dots \otimes g_k d\lambda \quad (5.2)$$

where  $\lambda$  is the Furstenberg joining.

*Proof that Theorem 5.2 implies Theorem 4.8.* Since  $\mu$  is a Borel measure on the compact metric space  $X$ , it is regular. Therefore, given an open set  $E \subset X$  with  $\mu(E) > 0$ , one can use Urysohn's lemma to find  $g \in C(X)$  with  $\int g d\mu > 0$  and  $0 \leq g \leq \mathbf{1}_E$ . Thus

$$\int_{X^{k+1}} T_\Delta^n(\mathbf{1} \otimes \mathbf{1}_E \otimes \dots \otimes \mathbf{1}_E) d\sigma \geq \int_{X^{k+1}} T_\Delta^n(\mathbf{1} \otimes g \otimes \dots \otimes g) d\sigma$$

for all  $n \in \mathbb{N}$ . Combining this observation with Theorem 5.2 gives that

$$\lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \int_{X^{k+1}} T_\Delta^n(\mathbf{1} \otimes \mathbf{1}_E \otimes \dots \otimes \mathbf{1}_E) d\sigma \geq \int \mathbf{1} \otimes g \otimes \dots \otimes g d\lambda.$$

By the definition of  $\lambda$ , this last expression is equal to

$$\lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \int_X T^n g \dots T^{kn} g d\mu$$

and this is positive by Furstenberg's multiple recurrence theorem [11, Theorem 11.13]. We have shown that

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \sigma(X \times T^{-t}E \times \dots \times T^{-t}E) > 0,$$

which is a stronger statement than what is required in the conclusion of Theorem 4.8.  $\square$

It remains to prove Theorem 5.2.

*Proof of Theorem 5.2.* Fix  $g_1, \dots, g_k \in C(X)$ . Our goal is to establish (5.2). It follows from [18, Theorem 12.1] that

$$\int g_1 \otimes \dots \otimes g_k d\lambda = \lim_{L \rightarrow \infty} \frac{1}{|\Phi_L|} \sum_{\ell \in \Phi_L} \int_{Z_{k-1}} T^\ell \mathbb{E}(g_1 | Z_{k-1}) \dots T^{k\ell} \mathbb{E}(g_k | Z_{k-1}) dm_{k-1}$$

where  $(Z_{k-1}, m_{k-1}, T)$  is the  $(k-1)$ -step pronilfactor of  $(X, \mu, T)$ . Combined with (4.3) it follows that, in order to establish (5.2), it suffices to prove

$$\lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \int_{Z_{k-1}^k} T^n h_1 \otimes \dots \otimes T^n h_k d\xi = \lim_{L \rightarrow \infty} \frac{1}{|\Phi_L|} \sum_{\ell \in \Phi_L} \int_{Z_{k-1}} T^\ell h_1 \dots T^{k\ell} h_k dm_{k-1}$$

whenever  $h_1, \dots, h_k : Z_{k-1} \rightarrow \mathbb{C}$  are bounded and measurable. By a standard approximation argument and Lemma 4.6 it suffices to establish this when the functions  $h_1, \dots, h_k$  are continuous, which we henceforth assume. Writing  $\tilde{a} = \pi_{k-1}(a)$  we have the following from the definition of  $\xi$ :

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \int_{Z_{k-1}^k} T^n h_1 \otimes \dots \otimes T^n h_k \, d\xi \\ &= \lim_{N \rightarrow \infty} \lim_{L \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \frac{1}{|\Phi_L|} \sum_{\ell \in \Phi_L} h_1(T^{m+\ell} \tilde{a}) \dots h_k(T^{m+k\ell} \tilde{a}). \end{aligned} \quad (5.3)$$

On the other hand, ergodicity of  $(X, \mu, T)$  ensures ergodicity, and hence unique ergodicity, of the pronilfactor  $(Z_{k-1}, m_{k-1}, T)$ , and so it follows that

$$\begin{aligned} & \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \int_{Z_{k-1}} T^\ell h_1 \dots T^{k\ell} h_k \, dm_{k-1} \\ &= \lim_{L \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{1}{|\Phi_L|} \sum_{\ell \in \Phi_L} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} h_1(T^{\ell+n} \tilde{a}) \dots h_k(T^{k\ell+n} \tilde{a}). \end{aligned} \quad (5.4)$$

Consider now the commuting nilrotations

$$\begin{aligned} S &= T \times T \times \dots \times T \\ R &= T \times T^2 \times \dots \times T^k \end{aligned}$$

on  $(Z_{k-1})^k$ . Together they induce an action of  $\mathbb{Z}^2$  on  $Z_{k-1}^k$ . Set

$$Y := \overline{\{S^n R^\ell(\tilde{a}, \dots, \tilde{a}) : n, \ell \in \mathbb{Z}\}} \subset Z_{k-1}^k$$

to be the orbit closure of  $(\tilde{a}, \dots, \tilde{a}) \in Z_{k-1}^k$  under  $S$  and  $R$ . Since  $(Z_{k-1}, m_{k-1}, T)$  is a pronilsystem, the  $\mathbb{Z}^2$ -system  $(Y, S, R)$  is also a pronilsystem and hence is uniquely ergodic [20, Theorem 17, Chapter 11]. Therefore, there is a unique invariant mean for these averages, which implies (see for example [2, Lemma 1.1]) that the expressions in (5.3) and (5.4) coincide, finishing the proof.  $\square$

## 6. Proof of Theorem 4.9

### 6.1. Ingredients in the proof of Theorem 4.9

We are left with proving Theorem 4.9. The main ingredients are the following two theorems. The first – a consequence of Furstenberg’s multiple recurrence theorem – is proved in Section 6.2. The second is proved in Section 6.4 and uses material about uniformity norms covered in Section 6.3.

**Theorem 6.1.** *Let  $(X, T)$  be a topological system, let  $\Phi = (\Phi_N)_{N \in \mathbb{N}}$  be a Følner sequence in  $\mathbb{N}$ , and let  $\nu \in \mathcal{M}(X^k)$  be  $T \times T^2 \times \dots \times T^k$  invariant. If  $G : X^k \rightarrow [0, 1]$  is continuous*

and satisfies  $\int_{X^k} G \, d\nu > 0$ , then

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \int_{X^k} G \cdot T_\Delta^n G \, d\nu > 0.$$

**Theorem 6.2.** Fix an ergodic measure preserving system  $(X, \mu, T)$  with topological pronilfactors,  $k \in \mathbb{N}$ , a Følner sequence  $(\Phi_N)_{N \in \mathbb{N}}$ , and a point  $a \in \text{gen}(\mu, \Phi)$ . Let  $\sigma$  be the measure given in Definition 4.5. For any continuous function  $G: X^k \rightarrow \mathbb{C}$  we have

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} T_\Delta^n(G \otimes \mathbf{1}) - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} T_\Delta^n(\mathbf{1} \otimes G) \right\|_{L^2(\sigma)} = 0.$$

Using these two theorems we can quickly prove Theorem 4.9.

*Proof that Theorems 6.1 and 6.2 imply Theorem 4.9.* Our goal is to show that the measure  $\sigma$  defined in Definition 4.5 is progressive. Fix open sets  $U_1, \dots, U_k$  with

$$\sigma(X \times U_1 \times \dots \times U_k) > 0$$

and set  $V = U_1 \times \dots \times U_k$ . Since  $V$  is open and  $\sigma(X \times V) > 0$ , there exists a continuous function  $G \in C(X^k)$  such that  $0 \leq G \leq \mathbf{1}_V$  and

$$\int_{X^{k+1}} \mathbf{1} \otimes G \, d\sigma > 0.$$

Applying Theorem 6.2, we obtain

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \sigma((X \times V) \cap T_\Delta^{-n}(V \times X)) \\ & \geq \liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \int_{X^{k+1}} (\mathbf{1} \otimes G) \cdot T_\Delta^n(G \otimes \mathbf{1}) \, d\sigma \\ & = \liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \int_{X^{k+1}} (\mathbf{1} \otimes G) \cdot T_\Delta^n(\mathbf{1} \otimes G) \, d\sigma. \end{aligned}$$

Let  $\nu \in \mathcal{M}(X^k)$  denote the projection of  $\sigma$  to the last  $k$  coordinates. In view of Lemma 4.7,  $\nu$  is invariant under  $T \times T^2 \times \dots \times T^k$  and hence we can use Theorem 6.1 to conclude that the last expression is positive. This shows that  $\sigma$  is progressive and completes the proof.  $\square$

## 6.2. An extension of Furstenberg's multiple recurrence theorem

We derive Theorem 6.1 from the following more general result, which is an extension of Furstenberg's multiple recurrence theorem that may be of independent interest.

Given a vector  $v = (v_1, \dots, v_k) \in \mathbb{N}^k$ , we write  $T_v = T^{v_1} \times \dots \times T^{v_k}$ .

**Theorem 6.3.** Let  $(X, T)$  be a topological system, let  $u, v \in \mathbb{N}^k$ , and let  $\Phi = (\Phi_N)_{N \in \mathbb{N}}$  be a Følner-sequence in  $\mathbb{N}$ . If  $\nu \in \mathcal{M}(X^k)$  is  $T_v$ -invariant and  $A_1, \dots, A_k \subset X$  are such that

$A = A_1 \times \cdots \times A_k \subset X^k$  satisfies  $\nu(A) > 0$ , then

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \nu(A \cap T_u^{-n} A) > 0. \quad (6.1)$$

When  $u = v$ , the conclusion of Theorem 6.3 follows quickly from the mean ergodic theorem.

The next example shows that in general one cannot drop the assumption that  $A$  is a product set in Theorem 6.3.

**Example 6.4.** Let  $T: \mathbb{T} \rightarrow \mathbb{T}$  be an irrational rotation  $T: x \mapsto x + \alpha$  and let  $\nu$  denote the Haar measure on the subtorus  $H := \{(x, 2x) : x \in \mathbb{T}\} \subset \mathbb{T}^2$ . With  $v = (1, 2)$  and  $u = (2, 1)$  we have that  $\nu$  is  $T_v$ -invariant, but  $T_u^{-n}H = H + (0, 3n\alpha)$  is disjoint from  $H$  for every  $n \in \mathbb{N}$ . In particular (6.1) does not hold with  $A = H$ .

Consider the case that  $(X, \mu, T)$  is a measure preserving system,  $u = (1, \dots, k)$ ,  $v = (1, \dots, 1)$ , and  $\nu$  is the diagonal measure on  $X^k$ , meaning that  $\nu$  is the push-forward of  $\mu$  under the diagonal embedding  $x \mapsto (x, x, \dots, x)$  of  $X$  into  $X^k$ . Taking  $A = B \times \cdots \times B$  in the theorem specializes (6.1) to

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \mu(B \cap T^{-n}B \cap \cdots \cap T^{-kn}B) > 0$$

which is Furstenberg's multiple recurrence theorem. We do not, however, give a new proof of Furstenberg's theorem, as it is an ingredient in our proof of Theorem 6.3.

When applying Theorem 6.3 to prove Theorem 6.1, we take  $u = (1, \dots, 1)$ ,  $v = (1, \dots, k)$ ; this interchanges the roles of  $u$  and  $v$  in Furstenberg's multiple recurrence.

*Proof of Theorem 6.3.* Define  $c = \text{LCM}(v_1, \dots, v_k)$  to be the least common multiple of  $v_1, \dots, v_k$  and let  $w_i = \frac{c u_i}{v_i}$ , so that

$$T_v^{-w_i}(X^{i-1} \times A_i \times X^{k-i}) = T_u^{-c}(X^{i-1} \times A_i \times X^{k-i}).$$

It follows that

$$\begin{aligned} T_u^{-cn} A &= \bigcap_{i=1}^k T_u^{-cn}(X^{i-1} \times A_i \times X^{k-i}) = \bigcap_{i=1}^k T_v^{-w_i n}(X^{i-1} \times A_i \times X^{k-i}) \\ &\supset \bigcap_{i=1}^k T_v^{-w_i n} A. \end{aligned}$$

It is clear that

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \nu(A \cap T_u^{-n} A) \geq \liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{\substack{n \in \Phi_N \\ c|n}} \nu(A \cap T_u^{-n} A)$$

and so taking  $\Psi_N = \Phi_N/c = \{n \in \mathbb{N} : cn \in \Phi_N\}$ , we have

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} \nu(A \cap T_u^{-n} A) \geq \liminf_{N \rightarrow \infty} \frac{1}{c|\Psi_N|} \sum_{n \in \Psi_N} \nu(A \cap T_u^{-cn} A)$$

$$\geq \liminf_{N \rightarrow \infty} \frac{1}{c|\Psi_N|} \sum_{n \in \Psi_N} \nu \left( A \cap \bigcap_{i=1}^k T_v^{-w_i n} A \right)$$

as in (6.10). Since  $\nu$  is  $T_v$ -invariant, the last expression is positive by applying Furstenberg's multiple recurrence theorem to the system  $(X^k, \nu, T_v)$ .  $\square$

*Proof of Theorem 6.1.* Since  $G: X^k \rightarrow [0, 1]$  is continuous and has positive integral, there exist open sets  $U_1, \dots, U_k$  and  $c > 0$  such that  $c \cdot \mathbf{1}_{U_1} \otimes \dots \otimes \mathbf{1}_{U_k} \leq G$  and  $\nu(U_1 \times \dots \times U_k) > 0$ . Applying Theorem 6.3 with  $u = (1, \dots, 1)$ ,  $v = (1, \dots, k)$ , the conclusion follows.  $\square$

### 6.3. Uniformity norms

For the proof of Theorem 6.2 we make use of the structure theory of measure preserving transformations via uniformity seminorms [18, 19]. Roughly speaking, these seminorms capture the idea that functions orthogonal to the pronilfactor do not contribute to multiple ergodic averages. We use this in the proof of Theorem 6.2 to, roughly speaking, replace the function  $G$  by a function that is measurable with respect to  $Z_{k-1}^k$  without changing the  $L^2(\sigma)$  norm. This reduces the proof of Theorem 6.2 to analyzing the behavior in pronilsystems, which have additional structure.

**Definition 6.5** (Uniformity Norms). Given a measure preserving system  $(X, \mu, T)$  and  $s \geq 0$ , the  $s$ -step uniformity seminorm  $\|f\|_{U^s(X, \mu, T)}$  of a function  $f \in L^\infty(\mu)$  is defined inductively as

$$\|f\|_{U^0(X, \mu, T)} = \int_X f \, d\mu \tag{6.2}$$

$$\|f\|_{U^{s+1}(X, \mu, T)}^{2^{s+1}} = \lim_{H \rightarrow \infty} \frac{1}{H} \sum_{h=1}^H \|T^h f \cdot \bar{f}\|_{U^s(X, \mu, T)}^{2^s}. \tag{6.3}$$

The fact that the limit always exists and that  $\|\cdot\|_{U^s(X, \mu, T)}$  defines a seminorm on  $L^\infty(\mu)$  for  $s \geq 1$  is proven in [18].

The mean ergodic theorem gives

$$\|f\|_{U^1(X, \mu, T)}^2 = \|\mathbb{E}(f \mid \mathcal{I})\|_2^2$$

where  $\mathcal{I}$  denotes the  $\sigma$ -algebra of invariant sets and so

$$\|f\|_{U^1(X, \mu, T)} = \left| \int_X f \, d\mu \right|, \tag{6.4}$$

when the system  $(X, \mu, T)$  is ergodic, in agreement with the standard definition. The main result in [18] states that given an ergodic system  $(X, \mu, T)$ , for each  $s \geq 1$ , all functions  $f \in L^\infty(\mu)$  satisfy

$$\|f\|_{U^{s+1}(X, \mu, T)} = 0 \iff \mathbb{E}(f \mid Z_s) = 0, \tag{6.5}$$

where  $(Z_s, m_s, T)$  denotes the  $s$ -step pronilfactor of  $(X, \mu, T)$ .

**Theorem 6.6.** *Let  $(X, \mu, T)$  be ergodic, let  $k \geq 2$ , let  $(\Phi_N)_{N \in \mathbb{N}}$  be a Følner sequence, and let  $\tau \in \mathcal{M}(X^{k+1})$  be invariant with respect to the transformation  $\text{Id} \times T \times T^2 \times \dots \times T^k$ . Assume that the marginals  $\tau_0, \tau_1, \dots, \tau_k$  of  $\tau$  satisfy  $\tau_i \leq i\mu$  for all  $1 \leq i \leq k-1$ . Then for any  $f_1, \dots, f_{k-1} \in L^\infty(\mu)$  and any bounded sequence  $b: \mathbb{N} \rightarrow \mathbb{C}$ , we have*

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} b(n) \cdot (\mathbf{1} \otimes T^n f_1 \otimes \dots \otimes T^n f_{k-1} \otimes \mathbf{1}) \right\|_{L^2(\tau)} \quad (6.6)$$

$$\leq C_k \cdot \|b\|_\infty \cdot \min\{\|f_i\|_{U^k(X, \mu, T)} : 1 \leq i \leq k-1\}$$

where  $C_k$  is a constant depending only on  $k$ .

For the proof of Theorem 6.6, we make use of various properties of uniformity seminorms, which we collect in the following lemma.

**Lemma 6.7.** *Let  $(X, \mu, T)$  be a measure preserving system.*

(i) *For all  $k \geq 1$  and  $f \in L^\infty(\mu)$ , we have*

$$\|f \otimes \bar{f}\|_{U^k(X \times X, \mu \times \mu, T \times T)} \leq \|f\|_{U^{k+1}(X, \mu, T)}^2.$$

(ii) *Let  $c \geq 1$  be an integer. For all  $f \in L^\infty(\mu)$  and all  $k \geq 1$  we have that*

$$\|f\|_{U^k(X, \mu, T)} \leq \|f\|_{U^k(X, \mu, T^c)},$$

and for all  $k \geq 2$  we have that

$$\|f\|_{U^k(X, \mu, T^c)} \leq c^{\frac{k}{2}} \|f\|_{U^k(X, \mu, T)}. \quad (6.7)$$

(iii) *For any  $k \geq 1$  and integers  $c_1, \dots, c_k \geq 1$ , there is a constant  $C$ , independent of the system, with the following property: for all  $f_1, \dots, f_k \in L^\infty(\mu)$  with  $\|f_i\|_{L^\infty(\mu)} \leq 1$  and all Følner sequences  $(\Phi_N)_{N \in \mathbb{N}}$ , we have*

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} T^{c_1 n} f_1 \dots T^{c_k n} f_k \right\|_{L^2(\mu)} \leq C \min\{\|f_i\|_{U^k(X, \mu, T)} : 1 \leq i \leq k\}.$$

(iv) *Let  $k \geq 2$ ,  $c \in \mathbb{N}$ , and let  $\rho \in \mathcal{M}(X)$  be invariant with respect to  $T^c$  and satisfy*

$$\rho \leq C\mu \quad (6.8)$$

for some constant  $C > 0$ . Then for any  $f \in L^\infty(\mu)$ , we have

$$\|f\|_{U^k(X, \rho, T^c)} \leq C^{1/2^k} \cdot c^{\frac{k}{2}} \|f\|_{U^k(X, \mu, T)}. \quad (6.9)$$

*Proof.* Part (i) is proved in the appendix of [10, Equation (A.5)]. Part (ii) is a special case of [9, Lemma 3.1]. Part (iii) follows by combining [16, Proposition 1] and [9, Lemma 3.1]. Finally, for part (iv), using (6.8) and that  $\rho$  is  $T^c$ -invariant, it follows quickly that

$$\|f\|_{U^k(X, \rho, T^c)}^{2^k} \leq C \|f\|_{U^k(X, \mu, T)}^{2^k}$$

using (6.4) for the case  $k = 1$  and (6.3) to prove the cases  $k > 1$  by induction. Thus (6.9) follows from (6.7).  $\square$

We include a short lemma to clarify how to decompose averages over Følner sequences

along residue classes.

**Lemma 6.8.** *Let  $a(n)$  be a bounded sequence taking values in a Banach space, let  $\Phi = (\Phi_N)_{N \in \mathbb{N}}$  be a Følner sequence in  $\mathbb{N}$ , let  $c \in \mathbb{N}$ , and let  $\Psi = (\Psi_N)_{N \in \mathbb{N}}$  be the Følner sequence defined by  $\Psi_N := \Phi_N/c$ . Then*

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} a(n) \right\| \leq \frac{1}{c} \sum_{j=0}^{c-1} \limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Psi_N|} \sum_{n \in \Psi_N} a(cn + j) \right\|.$$

*Proof.* Since  $\Phi$  is a Følner sequence, for every  $j \in \mathbb{N}$  we have

$$\frac{1}{|\Phi_N|} |(\Phi_N - j) \cap c\mathbb{N}| \rightarrow \frac{1}{c}$$

as  $N \rightarrow \infty$ . Hence we have that

$$\frac{1}{|\Phi_N|} \left| \Phi_N \cap \left( \bigcup_{j=0}^{c-1} c\Psi_N + j \right) \right| \rightarrow 1 \quad \text{and} \quad \frac{|\Psi_N|}{|\Phi_N|} \rightarrow \frac{1}{c} \quad (6.10)$$

as  $N \rightarrow \infty$ . It follows that as  $N \rightarrow \infty$ , we have that

$$\frac{1}{|\Phi_N|} \left\| \sum_{n \in \Phi_N} a(n) \right\| - \frac{1}{|\Phi_N|} \left\| \sum_{j=0}^{c-1} \sum_{n \in \Psi_N} a(cn + j) \right\| \rightarrow 0.$$

The conclusion is obtained by using the triangle inequality.  $\square$

On the way to proving Theorem 6.6, we first establish the following related result, corresponding roughly to the case  $b(n) = 1$  in Theorem 6.6.

**Theorem 6.9.** *Let  $(X, \mu, T)$  be a measure preserving system,  $k \in \mathbb{N}$ , and let  $\nu \in \mathcal{M}(X^k)$  be invariant under the transformation  $(T \times T^2 \times \dots \times T^k)$ . Assume that the marginals  $\nu_i$  of  $\nu$  satisfy*

$$\nu_i \leq i^2 \mu$$

for all  $1 \leq i \leq k$ . Then for all  $f_1, \dots, f_k \in L^\infty(\mu)$  with  $\|f_i\|_{L^\infty(\mu)} \leq 1$ , we have

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (T^n f_1 \otimes \dots \otimes T^n f_k) \right\|_{L^2(\nu)} \leq D_k \min\{\|f_i\|_{U^k(X, \mu, T)} : 1 \leq i \leq k\} \quad (6.11)$$

where  $D_k$  is a constant depending only on  $k$ .

*Proof.* Let  $c = \text{LCM}(2, \dots, k)$  be the least common multiple of the first  $k$  positive integers and set  $c_i = c/i$  for all  $1 \leq i \leq k$ . Let  $D_k$  be the constant in part (iii) of Lemma 6.7. Set

$$\Psi_N = \Phi_N/c = \{n \in \mathbb{N} : cn \in \Phi_N\}$$

which defines a Følner sequence. Then, in view of Lemma 6.8,

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} T^n f_1 \otimes \dots \otimes T^n f_k \right\|_{L^2(\nu)}$$

$$\leq \frac{1}{c} \sum_{j=0}^{c-1} \left( \limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Psi_N|} \sum_{n \in \Psi_N} T^{cn+j} f_1 \otimes \dots \otimes T^{cn+j} f_k \right\|_{L^2(\nu)} \right)$$

and so it suffices to show

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Psi_N|} \sum_{n \in \Psi_N} T^{cn+j} f_1 \otimes \dots \otimes T^{cn+j} f_k \right\|_{L^2(\nu)} \leq D_k \min\{\|f_i\|_{U^k(X, \mu, T)} : 1 \leq i \leq k\} \quad (6.12)$$

for each fixed integer  $j \leq c$ . For each  $1 \leq i \leq k$ , define  $g_i: X^k \rightarrow \mathbb{C}$  by

$$g_i(x_1, \dots, x_k) = (T^j f_i)(x_i).$$

Set  $S = T \times T^2 \times \dots \times T^k$  and note that

$$S^{cn} g_i = \mathbf{1} \otimes \dots \otimes \mathbf{1} \otimes T^{cn+j} f_i \otimes \mathbf{1} \otimes \dots \otimes \mathbf{1}$$

for all  $n \in \mathbb{N}$ . Thus we can rewrite (6.12) as

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Psi_N|} \sum_{n \in \Psi_N} S^{c_1 n} g_1 \dots S^{c_k n} g_k \right\|_{L^2(\nu)} \leq D_k \min\{\|f_i\|_{U^k(X, \mu, T)} : 1 \leq i \leq k\}.$$

By part (iii) of Lemma 6.7 applied to the system  $(X^k, \nu, S)$  and our choice of  $D_k$ , we have that

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Psi_N|} \sum_{n \in \Psi_N} S^{c_1 n} g_1 \dots S^{c_k n} g_k \right\|_{L^2(\nu)} \leq D_k \min\{\|g_i\|_{U^k(X^k, \nu, S)} : 1 \leq i \leq k\}.$$

For  $i = 1, \dots, k$ , we have

$$\|g_i\|_{U^k(X^k, \nu, S)} = \|T^j f_i\|_{U^k(X, \nu_i, T^i)}$$

and so the conclusion follows by applying part (iv) of Lemma 6.7.  $\square$

**Lemma 6.10.** *Let  $(Y, \tau)$  be a probability space,  $\Phi \subset \mathbb{N}$  be a finite set,  $b: \Phi \rightarrow \mathbb{C}$ , and  $(v_n)_{n \in \Phi}$  be a collection of vectors in  $L^2(\tau)$ . Then*

$$\left\| \frac{1}{|\Phi|} \sum_{n \in \Phi} b(n) v_n \right\|_{L^2(\tau)}^2 \leq \|b\|_\infty^2 \cdot \left\| \frac{1}{|\Phi|} \sum_{n \in \Phi} v_n \otimes \bar{v}_n \right\|_{L^2(\tau \times \tau)}.$$

*Proof.* We expand the square and obtain

$$\left\| \frac{1}{|\Phi|} \sum_{n \in \Phi} b(n) v_n \right\|_{L^2(\tau)}^2 = \frac{1}{|\Phi|^2} \sum_{n, m \in \Phi} b(n) \bar{b}(m) \int_Y v_n \bar{v}_m \, d\tau \leq \|b\|_\infty^2 \cdot \frac{1}{|\Phi|^2} \sum_{n, m \in \Phi} \left| \int_Y v_n \bar{v}_m \, d\tau \right|. \quad (6.13)$$

Using the Cauchy-Schwarz inequality, we estimate

$$\begin{aligned} \left( \frac{1}{|\Phi|^2} \sum_{n, m \in \Phi} \left| \int_Y v_n \bar{v}_m \, d\tau \right| \right)^2 &\leq \frac{1}{|\Phi|^2} \sum_{n, m \in \Phi} \left| \int_Y v_n \bar{v}_m \, d\tau \right|^2 \\ &= \frac{1}{|\Phi|^2} \sum_{n, m \in \Phi} \int_{Y \times Y} (v_n \otimes \bar{v}_n) \cdot (\bar{v}_m \otimes v_m) \, d(\tau \times \tau) \end{aligned}$$

$$= \int_{Y \times Y} \left| \frac{1}{|\Phi|} \sum_{n \in \Phi} v_n \otimes \bar{v}_n \right|^2 d(\tau \times \tau) = \left\| \frac{1}{|\Phi|} \sum_{n \in \Phi} v_n \otimes \bar{v}_n \right\|_{L^2(\tau \times \tau)}^2.$$

Combining this with (6.13), we obtain the conclusion.  $\square$

*Proof of Theorem 6.6.* Let  $\nu \in \mathcal{M}((X \times X)^{k-1})$  be the pushforward of  $\tau \times \tau$  under the map

$$\varphi: (x_0, \dots, x_k, y_0, \dots, y_k) \mapsto (x_1, y_1, x_2, y_2, \dots, x_{k-1}, y_{k-1}).$$

Note that  $\nu$  is invariant under  $(T \times T) \times (T^2 \times T^2) \times \dots \times (T^{k-1} \times T^{k-1})$  and that each of the  $k-1$  marginals  $\nu_i \in \mathcal{M}(X \times X)$  of  $\nu$  are  $\nu_i = \tau_i \times \tau_i$  and hence satisfy  $\nu_i \leq i^2(\mu \times \mu)$  for  $i = 1, \dots, k-1$ . Applying Theorem 6.9 to the system  $(X \times X, \mu \times \mu, T \times T)$  with  $k-1$  instead of  $k$  and  $f_i \otimes \bar{f}_i$  instead of  $f_i$ , it follows that

$$\begin{aligned} \limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (T \times T)^n (f_1 \otimes \bar{f}_1) \otimes \dots \otimes (T \times T)^n (f_{k-1} \otimes \bar{f}_{k-1}) \right\|_{L^2(\nu)} \\ \leq D_{k-1} \cdot \min \left\{ \|f_i \otimes \bar{f}_i\|_{U^{k-1}(X \times X, \mu \times \mu, T \times T)} : 1 \leq i \leq k-1 \right\}. \end{aligned}$$

Part (i) of Lemma 6.7 implies that

$$\|f_i \otimes \bar{f}_i\|_{U^{k-1}(X \times X, \mu \times \mu, T \times T)} \leq \|f_i\|_{U^k(X, \mu T)}^2$$

and so we deduce that

$$\begin{aligned} \limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (T \times T)^n (f_1 \otimes \bar{f}_1) \otimes \dots \otimes (T \times T)^n (f_{k-1} \otimes \bar{f}_{k-1}) \right\|_{L^2(\nu)} \\ \leq D_{k-1} \cdot \min \left\{ \|f_i\|_{U^k(X, \mu T)}^2 : 1 \leq i \leq k-1 \right\}. \end{aligned} \quad (6.14)$$

Setting

$$v_n = \mathbf{1} \otimes T^n f_1 \otimes \dots \otimes T^n f_{k-1} \otimes \mathbf{1} \in L^2(\tau),$$

we note that

$$v_n \otimes \bar{v}_n = ((T \times T)^n (f_1 \otimes \bar{f}_1) \otimes \dots \otimes (T \times T)^n (f_{k-1} \otimes \bar{f}_{k-1})) \circ \varphi.$$

Since  $\nu = \varphi_*(\tau \times \tau)$ , combining Lemma 6.10 (applied to each  $\Phi = \Phi_N$ ) with (6.14), it follows that

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} b(n) \cdot v_n \right\|_{L^2(\tau)}^2 \leq \|b\|_\infty^2 \cdot D_{k-1} \cdot \min \left\{ \|f_i\|_{U^k(X, \mu T)}^2 : 1 \leq i \leq k-1 \right\}.$$

Taking square roots on both sides, we obtain the desired conclusion with  $C_k = \sqrt{D_{k-1}}$ .  $\square$

#### 6.4. Proof of Theorem 6.2

For the proof of Theorem 6.2 we need one more lemma.

**Lemma 6.11.** *Let  $(X, \mu, T)$  be an ergodic system, let  $a \in \text{gen}(\mu, \Phi)$  for some Følner sequence  $\Phi$ , and denote by  $(Z_{k-1}, m_{k-1}, T)$  the  $(k-1)$ -step pronilfactor of  $(X, \mu, T)$ . Assume  $(Y, S)$*

is a  $(k-1)$ -step pronilsystem. Then for every  $g \in C(X)$ ,  $y \in Y$ , and  $F \in C(Y)$ ,

$$\limsup_{N \rightarrow \infty} \left| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} g(T^n a) F(S^n y) \right| \leq \|\mathbb{E}(g \mid Z_{k-1})\|_{L^1(m_{k-1})} \cdot \|F\|_\infty. \quad (6.15)$$

We stress that (6.15) requires the function  $g$  to be continuous, as we could otherwise modify it on the orbit of  $a$  without changing the right-hand side.

*Proof.* Pick an increasing sequence  $(N_j)_{j \in \mathbb{N}}$  of natural numbers such that

$$\limsup_{N \rightarrow \infty} \left| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} g(T^n a) F(S^n y) \right| = \lim_{j \rightarrow \infty} \left| \frac{1}{|\Phi_{N_j}|} \sum_{n \in \Phi_{N_j}} g(T^n a) F(S^n y) \right|.$$

By refining  $(N_j)_{j \in \mathbb{N}}$  if necessary, we can assume that  $(a, y)$  is generic along  $(\Phi_{N_j})_{j \in \mathbb{N}}$  for a measure  $\rho$  on  $X \times Y$  with respect to  $T \times S$ .

Since  $(Y, S)$  is a pronilsystem, there is an invariant measure  $\nu$  on  $Y$  for which  $y \in \text{gen}(\nu, \Phi)$ . Replacing  $Y$  with the support of  $\nu$  we may assume that  $(Y, \nu, S)$  is ergodic. Since  $a \in \text{gen}(\mu, \Phi)$  we have that the first marginal of  $\rho$  is  $\mu$  and the second marginal is  $\nu$ . Letting  $\tilde{g} = \mathbb{E}(g \mid Z_{k-1}) \circ \pi_{k-1}$ , we obtain

$$\begin{aligned} & \lim_{j \rightarrow \infty} \left| \frac{1}{|\Phi_{N_j}|} \sum_{n \in \Phi_{N_j}} g(T^n a) F(S^n y) \right| \\ &= \left| \int_{X \times Y} g \otimes F \, d\rho \right| \leq \left| \int_{X \times Y} (g - \tilde{g}) \otimes F \, d\rho \right| + \left| \int_{X \times Y} \tilde{g} \otimes F \, d\rho \right|. \end{aligned}$$

Note that

$$\|(g - \tilde{g}) \otimes \mathbf{1}\|_{U^k(X \times Y, \rho, T \times S)} = \|g - \tilde{g}\|_{U^k(X, \mu, T)} = 0.$$

Thus  $(g - \tilde{g}) \otimes \mathbf{1}$  is orthogonal to the maximal  $(k-1)$ -step pronilfactor of  $(X \times Y, \zeta, T \times S)$  for almost every ergodic component  $\zeta$  of  $\rho$  by [20, Proposition 18, Chapter 8]. On the other hand, since  $(Y, \nu, S)$  is an ergodic  $(k-1)$ -step nilsystem it is a factor of  $(X \times Y, \zeta, T \times S)$ . Thus  $(Y, \nu, S)$  is a factor of the maximal  $(k-1)$ -step pronilfactor of  $(X \times Y, \zeta, T \times S)$ . Therefore, as  $\mathbf{1} \otimes F$  is measurable with respect to the latter factor, the functions  $(g - \tilde{g}) \otimes \mathbf{1}$  and  $\mathbf{1} \otimes F$  are orthogonal in  $L^2(\zeta)$  for almost every ergodic component  $\zeta$  of  $\rho$ , and hence

$$\int_{X \times Y} (g - \tilde{g}) \otimes F \, d\rho = 0.$$

We are left with

$$\begin{aligned} \limsup_{N \rightarrow \infty} \left| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} g(T^n a) F(S^n y) \right| &\leq \left| \int_{X \times Y} \tilde{g} \otimes F \, d\rho \right| \\ &\leq \|F\|_\infty \cdot \int_X |\tilde{g}| \, d\mu = \|F\|_\infty \cdot \int_{Z_{k-1}} |\mathbb{E}(g \mid Z_{k-1})| \, dm_{k-1}, \end{aligned}$$

and the claim follows.  $\square$

*Proof of Theorem 6.2.* By a standard approximation argument, it suffices to show that for all  $f_1, \dots, f_k \in C(X)$  with  $\|f_i\|_\infty \leq 1$ ,

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} T_\Delta^n(f_1 \otimes \dots \otimes f_k \otimes \mathbf{1}) - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} T_\Delta^n(\mathbf{1} \otimes f_1 \otimes \dots \otimes f_k) \right\|_{L^2(\sigma)} = 0. \quad (6.16)$$

To ease the notation, we write  $g_i = \mathbb{E}(f_i \mid Z_{k-1}) \circ \pi_{k-1}$  for  $i = 1, \dots, k$ . Note that for every  $n \in \mathbb{N}$

$$T_\Delta^n(f_1 \otimes \dots \otimes f_k \otimes \mathbf{1}) = f_1(T^n a) \cdot T_\Delta^n(\mathbf{1} \otimes f_2 \otimes \dots \otimes f_k \otimes \mathbf{1})$$

holds  $\sigma$ -almost everywhere. Using (6.5), we have that  $\|f_i - g_i\|_{U^k(X, \mu, T)} = 0$  for each  $i$ . In view of Lemma 4.6,  $\sigma$  satisfies the assumptions of Theorem 6.6, which now implies that

$$\begin{aligned} \lim_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} f_1(T^n a) \cdot T_\Delta^n(\mathbf{1} \otimes f_2 \otimes \dots \otimes f_k \otimes \mathbf{1}) \right. \\ \left. - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} f_1(T^n a) \cdot T_\Delta^n(\mathbf{1} \otimes g_2 \otimes \dots \otimes g_k \otimes \mathbf{1}) \right\|_{L^2(\sigma)} = 0. \end{aligned}$$

On the other hand, Theorem 6.9 implies

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} T_\Delta^n(\mathbf{1} \otimes f_1 \otimes \dots \otimes f_k) - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} T_\Delta^n(\mathbf{1} \otimes g_1 \otimes \dots \otimes g_k) \right\|_{L^2(\sigma)} = 0.$$

This shows that (6.16) is equivalent to

$$\begin{aligned} \lim_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} f_1(T^n a) \cdot T_\Delta^n(\mathbf{1} \otimes g_2 \otimes \dots \otimes g_k \otimes \mathbf{1}) \right. \\ \left. - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} T_\Delta^n(\mathbf{1} \otimes g_1 \otimes \dots \otimes g_k) \right\|_{L^2(\sigma)} = 0. \end{aligned} \quad (6.17)$$

Using the definition of  $\sigma$ , we can rewrite (6.17) as

$$\begin{aligned} \lim_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} f_1(T^n a) \cdot (T^n \mathbb{E}(f_2 \mid Z_{k-1}) \otimes \dots \otimes T^n \mathbb{E}(f_k \mid Z_{k-1}) \otimes \mathbf{1}) \right. \\ \left. - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (T^n \mathbb{E}(f_1 \mid Z_{k-1}) \otimes \dots \otimes T^n \mathbb{E}(f_k \mid Z_{k-1})) \right\|_{L^2(\xi)} = 0, \end{aligned} \quad (6.18)$$

where  $\xi$  is as in (4.2). Fix  $1 > \varepsilon > 0$  and choose  $h_1, \dots, h_k \in C(Z_{k-1})$  such that

$$\|\mathbb{E}(f_i \mid Z_{k-1}) - h_i\|_{L^2(m_{k-1})} < \varepsilon. \quad (6.19)$$

As the functions  $f_1, \dots, f_k$  are uniformly bounded by 1, we can assume that the functions  $h_1, \dots, h_k$  are as well.

To prove (6.18), we establish the following three statements:

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (T^n \mathbb{E}(f_1 | Z_{k-1}) \otimes \dots \otimes T^n \mathbb{E}(f_k | Z_{k-1})) - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (T^n h_1 \otimes \dots \otimes T^n h_k) \right\|_{L^2(\xi)} \leq k^2 \varepsilon, \quad (6.20)$$

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} f_1(T^n a) \cdot (T^n \mathbb{E}(f_2 | Z_{k-1}) \otimes \dots \otimes T^n \mathbb{E}(f_k | Z_{k-1}) \otimes \mathbf{1}) - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} h_1(\pi_{k-1}(T^n a)) \cdot (T^n h_2 \otimes \dots \otimes T^n h_k \otimes \mathbf{1}) \right\|_{L^2(\xi)} \leq k^2 \varepsilon + \sqrt{\varepsilon}, \quad (6.21)$$

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} h_1(\pi_{k-1}(T^n a)) \cdot (T^n h_2 \otimes \dots \otimes T^n h_k \otimes \mathbf{1}) - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (T^n h_1 \otimes \dots \otimes T^n h_k) \right\|_{L^2(\xi)} = 0. \quad (6.22)$$

Since  $\varepsilon$  can be taken arbitrarily small, (6.18) follows by combining (6.20), (6.21), and (6.22) and using the triangle inequality.

To prove (6.20) we first combine (6.19) with the fact that  $m_{k-1}$  is  $T$ -invariant and (4.4) to deduce that for every  $n \in \mathbb{N}$ ,

$$\|T^n \mathbb{E}(f_i | Z_{k-1}) - T^n h_i\|_{L^2(\xi_i)} < \varepsilon \sqrt{i},$$

where  $\xi_i$  is the  $i$ -th marginal of  $\xi$ . This, in turn, implies that

$$\left\| (T^n \mathbb{E}(f_1 | Z_{k-1}) \otimes \dots \otimes T^n \mathbb{E}(f_k | Z_{k-1})) - (T^n h_1 \otimes \dots \otimes T^n h_k) \right\|_{L^2(\xi)} \leq k^2 \varepsilon$$

by estimating coordinate-wise, and hence (6.20) follows by averaging.

To prove (6.21) we first note that, by the same argument, we have

$$\left\| (T^n \mathbb{E}(f_2 | Z_{k-1}) \otimes \dots \otimes T^n \mathbb{E}(f_k | Z_{k-1}) \otimes \mathbf{1}) - (T^n h_2 \otimes \dots \otimes T^n h_k \otimes \mathbf{1}) \right\|_{L^2(\xi)} \leq k^2 \varepsilon. \quad (6.23)$$

Defining  $g(x) = f_1(x) - h_1(\pi_{k-1}(x))$ , we then have that

$$f_1(T^n a) - h_1(\pi_{k-1}(T^n a)) = g(T^n a).$$

Since  $\pi_{k-1}$  is continuous, the function  $g$  is continuous. Multiplying (6.23) by  $f_1(T^n a)$  and using the triangle inequality, we reduce (6.21) to

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} g(T^n a) \cdot (T^n h_2 \otimes \dots \otimes T^n h_k \otimes \mathbf{1}) \right\|_{L^2(\xi)} \leq \sqrt{\varepsilon}. \quad (6.24)$$

To establish (6.24) we use Lemma 6.11. Defining  $Y = Z_{k-1}^k$ ,  $S = T \times \dots \times T$ , and

$F(x_1, \dots, x_k) = h_2(x_1) \cdots h_k(x_{k-1})$ , we then have

$$\begin{aligned}
& \limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} g(T^n a) \cdot (T^n h_2 \otimes \cdots \otimes T^n h_k \otimes \mathbf{1}) \right\|_{L^2(\xi)}^2 \\
& \leq \limsup_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} g(T^n a) \cdot (T^n h_2 \otimes \cdots \otimes T^n h_k \otimes \mathbf{1}) \right\|_{L^1(\xi)} \\
& \leq \int_Y \limsup_{N \rightarrow \infty} \left| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} g(T^n a) \cdot F(S^n y) \right| d\xi(y) \\
& \leq \sup_{y \in Y} \limsup_{N \rightarrow \infty} \left| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} g(T^n a) \cdot F(S^n y) \right|.
\end{aligned}$$

Since  $\pi_{k-1}$  is continuous, the function  $g$  is continuous and therefore we can apply Lemma 6.11 to conclude that

$$\begin{aligned}
& \limsup_{N \rightarrow \infty} \left| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (f_1(T^n a) - h_1(\pi_{k-1}(T^n a))) \cdot F(S^n y) \right| \\
& \leq \|\mathbb{E}(g \mid Z_{k-1})\|_{L^1(m_{k-1})} \cdot \|F\|_\infty \leq \|\mathbb{E}(f_1 \mid Z_{k-1}) - h_1\|_{L^2(m_{k-1})} \cdot \|F\|_\infty \leq \varepsilon,
\end{aligned}$$

and the estimate in (6.24) follows.

To finish the proof, we are left with verifying (6.22). For this we use the definition (4.2) of  $\xi$  and the fact that orbit closures in pronilsystems are uniquely ergodic pronilsystems. We have

$$\begin{aligned}
& \lim_{N \rightarrow \infty} \left\| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} h_1(\pi_{k-1}(T^n a)) \cdot (T^n h_2 \otimes \cdots \otimes T^n h_k \otimes \mathbf{1}) - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (T^n h_1 \otimes \cdots \otimes T^n h_k) \right\|_{L^2(\xi)}^2 \\
& = \lim_{N \rightarrow \infty} \int_{Z_{k-1}^k} \left| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} h_1(\pi_{k-1}(T^n a)) \cdot (T^n h_2 \otimes \cdots \otimes T^n h_k \otimes \mathbf{1}) \right. \\
& \quad \left. - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} (T^n h_1 \otimes \cdots \otimes T^n h_k) \right|^2 d\xi \\
& = \lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} \frac{1}{|\Phi_M|} \sum_{m \in \Phi_M} \left| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} h_1(T^m(\pi_{k-1} a)) h_2(T^{n+m}(\pi_{k-1} a)) \cdots h_k(T^{n+(k-1)m}(\pi_{k-1} a)) \right. \\
& \quad \left. - \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} h_1(T^{n+m}(\pi_{k-1} a)) \cdots h_k(T^{n+km}(\pi_{k-1} a)) \right|^2.
\end{aligned}$$

Since  $Z_{k-1}$  is a pronilsystem, arguing as at the end of the proof of Theorem 5.2 we can interchange the order of the limits, and this last expression becomes

$$\lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{1}{|\Phi_M|} \sum_{m \in \Phi_M} \left| \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} h_1(T^m(\pi_{k-1} a)) h_2(T^{n+m}(\pi_{k-1} a)) \cdots h_k(T^{n+(k-1)m}(\pi_{k-1} a)) \right.$$

$$- \frac{1}{|\Phi_N|} \sum_{n \in \Phi_N} h_1(T^{m+m}(\pi_{k-1}a)) h_1(T^{m+2m}(\pi_{k-1}a)) \cdots h_k(T^{m+km}(\pi_{k-1}a)) \Big|^2.$$

Making the change of variables  $n \mapsto n - m$  in the second term shows that the two limits are identical, concluding the proof.  $\square$

## References

- [1] V. BERGELSON. Sets of recurrence of  $\mathbb{Z}^m$ -actions and properties of sets of differences in  $\mathbb{Z}^m$ . *J. London Math. Soc.* (2) **31** no. 2 (1985), 295–304.
- [2] V. BERGELSON AND A. LEIBMAN Cubic averages and large intersections. Recent trends in ergodic theory and dynamical systems, 5–19, *Contemp. Math.* **631**, Amer. Math. Soc., Providence, RI, 2015.
- [3] D. CHARAMARAS AND A. MOUNTAKIS. Finding product sets in some classes of amenable groups. *Forum Math. Sigma* **13** (2025), Paper No. e10.
- [4] M. DI NASSO, I. GOLDBRING, R. JIN, S. LETH, M. LUPINI, AND K. MAHLBURG. On a sunset conjecture of Erdős. *Canad. J. Math.* **67** (2015), no. 4, 795–809.
- [5] P. ERDŐS. Problems and results in combinatorial number theory. *Astérisque* **2–25** (1975), 295–309.
- [6] P. ERDŐS. Problems and results on combinatorial number theory. III. Number theory day (Proc. Conf., Rockefeller Univ., New York, 1976), pp. 43–72. *Lecture Notes in Math.*, Vol. 626, Springer, Berlin, 1977.
- [7] P. ERDŐS. A survey of problems in combinatorial number theory. *Ann. Discrete Math.* **6** (1980), 89–115.
- [8] P. ERDŐS AND R. L. GRAHAM. Old and new problems and results in combinatorial number theory. *Monographies de L’Enseignement Mathématique* **28**. *Université de Genève, L’Enseignement Mathématique*, Geneva, 1980. 128 pp.
- [9] N. FRANTZIKINAKIS AND B. KUČA. Joint ergodicity for commuting transformations and applications to polynomial sequences. *Invent. Math.* **239** (2025), no. 2, 621–706.
- [10] N. FRANTZIKINAKIS AND B. HOST. Weighted multiple ergodic averages and correlation sequences. *Ergodic Theory Dynam. Systems* **38** (2018), no. 1, 81–142.
- [11] H. FURSTENBERG. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.* **31** (1977), 204–256.
- [12] H. FURSTENBERG. *Recurrence in Ergodic Theory and Combinatorial Number Theory*. M. B. Porter Lectures. Princeton University Press, Princeton, N.J., 1981.
- [13] N. HINDMAN. Finite sums from sequences within cells of a partition of  $\mathbb{N}$ . *J. Combinatorial Theory Ser. A* **17** (1974), 1–11.
- [14] N. HINDMAN. Ultrafilters and combinatorial number theory, *Number theory, Carbondale 1979 (Proc. Southern Illinois Conf., Southern Illinois Univ., Carbondale, Ill., 1979)*, *Lecture Notes in Math.*, vol. **751**, Springer, Berlin, 1979, pp. 119–184.

- [15] N. HINDMAN. On density, translates, and pairwise sums of integers. *J. Combin. Theory Ser. A* **33** (1982), no. 2, 147–157.
- [16] B. HOST. Ergodic seminorms for commuting transformations and applications. *Studia Math.* **195** (2009), 31–49.
- [17] B. HOST. A short proof of a conjecture of Erdős proved by Moreira, Richter and Robertson. *Discrete Anal.* (2019), Paper No. 19, 10 pp.
- [18] B. HOST AND B. KRA. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)* **161** (2005), no. 1, 397–488.
- [19] B. HOST AND B. KRA. Uniformity norms on  $\ell^\infty$  and applications. *J. Anal. Math.* **108** (2009), 219–276.
- [20] B. KRA AND B. HOST. Nilpotent structures in ergodic theory. *Mathematical Surveys and Monographs*, **236**. American Mathematical Society, Providence, RI, 2018.
- [21] I. KOUSEK. Asymmetric infinite sumsets in large sets of integers. [arXiv:2502.03112](https://arxiv.org/abs/2502.03112).
- [22] I. KOUSEK AND T. RADIC. Infinite unrestricted sumsets of the form  $B + B$  in sets with large density. [arXiv:2404.12201](https://arxiv.org/abs/2404.12201).
- [23] B. KRA, J. MOREIRA, F. RICHTER, AND D. ROBERTSON. Infinite sumsets in sets with positive density. *J. Amer. Math. Soc.* **37** (2024), no. 3, 637–682.
- [24] B. KRA, J. MOREIRA, F. RICHTER, AND D. ROBERTSON. A proof of Erdős’s  $B+B+t$  conjecture. *Commun. Am. Math. Soc.* **4** (2024), 480–494.
- [25] B. KRA, J. MOREIRA, F. RICHTER, D. ROBERTSON. Problems on infinite sumset configurations in the integers and beyond. *Bull. Amer. Math. Soc. (N.S.)* [arXiv:2311.06197](https://arxiv.org/abs/2311.06197)
- [26] A. LEIBMAN. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold, *Ergodic Theory Dynam. Systems* **25** (2005), no. 1, 201–213.
- [27] E. LESIGNE. Sur une nilvariété  $\hat{e}$ , les parties minimales associées à une translation sont uniquement ergodiques. *Ergodic Theory Dynam. Systems* **11** (1991), no. 2, 379–391.
- [28] J. MOREIRA, F. RICHTER, AND D. ROBERTSON. A proof of a sumset conjecture of Erdős. *Ann. of Math. (2)* **189** (2019), no. 2, 605–652.
- [29] M. NATHANSON. Sumsets contained in infinite sets of integers. *J. Combin. Theory Ser. A* **28** (1980), no. 2, 150–155.
- [30] E. SZEMERÉDI. On sets of integers containing no  $k$  elements in arithmetic progression. *Acta Arith.* **27** (1975), 199–245,
- [31] T. TAO AND T. ZIEGLER. Infinite partial sumsets in the primes. *J. Anal. Math.* **151** (2023), no. 1, 375–389.
- [32] B. L. VAN DER WAERDEN. Beweis einer baudetschen vermutung. *Nieuw. Arch. Wisk.* **15** (1927), 212–216.

Bryna Kra  
NORTHWESTERN UNIVERSITY  
[kra@math.northwestern.edu](mailto:kra@math.northwestern.edu)

Joel Moreira  
UNIVERSITY OF WARWICK  
[joel.moreira@warwick.ac.uk](mailto:joel.moreira@warwick.ac.uk)

Florian K. Richter  
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL)  
[f.richter@epfl.ch](mailto:f.richter@epfl.ch)

Donald Robertson  
UNIVERSITY OF MANCHESTER  
[donald.robertson@manchester.ac.uk](mailto:donald.robertson@manchester.ac.uk)

---

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.