Ergod. Th. & Dynam. Sys. (2018), **38**, 81–142
doi:10.1017/etds.2016.19

Weighted multiple ergodic averages and correlation sequences

NIKOS FRANTZIKINAKIS† and BERNARD HOST‡

† *University of Crete, Department of Mathematics, Voutes University Campus, Heraklion 71003, Greece*
(e-mail: frantzikinakis@gmail.com)

‡ *Université Paris-Est Marne-la-Vallée, Laboratoire d'analyse et de mathématiques appliquées, UMR CNRS 8050, 5 Bd Descartes, 77454 Marne la Vallée Cedex, France*
(e-mail: bernard.host@u-pem.fr)

(*Received* 30 *November* 2015 *and accepted in revised form* 3 *February* 2016)

Abstract. We study mean convergence results for weighted multiple ergodic averages defined by commuting transformations with iterates given by integer polynomials in several variables. Roughly speaking, we prove that a bounded sequence is a good universal weight for mean convergence of such averages if and only if the average of this sequence times any nilsequence converges. Two decomposition results of independent interest play key roles in the proof. The first states that every bounded sequence in several variables satisfying some regularity conditions is a sum of a nilsequence and a sequence that has small uniformity norm (this generalizes a result of the second author and Kra); and the second states that every multiple correlation sequence in several variables is a sum of a nilsequence and a sequence that is small in uniform density (this generalizes a result of the first author). Furthermore, we use these results in order to establish mean convergence and recurrence results for a variety of sequences of dynamical and arithmetic origin and give some combinatorial implications.

Contents

1. *Introduction*

Since the early 1980s, a lot of effort has been put into the study of the limiting behavior of multiple ergodic averages. This study was partly motivated by combinatorial implications, since positiveness properties of such averages imply various far reaching extensions of the celebrated theorem of Szemerédi on arithmetic progressions. After a long series of partial results, most notably those in [**7**, **20**, **21**, **34**, **43–45**, **52**, **60**, **62**, **66**], Walsh [**64**], building on previous work of Tao [**62**], proved the following mean convergence result.

THEOREM 1.1. [**64**]† *Let $d, \ell, s \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space, and $T_1, \ldots, T_\ell \colon X \to X$ be invertible commuting measure preserving transformations. Then, for every Følner sequence $(I_k)_{k\in\mathbb{N}}$ of subsets of $\mathbb{N}^d$, polynomials $p_{i,j} \colon \mathbb{N}^d \to \mathbb{Z}$, $i = 1, \ldots, \ell$, $j = 1, \ldots, s$ and functions $f_1, \ldots, f_s \in L^\infty(\mu)$, the averages*

$$\frac{1}{|I_k|}\sum_{\mathbf{n}\in I_k} f_1\bigg(\prod_{i=1}^{\ell} T_i^{p_{i,1}(\mathbf{n})}x\bigg)\cdot\cdots\cdot f_s\bigg(\prod_{i=1}^{\ell} T_i^{p_{i,s}(\mathbf{n})}x\bigg) \tag{1.1}$$

converge in $L^2(\mu)$ as $k \to +\infty$.

Remark. In [**64**], the previous result was established under the weaker hypothesis that the transformations $T_1, \ldots, T_\ell$ generate a nilpotent group. We believe that our results can be extended to this more general set-up, but, in this article, we restrict our work to the case of commuting transformations.

One of the main purposes of this article is to study mean convergence for weighted versions of the averages (1.1): that is, averages of the form

$$\frac{1}{|I_k|}\sum_{\mathbf{n}\in I_k} w(\mathbf{n})\ f_1\bigg(\prod_{i=1}^{\ell} T_i^{p_{i,1}(\mathbf{n})}x\bigg)\cdot\cdots\cdot f_s\bigg(\prod_{i=1}^{\ell} T_i^{p_{i,s}(\mathbf{n})}x\bigg), \tag{1.2}$$

where $w \colon \mathbb{N}^d \to \mathbb{C}$ is a bounded sequence. A sequence w, for which the previous averages converge for all choices of systems, functions, and polynomials, is called a *good universal weight for mean convergence* of the averages (1.2).

When $d = 1$, examples of good universal weights for some multiple ergodic averages can be found in [**1–4**, **18**, **25**, **29**, **31**, **46**, **68**]. Most of these results deal with the case where $\ell = 1$ and are based on the theory of characteristic factors that was pioneered by Furstenberg. They depend, crucially, on the work of Host and Kra [**44**] and subsequent developments in [**45**, **52**], which, in the case where all the transformations are equal, gives a characterization in terms of nilsystems of the smallest factor of the system that controls the limiting behavior of the averages (1.2). Unfortunately, no such characterization is known in the case of general commuting transformations (but see [**8**, **9**] for related progress), which is the reason why this method is not applicable for our more general set-up. Moreover, the method used by Walsh in [**64**] does not seem applicable to weighted averages and no general criterion suitable for checking mean convergence of averages of the form (1.2) is known. We fill this gap by showing, in Theorem 2.2, that a bounded sequence $w \colon \mathbb{N}^d \to \mathbb{C}$

† The argument in [**64**] is given for Cesàro averages and $d = 1$ but the same proof works in this more general case (see [**67**] for details).

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

is a good universal weight for mean convergence of the averages (1.2) if and only if the averages

$$\frac{1}{|I_k|}\sum_{\mathbf{n}\in I_k} w(\mathbf{n})\cdot\psi(\mathbf{n}) \tag{1.3}$$

converge for every nilsequence ψ in d variables and Følner sequence $(I_k)_{k\in\mathbb{N}}$ in $\mathbb{N}^d$. Furthermore, when one replaces, throughout the averages, $(1/I_k)\sum_{\mathbf{n}\in I_k}$ by the Cesàro averages $(1/N^d)\sum_{\mathbf{n}\in[1,N]^d}$, we prove, in Theorem 2.4, that a similar criterion holds for weak convergence and that a condition somewhat stronger than (1.3) suffices for mean convergence. Even for single variable sequences, the mean convergence criterion is new and its proof (strangely) depends on decomposition results for sequences in two variables. Prior to this work, only the case $d=\ell=1$ was treated (in [**18**]) for mean convergence, while, for weak convergence, the case where $d=1$ and $\ell\in\mathbb{N}$ is arbitrary was treated in [**29**].

To prove these results, we use the mean convergence result of Walsh as a black box and two decomposition results of independent interest. These are Theorems 3.9 and 3.10, which extend similar results for single variable sequences from [**46**, Theorem 2.19] and [**29**, Theorem 1.2]. Roughly speaking, they state the following.

(i) If the averages (1.3) converge for every nilsequence $\psi\colon\mathbb{N}^d\to\mathbb{C}$ and every Følner sequence $(I_k)_{k\in\mathbb{N}}$ in $\mathbb{N}^d$, then the sequence $w\in\ell^\infty(\mathbb{N}^d)$ is a sum of a nilsequence and a sequence that has small uniformity norm.

(ii) Any sequence of the form

$$\int\prod_{i=1}^{\ell} f_1(T_i^{p_{i,1}(\mathbf{n})}x)\cdot\dots\cdot\prod_{i=1}^{\ell} f_s(T_i^{p_{i,s}(\mathbf{n})}x)\,d\mu,\qquad \mathbf{n}\in\mathbb{N}^d$$

is the sum of a nilsequence and a sequence that is small in uniform density.

Regarding the second decomposition, Theorem 2.6 gives more precise information when the iterates are linear: it implies, for example, that the sequences

$$\int f\cdot T_1^m f\cdot T_2^n f\cdot T_3^r f\,d\mu,\qquad \int f\cdot T_1^m f\cdot T_2^n f\cdot T_3^{m+n} f\,d\mu,$$
$$\int f\cdot T_1^n f\cdot T_2^n f\cdot T_3^n f\,d\mu \tag{1.4}$$

are 1-step, 2-step and 3-step nilsequences, respectively, modulo small errors in uniform density (simple examples show that the degree of nilpotency is optimal).

Using the previous criteria, we prove mean convergence results for weighted ergodic averages with weights given by various sequences of dynamical origin, bounded multiplicative functions, generalized polynomials and Hardy field sequences (see §§2.2, 2.3, 2.6). We deduce some multiple recurrence results and combinatorial consequences; showing for example that every set of integers with positive upper density contains arbitrarily long arithmetic progressions with common difference of the form m^2+n^2, where m, n have an odd (or an even) number of distinct prime factors (see Theorems 2.12–2.14) or m, n are taken from the set $\{k\in\mathbb{N}\colon \|k^a\|\in[1/2, 3/4]\}$, where a is any positive non-integer (see Theorems 9.1, 9.5). We also establish multidimensional variants of these results regarding patterns in positive density subsets of $\mathbb{Z}^\ell$.

In the next section, we give the precise formulation of our main results and define some concepts used throughout the article.

2. *Precise statement of main results*

2.1. *Notation and definitions.*

We first introduce some notation that is going to facilitate our presentation.

2.1.1. *Ergodic theory.*

Following, for example, [**35**] we say that a probability space $(X, \mathcal{X}, \mu)$ is a *Lebesgue space* if X can be given the structure of a Polish space (i.e. metrizable, separable, complete) such that $\mathcal{X}$ is its Borel σ-algebra. Throughout the article, we make the standard assumption that *all probability spaces considered are Lebesgue.*

By a system $(X, \mathcal{X}, \mu, T_1, \ldots, T_\ell)$ we mean a Lebesgue probability space $(X, \mathcal{X}, \mu)$ endowed with ℓ invertible commuting measure preserving transformations. For $\vec{n} = (n_1, \ldots, n_\ell) \in \mathbb{Z}^\ell$, we write $T_{\vec{n}} = T_1^{n_1} \cdot \cdots \cdot T_\ell^{n_\ell}$. Sometimes, we denote by $\vec{T}$ the action of $\mathbb{Z}^\ell$ on X and write the system as $(X, \mathcal{X}, \mu, \vec{T})$. In the subsequent work, we generally omit the σ-algebra $\mathcal{X}$ from our notation.

2.1.2. *Nilmanifolds and nilsequences.*

Let $s \in \mathbb{N}$, G be an s-step nilpotent Lie group and Γ be a discrete cocompact subgroup of G. Then the quotient space $X = G/\Gamma$ is called an *s-step nilmanifold*. We prefer to denote the elements of X as points $x, y, \ldots$, not as cosets. The point e_X is the image in X of the unit element of G. The natural action of G on X is written $(g, x) \mapsto g \cdot x$ and the unique measure on X invariant under this action is called the *Haar measure* of X and is denoted by m_X.

Let $\tau_1, \ldots, \tau_d$ be commuting elements of G. For $i = 1, \ldots, d$ let T_i be the translation $x \mapsto \tau_i \cdot x$ by τ_i on X. Then the system $(X, m_X, T_1, \ldots, T_d)$ is called an *s-step nilsystem*. Nilsystems have been extensively studied and basic properties were established by Auslander [**6**], Parry [**58**, **59**], Lesigne [**56**], Leibman [**50**, **51**] and others.

Definition. [**11**] If $X = G/\Gamma$ is an s-step nilmanifold, $\Psi \in C(X)$ and $\tau_1, \ldots, \tau_d \in G$ are commuting elements, then the sequence $(\Psi(\tau_1^{n_1} \cdot \cdots \cdot \tau_d^{n_d} \cdot e_X))_{n_1,\ldots,n_d \in \mathbb{N}}$ is called an *s-step nilsequence in d variables*. Also, for notational convenience, we define a 0*-step nilsequence* to be a constant sequence.

Remarks.

- In [**11**] the notion '*basic s*-step nilsequence' is used for what we call here an 's-step nilsequence'.
- By [**50**, Paragraph 1.11], the nilmanifold X is isomorphic to a sub-nilmanifold of a nilmanifold $\tilde{X} = \tilde{G}/\tilde{\Gamma}$, where $\tilde{G}$ is a connected and simply connected s-step nilpotent Lie group and all elements of G are represented in $\tilde{G}$. Hence, whenever needed, we can assume that the group G is connected and simply connected.

In recent years, nilsequences have played a key role in ergodic theory and additive combinatorics. They form the right substitute for linear exponential sequences needed to formalize certain inverse theorems which are used in the course of studying various multilinear expressions in analysis and number theory.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

2.1.3. *Følner sequences and related averages.* First, we recall some notions and introduce some notation.

Notation. We write $[N]$ for the interval $\{1, 2, \ldots, N\}$ in $\mathbb{N}$.

Definition. A *Følner sequence in* $\mathbb{N}^d$ is a sequence $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ of finite subsets of $\mathbb{N}^d$ that satisfies

$$\lim_{j\to\infty} \frac{|(I_j + \mathbf{k})\Delta I_j|}{|I_j|} = 0 \quad \text{for every } \mathbf{k} \in \mathbb{Z}^d,$$

where Δ denotes the symmetric difference and $I_j + \mathbf{k} := \{\mathbf{n} + \mathbf{k} \colon \mathbf{n} \in I_j\}$.

An example of a Følner sequence in $\mathbb{N}$ is a sequence of intervals whose lengths tend to infinity. If $N_j \to +\infty$ and $(\mathbf{k}_j)_{j\in\mathbb{N}}$ is a sequence in $\mathbb{N}^d$, then $I_j := \mathbf{k}_j + [N_j]^d$, $j \in \mathbb{N}$ is a Følner sequence in $\mathbb{N}^d$. In all subsequent results and proofs, we can replace the general Følner sequences by these particular examples.

If $a \colon \mathbb{N}^d \to \mathbb{C}$ is a sequence and $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ is a Følner sequence in $\mathbb{N}^d$, we let

$$\lim \mathrm{Av}_{\mathbf{I}}\, a(\mathbf{n}) := \lim_{j\to+\infty} \frac{1}{|I_j|} \sum_{\mathbf{n}\in I_j} a(\mathbf{n}),$$

assuming, of course, that the previous limit exists. If the previous limit exists for every Følner sequence $\mathbf{I}$, then it is independent of $\mathbf{I}$; we denote its common value with

$$\lim \mathrm{Av}\, a(\mathbf{n})$$

and say that *the averages of a converge*. When it is unclear with respect to which variable we take the averages, we use the notation

$$\lim \mathrm{Av}_{\mathbf{n},\mathbf{I}}\, a(\mathbf{n}), \qquad \lim \mathrm{Av}_{\mathbf{n}}\, a(\mathbf{n}).$$

Furthermore, we use the notation

$$\mathrm{limsup}\, |\mathrm{Av}_{\mathbf{I}}\, a(\mathbf{n})| := \limsup_{j\to+\infty} \left| \frac{1}{|I_j|} \sum_{\mathbf{n}\in I_j} a(\mathbf{n}) \right|$$

and

$$\mathrm{limsup}\, |\mathrm{Av}\, a(\mathbf{n})| := \sup_{\mathbf{I}} (\mathrm{limsup}\, |\mathrm{Av}_{\mathbf{I}}\, a(\mathbf{n})|),$$

where the sup is taken over all Følner sequences $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ of subsets of $\mathbb{N}^d$.

We use similar notation for limits in $L^2(\mu)$ involving averages of functions $(f_{\mathbf{n}})_{\mathbf{n}\in\mathbb{N}^d}$ in $L^2(\mu)$ and write

$$\lim \mathrm{Av}_{\mathbf{I}}\, f_{\mathbf{n}}, \quad \lim \mathrm{Av}\, f_{\mathbf{n}}, \quad \mathrm{limsup}\, \|\mathrm{Av}_{\mathbf{I}}\, f_{\mathbf{n}}\|_{L^2(\mu)}, \quad \mathrm{limsup}\, \|\mathrm{Av}\, f_{\mathbf{n}}\|_{L^2(\mu)}$$

for the corresponding limits, where, in the first two cases, convergence takes place in $L^2(\mu)$ and, in the last two cases, we use $L^2(\mu)$ norms in place of the absolute values.

2.2. *Convergence results for uniform averages.* First, we give convergence criteria for weighted ergodic averages which are defined using uniform averages, that is, averages over arbitrary Følner sequences in $\mathbb{N}^d$.

Definition. If $p_i : \mathbb{N}^d \to \mathbb{Z}$, $i = 1, \ldots, \ell$, are polynomials, we call the map $\vec{p} : \mathbb{N}^d \to \mathbb{Z}^\ell$ defined by $\vec{p} := (p_1, \ldots, p_\ell)$ a *polynomial mapping* from $\mathbb{N}^d$ to $\mathbb{Z}^\ell$. The *degree* $\deg(\vec{p})$ of $\vec{p}$ is $\max_{i=1,\ldots,\ell}(\deg(p_i))$.

We first state a strengthening of Theorem 1.1 that will be used frequently in this article. It is proved in §4.3.

PROPOSITION 2.1. *For every $d, \ell, s \in \mathbb{N}$, polynomial mappings $\vec{p_i} : \mathbb{N}^d \to \mathbb{Z}^\ell$, $i = 1, \ldots, s$, nilsequence $\psi : \mathbb{N}^d \to \mathbb{C}$, system $(X, \mu, T_1, \ldots, T_\ell)$ and functions $f_1, \ldots, f_s \in L^\infty(\mu)$, the limit*

$$\lim \text{Av}\, \psi(\mathbf{n}) \cdot T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s$$

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

Remark. For $d = 1$ this was proved in [**29**, §2.4].

Our main convergence criterion for uniform averages is the next result, which is proved in §7.1.

THEOREM 2.2. *Let $d, \ell, s, t \in \mathbb{N}$. Then there exists a positive integer $k = k(d, \ell, s, t)$, such that the following holds: if $w \in \ell^\infty(\mathbb{N}^d)$ is a sequence and, for every k-step nilsequence $\psi : \mathbb{N}^d \to \mathbb{C}$, the limit*

$$\lim \text{Av}\, w(\mathbf{n})\, \psi(\mathbf{n}) \text{ exists}, \tag{2.1}$$

then, for every system $(X, \mu, T_1, \ldots, T_\ell)$, functions $f_1, \ldots, f_s \in L^\infty(\mu)$ and polynomial mappings $\vec{p_i} : \mathbb{N}^d \to \mathbb{Z}^\ell$, $i = 1, \ldots, s$ of degree at most t, the limit

$$\lim \text{Av}\, w(\mathbf{n}) \cdot T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s \tag{2.2}$$

exists in $L^2(\mu)$. Furthermore, if the limit in (2.1) is zero for every k-step nilsequence ψ in d variables, then the limit in (2.2) is always zero. Lastly, if the polynomial mappings $\vec{p_i} : \mathbb{N}^d \to \mathbb{Z}^\ell$, $i = 1, \ldots, s$ are linear, then we can take $k = s$.

Remarks.

- For $d = \ell = t = 1$, this result was proved in [**46**] and, for $d = \ell = 1$ and $t \in \mathbb{N}$ arbitrary, in [**18**].
- In Theorem 3.9, we give a characterization using 'uniformity seminorms' of sequences satisfying the hypothesis of Theorem 2.2.

The next result shows that the hypothesis of Theorem 2.2 is necessary in order to have weak convergence of the averages in (2.2) for all linear polynomial mappings.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

PROPOSITION 2.3. *Let $d, s \in \mathbb{N}$ and $w \in \ell^\infty(\mathbb{N}^d)$ be a sequence. Suppose that for every system $(X, \mu, T_1, \ldots, T_d)$, functions $f_0, \ldots, f_s \in L^\infty(\mu)$ and linear forms $\overrightarrow{L_i} : \mathbb{N}^d \to \mathbb{Z}^d$, $i = 1, \ldots, s$, the limit*

$$\lim \operatorname{Av} w(\mathbf{n}) \int f_0 \cdot T_{\overrightarrow{L_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\overrightarrow{L_s}(\mathbf{n})} f_s \, d\mu$$

exists. Then the limit

$$\lim \operatorname{Av} w(\mathbf{n})\, \psi(\mathbf{n})$$

exists for every s-step nilsequence ψ in d variables.

Remark. For $d = 1$, this was proved in [**29**].

Next, for $d = 1$, we give some examples of sequences of weights in $\ell^\infty(\mathbb{N})$ for which Theorem 2.2 is applicable.

Examples. Let (Y, S) be a minimal uniquely ergodic system with invariant measure ν and, for every $s \in \mathbb{N}$, let (Z_s, ν_s, S) be the 'factor of order s' defined in [**44**]. Suppose that, for every $s \in \mathbb{N}$, the factor map $\pi_s : Y \to Z_s$ is continuous. Then, for every $\Psi \in C(Y)$ and every $y_0 \in Y$, the sequence $w : N \to \mathbb{C}$ defined by $w(n) := \Psi(S^n y_0)$, $n \in \mathbb{N}$ satisfies the hypothesis of Theorem 2.2 [**46**, Proposition 7.1]. Examples of this type include:

(i) the Thue–Morse sequence, which is the indicator function of those integers that have an odd sum of digits when expanded in base two (see [**46**, Proposition 2.21]); and

(ii) bounded generalized polynomials (see [**46**, Corollary 2.23])†. These include sequences of the form $(\{p(n)\})_{n\in\mathbb{N}}$ or $(e(p(n)))_{n\in\mathbb{N}}$, where $p : \mathbb{N} \to \mathbb{Z}$ is an arbitrary generalized polynomial, $\{x\}$ denotes the fractional part of x and $e(t) := e^{2\pi i t}$.

2.3. *Convergence results for Cesàro averages.* The assumptions of Theorem 2.2 are, in many cases, too strong to be of use (this is the case for the examples (i)–(v) below) and we would like to have a criterion that uses convergence assumptions of certain Cesàro averages instead of uniform averages. We obtain such a result by utilizing tools different from those used in the proof of Theorem 2.2 (Theorem 2.2 relies on Theorem 3.9 which necessitates the hypothesis (2.1)). The key new ingredients are decomposition results for multiple correlation sequences, which are stated in §2.4, below.

THEOREM 2.4. *Let $d, \ell, s, t \in \mathbb{N}$. Then there exists a positive integer $k = k(d, \ell, s, t)$, such that the following hold.*

(i) *If $w \in \ell^\infty(\mathbb{N}^d)$ and, for every k-step nilsequence $\psi : \mathbb{N}^d \to \mathbb{C}$, the limit*

$$\lim_{N\to+\infty} \frac{1}{N^d} \sum_{\mathbf{n}\in[N]^d} w(\mathbf{n})\, \psi(\mathbf{n}) \text{ exists,} \tag{2.3}$$

then, for every system $(X, \mu, T_1, \ldots, T_\ell)$, functions $f_0, \ldots, f_s \in L^\infty(\mu)$ and polynomial mappings $\overrightarrow{p_i} : \mathbb{N}^d \to \mathbb{Z}^\ell$, $i = 1, \ldots, s$, of degree at most t, the limit

† A *generalized polynomial* is a real-valued function that is obtained from the identity function and real constants by using the operations of addition, multiplication and taking the integer part.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

$$\lim_{N\to+\infty}\frac{1}{N^d}\sum_{\mathbf{n}\in[N]^d} w(\mathbf{n})\cdot\int f_0\cdot T_{\overrightarrow{p_1}(\mathbf{n})}f_1\cdot\cdots\cdot T_{\overrightarrow{p_s}(\mathbf{n})}f_s\,d\mu \tag{2.4}$$

exists.

(ii) *If $w\in\ell^\infty(\mathbb{N}^d)$ and, for every k-step nilsequence $\psi:\mathbb{N}^{2d}\to\mathbb{C}$, the limit*

$$\lim_{N,N'\to+\infty}\frac{1}{(NN')^d}\sum_{\mathbf{n}\in[N]^d,\ \mathbf{n}'\in[N']^d} w(\mathbf{n})\,\overline{w(\mathbf{n}')}\,\psi(\mathbf{n},\mathbf{n}') \tag{2.5}$$

exists, then, for every system $(X,\mu,T_1,\ldots,T_\ell)$, functions $f_1,\ldots,f_s\in L^\infty(\mu)$ and polynomial mappings $\overrightarrow{p_i}:\mathbb{N}^d\to\mathbb{Z}^\ell$, $i=1,\ldots,s$, of degree at most t, the limit

$$\lim_{N\to+\infty}\frac{1}{N^d}\sum_{\mathbf{n}\in[N]^d} w(\mathbf{n})\cdot T_{\overrightarrow{p_1}(\mathbf{n})}f_1\cdot\cdots\cdot T_{\overrightarrow{p_s}(\mathbf{n})}f_s \tag{2.6}$$

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

Furthermore, if the limit in (2.3) (respectively, (2.5)) is zero for every k-step nilsequence $\psi:\mathbb{N}^d\to\mathbb{C}$ (respectively, $\psi:\mathbb{N}^{2d}\to\mathbb{C}$), then the limit in (2.4) (respectively, (2.6)) is always zero.

Lastly, in (i) *(respectively,* (ii)*), if the polynomial mappings $\overrightarrow{p_i}$ are linear, then we can take $k=s$ (respectively, $k=2s-1$), and if, in addition, $\ell=s$ and $T_{\overrightarrow{p_i}(\mathbf{n})}=T_i^{L_i(\mathbf{n})}$ for $i=1,\ldots,s$, where $L_1,\ldots,L_s$ are linear forms spanning a subspace of dimension r, then we can take $k=s-r+1$ (respectively, $k=2s-2r+1$).*

Remarks.

- For $d=1$, the first part of this result was proved in [**29**].
- For single variable polynomials, in order to prove Part (ii) of this result, we rely on decomposition results of correlation sequences involving polynomials in two variables.

An analogue of Proposition 2.3, with Cesàro averages in place of uniform averages, holds with the same proof. This implies that the condition (2.3) is also necessary in order for the limit (2.4) to exist for all linear polynomial mappings.

Sequences $w\in\ell^\infty(\mathbb{N}^d)$ that satisfy the hypothesis (2.3) and (2.5) of Theorem 2.4 (but do not satisfy the condition (2.1) of Theorem 2.2, even for $\psi=1$) include the following.

(i) Any sequence of the form $(g(S_{\vec{n}}\,y))_{\vec{n}\in\mathbb{N}^d}$, where $(Y,\nu,S_1,\ldots,S_d)$ is a system and $g\in L^\infty(\nu)$, for $y\in Y$ belonging to a set of full measure that depends only on the system and the function g. For $d=1$, this was proved in [**46**, Theorem 2.22] for hypothesis (2.3) but a similar argument also gives hypothesis (2.5) and works (using Theorem 3.1) for general $d\in\mathbb{N}$.

(ii) Any 'good' multiplicative function $\phi:\mathbb{N}^d\to\mathbb{C}$ (see §2.6 and Theorem 8.1). For $d=1$, an alternate proof which depends on [**30**], and thus on deep results from [**38**] and [**40**], was given in [**31**].

(iii) The indicator function of all vectors of $\mathbb{N}^d$ whose coordinates have an even (or an odd) number of distinct prime factors; or more generally, the indicator function of any set S defined as in Theorem 2.12 below (this follows from Theorem 8.1 and the argument in §8.2).

(iv) Any sequence of the form $(e(\sum_{i=1}^{d} n_i^{a_i}))_{n_1,\ldots,n_d\in\mathbb{N}}$, where $a_1, \ldots, a_d$ are positive non-integers. In this case the limits in (2.3) and (2.5) are always zero (see Theorem 9.1 and Proposition 9.3). Moreover, for $d = 1$, we give necessary and sufficient conditions for a sequence of the form $(e(f(n)))_{n\in\mathbb{N}}$, where f is a Hardy field function of at most polynomial growth, to be a good universal weight for mean convergence of the averages (2.6) (see Corollary 9.2).

(v) The indicator function of any set of the form

$$S := \{n_1, \ldots, n_d \in \mathbb{N} \colon \|f_1(n_1)\| \in [a_1, b_1], \ldots, \|f_d(n_d)\| \in [a_d, b_d]\},$$

where $0 \le a_i < b_i \le 1/2$, $\|x\| := d(x, \mathbb{Z})$ and f_i are Hardy field functions of at most polynomial growth that stay away from polynomials (this follows from Proposition 9.3 and an approximation argument that uses the estimate (9.1)). Furthermore, the set S is good for multiple recurrence and mean convergence and the $L^2(\mu)$ limit

$$\lim_{N\to+\infty} \frac{1}{|S \cap [N]^d|} \sum_{\mathbf{n}\in S\cap[N]^d} T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s$$

is equal to the limit obtained when S is replaced by $\mathbb{N}^d$ (see Theorem 9.5).

2.4. *Multiple correlations in ergodic theory.* Multiple correlation sequences are well studied objects in ergodic theory and form an indispensable tool in the study of various multiple ergodic averages. For single variable sequences, structural results have been obtained in [**11**, **29**, **53**, **54**, **57**]; we extend some of these results to sequences in several variables. These extensions turn out to be key for the proof of the convergence criterion given in Theorem 2.4. Our argument follows closely the method used in [**29**] to obtain similar results for single variable sequences; but some refinements obtained (for example Theorem 2.6) require new methodology.

Notation. For a bounded sequence $a \colon \mathbb{N}^d \to \mathbb{C}$, we let

$$\|a\|_2 := \big(\limsup \mathrm{Av}\, |a(\mathbf{n})|^2\big)^{1/2}. \tag{2.7}$$

Definition. A bounded sequence $a \colon \mathbb{N}^d \to \mathbb{C}$ is an *approximate s-step nilsequence in d variables* if, for every $\varepsilon > 0$, it admits a decomposition as $a = a_{\mathrm{st}} + a_{\mathrm{er}}$, where:

(i) $a_{\mathrm{st}} \colon \mathbb{N}^d \to \mathbb{C}$ is an s-step nilsequence in d variables with $\|a_{\mathrm{st}}\|_\infty \le \|a\|_\infty$; and

(ii) $\|a_{\mathrm{er}}\|_2 \le \varepsilon$.

The subscripts 'st' and 'err' are used to indicate 'structured' and 'error', respectively. In §6.2 we show the following theorem.

THEOREM 2.5. *Let $d, \ell, s, t \in \mathbb{N}$. Then there exists a positive integer $k = k(d, \ell, s, t)$ such that, for every system $(X, \mu, T_1, \ldots, T_\ell)$, functions $f_0, \ldots, f_s \in L^\infty(\mu)$ and polynomial mappings $\vec{p_i} \colon \mathbb{N}^d \to \mathbb{Z}^\ell$, $i = 1, \ldots, s$, of degree at most t, the sequence $a \colon \mathbb{N}^d \to \mathbb{C}$, given by*

$$a(\mathbf{n}) := \int f_0 \cdot T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s \, d\mu, \quad \mathbf{n} \in \mathbb{N}^d, \tag{2.8}$$

is an approximate k-step nilsequence in d variables.

Furthermore, if the polynomial mappings $\overrightarrow{p_1}, \ldots, \overrightarrow{p_s}$ are linear, then we can take $k = s$.

Remark. For $d = 1$, this result was proved in [**29**].

The degree of nilpotency provided in the last part of Theorem 2.5 is not always optimal. In §6.3, we establish the following improvement for particular correlation sequences.

THEOREM 2.6. *For $d, \ell \in \mathbb{N}$, let $(X, \mu, T_1, \ldots, T_\ell)$ be a system, $f_0, \ldots, f_\ell \in L^\infty(\mu)$ be functions and $L_1, \ldots, L_\ell : \mathbb{N}^d \to \mathbb{Z}$ be linear forms spanning a space of dimension r. Then the sequence $a : \mathbb{N}^d \to \mathbb{C}$, given by*

$$a(\mathbf{n}) := \int f_0 \cdot T_1^{L_1(\mathbf{n})} f_1 \cdot \cdots \cdot T_\ell^{L_\ell(\mathbf{n})} f_\ell \, d\mu, \quad \mathbf{n} \in \mathbb{N}^d,$$

is an approximate $(\ell - r + 1)$-step nilsequence in d variables.

Remark. Examples of sequences for which this theorem applies and gives the optimal degree of nilpotency are the three sequences in (1.4).

A crucial ingredient in the proof of the previous two decomposition results is Theorem 3.10, which gives a characterization involving uniformity seminorms of approximate nilsequences. Furthermore, the proof of Theorem 2.6 uses a structural result for the generalized Kronecker factor of a not necessarily ergodic system that is of independent interest (see Theorem 5.2).

Lastly, we give an interesting corollary of Theorem 2.5. For $d \in \mathbb{N}$, we consider various subsets of $\ell^\infty(\mathbb{N}^d)$. The first is the set

$$\mathcal{N}_d := \{(\psi(\mathbf{n}))_{\mathbf{n} \in \mathbb{N}^d} : \psi \text{ is a nilsequence in } d \text{ variables}\}.$$

With $\mathcal{MC}_{\text{d,pol}}$ we denote the set that contains all sequences of the form

$$\left(\int f_0 \cdot T_{\overrightarrow{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\overrightarrow{p_s}(\mathbf{n})} f_s \, d\mu \right)_{\mathbf{n} \in \mathbb{N}^d}$$

for arbitrary systems $(X, \mu, T_1, \ldots, T_\ell)$, functions $f_0, \ldots, f_s \in L^\infty(\mu)$, polynomial mappings $\overrightarrow{p_1}, \ldots, \overrightarrow{p_s} : \mathbb{Z}^d \to \mathbb{Z}^\ell$ and $\ell, s \in \mathbb{N}$.

We also denote by $\mathcal{MC}_{\text{d,lin}}$ the set of multiple correlation sequences defined, as above, using linear polynomial mappings only.

THEOREM 2.7. *For every $d \in \mathbb{N}$, the sets $\mathcal{N}_d$, $\mathcal{MC}_{\text{d,lin}}$, $\mathcal{MC}_{\text{d,pol}}$ are subspaces of $\ell^\infty(\mathbb{N}^d)$ and*

$$\overline{\mathcal{N}_d} = \overline{\mathcal{MC}_{\text{d,lin}}} = \overline{\mathcal{MC}_{\text{d,pol}}},$$

where the closure is taken with respect to the seminorm $\|\cdot\|_2$, which was defined in (2.7).

2.5. *Multiple correlations for sequences in $\mathbb{N}^d$ and $\mathbb{Z}_N^d$.* We use the decomposition results of the previous subsection in order to deduce similar results for multiple correlations of bounded sequences in $\mathbb{N}^d$.

Definition. Let $\mathcal{A}$ be a finite collection of bounded complex valued sequences in ℓ variables and $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ be a Følner sequence in $\mathbb{N}^\ell$. We say that the collection $\mathcal{A}$ *admits correlations along* $\mathbf{I}$ if, for every $s \in \mathbb{N}$ and all $\mathbf{h}_1, \ldots, \mathbf{h}_s \in \mathbb{N}^\ell$, the limit

$$\lim \mathrm{Av}_{\mathbf{k},\mathbf{I}} \prod_{j=1}^{s} b_j(\mathbf{k} + \mathbf{h}_j)$$

exists, where, for $j = 1, \ldots, s$, the sequence b_j or the sequence $\overline{b_j}$ belongs to $\mathcal{A}$.

Combining Theorem 2.5 with the correspondence principle stated in Proposition 6.4 below, we deduce the following statement.

THEOREM 2.8. *Let $d, \ell, s, t \in \mathbb{N}$. Then there exists a positive integer $k = k(d, \ell, s, t)$ such that the following holds: if $a_1, \ldots, a_s : \mathbb{Z}^\ell \to \mathbb{C}$ are bounded sequences admitting correlations along a Følner sequence $\mathbf{I}$ in $\mathbb{N}^\ell$ and $\overrightarrow{p_i} : \mathbb{N}^d \to \mathbb{Z}^\ell$, $i = 1, \ldots, s$ are polynomial mappings of degree at most t, then the sequence $b : \mathbb{N}^d \to \mathbb{C}$, defined by*

$$b(\mathbf{n}) := \lim \mathrm{Av}_{\mathbf{k},\mathbf{I}} \prod_{i=1}^{s} a_i(\mathbf{k} + \overrightarrow{p_i}(\mathbf{n})), \quad \mathbf{n} \in \mathbb{N}^d,$$

is an approximate k-step nilsequence in d variables.

Moreover, if the polynomial mappings are linear, then we can take $k = s - 1$.

If A is a finite set, we let $\mathbb{E}_{n\in A} := (1/|A|)\sum_{n\in A}$. Decomposition results of a similar nature also hold in the finite world. For example, the following theorem is true.

THEOREM 2.9. *Let $d, \ell, s, t \in \mathbb{N}$. Then there exists a positive integer $k = k(d, \ell, s, t)$ such that the following holds: for every $\varepsilon > 0$ there exists a k-step nilmanifold $X = X(d, \ell, s, t, \varepsilon)$ such that, for every $N \in \mathbb{N}$, finite sequences $a_1, \ldots, a_s : \mathbb{Z}_N^\ell \to \mathbb{C}$ of modulus at most one and polynomial mappings $\overrightarrow{p_i} : \mathbb{N}^d \to \mathbb{Z}^\ell$, $i = 1, \ldots, s$, having integer coefficients and degree at most t, the sequence $b : \mathbb{N}^d \to \mathbb{C}$, defined by*

$$b(\mathbf{n}) := \mathbb{E}_{\mathbf{k}\in\mathbb{Z}_N^\ell} \prod_{i=1}^{s} a_i\big(\mathbf{k} + \overrightarrow{p_i}(\mathbf{n})\big), \quad \mathbf{n} \in \mathbb{Z}_N^d, \tag{2.9}$$

admits a decomposition of the form $b = b_{\mathrm{st}} + b_{\mathrm{er}}$ where:

(i) *$b_{\mathrm{st}} : \mathbb{N}^d \to \mathbb{C}$ is a convex combination of k-step nilsequences defined by functions on X with Lipschitz norm at most one; and*

(ii) $\mathbb{E}_{\mathbf{n}\in\mathbb{Z}_N^d} |b_{\mathrm{er}}(\mathbf{n})| \le \varepsilon$.

Furthermore, if the polynomial mappings are linear, then we can take $k = s - 1$.

Remark. It is important that the nilmanifold X and the Lipschitz norm of the function defining the nilsequence are independent of $N \in \mathbb{N}$.

2.6. *Applications to arithmetic.* Next we give some applications with number theoretic and combinatorial flavor.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

Definition. A function $\phi\colon \mathbb{N}\to\mathbb{C}$ is called *multiplicative* if

$$\phi(mn)=\phi(m)\,\phi(n)\quad \text{whenever } (m,n)=1.$$

It is called *completely multiplicative* if this relation holds for all $m, n\in\mathbb{N}$.

We say that a multiplicative function $\phi\colon \mathbb{N}\to\mathbb{C}$ that is bounded by one is *good* if the limit

$$\lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}\phi(an+b) \tag{2.10}$$

exists for all $a\in\mathbb{N}$, $b\in\mathbb{Z}_+$. It is called *aperiodic* if all these limits are equal to zero.

For $d\in\mathbb{N}$, a function $\phi\colon\mathbb{N}^d\to\mathbb{C}$ is called *multiplicative* if it is of the form

$$\phi(n_1,\ldots,n_d)=\phi_1(n_1)\cdots\phi_d(n_d),\qquad n_1,\ldots,n_d\in\mathbb{N}$$

for some multiplicative functions $\phi_i\colon\mathbb{N}\to\mathbb{C}$, $i=1,\ldots,d$, which we call the *components of* ϕ. We call a multiplicative function $\phi\colon\mathbb{N}^d\to\mathbb{C}$ *good* if all its component functions are good and *aperiodic* if at least one of its component functions is aperiodic.

By a classical result of Wirsing [**65**], every real-valued multiplicative function that is bounded by one is good. A result of Halász [**42**] allows us to characterize good and aperiodic multiplicative functions. Let $\mathbb{P}$ be the set of primes. A Dirichlet character is a periodic completely multiplicative function which takes the value one at one.

Notation. [**36**] If $\phi_1,\phi_2\colon\mathbb{N}\to\mathbb{C}$ are multiplicative functions, bounded by one, we define $\mathbb{D}(\phi_1,\phi_2)\in[0,+\infty]$ by

$$\mathbb{D}(\phi_1,\phi_2)^2:=\sum_{p\in\mathbb{P}}\frac{1}{p}\,(1-\Re(\phi_1(p)\overline{\phi_2(p)})).$$

Remark. Note that if $|\phi_1|=|\phi_2|=1$, then $\mathbb{D}(\phi_1,\phi_2)^2=\sum_{p\in\mathbb{P}}(1/2p)\,|\phi_1(p)-\phi_2(p)|^2$.

The next result can be deduced from [**27**, Theorem 6.3].

THEOREM 2.10. *Let $\phi\colon\mathbb{N}\to\mathbb{C}$ be a multiplicative function that is bounded by* 1. *Then ϕ is good if and only if, for every Dirichlet character χ, we either have:*

(i) $\mathbb{D}(\phi\,\chi,n^{it})=+\infty$ *for every $t\in\mathbb{R}$; or*

(ii) *for some $t\in\mathbb{R}$, $\mathbb{D}(\phi\chi,n^{it})<\infty$ and $\chi(2)^k\phi(2^k)=-2^{ikt}$ for all $k\in\mathbb{N}$; or*

(iii) $\sum_{p\in\mathbb{P}}1/p(1-\phi(p)\chi(p))$ *converges.*

Moreover, ϕ is aperiodic if and only if either the condition (i) or (ii) is satisfied for every Dirichlet character χ.

For a more complete discussion of these notions, see [**31**, §2.5]. The next result is proved in §8.1.

THEOREM 2.11. *Let $d\in\mathbb{N}$ and $\phi\colon\mathbb{N}^d\to\mathbb{C}$ be a good multiplicative function. Then,for every $\ell,s\in\mathbb{N}$, system $(X,\mu,T_1,\ldots,T_\ell)$, functions $f_1,\ldots,f_s\in L^\infty(\mu)$ and polynomial mappings $\vec{p_i}\colon\mathbb{N}^d\to\mathbb{Z}^\ell$, $i=1,\ldots,s$, the limit*

$$\lim_{N\to+\infty}\frac{1}{N^d}\sum_{\mathbf{n}\in[N]^d}\phi(\mathbf{n})\cdot T_{\vec{p_1}(\mathbf{n})}f_1\cdot\cdots\cdot T_{\vec{p_s}(\mathbf{n})}f_s\quad \textit{exists in } L^2(\mu). \tag{2.11}$$

Furthermore, if the multiplicative function ϕ is aperiodic, then the limit is equal to zero.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

Remarks.

- This result and its consequences below were proved in [**31**] for $d = 1$ using a deep structural result for multiplicative functions from [**30**]. The current argument relies on the convergent criterion of Theorem 2.4 and uses much softer number theoretic input (we only use Theorem 8.1).
- Conversely, if for $d = \ell = s = 1$ the averages in (2.11) converge weakly, then examples of periodic systems show that ϕ has to be good, and if the averages in (2.11) converge weakly to zero, then ϕ has to be aperiodic.
- Similar statements, with similar proofs, hold if, in (2.11), we use averages of the form $(1/(N_1 \cdots N_d)) \sum_{\mathbf{n}\in[N_1]\times\cdots\times[N_d]}$ and take the limit as $N_1, \ldots, N_d \to +\infty$. A similar comment applies for the next two results.

Definition. For $a \in \mathbb{Z}_+$ and $b \in \mathbb{N}$ we let $S_{a,b}$ consist of those $n \in \mathbb{N}$ whose number of distinct prime factors is congruent to $a \mod b$.

We can also define $S_{a,b}$ by counting prime factors with multiplicity; then all results stated below continue to hold with similar proofs. The next result is proved in §8.2.

THEOREM 2.12. *Let $d \in \mathbb{N}$, $a_i, c_i \in \mathbb{Z}_+$, $b_i \in \mathbb{N}$, $i = 1, \ldots, d$ and let*

$$S := (S_{a_1,b_1} + c_1) \times \cdots \times (S_{a_d,b_d} + c_d). \tag{2.12}$$

Then, for all $\ell, s \in \mathbb{N}$, polynomial mappings $\vec{p_1}, \ldots, \vec{p_s} : \mathbb{N}^d \to \mathbb{Z}^\ell$, system $(X, \mu, T_1, \ldots, T_\ell)$ and functions $f_1, \ldots, f_s \in L^\infty(\mu)$, the limit

$$\frac{1}{|S \cap [N]^d|} \sum_{\mathbf{n}\in S\cap[N]^d} T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s$$

exists in $L^2(\mu)$ and is equal to the limit obtained when one replaces S with $\mathbb{N}^d$.

Remark. It follows, from our argument, that $\lim_{N\to+\infty} |S \cap [N]^d|/N^d = (\prod_{i=1}^d b_i)^{-1}$.

In §8.2, we deduce from this result the following multiple recurrence statement.

THEOREM 2.13. *We use the notation of Theorem 2.12 and assume, in addition, that $\vec{p_i}(\mathbf{0}) = \vec{0}$ for $i = 1, \ldots, s$. Then, for S as in (2.12), for every $A \in \mathcal{X}$ with $\mu(A) > 0$,*

$$\lim_{N\to+\infty} \frac{1}{|S \cap [N]^d|} \sum_{\mathbf{n}\in S\cap[N]^d} \mu(A \cap T_{-\vec{p_1}(\mathbf{n})} A \cap \cdots \cap T_{-\vec{p_s}(\mathbf{n})} A) > 0.$$

Lastly, we give some combinatorial implications of the previous multiple recurrence result. We define the *upper Banach density* $d^*(E)$ of a set $E \subset \mathbb{Z}^\ell$ by $d^*(E) := \limsup_{|I|\to+\infty} |E \cap I|/|I|$, where the lim sup is taken over all parallelepipeds $I \subset \mathbb{Z}^\ell$ whose side lengths tend to infinity, and we define the *lower natural density* as $\liminf_{N\to\infty} |E \cap [-N, N]^\ell|/(2N + 1)^\ell$. Using a modification of the correspondence principle of Furstenberg ([**33**], proved as in [**12**]), we deduce, from Theorem 2.13, the following result.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

THEOREM 2.14. *We use the notation of Theorem 2.12 and assume, in addition, that* $\overrightarrow{p_i}(\mathbf{0}) = \overrightarrow{0}$ *for* $i = 1, \ldots, s$. *Then, for S as in (2.12) and for every set* $E \subset \mathbb{Z}^\ell$ *with* $d^*(E) > 0$, *the set*

$$\{\mathbf{n} \in S : d^*(E \cap (E - \overrightarrow{p_1}(\mathbf{n})) \cap \cdots \cap (E - \overrightarrow{p_s}(\mathbf{n}))) > 0\}$$

has positive lower natural density.

Applying this for $d = 2$, $a_i = 2$, $b_i = 0$ or 1, $c_i = 0$ and $\overrightarrow{p_i}(n_1, n_2) = i(n_1^2 + n_2^2)$ for $i = 1, \ldots, s$, we obtain the refinement of Szemerédi's theorem, which was mentioned towards the end of the introduction.

2.7. *Open problems.* When all the maps $T_1, \ldots, T_\ell$ are powers of the same transformation and $d = 1$, a strengthening of Theorem 2.5 holds which shows that the error term can be taken to converge to zero in uniform density (see [**11**, **53–55**]). It is not clear whether a similar result holds for arbitrary commuting transformations, even when $d = 1$, $\ell = 2$, and the polynomial mappings are linear.

Problem 1. Let (X, μ, T_1, T_2) be a system and $f_0, f_1, f_2 \in L^\infty(\mu)$. Is it true that the sequence $a : \mathbb{N} \to \mathbb{C}$, defined by

$$a(n) := \int f_0 \cdot T_1^n f_1 \cdot T_2^n f_2 \, d\mu, \quad n \in \mathbb{N},$$

can be decomposed as $a = a_{\mathrm{st}} + a_{\mathrm{er}}$, where a_{st} is a uniform limit of 2-step nilsequences and $\|a_{\mathrm{er}}\|_2 = 0$?

When $T_2 = T_1^2$ this is shown to be the case in [**11**, **54**, **55**].

Theorem 2.4 shows that condition (2.3) is sufficient for weak convergence of the averages in (2.6), but we needed the stronger hypothesis (2.5) in order to guarantee mean convergence. This is probably an artefact of our proof.

Problem 2. Show that condition (2.3) is sufficient for mean convergence of the averages in (2.6).

When $d = 1$ and all the maps $T_1, \ldots, T_\ell$ are powers of the same transformation, this is shown to be the case, in [**18**, Theorem 1.3].

In Theorem 2.5, even in seemingly simple cases, it is not clear what the optimal dependence of k on d, ℓ, s, t is, even when the polynomial mappings are linear. It is expected (but we are unable to verify this) that this optimal dependence can already be inferred from the case where all the $T_1, \ldots, T_\ell$ are powers of the same transformation (a case which is much more tractable using the theory of characteristic factors). We record here a relevant open problem.

Problem 3. Let (X, μ, T_1, T_2) be a system and $f, g, h \in L^\infty(\mu)$. Show that the sequence $a : \mathbb{N}^2 \to \mathbb{C}$, defined by

$$a(m, n) := \int f \cdot T_1^m T_2^n g \cdot T_1^n T_2^m h \, d\mu, \quad m, n \in \mathbb{N},$$

is an approximate 1-step nilsequence in two variables.

When T_1, T_2 are powers of the same transformation, this can be verified by an argument similar to the one used in the proof of Theorem 2.6. Note, also, that Theorem 2.5 gives that the sequence a is an approximate 2-step nilsequence in two variables.

2.8. *Notation and conventions.* For the convenience of the reader, we gather here some notation used throughout the article.

- We denote the set of positive integers by $\mathbb{N}$, the set of non-negative integers by $\mathbb{Z}_+$ and the set of non-negative real numbers by $\mathbb{R}_+$.
- For $N \in \mathbb{N}$, we denote the set $\{1, \ldots, N\}$ by $[N]$.
- With $\ell^\infty(\mathbb{N}^d)$, we denote the space of all bounded sequences $a\colon \mathbb{N}^d \to \mathbb{C}$.
- If A is a finite set, we let $\mathbb{E}_{n\in A} := (1/|A|) \sum_{n\in A}$.
- We write $\mathcal{C}\colon \mathbb{C} \to \mathbb{C}$ for the complex conjugation.
- If x is a real, $\mathrm{e}(x)$ denotes the number $e^{2\pi i x}$ and $\|x\|$ denotes the distance between x and the nearest integer.
- Given $d \in \mathbb{N}$, we write $\mathbf{n} = (n_1, \ldots, n_d)$ for a point of $\mathbb{Z}^d$.
- We typically use the letter ψ to denote nilsequences.

3. *Uniformity seminorms and decomposition results*

In this section, we extend to sequences in $\ell^\infty(\mathbb{N}^d)$ some results established in [**46**] and [**29**] for sequences in $\ell^\infty(\mathbb{N})$. The statements and the proofs are analogous and we only give the necessary definitions and sketch the main steps of the proofs.

Some definitions and notation. We write $\mathcal{C}\colon \mathbb{C} \to \mathbb{C}$ for the complex conjugation; then $\mathcal{C}^k z = z$, if k is even, and $\mathcal{C}^k z = \overline{z}$, if k is odd. We let $[\![k]\!] := \{0, 1\}^k$ and $[\![k^*]\!] := [\![k]\!] \setminus \{\mathbf{0}\}$. Elements of $[\![k]\!]$ are written as $\underline{\epsilon} = (\epsilon_1, \ldots, \epsilon_k)$. We let $|\underline{\epsilon}| := \epsilon_1 + \cdots + \epsilon_k$. Elements of $(\mathbb{N}^d)^k$ are written as $\underline{\mathbf{h}} = (\mathbf{h}_1, \ldots, \mathbf{h}_k)$, where $\mathbf{h}_i \in \mathbb{N}^d$ for $i = 1, \ldots, k$. For $\underline{\mathbf{h}} \in (\mathbb{N}^d)^k$ and $\underline{\epsilon} \in [\![k]\!]$, we let $\underline{\epsilon} \cdot \underline{\mathbf{h}} := \epsilon_1\mathbf{h}_1 + \cdots + \epsilon_k\mathbf{h}_k \in \mathbb{N}^d$.

3.1. *The definition of the seminorms.* We follow [**46**, §2].

We say that a finite or countable family $\mathcal{F}$ of bounded sequences in $\ell^\infty(\mathbb{N}^d)$ *admits correlations* along a Følner sequence $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ in $\mathbb{N}^d$ if the limit

$$\lim \mathrm{Av}_{\mathbf{n},\mathbf{I}} \left(\prod_{i=1}^{m} b_i(\mathbf{n} + \mathbf{h}_i) \right)$$

exists for every $m \in \mathbb{N}$, all $\mathbf{h}_1, \ldots, \mathbf{h}_m \in \mathbb{N}^d$ and all sequences $b_1, \ldots, b_m \in \ell^\infty(\mathbb{N}^d)$ such that either b_i or $\overline{b_i}$ belongs to $\mathcal{F}$ for $i = 1, \ldots, m$. We remark that, from every Følner sequence $\mathbf{I}$, we can extract a subsequence $\mathbf{I}'$ so that a given family of sequences admits correlations along $\mathbf{I}'$.

Suppose that the sequence $a \in \ell^\infty(\mathbb{N}^d)$ admits correlations along $\mathbf{I}$. Then, for $k \in \mathbb{N}$ and $\underline{\mathbf{h}} = (\mathbf{h}_1, \ldots, \mathbf{h}_k) \in (\mathbb{N}^d)^k$, we write

$$\mathrm{Corr}_{\mathbf{I}}(a; \underline{\mathbf{h}}) := \lim \mathrm{Av}_{\mathbf{n},\mathbf{I}} \left(\prod_{\underline{\epsilon}\in[\![k]\!]} \mathcal{C}^{|\underline{\epsilon}|} a(\mathbf{n} + \underline{\epsilon} \cdot \underline{\mathbf{h}}) \right)$$

and

$$\|a\|_{\mathbf{I},k} := \left(\lim_{H\to+\infty} \frac{1}{H^{dk}} \sum_{\mathbf{h}_1,\ldots,\mathbf{h}_k\in[H]^d} \mathrm{Corr}_{\mathbf{I}}(a; \underline{\mathbf{h}}) \right)^{1/2^k}.$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

In [**46**, Proposition 2.4] it is shown that, for $d = 1$, the previous limit exists and is non-negative; the proof is similar for general $d \in \mathbb{N}$. Furthermore, the map $a \mapsto \|a\|_{\mathbf{I},k}$ is subadditive ([**46**, Proposition 2.5] for $d = 1$): that is, if the sequences a, b, and $a + b$ admit correlations along $\mathbf{I}$, then $\|a + b\|_{\mathbf{I},k} \le \|a\|_{\mathbf{I},k} + \|b\|_{\mathbf{I},k}$.

For $a \in \ell^\infty(\mathbb{N}^d)$, we define

$$\|a\|_{U^k(\mathbb{N}^d)} := \sup_{\mathbf{I}} \|a\|_{\mathbf{I},k},$$

where the supremum is taken over all Følner sequences $\mathbf{I}$ in $\mathbb{N}^d$ for which the sequence a admits correlations. Then the map $a \mapsto \|a\|_{U^k(\mathbb{N}^d)}$ is a seminorm on $\ell^\infty(\mathbb{N}^d)$; we call it the *uniformity seminorm of order k* of a.

3.2. *Interpretation.* Next, we interpret the previous definitions and results in dynamical terms. We use a variant of Furstenberg's correspondence principle that enables to transfer results from ergodic theory to results about bounded sequences of complex numbers. We follow the method used in [**46**, §6.1] when $d = 1$; similar arguments work for general $d \in \mathbb{N}$ and we summarize them here. For notational convenience, we restrict to the case where the family $\mathcal{F}$ contains only a single sequence $a\colon \mathbb{N}^d \to \mathbb{C}$; the general case being completely similar. Let D be the closed disk in $\mathbb{C}$ of radius $\|a\|_\infty$ and let $D^{\mathbb{Z}^d}$ be endowed with the product topology and with the natural shifts $T_1, \ldots, T_d$ given by

$$(T_i x)(n_1, \ldots, n_d) = x(n_1, \ldots, n_{i-1}, n_i + 1, n_{i+1}, \ldots, n_d)$$

for $i = 1, \ldots, d$, where we use the notation $\mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{Z}^d$ and $x = (x(\mathbf{n}))_{\mathbf{n} \in \mathbb{Z}^d} \in D^{\mathbb{Z}^d}$. We define the continuous function $f\colon D^{\mathbb{Z}^d} \to \mathbb{C}$ by $f(x) := x(\mathbf{0})$. Furthermore, we define the point ω in $D^{\mathbb{Z}^d}$ by $\omega(\mathbf{n}) := a(\mathbf{n})$ for $\mathbf{n} \in \mathbb{N}^d$ and $\omega(\mathbf{n}) = 0$ otherwise. Then

$$f(T_{\mathbf{n}}\omega) = a(\mathbf{n}), \qquad \mathbf{n} \in \mathbb{N}^d.$$

Let X be the closed orbit of ω under $T_1, \ldots, T_d$. Then $(X, T_1, \ldots, T_d)$ is a topological dynamical system and ω is a transitive point of this system, meaning that it has a dense orbit in X.

Let $\mathbf{I} = (I_j)_{j \in \mathbb{N}}$ be a Følner sequence in $\mathbb{N}^d$ for which the sequence a admits correlations. Let μ be a w^*-limit point for the sequence of measures

$$\mu_j := \frac{1}{|I_j|} \sum_{\mathbf{n} \in I_j} \delta_{T_{\mathbf{n}}\omega}, \qquad j \in \mathbb{N}.$$

Then μ is a probability measure on X, invariant under $T_1, \ldots, T_d$ and, by construction, for every $m \in \mathbb{N}$, all $\eta_1, \ldots, \eta_m \in \{0, 1\}$ and all $\mathbf{h}_1, \ldots, \mathbf{h}_m \in \mathbb{Z}^d$,

$$\lim \mathrm{Av}_{\mathbf{n},\mathbf{I}} \left(\prod_{i=1}^m \mathcal{C}^{\eta_i} a(\mathbf{n} + \mathbf{h}_i) \right) = \lim \mathrm{Av}_{\mathbf{n},\mathbf{I}} \left(\prod_{i=1}^m \mathcal{C}^{\eta_i} T_{\mathbf{h}_i} f(T_{\mathbf{n}}\omega) \right) = \int \prod_{i=1}^m \mathcal{C}^{\eta_i} T_{\mathbf{h}_i} f \, d\mu.$$

In particular, for $\underline{\mathbf{h}} = (\mathbf{h}_1, \ldots, \mathbf{h}_k) \in (\mathbb{N}^d)^k$,

$$\mathrm{Corr}_{\mathbf{I}}(a; \underline{\mathbf{h}}) = \int \prod_{\underline{\epsilon} \in [\![k]\!]} \mathcal{C}^{|\underline{\epsilon}|} T_{\underline{\epsilon} \cdot \underline{\mathbf{h}}} f \, d\mu$$

and thus

$$\|a\|_{\mathbf{I},k} = |||f|||_{\mu,k},$$

where $|||\cdot|||_{\mu,k}$ is the seminorm on $L^\infty(\mu)$ defined in [**44**] in the ergodic case and in [**19**] in the general case†. We recall the definition and some properties of these seminorms in Appendix A. Note, also, that if $\psi : \mathbb{N}^d \to \mathbb{C}$ is a nilsequence of the form $(\Phi(T_{\mathbf{n}}x))_{\mathbf{n}\in\mathbb{N}^d}$, then ψ admits correlations along every Følner sequence $\mathbf{I}$ and

$$\|\psi\|_{\mathbf{I},k} = \|\psi\|_{U^k(\mathbb{N}^d)} = \|\Phi\|_{\mu,k}, \tag{3.1}$$

where the last seminorm is defined with respect to the action of $T_{\mathbf{n}}$, $\mathbf{n} \in \mathbb{N}^d$, on X. From the properties (A.1) and (A.4), we deduce

$$\limsup |\mathrm{Av}_{\mathbf{I}}\, a(\mathbf{n})| \le \|a\|_{\mathbf{I},1}, \tag{3.2}$$

$$\|a\|_{\mathbf{I},k+1}^{2^k+1} = \lim_{H\to+\infty} \frac{1}{H^d} \sum_{\mathbf{h}\in[H]^d} \|\sigma_{\mathbf{h}}a \cdot \overline{a}\|_{\mathbf{I},k}^{2^k}, \quad k \in \mathbb{N}, \tag{3.3}$$

where $\sigma_{\mathbf{h}}a(\mathbf{n}) := a(\mathbf{n}+\mathbf{h})$ for $\mathbf{h}, \mathbf{n} \in \mathbb{N}^d$.

Then (see [**46**, Proposition 4.5] for $d = 1$)

$$\begin{aligned}\|a\|_{U^k(\mathbb{N}^d)} &= \sup_{\mu \text{ invariant probability measure on } X} |||f|||_{\mu,k} \\ &= \sup_{\mu \text{ invariant ergodic probability measure on } X} |||f|||_{\mu,k},\end{aligned} \tag{3.4}$$

where the last equality follows by using the ergodic decomposition of the measure μ.

3.3. *Tools.* For an ergodic system (X, μ, T), the structure theorem of [**44**] links the seminorms $|||\cdot|||_{\mu,k}$ with the factors of the system (X, μ, T) that are $(k-1)$-step nilsystems. This result was generalized to $\mathbb{Z}^d$-actions by Griesmer [**41**, Lemma 4.4.3 and Theorem 4.10.1] and an alternate proof, based on finitestic inverse theorems, was recently given by Tao [**63**, Remark 4]. We record an immediate corollary of this result that is more convenient for our purposes.

THEOREM 3.1. ([**44**] for $d = 1$, [**41**, **63**] for general d) *Let $d, k \in \mathbb{N}$, $(X, \mu, T_1, \ldots, T_d)$ be an ergodic system, $f \in L^\infty(\mu)$ and $\varepsilon > 0$. Then there exists a $(k-1)$-step nilsystem $(Y, \nu, T_1, \ldots, T_d)$, a factor map $\pi : X \to Y$ and a continuous function Φ on Y such that*

$$\|\Phi\|_\infty \le \|f\|_\infty \quad \textit{and} \quad |||f - \Phi\circ\pi|||_{\mu,k} \le \varepsilon.$$

The next result can be considered as a strengthening of the correspondence principle of Furstenberg. We recall that a topological dynamical system $(Y, T_1, \ldots, T_d)$ is *distal* if, for all $y \ne y' \in Y$, $\inf_{\mathbf{n}\in\mathbb{Z}^d} d_Y(T_{\mathbf{n}}, y, T_{\mathbf{n}}y') > 0$, where d_Y is the distance on Y defining its topology. It is known that every nilsystem is distal [**6**].

PROPOSITION 3.2. ([**46**, Proposition 6.1] for $d = 1$) *Let $d \in \mathbb{N}$, $(X, T_1, \ldots, T_d)$ be a topological dynamical system, $\omega \in X$ be a transitive point and μ be an invariant ergodic*

† The seminorms were defined for a single transformation but the properties we use extend immediately to the case of several commuting transformations.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

measure on X. Moreover, let $(Y, T_1, \dots, T_d)$ be a distal topological dynamical system, ν be an invariant measure on Y and $\pi : X \to Y$ be a measure theoretic factor map. Then there exists a point $y_0 \in Y$ and a Følner sequence $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ on $\mathbb{N}^d$ such that

$$\lim_{j\to+\infty} \frac{1}{|I_j|} \sum_{\mathbf{n}\in I_j} f(T_{\mathbf{n}}\omega)\, g(T_{\mathbf{n}} y_0) = \int_X f \cdot g \circ \pi \, d\mu$$

for every $f \in C(X)$ and every $g \in C(Y)$.

The important point in this statement is that we do not assume that the map π is continuous. The proof is exactly the same as in the case of a single transformation. The following theorem states the classical property of distal systems that we use.

THEOREM 3.3. [**5**, Ch. 5] *Let $d \in \mathbb{N}$ and $(Y, T_1, \dots, T_d)$ be a distal system. Then, for every $y_1 \in Y$ and every sequence $(\mathbf{m}_i)_{i\in\mathbb{N}}$ with values in $\mathbb{N}^d$, there exists $y_0 \in Y$ and a subsequence $(\mathbf{m}'_i)_{i\in\mathbb{N}}$ of $(\mathbf{m}_i)_{i\in\mathbb{N}}$ such that $T_{\mathbf{m}'_i} y_0$ converges to y_1.*

From Theorem 3.1 and the discussion of §3.1 we deduce the following proposition.

PROPOSITION 3.4. [**46**, Proposition 6.2 for $d = 1$] *Let $d, k \in \mathbb{N}$, $a \in \ell^\infty(\mathbb{N}^d)$ be a sequence and let $\varepsilon > 0$. Then there exists a Følner sequence $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ and a $(k-1)$-step nilsequence ψ_1 in d variables such that the sequences a and $a - \psi_1$ admit correlations along $\mathbf{I}$ and*

$$\|a\|_{\mathbf{I},k} \ge \|a\|_{U^k(\mathbb{N}^d)} - \varepsilon, \quad \|\psi_1\|_\infty \le \|a\|_\infty \quad \text{and} \quad \|a - \psi_1\|_{\mathbf{I},k} \le \varepsilon.$$

Proof. As explained above, there exists a system $(X, T_1, \dots, T_d)$, a transitive point $\omega \in X$ and a continuous function f on X such that $a(\mathbf{n}) = f(T_{\mathbf{n}}\omega)$ for every $\mathbf{n} \in \mathbb{N}^d$. Note that then $\|a\|_\infty = \|f\|_\infty$. Moreover, by (3.4), there exists an invariant ergodic probability measure μ on X with $|||f|||_{\mu,k} \ge \|a\|_{U^k(\mathbb{N}^d)} - \varepsilon$.

Let the nilsystem $(Y, \nu, T_1, \dots, T_d)$, the factor map π and the function Φ be defined as in Theorem 3.1. Let σ be the measure on $X \times Y$ which is the image of μ under the map $\mathrm{id} \times \pi$. This measure is ergodic under the product action and thus admits a generic point (x_1, y_1). Since ω is a transitive point of X, there exists a sequence $(\mathbf{m}_j)_{j\in\mathbb{N}}$ in $\mathbb{N}^d$ such that $T_{\mathbf{m}_j}\omega \to x_1$. By Theorem 3.3, substituting a subsequence for the sequence $(\mathbf{m}_j)_{j\in\mathbb{N}}$, we can assume that there exists a point $y_0 \in Y$ such that $T_{\mathbf{m}_j} y_0 \to y_1$ and thus $(T \times T)_{\mathbf{m}_j}(\omega, x_0) \to (x_1, y_1)$. Since the point (x_1, y_1) is generic, substituting, again, a subsequence for $(\mathbf{m}_j)_{j\in\mathbb{N}}$ and defining the Følner sequence $\mathbf{I}$ by $I_j = \mathbf{m}_j + [j]^d$, $j \in \mathbb{N}$, we obtain that the sequence of probability measures

$$\frac{1}{|I_j|} \sum_{\mathbf{n}\in I_j} \delta_{T_{\mathbf{n}}\omega} \times \delta_{T_{\mathbf{n}} y_0}, \qquad j \in \mathbb{N}$$

on $X \times Y$ converges weak* to a probability measure σ. Let the nilsequence ψ_1 be defined by $\psi_1(\mathbf{n}) := \Phi(T_{\mathbf{n}} y_0)$, $\mathbf{n} \in \mathbb{N}^d$. We have $\|\psi_1\|_\infty \le \|\Phi\|_\infty \le \|f\|_\infty = \|a\|_\infty$. By applying the preceding discussion to the product system on $X \times Y$, for the point (ω, y_0) (the Følner sequence $\mathbf{I}$ and the function given by $F(x, y) = f(x)$), we obtain $\|a\|_{\mathbf{I},k} = |||F|||_{\sigma,k}$. Letting $G(x, y) = f(x) - \Phi(y)$, gives $\|a - \psi_1\|_{\mathbf{I},k} = |||G|||_{\sigma,k}$. By the definition of σ and of the seminorms $|||\cdot|||_k$, we have $|||F|||_{\sigma,k} = |||f|||_{\mu,k}$ and, by the definition of Φ, $|||G|||_{\sigma,k} = |||f - \Phi \circ \pi|||_{\mu,k} \le \varepsilon$. This completes the proof. □

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

3.4. *Anti-uniformity.* We introduce certain classes of sequences that are asymptotically approximately orthogonal to k-uniform sequences; in Theorem 3.1 we give a characterization of such sequences in terms of $(k-1)$-step nilsequences.

Definition. Let $a \in \ell^\infty(\mathbb{N}^d)$.

- We say that the sequence a is *strongly k-anti-uniform* if there exists a constant $C \geq 0$ such that, for every $b \in \ell^\infty(\mathbb{N}^d)$,
$$\limsup |\text{Av}\, a(\mathbf{n})\, b(\mathbf{n})| \leq C \|b\|_{U^k(\mathbb{N}^d)}. \tag{3.5}$$
In this case, we write $\|a\|^*_{U^k(\mathbb{N}^d)}$ for the smallest constant C such that (3.5) holds.
- We say that the sequence a is *k-anti-uniform* if, for every $\varepsilon > 0$, there exists $C = C(\varepsilon) \geq 0$ such that, for every $b \in \ell^\infty(\mathbb{N}^d)$,
$$\limsup |\text{Av}\, a(\mathbf{n})\, b(\mathbf{n})| \leq C \|b\|_{U^k(\mathbb{N}^d)} + \varepsilon \|b\|_\infty.$$

PROPOSITION 3.5. ([**46**, §5] for $d = 1$) *Let $d, k \in \mathbb{N}$, $(X, T_1, \ldots, T_d)$ be an ergodic $(k-1)$-step nilsystem and let $f_{\underline{\epsilon}} \in C(X)$ for $\underline{\epsilon} \in [\![k^*]\!]$. Then the limit*
$$\Phi(x) := \lim_{H \to +\infty} \frac{1}{H^{dk}} \sum_{\mathbf{h}_1, \ldots, \mathbf{h}_k \in [0,H)^d} \prod_{\underline{\epsilon} \in [\![k^*]\!]} f_{\underline{\epsilon}}(T_{\underline{\epsilon} \cdot \underline{\mathbf{h}}}\, x) \tag{3.6}$$
exists for every $x \in X$ and the convergence is uniform in $x \in X$ (and hence $\Phi \in C(X)$). If $x_0 \in X$ and $a \in \ell^\infty(\mathbb{N}^d)$ is the sequence defined by
$$a(\mathbf{n}) := \Phi(T_{\mathbf{n}} x_0), \quad \mathbf{n} \in \mathbb{N}^d,$$
then a is strongly k-anti-uniform and
$$\|a\|^*_{U^k(\mathbb{N}^d)} \leq \prod_{\epsilon \in [\![k^*]\!]} |\!|\!| f_\epsilon |\!|\!|_k \leq \prod_{\epsilon \in [\![k^*]\!]} \|f_\epsilon\|_\infty.$$

Sketch of the proof. The first part of the result is proved in [**46**, Corollary 5.2].

To prove the second part, let $b \in \ell^\infty(\mathbb{N}^d)$ and $\mathbf{I} = (I_j)_{j \in \mathbb{N}}$ be a Følner sequence such that b admits correlations along $\mathbf{I}$. We use (3.6) for $x := T_{\mathbf{n}} x_0$, take the averages for $\mathbf{n} \in I_j$ and exchange the limits in j and in H; this can be achieved because of the uniform convergence in (3.6) (see [**46**, Theorem 5.4]). By an iterated use of the Cauchy–Schwarz inequality (this is estimate (12) in [**46**]), we obtain that
$$\limsup |\text{Av}_{\mathbf{I}}\, a(\mathbf{n})\, b(\mathbf{n})| \leq \|b\|_{\mathbf{I},k} \cdot \prod_{\epsilon \in [\![k^*]\!]} |\!|\!| f_\epsilon |\!|\!|_k. \tag{3.7}$$
Taking the supremum over all Følner sequences $\mathbf{I}$ in the left-hand side, we obtain the announced bound. For $d = 1$ the details can be found in [**46**, §5.4]; the proof for general $d \in \mathbb{N}$ is similar. □

Let $(X, T_1, \ldots, T_d)$ be a $(k-1)$-step nilsystem defining the nilsequence $\psi\colon \mathbb{N}^d \to \mathbb{C}$. By [**46**, Proposition 5.6] (see also [**19**, Proposition 3.2]), the linear span of the functions defined as in (3.6) is dense in $C(X)$ with the uniform norm. By Proposition 3.5, the sequence ψ is a uniform limit of strongly k-anti-uniform sequences. We deduce the following corollary.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

COROLLARY 3.6. *Every $(k-1)$-step nilsequence is k-anti-uniform.*

Remark. Alternatively, this follows by combining Propositions 4.2 and 6.1.

For $d=1$, the first statement of the next Proposition is [**46**, Theorem 2.16] and the second statement is [**29**, Theorem 2.1].

PROPOSITION 3.7. *Let $d, k \in \mathbb{N}$, $a \in \ell^\infty(\mathbb{N}^d)$ be a sequence and $\delta > 0$. Then there exists a Følner sequence $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ such that the following hold.*

(i) *There exists a $(k-1)$-step nilsequence $\psi_2 \colon \mathbb{N}^d \to \mathbb{C}$ such that*

$$\|\psi_2\|^*_{U^k(\mathbb{N}^d)} \le 1 \quad \text{and} \quad \limsup |\mathrm{Av}_{\mathbf{I}}\, a(\mathbf{n})\, \psi_2(\mathbf{n})| \ge \|a\|_{U^k(\mathbb{N}^d)} - \delta.$$

Moreover, if $b \in \ell^\infty(\mathbb{N}^d)$ is a sequence that admits correlations along a Følner sequence $\mathbf{J}$, then

$$\limsup |\mathrm{Av}_{\mathbf{J}}\, \psi_2(\mathbf{n})\, b(\mathbf{n})| \le \|b\|_{\mathbf{J},k}. \tag{3.8}$$

(ii) *If $\|a\|_\infty \le 1$, then there exists a $(k-1)$-step nilsequence $\psi_3 \colon \mathbb{N}^d \to \mathbb{C}$ such that*

$$\|\psi_3\|_\infty \le 1 \quad \text{and} \quad \limsup |\mathrm{Av}_{\mathbf{I}}\, a(\mathbf{n})\, \psi_3(\mathbf{n})| \ge \|a\|^{2^k}_{U^k(\mathbb{N}^d)} - \delta.$$

Proof. Let the $(k-1)$-step nilsequence $\psi_1 \colon \mathbb{N}^d \to \mathbb{C}$ and the Følner sequence $\mathbf{I}$ be given by Proposition 3.4 for an $\varepsilon > 0$ that will be specified later. Then $a - \psi_1$ admits correlations along $\mathbf{I}$ and

$$\|a-\psi_1\|_{\mathbf{I},k} \le \varepsilon, \quad \|\psi_1\|_{\mathbf{I},k} \ge \|a\|_{U^k(\mathbb{N}^d)} - 2\varepsilon \quad \text{and} \quad \|\psi_1\|_\infty \le \|a\|_\infty. \tag{3.9}$$

The sequence ψ_1 has the form $\psi_1(\mathbf{n}) = \Phi_1(T_{\mathbf{n}}x_0)$, $\mathbf{n} \in \mathbb{N}^d$, for some $(k-1)$-step ergodic nilsystem $(X, \mu, T_1, \dots, T_d)$, point $x_0 \in X$ and function $\Phi_1 \in C(X)$ with $\|\Phi_1\|_\infty \le \|a\|_\infty$. For $x \in X$, we define the function Φ on X as in Proposition 3.5, taking $f_{\underline{\epsilon}} := \mathcal{C}^{|\underline{\epsilon}|}\Phi_1$ for $\underline{\epsilon} \in [\![k^*]\!]$. Then $\Phi \in C(X)$ and $\int \Phi\, \Phi_1\, d\mu = |\!|\!|\Phi_1|\!|\!|^{2^k}_{\mu,k}$. Let $\psi_3 \colon \mathbb{N}^d \to \mathbb{C}$ be the nilsequence defined by

$$\psi_3(\mathbf{n}) := \Phi(T_{\mathbf{n}}x_0), \quad \mathbf{n} \in \mathbb{N}^d.$$

If we assume that $\|a\|_\infty \le 1$, then $\|\Phi_1\|_\infty \le 1$, and thus $\|\Phi\|_\infty \le 1$ and $\|\psi_3\|_\infty \le 1$. By unique ergodicity of $(X, \mu, T_1, \dots, T_d)$,

$$\begin{aligned}\lim \mathrm{Av}_{\mathbf{I}}\, \psi_1(\mathbf{n})\, \psi_3(\mathbf{n}) = \int \Phi_1\, \Phi\, d\mu = |\!|\!|\Phi_1|\!|\!|^{2^k}_{\mu,k} = \|\psi_1\|^{2^k}_{\mathbf{I},k} \quad \text{by (3.1)} \\ \ge (\|a\|_{U^k(\mathbb{N}^d)} - 2\varepsilon)^{2^k} \ge \|a\|^{2^k}_{U^k(\mathbb{N}^d)} - \delta/2,\end{aligned}$$

by (3.9), if ε is chosen sufficiently small. Moreover, since $a - \psi_1$ admits correlations along $\mathbf{I}$, by (3.7),

$$\limsup |\mathrm{Av}_{\mathbf{I}}\, \psi_3(\mathbf{n})\, (a-\psi_1)(\mathbf{n})| \le \|a-\psi_1\|_{\mathbf{I},k}\, \|\Phi_1\|^{2^k-1}_\infty \le \varepsilon \le \delta/2,$$

by (3.9), if $\varepsilon \le \delta/2$. Combining the last two estimates we get Part (ii).

We move now to the proof of Part (i). First, notice that, using Proposition 3.5 and (3.1), we get that

$$\|\psi_3\|^*_{U^k(\mathbb{N}^d)} \le |\!|\!|\Phi_1|\!|\!|^{2^k-1}_{\mu,k} = \|\psi_1\|^{2^k-1}_{U^k(\mathbb{N}^d)}.$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

Moreover, if the sequence $b \in \ell^\infty(\mathbb{N}^d)$ admits correlations along a Følner sequence $\mathbf{J}$, then (3.7) and (3.1) give

$$\limsup |\mathrm{Av}_{\mathbf{J}}\ \psi_3(\mathbf{n})\ b(\mathbf{n})| \le \|b\|_{\mathbf{J},k}\ |||\Phi_1|||_{\mu,k}^{2^k-1} = \|b\|_{\mathbf{J},k}\ \|\psi_1\|_{U^k(\mathbb{N}^d)}^{2^k-1}.$$

Defining

$$\psi_2 := (\|\psi_1\|_{U^k(\mathbb{N}^d)}^{2^k-1})^{-1} \cdot \psi_3,$$

we deduce that $\|\psi_2\|_{U^k(\mathbb{N}^d)}^* \le 1$ and estimate (3.8) is satisfied. Furthermore, as before, we get

$$\lim \mathrm{Av}_{\mathbf{I}}\ \psi_1(\mathbf{n})\ \psi_2(\mathbf{n}) = \|\psi_1\|_{U^k(\mathbb{N}^d)}^{2^k} / \|\psi_1\|_{U^k(\mathbb{N}^d)}^{2^k-1} = \|\psi_1\|_{U^k(\mathbb{N}^d)} \ge \|a\|_{U^k(\mathbb{N}^d)} - \delta/2,$$

by (3.9), if $\varepsilon \le \delta/4$. Using (3.7) and (3.1), we get

$$\limsup |\mathrm{Av}_{\mathbf{I}}\ \psi_2(\mathbf{n})\ (a - \psi_1)(\mathbf{n})| \le \|a - \psi_1\|_{\mathbf{I},k}\ \frac{|||\Phi_1|||_{\mu,k}^{2^k-1}}{\|\psi_1\|_{U^k(\mathbb{N}^d)}^{2^k-1}} = \|a - \psi_1\|_{\mathbf{I},k} \le \varepsilon \le \delta/2,$$

by (3.9), if $\varepsilon \le \delta/2$. Combining the last two estimates finishes the proof of Part (i). □

COROLLARY 3.8. *Let $d, k \in \mathbb{N}$ and $a \in \ell^\infty(\mathbb{N}^d)$ be such that the averages of $a(\mathbf{n})\ \psi(\mathbf{n})$ converge to zero for every $(k-1)$-step nilsequence $\psi : \mathbb{N}^d \to \mathbb{C}$. Then $\|a\|_{U^k(\mathbb{N}^d)} = 0$.*

3.5. *Regular sequences and their structure.* Next, we introduce certain classes of sequences for which we are able to prove the two decomposition results of this section.

Definition. Let $d \in \mathbb{N}$, $k \in \mathbb{Z}_+$. We say that a sequence $a \in \ell^\infty(\mathbb{N}^d)$ is *k-regular* if the limit

$$\lim \mathrm{Av}\ a(\mathbf{n})\ \psi(\mathbf{n})$$

exists for every k-step nilsequence $\psi : \mathbb{N}^d \to \mathbb{C}$.

Remarks.

- Every nilsequence is k-regular for every $k \in \mathbb{N}$.
- The product of two k-regular sequences may not be k-regular†.

THEOREM 3.9. ([**46**, Theorem 2.19] for $d = 1$) *Let $d, k \in \mathbb{N}$ and $a \in \ell^\infty(\mathbb{N}^d)$ be a sequence. Then the following are equivalent.*

(i) *a is $(k-1)$-regular.*

(ii) *For every $\delta > 0$, the sequence a can be written as $a = \psi + u$, where ψ is a $(k-1)$-step nilsequence with $\|\psi\|_\infty \le \|a\|_\infty$, and $\|u\|_{U^k(\mathbb{N}^d)} \le \delta$.*

Proof. The implication (ii) $\Longrightarrow$ (i) is a simple consequence of Corollary 3.6 and the fact that the product of two nilsequences is a nilsequence and thus has convergent averages.

Now, we establish the converse implication. For $\delta > 0$, let $\psi := \psi_1$ and $\mathbf{I}$ be as in Proposition 3.4, with $\varepsilon := \delta/2$. Then $\|a - \psi\|_{\mathbf{I},k} \le \delta/2$. Since $a = \psi + (a - \psi)$, it

† Let $a(n) := e(n^{k+1}\alpha)$, $n \in \mathbb{N}$, where $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and $b := \bar{a} \cdot \sum_{k=0}^{\infty} \mathbf{1}_{[2^{2k}, 2^{2k+1})}$. The sequences a, b are k-regular for every $k \in \mathbb{N}$, but the sequence $a \cdot b$ is not even 0-regular since the averages $1/N \sum_{n=1}^{N} a(n)\ b(n)$ do not converge as $N \to +\infty$.

remains to show that $\|a-\psi\|_{U^k(\mathbb{N}^d)} \le \delta$. Suppose that this is not the case. Then Part (i) of Proposition 3.7 provides a Følner sequence $\mathbf{J}$ and a $(k-1)$-step nilsequence ψ_2 with

$$\limsup |\mathrm{Av}_{\mathbf{J}}(a(\mathbf{n})-\psi(\mathbf{n}))\,\psi_2(\mathbf{n})| \ge \|a-\psi\|_{U^k(\mathbb{N}^d)} - \delta/3 \ge 2\delta/3.$$

Since a is a $(k-1)$-regular sequence and $\psi\psi_2$ is a $(k-1)$-step nilsequence, the sequence $(a-\psi)\psi_2$ has convergent averages. We deduce that

$$|\lim \mathrm{Av}_{\mathbf{I}}\,(a(\mathbf{n})-\psi(\mathbf{n}))\,\psi_2(\mathbf{n})| = |\lim \mathrm{Av}_{\mathbf{J}}\,(a(\mathbf{n})-\psi(\mathbf{n}))\,\psi_2(\mathbf{n})| \ge 2\delta/3.$$

On the other hand,

$$|\lim \mathrm{Av}_{\mathbf{I}}\,(a(\mathbf{n})-\psi(\mathbf{n}))\,\psi_2(\mathbf{n})| \le \|a-\psi\|_{\mathbf{I},k} \le \delta/2,$$

where the second estimate follows from our data and the first estimate follows from the estimate (3.8), since the sequence $a-\psi$ admits correlations along $\mathbf{I}$. Combining the above, we get a contradiction. □

3.6. *The structure of regular anti-uniform sequences.* The next result gives a characterization of regular anti-uniform sequences in $\ell^\infty(\mathbb{N}^d)$. Assuming the multiparameter inverse theorem of Proposition 3.7, its proof is identical to the proof of Theorem 1.2 in [**29**], where the case $d=1$ was treated; we only sketch the main idea of the proof below.

THEOREM 3.10. ([**29**, Theorem 1.2] for $d=1$) *Let $d, k \in \mathbb{N}$ and $a \in \ell^\infty(\mathbb{N}^d)$ be a sequence. Then the following properties are equivalent.*
(i) *a is $(k-1)$-regular and k-anti-uniform.*
(ii) *a is an approximate $(k-1)$-step nilsequence.*

Idea of the proof. The implication (ii) $\Longrightarrow$ (i) follows from Corollary 3.6 and the fact that a product of two nilsequences is a nilsequence, and hence a regular sequence.

We explain now the main idea of the converse implication. Let $a \in \ell^\infty(\mathbb{N}^d)$ be $(k-1)$-regular and k-anti-uniform. Let $\mathcal{N}$ be the linear space of $(k-1)$-step nilsequences in d variables and let $\mathcal{H}$ be the linear span of $\mathcal{N}$ and a. Then it follows from our $(k-1)$-regularity assumption and Theorem 3.9 that, for all $c, c' \in \mathcal{H}$, the averages of $c(\mathbf{n})\,\overline{c'(\mathbf{n})}$ converge and we write $\langle c, c'\rangle$ for this limit. We remark that $\|c\|_2^2 = \langle c, c\rangle$.

If the space $\mathcal{H}$ endowed with the 'scalar product' $\langle\cdot,\cdot\rangle$ was a Hilbert space and $\mathcal{N}$ was a closed subspace, then we could define ψ to be the orthogonal projection of a on $\mathcal{N}$. Then $\langle a-\psi, \psi'\rangle = 0$ for all $\psi' \in \mathcal{N}$ and Corollary 3.8 would imply that $\|a-\psi\|_{U^k(\mathbb{N}^d)} = 0$. Since, by assumption, a is k-anti-uniform, we deduce that $\|a-\psi\|_2 = 0$: that is, a is a $(k-1)$-step nilsequence plus a sequence that converges to zero in uniform density. In our present set-up, there is lack of completeness, so we choose ψ to be an 'approximate orthogonal projection', in the sense that $\|a-\psi\|_2$ is sufficiently close to the distance of a to $\mathcal{N}$, and we obtain the announced decomposition, using Part (i) of Proposition 3.7. For the details, see the proof of Theorem [**29**, Theorem 1.2], which contains a proof for $d=1$ under the assumption of strong k-anti-uniformity (it is called k-anti-uniformity there); the same argument works without change for general $d \in \mathbb{N}$ under the weaker assumption of k-anti-uniformity. □

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

4. *Correlations are regular sequences*
The goal of this section is to show that modulo small ℓ^∞-errors, nilsequences can be represented as multiple correlation sequences, and then use the mean convergence result of Walsh (Theorem 1.1) in order to show that multiple correlation sequences are regular sequences.

4.1. *Producing nilsequences as correlations.* The argument we use below is analogous to the one used in [**29**] to handle single variable nilsequences.

LEMMA 4.1. [**37**, Lemma 14.2] *Let $d, k \in \mathbb{N}$ and $X = G/\Gamma$ be a $(k-1)$-step nilmanifold. Then there exists a continuous map $P: X^k \to X$ such that*

$$P(hg\cdot e_X, h^2g\cdot e_X, \ldots, h^kg\cdot e_X) = g\cdot e_X \quad \text{for all } g, h \in G. \tag{4.1}$$

Remark. The result in [**37**, Lemma 14.2] gives $P(g\Gamma, hg\Gamma, h^2g\Gamma, \ldots, h^{\ell-1}g\Gamma) = h^\ell g\Gamma$. Inserting $h^{-\ell}g$ in place of g, then h^{-1} in place of h and rearranging coordinates, we get (4.1).

PROPOSITION 4.2. *Let $d, k \in \mathbb{N}$ and $\psi: \mathbb{N}^d \to \mathbb{C}$ be a $(k-1)$-step nilsequence. Then, for every $\varepsilon > 0$, there exists a system $(Y, \nu, S_1, \ldots, S_d)$ and functions $F_1, \ldots, F_k \in L^\infty(\nu)$, such that the sequence $b: \mathbb{N}^d \to \mathbb{C}$, defined by*

$$b(\mathbf{n}) := \int \prod_{j=1}^{k}\bigg(\prod_{i=1}^{d} S_i^{\ell_j n_i}\bigg)F_j\, d\nu, \quad \mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d, \tag{4.2}$$

where $\ell_j = k!/j$ for $j = 1, \ldots, k$, satisfies

$$\|\psi - b\|_{\ell^\infty(\mathbb{N}^d)} \le \varepsilon.$$

Proof. The sequence $\psi: \mathbb{N}^d \to \mathbb{C}$ has the form

$$\psi(\mathbf{n}) = \Psi\bigg(\prod_{i=1}^{d} \tau_i^{n_i}\cdot e_X\bigg), \quad \mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d,$$

for some $(k-1)$-step nilmanifold $X = G/\Gamma$, commuting elements $\tau_1, \ldots, \tau_d$ of G and function $\Psi \in C(X)$. Let $\varepsilon > 0$. As remarked after the definition of the nilsequence in §2.1.2, we can assume that the group G is connected.

For $i = 1, \ldots, d$, let $g_i \in G$ be such that $g_i^{k!} = \tau_i$ (such elements exist since G is connected, and hence divisible). Let $\mathbf{m} = (m_1, \ldots, m_d) \in \mathbb{N}^d$. Using Lemma 4.1, with $g := \prod_{i=1}^d g_i^{k!n_i}$ and $h := \prod_{i=1}^d g_i^{m_i}$ and writing $H = \Psi \circ P$, we have $H \in C(X^k)$ and we obtain

$$\psi(\mathbf{n}) = H\bigg(\bigg(\bigg(\prod_{i=1}^{d} g_i^{m_i+k!n_i}\bigg), \bigg(\prod_{i=1}^{d} g_i^{2m_i+k!n_i}\bigg), \ldots, \bigg(\prod_{i=1}^{d} g_i^{km_i+k!n_i}\bigg)\bigg)\cdot e_{X^k}\bigg)$$

for all $\mathbf{m}$ and $\mathbf{n} \in \mathbb{N}^d$. Letting

$$\widetilde{\alpha}_i := (g_i, g_i^2, \ldots, g_i^k) \in G^k, \quad i = 1, \ldots, d,$$

averaging with respect to $\mathbf{m} \in \mathbb{N}^d$ and using the equidistribution results for sequences on nilmanifolds (see [**56**] or [**50**]), we get that, for every $\mathbf{n} \in \mathbb{N}^d$,

$$\psi(\mathbf{n}) = \lim_{M\to+\infty} \frac{1}{M^d} \sum_{\mathbf{m}\in[M]^d} H\bigg(\bigg(\bigg(\prod_{i=1}^d g_i^{k!n_i}\bigg), \bigg(\prod_{i=1}^d g_i^{k!n_i}\bigg), \\ \ldots, \bigg(\prod_{i=1}^d g_i^{k!n_i}\bigg)\bigg) \cdot \widetilde{\alpha}_1^{m_1} \cdot \cdots \cdot \widetilde{\alpha}_d^{m_d} \cdot e_{X^k}\bigg) \\ = \int_{\widetilde{Y}} H\bigg(\bigg(\prod_{i=1}^d g_i^{k!n_i}\bigg)\cdot x_1, \bigg(\prod_{i=1}^d g_i^{k!n_i}\bigg)\cdot x_2, \ldots, \bigg(\prod_{i=1}^d g_i^{k!n_i}\bigg)\cdot x_k\bigg)\, dm_{\widetilde{Y}}(x_1, \ldots, x_k),$$

where $\widetilde{Y}$ is the closure of the sequence $\{\widetilde{\alpha}_1^{m_1} \cdot \cdots \cdot \widetilde{\alpha}_d^{m_d} \cdot e_{X^k} : m_1, \ldots, m_d \in \mathbb{N}\}$ in X^k and $m_{\widetilde{Y}}$ is the Haar measure of this sub-nilmanifold of X^k. For $i = 1, \ldots, d$, let $\widetilde{S}_i : \widetilde{Y} \to \widetilde{Y}$ be the translation by $\widetilde{\alpha}_i$. Note that $(\widetilde{Y}, m_{\widetilde{Y}}, \widetilde{S}_1, \ldots, \widetilde{S}_d)$ is a nilsystem.

The continuous function H on X^k can be approximated uniformly by linear combinations of functions of the form $f_1 \otimes \cdots \otimes f_k$, where $f_j \in C(X)$ for $j = 1, \ldots, k$. Since finite linear combinations of sequences of the form (4.2) have the same form (see the proof of Theorem 2.7 in §6.6 below), it remains to show that any sequence $\psi' : \mathbb{N}^d \to \mathbb{C}$, given by

$$\psi'(\mathbf{n}) := \int_{\widetilde{Y}} \prod_{j=1}^k f_j\bigg(\bigg(\prod_{i=1}^d g_i^{k!n_i}\bigg)\cdot x_j\bigg)\, dm_{\widetilde{Y}}(x_1, \ldots, x_k), \quad \mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d,$$

has the form (4.2). To this end, for $j = 1, \ldots, k$, let $F_j \in C(X^k)$ be given by $F_j(\widetilde{x}) := f_j(x_j)$ for $\widetilde{x} = (x_1, \ldots, x_k) \in X^k$. Recall that $\ell_j = k!/j$ for $j = 1, \ldots, k$. Since $\widetilde{S}_i\widetilde{x} = \widetilde{\alpha}_i\widetilde{x} = (g_i, g_i^2, \ldots, g_i^k)\widetilde{x}$, $i = 1, \ldots, d$, the jth coordinate of the element $(\prod_{i=1}^d \widetilde{S}_i^{\ell_j n_i})\widetilde{x}$ is $\prod_{i=1}^d g_i^{j\ell_j n_i} \cdot x_j = \prod_{i=1}^d g_i^{k!n_i} \cdot x_j$ and thus, for $j = 1, \ldots, k$,

$$F_j\bigg(\bigg(\prod_{i=1}^d \widetilde{S}_i^{\ell_j n_i}\bigg)\widetilde{x}\bigg) = f_j\bigg(\prod_{i=1}^d g_i^{k!n_i} \cdot x_j\bigg), \quad n_1, \ldots, n_d \in \mathbb{N}.$$

Therefore,

$$\psi'(\mathbf{n}) = \int_{\widetilde{Y}} \prod_{j=1}^k F_j\bigg(\bigg(\prod_{i=1}^d \widetilde{S}_i^{\ell_j n_i}\bigg)\widetilde{x}\bigg)\, dm_{\widetilde{Y}}(\widetilde{x}), \quad \mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d.$$

This completes the proof. □

4.2. *Proof of Proposition 2.3.* Let $d, s \in \mathbb{N}$ and $w \in \ell^\infty(\mathbb{N}^d)$ satisfy the hypothesis of Proposition 2.3. Let ψ be an s-step nilsequence in d variables.

We set $k := s + 1$. Let $\varepsilon > 0$ and let the system $(Y, \nu, S_1, \ldots, S_d)$ and the functions $F_1, \ldots, F_k \in L^\infty(\nu)$ be as in Proposition 4.2. Letting $h_j := F_{j+1}$, for $j = 0, \ldots, s$, the sequence $b : \mathbb{N}^d \to \mathbb{C}$, defined by (4.2), can be rewritten as

$$b(\mathbf{n}) = \int h_0 \cdot \prod_{j=0}^s S_{(\ell_{j+1}-\ell_1)\cdot\mathbf{n}} h_j \, d\nu, \quad \mathbf{n} \in \mathbb{N}^d.$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

By hypothesis, the averages of $w(\mathbf{n})\, b(\mathbf{n})$ converge. Since $|b(\mathbf{n}) - \psi(\mathbf{n})|$ is uniformly bounded by ε, the oscillations of the averages of $w(\mathbf{n})\, \psi(\mathbf{n})$ are uniformly bounded by 2ε. Since this holds for every $\varepsilon > 0$, the averages of $w(\mathbf{n})\, \psi(\mathbf{n})$ converge, which completes the proof. □

4.3. *Proof of Proposition 2.1.* By Proposition 4.2, it suffices to prove that the limit

$$\lim \operatorname{Av} b(\mathbf{n})\, T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s$$

exists in $L^2(\mu)$ for every sequence $(b(\mathbf{n}))_{\mathbf{n}\in\mathbb{N}^d}$, defined as in (4.2).

By Theorem 1.1, the limit

$$\lim \operatorname{Av}_{\mathbf{n}} \bigg(\prod_{j=1}^{k} F_j \bigg(\prod_{i=1}^{d} S_i^{\ell_j n_i} y \bigg) \cdot \prod_{m=1}^{s} f_m \big(T_{\vec{p_m}(\mathbf{n})} x \big) \bigg)$$

exists in $L^2(\nu \times \mu)$. Taking the integral over Y with respect to ν we obtain the announced result. □

5. *The structure of systems of order one*

In this section, we prove a structural result which is going to be used in the proof of Theorem 2.6 in the next section. Here we work only with systems (X, μ, T) with a single transformation. We denote the σ-algebra of T-invariant subsets of X by $\mathcal{I}(T)$ and write the ergodic decomposition of μ under T as

$$\mu = \int \mu_x \, d\mu(x),$$

where for μ-almost every $x \in X$, the measure μ_x is invariant and ergodic under T and the map $x \mapsto \mu_x$ is invariant under T. Let $f \in L^1(\mu)$. Then, for μ-almost every $x \in X$, the function f is defined μ_x-almost everywhere and belongs to $L^1(\mu_x)$; the map $x \mapsto \int f \, d\mu_x$ is measurable with respect to $\mathcal{I}(T)$, and

$$\mathbb{E}_\mu(f \mid \mathcal{I}(T))(x) = \int f \, d\mu_x \quad \text{for } \mu\text{-almost every } x \in X,$$

which means that, for every set $A \in \mathcal{I}(T)$,

$$\int_A f \, d\mu = \int_A \bigg(\int f \, d\mu_x \bigg) \, d\mu(x).$$

The factor $\mathcal{Z}_1$ is defined in Appendix A.2 and reduces to the Kronecker factor for ergodic systems.

Definition. We say that a system (X, μ, T) has *order* one if $\mathcal{Z}_1 = \mathcal{X}$.

Throughout this section, we work only with systems of order one. Generalizing the construction of a Fourier basis of a compact Abelian group, our goal is to construct a 'relative orthonormal' basis for systems of order one consisting of 'relative' eigenfunctions. This is the context of Theorem 5.2, below.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

5.1. *Relative orthonormal basis.*

Definition. A *relative orthonormal system* (with respect to the T-invariant σ-algebra $\mathcal{I}(T)$) is a countable family $(\phi_j)_{j\in\mathbb{N}}$ of functions belonging to $L^2(\mu)$ such that:

(i) $\mathbb{E}_\mu(|\phi_j|^2 \mid \mathcal{I}(T))$ has value zero or one μ-almost everywhere for every $j \in \mathbb{N}$; and

(ii) $\mathbb{E}_\mu(\phi_j\overline{\phi_k} \mid \mathcal{I}(T)) = 0$ μ-almost everywhere for all $j, k \in \mathbb{N}$ with $j \neq k$.

The family $(\phi_j)_{j\in\mathbb{N}}$ is a *relative orthonormal basis* if it also satisfies:

(iii) the linear space spanned by all functions of the form $\phi_j\psi$, is dense in $L^2(\mu)$, where $j \in \mathbb{N}$ and $\psi \in L^\infty(\mu)$ varies over all T-invariant functions.

We do not make the apparently natural assumption that $\mathbb{E}_\mu(|\phi_j|^2 \mid \mathcal{I}(T)) = 1$ μ-almost everywhere, as there does not exist, in general, a relative orthonormal basis satisfying this additional condition (consider a system with ergodic components given by rotations on cyclic groups of different order). This creates a few minor complications in the statements and the proofs below. We remark that the definition allows some of the elements of the base to be identically zero. This explains why we can assume, without loss of generality, that all relative orthonormal systems are countably infinite, and thus indexed by $\mathbb{N}$.

Definition. Given a relative orthonormal system $(\phi_j)_{j\in\mathbb{N}}$ and $f \in L^2(\mu)$, we let

$$f_j := \mathbb{E}_\mu(f\overline{\phi_j} \mid \mathcal{I}(T)), \quad j \in \mathbb{N}.$$

If $(\phi_j)_{j\in\mathbb{N}}$ is a relative orthonormal basis, we say that the T-invariant functions f_j, $j \in \mathbb{N}$, are the *coordinates* of f in this basis.

Example 5.1. On $\mathbb{T}^2$ with the Haar measure $m_{\mathbb{T}^2}$, let $T : \mathbb{T}^2 \to \mathbb{T}^2$ be given by $T(x, y) = (x, y + x)$. Then $(e(jy))_{j\in\mathbb{Z}}$ is a relative orthonormal basis for $L^2(m_{\mathbb{T}^2})$. The coordinates $(f_j)_{j\in\mathbb{Z}}$ of a function $f \in L^2(m_{\mathbb{T}^2})$ are given by $f_j(x) := \int f(x, y)\, e(-jy)\, dy$, $j \in \mathbb{Z}$.

We remark that if $\|f\|_{L^\infty(\mu)} \le 1$, then $\|f_j\|_{L^\infty(\mu)} \le 1$ for every $j \in \mathbb{N}$. We will use, repeatedly, that condition (i) implies the identity

$$f_j = f_j \cdot \mathbb{E}_\mu(|\phi_j|^2 \mid \mathcal{I}(T)) \text{ } \mu\text{-almost everywhere}, \quad j \in \mathbb{N}; \tag{5.1}$$

in particular, $f_j(x) = 0$ μ-almost everywhere on the set where $\mathbb{E}(|\phi_j|^2 \mid \mathcal{I}(T))(x) = 0$. We remark, also, that for μ-almost every $x \in X$ for every $j \in \mathbb{N}$, we have the identities

$$f_j(x) = \int f\, \overline{\phi_j}\, d\mu_x \quad \text{and} \quad \mathbb{E}(|f|^2 \mid \mathcal{I}(T))(x) = \|f\|^2_{L^2(\mu_x)}.$$

Next, we establish a relative version of the Parseval and Fourier identity and give necessary and sufficient conditions for a relative orthonormal system to be a relative orthonormal basis.

PROPOSITION 5.1. *Let (X, μ, T) be a system with ergodic decomposition $\mu = \int \mu_x\, d\mu(x)$ and let $(\phi_j)_{j\in\mathbb{N}}$ be a relative orthonormal system for $L^2(\mu)$.*

(i) *For every $f \in L^2(\mu)$ and for μ-almost every $x \in X$, the series*

$$Pf := \sum_{j\in\mathbb{N}} f_j\phi_j \tag{5.2}$$

converges in $L^2(\mu_x)$ and

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

$$\|Pf\|^2_{L^2(\mu_x)} = \sum_{j\in\mathbb{N}} |f_j(x)|^2 \quad \textit{and} \quad \|f\|^2_{L^2(\mu_x)} = \|Pf\|^2_{L^2(\mu_x)} + \|f - Pf\|^2_{L^2(\mu_x)}. \tag{5.3}$$

(ii) *$(\phi_j)_{j\in\mathbb{N}}$ is a relative orthonormal basis if and only if, for every $f \in L^2(\mu)$ and for μ-almost every $x \in X$, the series (5.2) converges to f in $L^2(\mu_x)$. In this case,*

$$\|f\|^2_{L^2(\mu_x)} = \sum_{j\in\mathbb{N}} |f_j(x)|^2 \quad \mu\textit{-almost everywhere.} \tag{5.4}$$

(iii) *$(\phi_j)_{j\in\mathbb{N}}$ is a relative orthonormal basis if and only if, for every $f \in L^2(\mu)$,*

$$\|f\|^2_{L^2(\mu)} = \sum_{j\in\mathbb{N}} \|f_j\|^2_{L^2(\mu)}. \tag{5.5}$$

Proof. We prove (i). Let $f \in L^2(\mu)$; then $f \in L^2(\mu_x)$ for μ-almost every $x \in X$. Recall that $\int |\phi_j|^2\, d\mu_x = \mathbb{E}_\mu(|\phi_j|^2 \mid \mathcal{I}(T))(x) \in \{0, 1\}$ for μ-almost every $x \in X$. Let $E_x = \{j \in \mathbb{N} : \int |\phi_j|^2\, d\mu_x = 1\}$. Then for μ-almost every $x \in X$, the functions $(\phi_j)_{j\in E_x}$ form an orthonormal system in the Hilbert space $L^2(\mu_x)$ and, viewing f as an element of $L^2(\mu_x)$, the functions $(f_j)_{j\in E_x}$ are the coordinates of f in this system. Note, also, that by (5.1), $f_j = 0$ if $j \notin E_x$. This establishes (i).

We prove (ii). Suppose, first, that, for every $f \in L^2(\mu)$ for μ-almost every $x \in X$ we have $Pf = f$ μ_x-almost everywhere. It follows that the series in (5.2) converges to f in $L^2(\mu)$. Therefore, f belongs to the closed linear subspace of $L^2(\mu)$ spanned by all functions of the form $\phi_j \psi$, where $j \in \mathbb{N}$ and $\psi \in L^\infty(\mu)$ varies over all T-invariant functions. By definition, $(\phi_j)_{j\in\mathbb{N}}$ is a relative orthonormal system.

Now we establish the converse implication. Suppose that $(\phi_j)_{j\in\mathbb{N}}$ is a relative orthonormal basis and let $f \in L^2(\mu)$. For μ-almost every $x \in X$, the function Pf, defined by (5.2), satisfies $\int Pf\, \overline{\phi_j}\, d\mu_x = f_j$ for every $j \in \mathbb{N}$, and hence $\mathbb{E}_\mu((f - Pf)\, \overline{\phi_j} \mid \mathcal{I}(T)) = 0$ and $\int (f - Pf)\, \overline{\phi_j}\, \overline{\psi}\, d\mu = 0$ for every T-invariant function $\psi \in L^\infty(\mu)$. Therefore, the function $f - Pf$ is orthogonal in $L^2(\mu)$ to the linear space spanned by all functions of the form $\phi_j \psi$, where $j \in \mathbb{N}$ and $\psi \in L^\infty(\mu)$ varies over all T-invariant functions. Thus $f - Pf = 0$, by hypothesis. Inserting f in place of Pf in (5.3) gives identity (5.4).

We prove (iii). If $(\phi_j)_{j\in\mathbb{N}}$ is a relative orthonormal basis, then (5.5) follows by integrating the identity in (5.4) over X with respect to μ. To prove the converse implication, let $f \in L^2(\mu)$. Using (5.5), inserting the first identity of (5.3) in the second and integrating over X with respect to μ, we deduce that $f = Pf$ for μ-almost every $x \in X$. Hence, Property (iii) of the definition of a relative orthonormal basis is satisfied. □

5.2. *Relative orthonormal basis of eigenfunctions.*

Definition. Let $\lambda \in L^\infty(\mu)$ be a T-invariant function and $\phi \in L^\infty(\mu)$. We say that ϕ is an *eigenfunction* with eigenvalue λ if:

(i) $|\phi(x)|$ has value zero or one for μ-almost every $x \in X$;
(ii) $\lambda(x) = 0$ for μ-almost every $x \in X$ such that $\phi(x) = 0$; and
(iii) $\phi \circ T = \lambda \cdot \phi$ μ-almost everywhere.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

The role of Property (ii) is to avoid ambiguities. Note, also, that Property (iii) does not imply anything about the value of $\lambda(x)$ at the points $x \in X$ where $\phi(x) = 0$.

Property (iii) gives that, for μ-almost every $x \in X$ such that $\phi(x) \neq 0$, we have $\phi(T^{-1}x) \neq 0$ and thus $|\phi(x)| = |\phi(T^{-1}x)| = 1$, by Property (i) above, and $|\lambda(x)| = 1$, by Property (iii). On the other hand, for μ-almost every $x \in X$ such that $\phi(x) = 0$, we have $\phi(T^{-1}x) = 0$, by Property (iii), and $\lambda(x) = 0$, by Property (ii). Therefore, the function $|\phi|$ is T-invariant and

$$|\phi| = |\lambda|. \tag{5.6}$$

Next, we state the main result of this section and we prove it in §5.4.

THEOREM 5.2. *Let (X, μ, T) be a system of order one. Then $L^2(\mu)$ admits a relative orthonormal basis of eigenfunctions.*

Remarks.

- It is true, and not difficult to prove, that if a system has a relative orthonormal basis of eigenfunctions, then it has order one.
- For ergodic systems, Theorem 5.2 is well known, but it is not easy to deduce the general case from the ergodic one. The reason is that, although the ergodic components of a system of order one are ergodic rotations (Proposition A.1), we cannot simply 'glue' together their eigenfunctions, because of measurability issues.

Example 5.2. For the system described in the Example 5.1, $(e(jy))_{j\in\mathbb{Z}}$ is a relative orthonormal basis of eigenfunctions with eigenvalues $(e(jx))_{j\in\mathbb{Z}}$.

The next proposition will be used in the proof of Theorem 2.6.

PROPOSITION 5.3. *Let (X, μ, T) be a system of order one with ergodic decomposition $\mu = \int \mu_x \, d\mu(x)$. Suppose that $(\phi_j)_{j\in\mathbb{N}}$ is an orthonormal basis of eigenfunctions and that $f \in L^\infty(\mu)$, and let $(f_j)_{j\in\mathbb{N}}$ be the coordinates of f in this base. Then, for μ-almost every $x \in X$,*

$$|||f|||^4_{T,\mu_x,2} = \sum_{j\in\mathbb{N}} |f_j(x)|^4 \quad \mu\text{-almost everywhere.} \tag{5.7}$$

Remark. After integrating (5.7) over X with respect to μ, it follows, from (A.3), that $|||f|||^4_{T,\mu,2} = \sum_{j\in\mathbb{N}} \|f_j\|^4_{L^4(\mu)}$.

In the proof of Proposition 5.3, we will use the following basic fact.

LEMMA 5.4. *Let (X, μ, T) be a system, let $(\phi_j)_{j\in\mathbb{N}}$ be a relative orthonormal system of eigenfunctions and let $(\lambda_j)_{j\in\mathbb{N}}$ be the corresponding eigenvalues. Then, for μ-almost every $x \in X$, $\lambda_j(x)\overline{\lambda_k(x)} \neq 1$ for all $j, k \in \mathbb{N}$ with $j \neq k$.*

Proof of Lemma 5.4. Note that the set $A = \{x : \lambda_j(x)\overline{\lambda_k(x)} = 1\}$ is T-invariant and the function $\mathbf{1}_A \phi_j \overline{\phi_k}$ is T-invariant, by Part (iii) of the definition of an eigenfunction, and thus equal to zero, by Part (ii) of the definition of a relative orthonormal system. On the other hand, for μ-almost every $x \in A$, $|\phi_j(x)\overline{\phi_k(x)}| = 1$, by (5.6). Thus, $\mu(A) = 0$ and the claim follows. □

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

Proof of Proposition 5.3. By Part (ii) of Proposition 5.1, for μ-almost every $x \in X$,

$$f = \sum_{j \in \mathbb{N}} f_j \cdot \phi_j,$$

where $f_j = \int f \overline{\phi_j}\, d\mu_x$ and the convergence takes place in $L^2(\mu_x)$. It follows that, for μ-almost every $x \in X$ and for every $n \in \mathbb{N}$,

$$T^n f \cdot \overline{f} = \sum_{j,k \in \mathbb{N}} \lambda_j^n\, f_j\, \overline{f_k}\, \phi_j\, \overline{\phi_k},$$

where convergence takes place in $L^1(\mu_x)$. Using this, identity (5.1) and that $\int \phi_j \overline{\phi_k}\, d\mu_x = 0$ for $j, k \in \mathbb{N}$ with $j \neq k$, we deduce that, for μ-almost every $x \in X$,

$$\int T^n f \cdot \overline{f}\, d\mu_x = \sum_{j \in \mathbb{N}} \lambda_j^n |f_j|^2.$$

Note that, by (5.4), the above series converges absolutely for μ-almost every $x \in X$. Hence, for μ-almost every $x \in X$,

$$\left| \int T^n f \cdot \overline{f}\, d\mu_x \right|^2 = \sum_{j,k \in \mathbb{N}} (\lambda_j \overline{\lambda_k})^n |f_j|^2 |f_k|^2.$$

Averaging in $n \in \mathbb{N}$ and using that, by (A.2),

$$|||f|||_{T,\mu_x,2}^4 = \lim_{N \to +\infty} \frac{1}{N} \sum_{n=1}^{N} \left| \int T^n f \cdot \overline{f}\, d\mu_x \right|^2,$$

we obtain, for μ-almost every $x \in X$, that

$$|||f|||_{T,\mu_x,2}^4 = \sum_{j,k \in \mathbb{N}} \left(\lim_{N \to +\infty} \frac{1}{N} \sum_{n=1}^{N} (\lambda_j \overline{\lambda_k})^n \right) \cdot |f_j|^2 |f_k|^2,$$

where the interchange of limits is justified because $\sum_{j,k \in \mathbb{N}} |f_j|^2 |f_k|^2$ converges μ-almost everywhere, by (5.4).

If $j \neq k$, since $|\lambda_j \overline{\lambda_k}| \leq 1$ and $\lambda_j \overline{\lambda_k} \neq 1$, by Lemma 5.4, for μ-almost every $x \in X$,

$$\lim_{N \to +\infty} \frac{1}{N} \sum_{n=1}^{N} (\lambda_j \overline{\lambda_k})^n = 0.$$

Suppose now that $j = k$. For μ-almost every $x \in X$, the following holds: if $\lambda_j(x) = 0$, then $\phi_j(x) = 0$, by (5.6), and $f_j(x) = 0$, by (5.1). If $\lambda_j(x) \neq 0$, then $|\lambda_j(x)| = 1$, by (5.6) and Part (i) of the definition of an eigenfunction. In both cases, $|\lambda_j(x)|^{2n} |f_j(x)|^4 = |f_j(x)|^4$. Combining the above we get (5.7). □

5.3. *A Borel selection result.* The proof of Theorem 5.2 needs some technical preliminaries. We will need the following selection theorem of Lusin–Novikov.

THEOREM 5.5. (See, for example, [**49**, Theorem 18.10]) *Let X, Y be Polish (i.e. complete, separable, metric) spaces and $P \subset X \times Y$ be a Borel set such that every vertical section $P_x = \{y : (x, y) \in P\}$ is a countable set. Then the vertical projection A of P on X is Borel and there exists a Borel function $f : A \to Y$ such that $f(x) \in P_x$ for every $x \in A$.*

PROPOSITION 5.6. *Let X, K be Polish spaces and $F: X \times K \to \mathbb{R}_+$ be a bounded Borel function. Suppose that, for every $x \in X$, the set*

$$P_x := \{y \in K : F(x, y) > 0\}$$

is countable and

$$\Sigma(x) := \sum_{y \in P_x} F(x, y) < +\infty. \tag{5.8}$$

Let $N(x) := |P_x| \in [0, +\infty]$. Then there exists a sequence $(t_j)_{j\in\mathbb{N}}$ of Borel maps $X \to K$ such that, for every $x \in X$, the values $t_j(x)$, $1 \le j < 1 + N(x)$, are pairwise distinct and $P_x = \{t_j(x) : 1 \le j \le N(x)\}$.

Remarks.

- The condition $j < 1 + N(x)$ means that $j \le N(x)$ if $N(x) < +\infty$ and j is arbitrary if $N(x) = +\infty$.
- Note that if $P_x = \emptyset$, then the values of $t_j(x)$ are not determined by the statement.

Proposition 5.6 is a consequence of the next lemma.

LEMMA 5.7. *Let F, K, X be as in Proposition 5.6 and, for $x \in X$, let*

$$S(x) := \sup_{y \in K} F(x, y).$$

Then there exists a Borel map $t: X \to K$ such that $F(x, t(x)) = S(x)$ for every $x \in X$.

Proof of Lemma 5.7. Note that (5.8) implies that this supremum $S(x)$ is attained. Let $\mathcal{X}$ and $\mathcal{K}$ be the Borel σ-algebras of the spaces X and K, respectively.

We first claim that the function S is Borel. Indeed, let $s \ge 0$. The set $\{(x, y) \in X \times K : F(x, y) > s\}$ belongs to $\mathcal{X} \otimes \mathcal{K}$ and has countable fibers. By Theorem 5.5, its projection on X belongs to $\mathcal{X}$. Since this projection is the set $\{x \in X : S(x) > s\}$, the map S is Borel.

For $m \in \mathbb{N}$, let

$$A_m := \{(x, y) \in X \times K : F(x, y) \ge (1 - 2^{-m})S(x)\}.$$

Then A_m belongs to $\mathcal{X} \otimes \mathcal{K}$, the projection of A_m on X is onto and this projection is countable to one. By Theorem 5.5, for every $m \in \mathbb{N}$, there exists a Borel map $t_m : X \to K$ such that $(x, t_m(x)) \in A_m$ for every $x \in X$: that is,

$$F(x, t_m(x)) \ge (1 - 2^{-m})S(x) \quad \text{for every } x \in X.$$

If x is such that $S(x) = 0$, then $t_m(x) = 0$ for every $m \in \mathbb{N}$. If not, then $F(x, t_m(x)) \ge S(x)/2$ for every $m \in \mathbb{N}$ and thus the set $\{t_m(x) : m \ge 1\}$ contains at most $2\Sigma(x)S(x)^{-1}$ distinct elements. It follows that, for every $x \in X$, the sequence $(t_m(x))_{m\in\mathbb{N}}$ is eventually constant. The limit value $t(x)$ of this sequence is therefore well defined, is a Borel map from X to K and satisfies $F(x, t(x)) = S(x)$. This completes the proof. □

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

Proof of Proposition 5.6. We build, by induction, the family of Borel maps $t_j : X \to K$, $j \in \mathbb{N}$. Let $t_1 : X \to K$ be given by Lemma 5.7. We have that $F(x, t_1(x)) = 0$ if and only if P_x is empty, that is, if $N(x) = 0$. We define

$$F'(x, t) := \begin{cases} 0 & \text{if } t = t_1(x), \\ F(x, t) & \text{otherwise,} \end{cases}$$
$$P'_x := \{t : F'(x, t) > 0\} \quad \text{for } x \in X.$$

The function F' is Borel and, for every $x \in X$ for which P_x is non-empty, this set is the disjoint union of P'_x and $\{t_1(x)\}$. We replace the function F with F', and Lemma 5.7 provides a map $t_2 : X \to K$. Iterating, we obtain a sequence $(t_j)_{j\in\mathbb{N}}$ of Borel maps $X \to K$ that satisfy

$$F(x, t_1(x)) \geq F(x, t_2(x)) \geq \cdots \geq F(x, t_j(x)),$$
$$F(x, t_j(x)) = 0 \text{ if and only if } j > N(x),$$
$$\text{if } N(x) \geq j \text{ then } t_{j+1}(x) \notin \{t_1(x), \ldots, t_j(x)\},$$
$$F(x, t_{j+1}(x)) = \sup\big\{F(x, t) : t \notin \{t_1(x), \ldots, t_j(x)\}\big\}.$$

We have $\{t_j(x) : 1 \leq j \leq N(x)\} \subset P_x$ and we claim that equality holds. Suppose that this is not the case and let $t \in P_x \setminus \{t_j(x) : 1 \leq j \leq N(x)\}$. Then, by construction, $t_j(x) \geq t$ for $j \leq N(x)$ and thus $\sigma(x) \geq N(x)t$. It follows that $N(x)$ is finite. By construction, $|\{t_j(x) : 1 \leq j \leq N(x)\}| = N(x) = |P_x|$, which is a contradiction. This completes the proof. □

5.4. *Proof of Theorem 5.2.* In this subsection we use a different presentation of the ergodic decomposition of a system that is more convenient for our purposes†. Recall that we assume that $(X, \mathcal{X}, \mu)$ is a Lebesgue space. It is well known (see, for example, [**35**, Theorems 8.7 and A.7]) that there exists a Lebesgue space $(Y, \mathcal{Y}, \nu)$, a measure preserving map $\pi : X \to Y$ that satisfies $\pi \circ T = \pi$, $\mathcal{I}(T) = \pi^{-1}(\mathcal{Y})$ up to μ-null sets and, for $y \in Y$, a probability measure μ_y on X such that the following hold.

(i) The map $y \mapsto \mu_y$ is Borel, meaning that, for every bounded Borel function on X, the function $y \mapsto \int f \, d\mu_y$ is Borel.

(ii) For every bounded Borel function on X,

$$\mathbb{E}_\mu(f \mid (\mathcal{I}(T))(x) = \int f \, d\mu_{\pi(x)} \quad \text{for } \mu\text{-almost every } x \in X.$$

(iii) For ν-almost every $y \in Y$, the measure μ_y on X is concentrated on $\pi^{-1}(\{y\})$ and is invariant and ergodic under T.

Taking the integral in (ii), we obtain

$$\mu = \int_Y \mu_y \, d\nu(y).$$

By density,

† The measures μ_x of the top of §5 are written as $\mu_{\pi(x)}$ here.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

(iv) for $f \in L^1(\mu)$, $f \in L^1(\mu_y)$ for ν-almost every $y \in Y$, $\|f\|_{L^1(\mu)} = \int \|f\|_{L^1(\mu_y)}\, d\nu(y)$ and the equality in (ii) remains valid.

Lastly, since our standing assumption, in this section, is that the system (X, μ, T) has order one, by Proposition A.1,

(v) for ν-almost every $y \in Y$, the system (X, μ_y, T) is an ergodic rotation.

We first prove the following intermediate result.

LEMMA 5.8. *Let (X, μ, T) be a system of order one and $f \in L^2(\mu)$. Then there exists a relative orthonormal system of eigenfunctions $(\phi_j)_{j\in\mathbb{N}}$ such that*

$$f = \sum_{j\in\mathbb{N}} \mathbb{E}_\mu(f\overline{\phi_j} \mid \mathcal{I}(T)) \cdot \phi_j, \tag{5.9}$$

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

Moreover, for every $j \in \mathbb{N}$, the eigenfunction ϕ_j belongs to the smallest closed T-invariant subspace of $L^2(\mu)$ containing the set $\{f\phi\colon \phi \in L^\infty(\mu)$ is T-invariant$\}$.

Remark. For notational convenience, the orthonormal system we build below is indexed by $\mathbb{Z}_+$ instead of $\mathbb{N}$.

Proof. We can assume that f is a Borel function defined everywhere. The set Y_0 of $y \in Y$ such that μ_y is invariant under T is Borel and has full measure. Since the map $y \mapsto \|f\|_{L^2(\mu_y)}$ is Borel, and since $\int \|f\|^2_{L^2(\mu_y)}\, d\nu(y) = \|f\|_{L^2(\mu)} < +\infty$, the set Y_1 of points $y \in Y_0$ such that $f \in L^2(\mu_y)$ is Borel and has full measure. Substituting Y_1 for Y, we are reduced to the case where the measure μ_y is T-invariant and $f \in L^2(\mu_y)$ for every $y \in Y$. Since $(Y, \mathcal{Y}, \nu)$ is a Lebesgue space, we can assume that Y is a Polish space and $\mathcal{Y}$ is its Borel σ-algebra.

For $y \in Y$, we write σ_y for the spectral measure of f with respect to the system (X, μ_y, T); it is the finite positive measure on $\mathbb{T}$ defined by

$$\widehat{\sigma_y}(n) := \int T^n f \cdot \overline{f}\, d\mu_y, \quad n \in \mathbb{Z}.$$

For ν-almost every $y \in Y$ this measure is atomic because (X, μ_y, T) is a rotation. For every $t \in \mathbb{T}$ and every $y \in Y$, the limit

$$F(y, t) := \lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \widehat{\sigma_y}(n)\, \mathrm{e}(-nt)$$

exists and

$$F(y, t) = \sigma_y(\{t\}).$$

Since, for every $n \in \mathbb{N}$, the function $T^n f \cdot \overline{f}$ is Borel, we get that the map $y \mapsto \widehat{\sigma_y}(n)$ is Borel on Y. Thus, the function F satisfies the hypothesis of Proposition 5.6 for the Polish space $Y \times \mathbb{T}$. Henceforth, we use the notation of this proposition and let $t_j\colon Y \to \mathbb{T}$, $j \in \mathbb{N}$, be the Borel maps obtained. For $j \in \mathbb{Z}_+$, let

$$A_j := \{y \in Y\colon N(y) > j\} = \{y \in Y : \sigma_y(\{t_j(y)\}) > 0\},$$
$$\lambda_j(y) := \mathbf{1}_{A_j}(y)\, \mathrm{e}(t_j(y)) \quad \text{for} \quad y \in Y.$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

For $y \in A_j$, $j \in \mathbb{Z}_+$, $\lambda_j(y) \in P_y$: that is, $\sigma_y(\{t_j(y)\}) > 0$. By the Wiener–Wintner theorem, the limit

$$\psi_j(x) := \lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} f(T^n x)\,\overline{\lambda_j(\pi(x))}^n$$

exists in $L^2(\mu)$ (and for μ-almost every $x \in X$). We remark that if $\nu(A_j) = 0$, then the function ψ_j is equal to zero μ-almost everywhere, and the same holds for the functions θ_j and ϕ_j defined below. For ν-almost every $y \in A_j$,

$$\begin{aligned}\int f\,\overline{\psi_j}\,d\mu_y &= \lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}\int f\cdot T^n\overline{f}d\mu_y\cdot\lambda_j(y)^n\\ &= \lim_{N\to+\infty}\frac{1}{N}\sum_{n=1}^{N}\widehat{\sigma_y}(-n)\,\lambda_j(y)^n = \sigma_y(\{t_j(y)\}).\end{aligned} \tag{5.10}$$

Furthermore, for $j \in \mathbb{Z}_+$, the function ψ_j satisfies

$$\psi_j(Tx) = \psi_j(x)\,\lambda_j(\pi(x)). \tag{5.11}$$

It follows that $|\psi_j|$ is T-invariant. Expressing $|\psi_j|^2$ as

$$|\psi_j|^2 = \theta_j\circ\pi \quad \text{for some } \theta_j \in L^\infty(Y,\nu), \tag{5.12}$$

a similar computation gives

$$\begin{aligned}\theta_j(y) = \int |\psi_j|^2\,d\mu_y &= \lim_{N\to+\infty}\frac{1}{N^2}\sum_{m,n\in[N]}\int T^n f\cdot T^m\overline{f}d\mu_y\cdot\lambda_j(y)^m\,\overline{\lambda_j(y)}^n\\ &= \sigma_y(\{t_j(y)\}) > 0 \quad \text{for } \nu\text{-almost every } y \in A_j.\end{aligned} \tag{5.13}$$

We let

$$\phi_j(x) := \begin{cases} |\psi_j(x)|^{-1}\psi_j(x) & \text{for } x \in \pi^{-1}(A_j),\\ 0 & \text{for } x \notin \pi^{-1}(A_j).\end{cases}$$

It is immediate to check that ϕ_j is an eigenfunction for the eigenvalue λ_j. By construction, for $i \neq j$, we have $\lambda_i(x) \neq \lambda_j(x)$ for μ-almost every $x \in X$, except when $\lambda_i(x) = \lambda_j(x) = 0$: thus $\lambda_i(x)\overline{\lambda_j(x)} \neq 1$ μ-almost everywhere. On the other hand, $\mathbb{E}_\mu(\phi_i\overline{\phi_j} \mid \mathcal{I}(T)) = \lambda_i\overline{\lambda_j}\mathbb{E}_\mu(\phi_i\overline{\phi_j} \mid \mathcal{I}(T))$. Therefore,

$$\mathbb{E}_\mu(\phi_i\cdot\overline{\phi_j} \mid \mathcal{I}(T)) = 0 \quad \text{for all } i \neq j,\ i, j \in \mathbb{Z}_+.$$

Furthermore, for every $j \in \mathbb{Z}_+$, by the definition of the ϕ_j, $\mathbb{E}_\mu(|\phi_j|^2|\mathcal{I}(T))$ takes the values zero or one. Hence $(\phi_j)_{j\in\mathbb{Z}_+}$ is a relative orthonormal system.

Next, we establish identity (5.9). Arguing as in Part (iii) of Proposition 5.1, it suffices to show that

$$\|f\|^2_{L^2(\mu)} = \sum_{j\in\mathbb{Z}_+}\|\mathbb{E}_\mu(f\cdot\overline{\phi_j} \mid \mathcal{I}(T))\|^2_{L^2(\mu)}. \tag{5.14}$$

Using the definition of the function ϕ_j and then (5.12) combined with (5.13), we get

$$\begin{aligned}\sum_{j\in\mathbb{Z}_+}\left|\int f\cdot\overline{\phi_j}\,d\mu_y\right|^2 &= \sum_{j\in\mathbb{Z}_+}\mathbf{1}_{A_j}(y)\frac{1}{|\psi_j|^2}\left|\int f\cdot\overline{\psi_j}\,d\mu_y\right|^2\\ &= \sum_{j\in\mathbb{Z}_+}\mathbf{1}_{A_j}(y)\frac{1}{\sigma_y(\{t_j(y)\})}\left|\int f\cdot\overline{\psi_j}\,d\mu_y\right|^2.\end{aligned}$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

By (5.10), for ν-almost every $y \in Y$, the last sum is equal to

$$\sum_{j \in \mathbb{Z}_+} \mathbf{1}_{A_j}(y)\, \sigma_y(\{t_j(y)\}) = \sum_{j \in \mathbb{Z}_+} \sigma_y(\{t_j(y)\}) = \sigma_y(\mathbb{T}) = \widehat{\sigma_y}(0) = \int |f|^2 \, d\mu_y,$$

where we used the definition of the set A_j to get the first identity and, to get the second identity, we used the defining property of the maps t_j, $j \in \mathbb{Z}_+$, and that, for ν-almost every $y \in Y$, the measure σ_y is atomic. Integrating the established identity over Y with respect to ν, we obtain (5.14).

Finally, the last claim of the lemma follows by the construction of the functions ψ_j and ϕ_j for $j \in \mathbb{Z}_+$. This completes the proof. □

We are now ready to complete the proof of Theorem 5.2.

End of proof of Theorem 5.2. Let $(f_k)_{k \in \mathbb{N}}$ be a dense sequence in $L^2(\mu)$. For every $k \in \mathbb{N}$, we build, by induction, a countable family $\mathcal{F}_k$ of functions in $L^2(\mu)$ such that, for every $k \in \mathbb{N}$:

(i) $\mathcal{F}_1 \cup \cdots \cup \mathcal{F}_k$ is a relative orthonormal system of eigenfunctions; and

(ii) the function f_k belongs to the closed subspace $\mathcal{H}_k$ of $L^2(\mu)$ spanned by all functions of the form $\phi\, w$, where $\phi \in \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_k$ and $w \in L^\infty(\mu)$ varies over all T-invariant functions.

For $k = 1$, the result is given by Lemma 5.8 with $f = f_1$. Suppose that the result holds for $k \in \mathbb{N}$; we shall show that it holds for $k+1$. We can decompose f_{k+1} as

$$f_{k+1} = g + f \quad \text{where } g \in \mathcal{H}_k \text{ and } f \perp \mathcal{H}_k.$$

The space $\mathcal{H}_k$ is invariant under multiplication by bounded T-invariant functions and thus $\mathbb{E}_\mu(f \cdot \overline{h} \mid \mathcal{I}(T)) = 0$ for every $h \in \mathcal{H}_k$. On the other hand, every $\phi \in \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_k$ is an eigenfunction, and thus, by definition, $T\phi$ belongs to the space $\mathcal{H}_k$. It follows that $\mathcal{H}_k$ is invariant under T and thus

$$\mathbb{E}_\mu(T^n f \cdot \overline{h} \mid \mathcal{I}(T)) = 0 \quad \text{for every } n \in \mathbb{N} \text{ and every } h \in \mathcal{H}_k. \tag{5.15}$$

Applying Lemma 5.8 to the function f, we obtain a relative orthonormal system $\mathcal{F}_{k+1} = (\phi_j)_{j \in \mathbb{N}}$ of eigenfunctions such that f belongs to the closed linear span of all functions of the form $\phi_j w$, where $j \in \mathbb{N}$ is arbitrary and $w \in L^\infty(\mu)$ varies over all T-invariant functions. Moreover, for every $j \in \mathbb{N}$, ϕ_j belongs to the smallest T-invariant subspace of $L^2(\mu)$ containing all functions of the form $f w$, where w varies over all T-invariant functions in $L^\infty(\mu)$. Hence, by (5.15), for every $j \in \mathbb{N}$ and every $h \in \mathcal{H}_k$, $\mathbb{E}_\mu(\phi_j \cdot \overline{h} \mid \mathcal{I}(T)) = 0$; this holds, in particular, for all functions h belonging to $\mathcal{F}_1 \cup \cdots \cup \mathcal{F}_k$. Therefore, $\mathcal{F}_1 \cup \cdots \cup \mathcal{F}_k \cup \mathcal{F}_{k+1}$ is a relative orthonormal system.

The closed subspace of $L^2(\mu)$ spanned by functions of the form $\phi\, w$, where $\phi \in \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_k \cup \mathcal{F}_{k+1}$ and $w \in L^\infty(\mu)$, varies over all T-invariant functions, contains f and g, and thus f_{k+1}. This completes the induction.

We choose an enumeration of the countable set $\mathcal{F} := \bigcup_{k=1}^\infty \mathcal{F}_k$ and write it as $(\phi'_j)_{j \in \mathbb{N}}$; then $(\phi'_j)_{j \in \mathbb{N}}$ is a relative orthonormal system. The closed subspace of $L^2(\mu)$, spanned by all functions of the form $\phi'_j\, w$, where $j \in \mathbb{N}$ and $w \in L^\infty(\mu)$, varies over all T-invariant functions, contains the functions f_k for every $k \in \mathbb{N}$, and thus is equal to $L^2(\mu)$. Hence, $(\phi'_j)_{j \in \mathbb{N}}$ is a relative orthonormal basis of eigenfunctions, which completes the proof. □

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

6. *Decomposition of correlation sequences*

In this section, we prove the decomposition results stated in §§2.4 and 2.5.

6.1. *Anti-uniformity in norm.*

We start with some preparatory results. The main tool in verifying anti-uniformity properties is the following inner product space variant of a well-known estimate of van der Corput (for a proof, see [**10**]).

VAN DER CORPUT LEMMA *Let $d \in \mathbb{N}$, $\mathcal{H}$ be an inner product space, let $\xi : \mathbb{N}^d \to \mathcal{H}$ be a bounded sequence and let* $\mathbf{I}$ *be a Følner sequence in $\mathbb{N}^d$. Then*

$$\limsup \|\mathrm{Av}_{\mathbf{I}}\, \xi_{\mathbf{n}}\|^2 \le 4 \limsup_{H\to+\infty} \frac{1}{H^d} \sum_{\mathbf{h}\in[H]^d} \limsup |\mathrm{Av}_{\mathbf{n},\mathbf{I}}\, \langle \xi_{\mathbf{n}+\mathbf{h}}, \xi_{\mathbf{n}}\rangle|.$$

PROPOSITION 6.1. *Let $d, \ell, s, t \in \mathbb{N}$. Then there exists a positive integer $k = k(d, \ell, s, t)$ such that, for every system $(X, \mu, T_1, \ldots, T_\ell)$, functions $f_1, \ldots, f_s \in L^\infty(\mu)$ bounded by one, polynomial mappings $\vec{p_1}, \ldots, \vec{p_s} : \mathbb{N}^d \to \mathbb{Z}^\ell$ of degree at most t and sequence $w \in \ell^\infty(\mathbb{N}^d)$,*

$$\limsup \|\mathrm{Av}\, w(\mathbf{n})\, T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s\|_{L^2(\mu)} \le 4\|w\|_{U^{k+1}(\mathbb{N}^d)}.$$

Furthermore, if the polynomial mappings are linear, then we can take $k = s$.

Sketch of the proof. Let $\mathbf{I} = (I_i)_{j\in\mathbb{N}}$ be a Følner in $\mathbb{N}^d$. After passing to a subsequence, we can assume that the sequence w admits correlations along $\mathbf{I}$. It suffices to show that

$$\limsup \|\mathrm{Av}_{\mathbf{I}}\, w(\mathbf{n}) \cdot T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s\|_{L^2(\mu)} \le 4\|w\|_{\mathbf{I},k+1}$$

for some $k = k(d, \ell, s, t)$. To verify this, one applies an inductive argument, often called PET (polynomial exhaustion technique) induction, introduced by Bergelson in [**10**]. Each step uses the van der Corput lemma in $L^2(\mu)$, invariance of the measure under some of the transformations and the Cauchy–Schwarz inequality. The details are similar to several other arguments in the literature (see, for example, the proof of [**32**, Lemma 3.5]) and so we do not give the proof.

In the case of linear polynomials, it can be shown, by induction, on $s \in \mathbb{Z}_+$ that one can take $k = s$; for $s = 0$, the statement is trivial and the inductive step can be carried out as the inductive step in the proof of Theorem 2.6, below. □

6.2. *Proof of Theorem 2.5.*

By Theorem 3.10, it suffices to show that the sequence $a : \mathbb{N}^d \to \mathbb{C}$, given by

$$a(\mathbf{n}) := \int f_0 \cdot T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s \, d\mu, \quad \mathbf{n} \in \mathbb{N}^d,$$

is k-regular and $(k+1)$-anti-uniform for some $k \in \mathbb{N}$ that depends only on the integers d, ℓ, s, t. The regularity is given by Proposition 2.1 and the anti-uniformity by Proposition 6.1. □

6.3. *Proof of Theorem 2.6.* We start with some preparatory results.

LEMMA 6.2. *Let $d, \ell \in \mathbb{N}$ and $L_1, \ldots, L_\ell : \mathbb{N}^d \to \mathbb{Z}$ be linearly independent linear forms. Then there exists a constant $C := C(d, L_1, \ldots, L_\ell)$ such that the following holds: if $(X, \mu, T_1, \ldots, T_\ell)$ is a system and $f_0, \ldots, f_\ell \in L^\infty(\mu)$ are functions bounded by one, then*

$$\lim \text{Av} \left| \int f_0 \cdot T_1^{L_1(\mathbf{n})} f_1 \cdot \cdots \cdot T_\ell^{L_\ell(\mathbf{n})} f_\ell \, d\mu \right|^2 \leq C \min_{1 \leq i \leq \ell} |||f_i|||^2_{T_i,\mu,2}. \tag{6.1}$$

Proof. We first note that the limit on the left-hand side of (6.1) can be rewritten as

$$\lim \text{Av} \left(\int (f_0 \otimes \overline{f_0}) \cdot \prod_{i=1}^{\ell} (T_i \times T_i)^{L_i(\mathbf{n})} (f_i \otimes \overline{f_i}) \, d(\mu \times \mu) \right)$$

and thus exists, by Theorem 1.1. Therefore, in (6.1), we can restrict to averages taken on the cubes $[N]^d$, $N \in \mathbb{N}$: that is, it suffices to obtain bounds for the limit

$$\lim_{N \to +\infty} \frac{1}{N^d} \sum_{\mathbf{n} \in [N]^d} \left| \int f_0 \cdot T_1^{L_1(\mathbf{n})} f_1 \cdot \cdots \cdot T_\ell^{L_\ell(\mathbf{n})} f_\ell \, d\mu \right|^2. \tag{6.2}$$

Next, we claim that it suffices to consider the case where $d = \ell$. Indeed, since the linear forms are linearly independent, $d \geq \ell$ and, if $d > \ell$, then there exist linear forms $L_{\ell+1}, \ldots, L_d : \mathbb{N}^d \to \mathbb{Z}$ such that the linear forms $L_1, \ldots, L_d$ are linearly independent. Then applying the $d = \ell$ case of the result for this set of linear forms and the functions $f_0', \ldots, f_d'$, defined by $f_i' := f_i$ for $i = 0, \ldots, \ell$ and $f_i' := 1$ for $i = \ell + 1, \ldots, d$, we get the asserted estimate. Henceforth, we assume that $d = \ell$.

Let $\overrightarrow{L} : \mathbb{N}^d \to \mathbb{N}^d$ be defined by $\overrightarrow{L}(\mathbf{n}) := (L_1(\mathbf{n}), \ldots, L_d(\mathbf{n}))$, $\mathbf{n} \in \mathbb{N}^d$. Since the linear forms $L_1, \ldots, L_d$ are linearly independent, the linear map $\overrightarrow{L}$ is injective. Furthermore, there exists a positive integer $M = M(L_1, \ldots, L_d)$ such that $\overrightarrow{L}([N]^d) \subset [-MN, MN]^d$ for every $N \in \mathbb{N}$. This easily implies that the limit in (6.2) is bounded by

$$\begin{aligned} &(3M)^d \lim_{N \to +\infty} \frac{1}{(2N+1)^d} \sum_{\mathbf{n} \in [-N,N]^d} \left| \int f_0 \cdot T_1^{n_1} f_1 \cdot \cdots \cdot T_d^{n_d} f_d \, d\mu \right|^2 \\ &= (3M)^d \lim_{N \to +\infty} \frac{1}{(2N+1)^d} \sum_{\mathbf{n} \in [-N,N]^d} \int (f_0 \otimes \overline{f_0}) \cdot \prod_{i=1}^{d} (T_i \times T_i)^{n_i} (f_i \otimes \overline{f_i}) \, d(\mu \times \mu). \end{aligned}$$

By the ergodic theorem, the last limit is equal to

$$\int (f_0 \otimes \overline{f_0}) \cdot \prod_{i=1}^{d} \mathbb{E}_\mu(f_i \otimes \overline{f_i} \mid (\mathcal{I}(T_i \times T_i)) \, d(\mu \times \mu).$$

For $i = 1, \ldots, d$, this quantity is bounded by

$$\left\| \mathbb{E}_\mu \big(f_i \otimes \overline{f_i} \mid (\mathcal{I}(T_i \times T_i) \big) \right\|_{L^2(\mu)} = |||f_i \otimes \overline{f_i}|||_{T_i \times T_i, \mu \times \mu, 1} \leq |||f_i|||^2_{T_i,\mu,2},$$

where we used that all the functions are bounded by one and the estimate (A.5). This completes the proof. □

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

PROPOSITION 6.3. *Let $d, \ell \in \mathbb{N}$ and $L_1, \dots, L_\ell : \mathbb{N}^d \to \mathbb{Z}$ be linearly independent linear forms. Then, for every $\varepsilon > 0$, there exists a constant $C := C(d, L_1, \dots, L_\ell, \varepsilon) > 0$ such that the following holds: if $(X, \mu, T_1, \dots, T_\ell)$ is a system and $f_0, \dots, f_\ell \in L^\infty(\mu)$ are functions bounded by one, then, for every sequence $w \in \ell^\infty(\mathbb{N}^d)$,*

$$\limsup \left| \mathrm{Av}\left(w(\mathbf{n}) \cdot \int f_0 \cdot T_1^{L_1(\mathbf{n})} f_1 \cdot \dots \cdot T_\ell^{L_\ell(\mathbf{n})} f_\ell \, d\mu \right) \right| \le C \|w\|_{U^2(\mathbb{N}^d)} + \varepsilon \|w\|_\infty. \tag{6.3}$$

Remark. This proves that the correlation sequence defined by the integral is 2-anti-uniform with constants C that do not depend on the functions $f_0, \dots, f_\ell$ as long as they are bounded by one. This condition is essential in the proof of Theorem 2.6, and Theorem 5.2 is key in establishing the condition.

Proof. Let $\varepsilon \in (0, 1)$, $w \in \ell^\infty(\mathbb{N}^d)$ and $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ be a Følner sequence in $\mathbb{N}^d$. By passing to a Følner subsequence, we can assume that w admits correlations along $\mathbf{I}$.

By the defining property of the factor $\mathcal{Z}_1$ (see §A.2), $|||f_i - \mathbb{E}_\mu(f \mid \mathcal{Z}_1(X, \mu, T_i))|||_{\mu,T_i,2} = 0$ for $i = 1, \dots, \ell$. Hence, by Lemma 6.2, the lim sup in (6.3) remains unchanged if we replace each function f_i with $\mathbb{E}_\mu(f_i \mid \mathcal{Z}_1(X, \mu, T_i))$. Therefore, we can, and will, assume that for $i = 1, \dots, \ell$, the function f_i is measurable with respect to $\mathcal{Z}_1(X, \mu, T_i)$. For $i = 1, \dots, \ell$, let $\mu = \int \mu_{i,x} \, d\mu(x)$ be the ergodic decomposition of the system (X, μ, T_i).

By Theorem 5.2, for $i = 1, \dots, \ell$, the space $L^2(\mathcal{Z}_1(X, \mu, T_i), \mu)$ admits a relative orthonormal basis $(\phi_{i,j})_{j\in\mathbb{N}}$ such that $\phi_{i,j}$ is an eigenfunction of (X, μ, T_i) with eigenvalue $\lambda_{i,j}$ for $j \in \mathbb{N}$. We write $(f_{i,j})_{j\in\mathbb{N}}$ for the coordinates of f_i in this base. We recall that $f_{i,j}$ is invariant under T_i and that

$$f_{i,j} = \mathbb{E}_\mu(f_i \, \overline{\phi_{i,j}} \mid \mathcal{I}(T_i)).$$

Then $\|f_{i,j}\|_{L^\infty(\mu)} \le 1$ for all $i, j \in \mathbb{N}$. Moreover, by part (ii) of Proposition 5.1,

$$\mathbb{E}_\mu(|f_i|^2 \mid \mathcal{I}(T_i)) = \sum_{j\in\mathbb{N}} |f_{i,j}|^2, \quad \mu\text{-almost everywhere,} \tag{6.4}$$

$$f_i = \sum_{j\in\mathbb{N}} f_{i,j} \, \phi_{i,j}, \tag{6.5}$$

where convergence in (6.4) is pointwise and in (6.5) is in $L^2(\mu)$ and in $L^2(\mu_{i,x})$ μ-almost everywhere.

For $i = 1, \dots, \ell$, we separate the series (6.5) in two parts. For $x \in X$, we let

$$E_i(x) := \{j \in \mathbb{N} : |f_{i,j}(x)|^2 \ge \varepsilon^{10^i}\},$$

$$g_i(x) := \sum_{j\in E_i(x)} f_{i,j}(x) \, \phi_{i,j}(x) \quad \text{and} \quad h_i := f_i - g_i.$$

By (6.4) and since all functions are bounded by one,

$$|E_i(x)| \le \varepsilon^{-10^i} \quad \mu\text{-almost everywhere,} \quad \|g_i\|_{L^\infty(\mu)} \le \varepsilon^{-10^i}, \quad i = 1, \dots, \ell. \tag{6.6}$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

Furthermore, since $f_{i,j}$ are T_i-invariant, $E_i(T_i x) = E_i(x)$ μ-almost everywhere and the set $A_{i,j} = \{x \in X : j \notin E_i(x)\}$ is invariant under T_i.

$$h_i = \sum_{j \in \mathbb{N}} \mathbf{1}_{A_{i,j}} \, f_{i,j} \, \phi_{i,j}$$

and thus the coordinates of the function h_i in the base $(\phi_{i,j})_{j \in \mathbb{N}}$ are the functions $\mathbf{1}_{A_{i,j}} f_{i,j}$. By Proposition 5.3, we obtain

$$\begin{aligned} |||h_i|||^4_{T_i, \mu_{i,x}, 2} &= \sum_{j \in \mathbb{N}} |\mathbf{1}_{A_{i,j}}(x) f_{i,j}(x)|^4 \\ &= \sum_{j \notin E_i(x)} |f_{i,j}(x)|^4 \quad \mu\text{-almost everywhere}, \ i = 1, \ldots, \ell. \end{aligned} \tag{6.7}$$

Therefore,

$$\begin{aligned} |||h_i|||^4_{T_i, \mu, 2} &= \int |||h_i|||^4_{T_i, \mu_{i,x}, 2} \, d\mu \\ &= \int \sum_{j \notin E_i(x)} |f_{i,j}(x)|^4 \, d\mu \le \varepsilon^{10^i} \int \sum_{j \in \mathbb{N}} |f_{i,j}(x)|^2 \, d\mu \le \varepsilon^{10^i}, \end{aligned}$$

where we used (A.3) in the appendix to get the first identity, (6.7) to get the second identity, the definition of the sets $E_i(x)$ to get the first estimate and (6.4) combined with the fact that the functions f_i are bounded by one to get the last estimate.

Let C be the constant defined in Lemma 6.2. Combining this lemma with the preceding estimates we deduce, for $m = 1, \ldots, \ell$, that

$$\begin{aligned} &\limsup \operatorname{Av}_{\mathbf{I}} \left| \int f_0 \cdot \left(\prod_{i=1}^{m-1} T_i^{L_i(\mathbf{n})} g_i \right) \cdot T_m^{L_m(\mathbf{n})} h_m \cdot \left(\prod_{i=m+1}^{\ell} T_i^{L_i(\mathbf{n})} f_i \right) d\mu \right|^2 \\ &\quad \le C \prod_{i=1}^{m-1} \|g_i\|^2_{L^\infty(\mu)} \cdot |||h_m|||^2_{T_m, \mu, 2} \le C \varepsilon^{-2 \sum_{i=1}^{m-1} 10^i} \, \varepsilon^{5 \cdot 10^{m-1}} \le C \varepsilon^2. \end{aligned}$$

Using the Cauchy–Schwarz inequality and telescoping, we obtain

$$\begin{aligned} &\limsup \left| \operatorname{Av}_{\mathbf{I}} \left(w(\mathbf{n}) \int f_0 \cdot \prod_{i=1}^{\ell} T_i^{L_i(\mathbf{n})} f_i \, d\mu - w(\mathbf{n}) \int f_0 \cdot \prod_{i=1}^{\ell} T_i^{L_i(\mathbf{n})} g_i \, d\mu \right) \right| \\ &\quad \le \ell C^{1/2} \varepsilon \|w\|_\infty. \end{aligned}$$

On the other hand, using the definition of the functions $g_1, \ldots, g_\ell$ and recalling that, for $i = 1, \ldots, \ell$ and $j \in \mathbb{N}$, the function $\phi_{i,j}$ is a T_i-eigenfunction with eigenvalue $\lambda_{i,j}$, we get

$$\begin{aligned} \int f_0 \cdot \prod_{i=1}^{\ell} T_i^{L_i(\mathbf{n})} g_i \, d\mu &= \int \sum_{j_1 \in E_1, \ldots, j_\ell \in E_\ell} f_0 \cdot \prod_{i=1}^{\ell} f_{i,j_i} \cdot T_i^{L_i(\mathbf{n})} \phi_{i,j_i} \, d\mu \\ &= \int \sum_{j_1 \in E_1, \ldots, j_\ell \in E_\ell} g_{j_1, \ldots, j_\ell} \cdot \prod_{i=1}^{\ell} \lambda_{i,j_i}^{L_i(\mathbf{n})} \, d\mu, \end{aligned}$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

where

$$g_{j_1,\dots,j_\ell} := f_0 \cdot \prod_{i=1}^{\ell} f_{i,j_i} \cdot \phi_{i,j_i}.$$

Since for μ-almost every $x \in X$ we have, by (6.6), that $|E_i(x)| \le \varepsilon^{-10^i}$ for $i = 1, \dots, \ell$, we deduce that, for μ-almost every $x \in X$, the sum contains at most $\varepsilon^{-10^{\ell+1}}$ terms. Moreover, since $L_1, \dots, L_\ell$ are linear, using the van der Corput lemma on $\mathbb{C}$ and the fact that the functions $g_{j_1,\dots,j_\ell}$ are bounded by one, we have the pointwise estimate

$$\begin{aligned}\text{limsup}\,\Big|\text{Av}_{\mathbf{I}}\, w(\mathbf{n}) \cdot \prod_{i=1}^{\ell} \lambda_{i,j_i}^{L_i(\mathbf{n})}\Big|^2 &\le 4 \limsup_{H\to+\infty} \frac{1}{H^d} \sum_{\mathbf{h}\in[H]^d} \Big|\lim \text{Av}_{\mathbf{n},\mathbf{I}}\; w(\mathbf{n}+\mathbf{h})\overline{w(\mathbf{n})}\Big| \\ &\le 4 \lim_{H\to+\infty} \frac{1}{H^d} \sum_{\mathbf{h}\in[H]^d} \|\sigma_{\mathbf{h}}\, w \cdot \overline{w}\|_{\mathbf{I},1} \\ &\le 4\|w\|_{\mathbf{I},2}^2 \le 4\|w\|_{U^2(\mathbb{N}^d)}^2,\end{aligned}$$

where we used (3.2) and (3.3). Combining the above estimates, we deduce that the left-hand side of (6.3) is bounded by $2\varepsilon^{-10^{\ell+1}}\|w\|_{U^2(\mathbb{N}^d)} + \ell C^{1/2}\varepsilon\|w\|_\infty$. This completes the proof. □

End of proof of Theorem 2.6. We can assume that the functions $f_0, \dots, f_\ell$ are bounded by one. Furthermore, we can extract a linearly independent subset of r elements of $\{L_1, \dots, L_\ell\}$; hence, after reordering the linear forms, we can assume that the first r ones are linearly independent.

Let $k := \ell - r + 1$. In order to prove that the sequence $a\colon \mathbb{N}^d \to \mathbb{C}$, defined by

$$a(\mathbf{n}) := \int f_0 \cdot T_1^{L_1(\mathbf{n})} f_1 \cdot \dots \cdot T_\ell^{L_\ell(\mathbf{n})} f_\ell \, d\mu, \quad \mathbf{n} \in \mathbb{N}^d,$$

admits a decomposition of the announced type, it suffices, by Theorem 3.10, to show that it is k-regular and $(k+1)$-anti-uniform.

The regularity follows from Proposition 2.1.

Next, we verify $(k+1)$-anti-uniformity. Let $C = C(d, L_1, \dots, L_r, \varepsilon)$ be the constant defined by Proposition 6.3. We can assume that $C > 1$. For fixed $d \in \mathbb{N}$, we will prove, by induction on $k \in \mathbb{N}$, that if the functions $f_0, \dots, f_\ell$ are bounded by one, then, for every $\varepsilon > 0$ and every $w \in \ell^\infty(\mathbb{N}^d)$ with $\|w\|_\infty = 1$,

$$\text{limsup}\,|\text{Av}\, a(\mathbf{n})\, w(\mathbf{n})| \le 4C\|b\|_{U^{k+1}(\mathbb{N}^d)} + 4\varepsilon^{1/2^{k-1}}.$$

It will then follow that the sequence a is $(k+1)$-anti-uniform with anti-uniformity constant $C' := 4C(d, L_1, \dots, L_r, \frac{1}{4}\varepsilon^{2^{k-1}})$.

For $k = 1$, the statement follows from Proposition 6.3.

Suppose that $k \ge 2$ and that the statement holds for $k - 1$. Let $\varepsilon > 0$ and $w \in \ell^\infty(\mathbb{N}^d)$. Let also $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$ be a Følner sequence in $\mathbb{N}^d$. By passing to a Følner subsequence, we can assume that w admits correlations along $\mathbf{I}$.

Composing with $T_\ell^{-L_\ell(\mathbf{n})}$, we rewrite the sequence $a\colon \mathbb{N}^d \to \mathbb{C}$ as†

$$a(\mathbf{n}) = \int T_\ell^{-L_\ell(\mathbf{n})} f_0 \cdot T_\ell^{-L_\ell(\mathbf{n})} T_1^{L_1(\mathbf{n})} f_1 \cdot \dots \cdot T_\ell^{-L_\ell(\mathbf{n})} T_{\ell-1}^{L_{\ell-1}(\mathbf{n})} f_{\ell-1} \cdot f_\ell \, d\mu, \quad \mathbf{n} \in \mathbb{N}^d.$$

† This maneuver is necessary; a direct application of the van der Corput lemma produces a weaker estimate involving a seminorm of order $k+2$.

Using the Cauchy–Schwarz inequality and that $\|f_\ell\|_{L^\infty(\mu)} \le 1$, we get that

$$\limsup |\mathrm{Av}_{\mathbf{I}}\, a(\mathbf{n})\, w(\mathbf{n})|^2 \le \limsup \|\mathrm{Av}_{\mathbf{I}}\, \xi_{\mathbf{n}}\|^2_{L^2(\mu)}, \tag{6.8}$$

where $\xi : \mathbb{N}^d \to L^2(\mu)$ is defined by

$$\xi_{\mathbf{n}} := w(\mathbf{n}) \cdot T_\ell^{-L_\ell(\mathbf{n})} f_0 \cdot T_\ell^{-L_\ell(\mathbf{n})} T_1^{L_1(\mathbf{n})} f_1 \cdot \cdots \cdot T_\ell^{-L_\ell(\mathbf{n})} T_{\ell-1}^{-L_{\ell-1}(\mathbf{n})} f_{\ell-1}, \quad \mathbf{n} \in \mathbb{N}^d.$$

Using van der Corput's lemma in $L^2(\mu)$ for the sequence ξ, we obtain that the right-hand side of (6.8) is bounded by

$$4 \limsup_{H\to+\infty} \frac{1}{H^d} \sum_{\mathbf{h}\in[H]^d} \limsup |\mathrm{Av}_{\mathbf{n},\mathbf{I}}\, \langle \xi_{\mathbf{n}+\mathbf{h}}, \xi_{\mathbf{n}}\rangle|. \tag{6.9}$$

Recall that $\sigma_{\mathbf{h}} w(\mathbf{n}) = w(\mathbf{n}+\mathbf{h})$ for all $\mathbf{h}, \mathbf{n} \in \mathbb{N}^d$. A simple computation gives that, for every $\mathbf{h} \in \mathbb{N}^d$,

$$\frac{1}{|I_j|} \sum_{\mathbf{n}\in I_j} \langle \xi_{\mathbf{n}+\mathbf{h}}, \xi_{\mathbf{n}}\rangle = \frac{1}{|I_j|} \sum_{\mathbf{n}\in I_j} \sigma_{\mathbf{h}} w(\mathbf{n}) \cdot \overline{w(\mathbf{n})} \times \int \widetilde{f}_{0,\mathbf{h}} \cdot T_1^{L_1(\mathbf{n})} \widetilde{f}_{1,\mathbf{h}} \cdot \cdots \cdot T_{\ell-1}^{L_{\ell-1}(\mathbf{n})} \widetilde{f}_{\ell-1}\, d\mu, \tag{6.10}$$

where $\widetilde{f}_{j,\mathbf{h}} := T_\ell^{-L_\ell(\mathbf{h})} T_j^{L_j(\mathbf{h})} f_j \cdot \overline{f_j}$ for $j = 0, \ldots, \ell-1$ and $L_0 := 0$.

Note that, for every $\mathbf{h} \in \mathbb{N}^d$, the sequence $(w(\mathbf{n}+\mathbf{h})\, \overline{w(\mathbf{n})})_{\mathbf{n}\in\mathbb{N}^d}$ admits correlations along $\mathbf{I}$ and that $\|\widetilde{f}_{j,\mathbf{h}}\|_{L^\infty(\mu)} \le 1$ for $j = 0, \ldots, \ell-1$. The expression on the right-hand side of (6.10) is thus of the type studied, with $(\ell-1)$ in place of ℓ. Hence, the induction hypothesis applies and gives that, for every $\mathbf{h} \in \mathbb{N}^d$,

$$\limsup |\mathrm{Av}_{\mathbf{n},\mathbf{I}}\, \langle \xi_{\mathbf{n}+\mathbf{h}}, \xi_{\mathbf{n}}\rangle| \le 4\, C\, \|\sigma_{\mathbf{h}} w \cdot w\|_{\mathbf{I},k} + 4\, \varepsilon^{1/2^{k-2}}.$$

It is important, for the last part of the argument, that the constant C is independent of the parameter $\mathbf{h}$. Combining the above, we deduce that the expression (6.9) is bounded by

$$\begin{aligned} &16\, C \limsup_{H\to+\infty} \frac{1}{H^d} \sum_{\mathbf{h}\in[H]^d} \|\sigma_{\mathbf{h}} w \cdot \overline{w}\|_{\mathbf{I},k} + 16\, \varepsilon^{1/2^{k-2}} \\ &\le 16\, C \limsup_{H\to+\infty} \bigg(\frac{1}{H^d} \sum_{\mathbf{h}\in[H]^d} \|\sigma_{\mathbf{h}} w \cdot \overline{w}\|^{2^k}_{\mathbf{I},k} \bigg)^{1/2^k} + 16\, \varepsilon^{1/2^{k-2}} \\ &= 16\, C\, \|w\|^2_{\mathbf{I},k+1} + 16\, \varepsilon^{1/2^{k-2}}, \end{aligned}$$

where the last identity follows from the inductive property (3.3) of the uniformity seminorms. Putting together the previous estimates and taking square roots, we get the announced bound. This completes the induction and the proof. □

6.4. *Proof of Theorem 2.8.* We use the variant of Furstenberg's correspondence principle, which we have already used in §3.1 for a single sequence. In the case of several sequences, it is proved in a similar fashion and gives the following proposition.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

PROPOSITION 6.4. *Let $\ell, s \in \mathbb{N}$ and $a_1, \ldots, a_s : \mathbb{Z}^\ell \to \mathbb{C}$ be bounded sequences such that the family $\mathcal{F} = \{a_1, \ldots, a_s\}$ admits correlations along the Følner sequence* $\mathbf{I}$ *in* $\mathbb{N}^\ell$. *Then there exist a topological dynamical system $(X, T_1, \ldots, T_\ell)$, where $T_1, \ldots, T_\ell$ are commuting homeomorphisms, functions $f_1, \ldots, f_s \in C(X)$ and a Borel probability measure μ on X that is T_i-invariant for $i = 1, \ldots, \ell$, such that*

$$\int T_{\vec{n_1}} f_1 \cdot \cdots \cdot T_{\vec{n_s}} f_s \, d\mu = \lim \mathrm{Av}_{\mathbf{k},\mathbf{I}} \left(\prod_{i=1}^{s} a_i(\mathbf{k} + \vec{n_i}) \right)$$

for every $\vec{n_1}, \ldots, \vec{n_s} \in \mathbb{Z}^\ell$.

Combining this with Theorem 2.5, we immediately deduce Theorem 2.8. □

6.5. *Proof of Theorem 2.9.* We start by recalling the definition of the Gowers norms in $\mathbb{Z}_N^d$.

Definition. Let $d, N \in \mathbb{N}$ and $f : \mathbb{Z}_N^d \to \mathbb{C}$ be a function. For every $\mathbf{h} \in \mathbb{Z}_N^d$, we write $f_{\mathbf{h}}(\mathbf{n}) := f(\mathbf{n} + \mathbf{h})$, $\mathbf{n} \in \mathbb{Z}_N^d$. For $s \in \mathbb{N}$, we denote by $\|f\|_{U^s(\mathbb{Z}_N^d)}$ the *Gowers* $U^s(\mathbb{Z}_N^d)$-*norm* of f that is defined inductively as follows: we let

$$\|f\|_{U^1(\mathbb{Z}_N^d)} := |\mathbb{E}_{\mathbf{n} \in \mathbb{Z}_N^d} f(\mathbf{n})| = (\mathbb{E}_{\mathbf{n},\mathbf{h} \in \mathbb{Z}_N^d} f(\mathbf{n}) \, \overline{f}(\mathbf{n} + \mathbf{h}))^{1/2},$$

and, for every $s \geq 1$, we let

$$\|f\|_{U^{s+1}(\mathbb{Z}_N^d)} := (\mathbb{E}_{\mathbf{h} \in \mathbb{Z}_N^d} \|f \cdot \overline{f}_{\mathbf{h}}\|_{U^s(\mathbb{Z}_N^d)}^{2^s})^{1/2^{s+1}}.$$

For $d = 1$, the next result is deduced in [**47**, §3.1] from the inverse theorem for the Gowers norms in $\mathbb{Z}_N$ [**40**]. A multidimensional extension of this inverse theorem was established in [**17**, **61**] (for an alternate proof see [**63**]) and the argument in [**47**, §3.1] allows us to deduce the following result.

PROPOSITION 6.5. ([**47**, §3.1] for $d = 1$) *Let $d, k \in \mathbb{N}$, $C > 0$ and $\varepsilon > 0$. Then there exists a k-step nilmanifold X such that the following holds: if $N \in \mathbb{N}$ and $a : \mathbb{Z}_N^d \to \mathbb{C}$ satisfies*

$$|\mathbb{E}_{\mathbf{n} \in \mathbb{Z}_N^d} a(\mathbf{n}) \, b(\mathbf{n})| \leq C \|b\|_{U^{k+1}(\mathbb{Z}_N^d)} \quad \textit{for every } b : \mathbb{Z}_N \to \mathbb{C}, \tag{6.11}$$

then we have the decomposition $a = a_{\mathrm{st}} + a_{\mathrm{er}}$ where:

(i) *a_{st} is a convex combination of k-step nilsequences defined by functions on X with Lipschitz norm at most one; and*

(ii) $\mathbb{E}_{\mathbf{n} \in \mathbb{Z}_N^d} |a_{\mathrm{er}}(\mathbf{n})| \leq \varepsilon$.

It follows, from the previous result, that, in order to complete the proof of Theorem 2.9, it suffices to show that the sequence $b : \mathbb{Z}_N^d \to \mathbb{C}$, defined by (2.9), satisfies the property (6.11) for some $k \in \mathbb{N}$ and $C > 0$ that depend only on the integers d, ℓ, s, t. This can be verified directly with $C = 1$ by using a PET-induction argument (as in [**32**, Lemma 3.5]) and the estimate $|\mathbb{E}_{\mathbf{n} \in \mathbb{Z}_N^d} a(\mathbf{n})|^2 \leq \mathbb{E}_{\mathbf{h} \in \mathbb{Z}_N^d} |\mathbb{E}_{\mathbf{n} \in \mathbb{Z}_N^d} a(\mathbf{n} + h) \, \overline{a(\mathbf{n})}|$, but it turns out to be simpler to deduce (6.11) directly from Proposition 6.1 as follows.

For fixed $N \in \mathbb{N}$, we interpret $b : \mathbb{Z}_N^d \to \mathbb{C}$ as a periodic sequence in $\mathbb{N}^d$, and note that b can be represented as

$$b(\mathbf{n}) = \int T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s \, d\mu, \quad \mathbf{n} \in \mathbb{N}^d,$$

where $X := \mathbb{Z}_N^\ell$, μ is the Haar measure on X and, for $i = 1, \ldots, s$, $f_i := a_i$ and T_i is the measure preserving transformation on $\mathbb{Z}_N^\ell$ (with the Haar measure) that shifts the ith coordinate of an element of $\mathbb{Z}_N^\ell$ by one and leaves the other coordinates unchanged. Notice, also, that if the sequence $c \colon \mathbb{N}^d \to \mathbb{N}$ is N-periodic on every coordinate direction, then

$$\lim \operatorname{Av} c(\mathbf{n}) = \mathbb{E}_{\mathbf{n}\in\mathbb{Z}_N^d} c(\mathbf{n}) \quad \text{and} \quad \|c\|_{U^k(\mathbb{Z}^d)} = \|c\|_{U^k(\mathbb{Z}_N^d)}$$

for every $k \in \mathbb{N}$. Keeping all these in mind and using Proposition 6.1, we deduce that the estimate (6.11) holds with $C = 4$ for some $k = k(d, \ell, s, t)$, and we can take $k = s$ if the polynomial mappings are linear. This completes the proof of Theorem 2.9. □

6.6. *Proof of Theorem 2.7.* First, we check that $\mathcal{N}_d$ is a linear subspace of $\ell^\infty(\mathbb{N}^d)$. For $i = 1, 2$, let $(\psi_i(\mathbf{n}))_{\mathbf{n}\in\mathbb{N}^d}$ be a k_i-step nilsequence given by $\psi_i(\mathbf{n}) = \Psi_i(\prod_{j=1}^d \tau_{i,j}^{n_i} \cdot e_{X_i})$, where $X_i = G_i/\Gamma_i$ is a k_i-step nilmanifold and $\tau_i \in G_i$. Then, for $k := \max(k_1, k_2)$, their sum is the k-step nilsequence $(\Psi(\prod_{i=1}^d \tau_i^{n_i} \cdot e_X))_{n_1,\ldots,n_d\in\mathbb{N}}$, where $X := X_1 \times X_2 = (G_1 \times G_2)/(\Gamma_1 \times \Gamma_2)$, $\Psi(x_1, x_2) := \Psi_1(x_1) + \Psi_2(x_2)$, $e_X := (e_{X_1}, e_{X_2})$ and $\tau_j := (\tau_{1,j}, \tau_{2,j})$ for $j = 1, \ldots, d$.

Next, we show that the space $\mathcal{MC}_{\rm d,pol}$ is linear; a similar argument works for the space $\mathcal{MC}_{\rm d,lin}$. Let $a \colon \mathbb{N}^d \to \mathbb{C}$ be given by

$$a(\mathbf{n}) = \int f_0 \cdot T_{\vec{p}_1(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p}_s(\mathbf{n})} f_s \, d\mu, \qquad \mathbf{n} \in \mathbb{N}^d,$$

for some system $(X, \mu, T_1, \ldots, T_\ell)$, functions $f_0, \ldots, f_s \in L^\infty(\mu)$ and polynomial mappings $\vec{p}_1, \ldots, \vec{p}_s \colon \mathbb{N}^d \to \mathbb{Z}^\ell$. Also, let $a' \colon \mathbb{N}^d \to \mathbb{C}$ be defined by a similar formula, with ℓ' in place of ℓ, $(X', \mu', T_1', \ldots, T_{\ell'}')$ in place of $(X, \mu, T_1, \ldots, T_\ell)$, s' in place of s, $\vec{p}_1', \ldots, \vec{p}_{s'}' \colon \mathbb{Z}^d \to \mathbb{Z}^{\ell'}$ in place of $\vec{p}_1, \ldots, \vec{p}_s$ and $f_0', \ldots, f_{s'}' \in L^\infty(\mu')$ in place of $f_0, \ldots, f_s$. We define a system $(Y, \nu, S_1, \ldots, S_{\ell+\ell'})$ by letting Y be the disjoint union $X \uplus X'$ of X and X', $\nu = \frac{1}{2}(\mu + \mu')$ and defining the transformations $S_1, \ldots, S_{\ell+\ell'}$ of Y by

$$\begin{aligned} S_j|_X &:= T_j \quad \text{and} \quad S_j|_{X'} := \mathrm{Id} \quad \text{for } 1 \le j \le \ell,\\ S_j|_X &:= \mathrm{Id} \quad \text{and} \quad S_j|_{X'} := T_j' \quad \text{for } \ell < j \le \ell + \ell'. \end{aligned}$$

We also define the polynomial mappings $\vec{q}_1, \ldots, \vec{q}_s \colon \mathbb{N}^d \to \mathbb{Z}^{\ell+\ell'}$ and $\vec{q}_1', \ldots, \vec{q}_{s'}' \colon \mathbb{N}^d \to \mathbb{Z}^{\ell+\ell'}$ as

$$\begin{aligned} \vec{q}_i(\mathbf{n}) &:= (p_{i,1}(\mathbf{n}), \ldots, p_{i,\ell}(\mathbf{n}), 0, \ldots, 0) \quad \text{for } 1 \le i \le s,\\ \vec{q}_i'(\mathbf{n}) &:= (0, \ldots, 0, p_{i,1}'(\mathbf{n}), \ldots, p_{i,\ell'}'(\mathbf{n})) \quad \text{for } 1 \le i \le s'. \end{aligned}$$

We also let $\vec{q}_0 = \vec{q}_0' := 0$. Finally, we define the functions $g_i \in L^\infty(\nu)$, $0 \le i \le s$, and $g_i' \in L^\infty(\nu)$, $0 \le i \le s'$ by

$$g_i := \mathbf{1}_X f_i + \mathbf{1}_{X'} \quad \text{and} \quad g_i' := \mathbf{1}_X + \mathbf{1}_{X'} f_i'.$$

Then

$$\int_Y \prod_{i=0}^{s} S_{\vec{q}_i(\mathbf{n})} g_i \cdot \prod_{i=0}^{s'} S_{\vec{q}_i'(\mathbf{n})} g_i' \, d\nu = \frac{1}{2}(a(\mathbf{n}) + a'(\mathbf{n})), \qquad \mathbf{n} \in \mathbb{N}^d.$$

This completes the proof of the linearity of the space $\mathcal{MC}_{\rm d,pol}$.

The inclusion $\overline{\mathcal{N}_d} \subset \overline{\mathcal{MC}_{\rm d,lin}}$ follows from Proposition 4.2. The inclusion $\overline{\mathcal{MC}_{\rm d,lin}} \subset \overline{\mathcal{MC}_{\rm d,pol}}$ is obvious. The inclusion $\overline{\mathcal{MC}_{\rm d,pol}} \subset \overline{\mathcal{N}_d}$ follows from Theorem 2.5. This completes the proof of Theorem 2.7. □

7. *Convergence criteria for weighted averages*

In this short section, we use the machinery developed in the previous sections in order to prove the convergence criteria stated in §§2.2 and 2.3.

7.1. *Proof of Theorem 2.2.* Let $d, \ell, s, t \in \mathbb{N}$, $(X, \mu, T_1, \ldots, T_\ell)$ be a system, $f_1, \ldots, f_s$ be functions in $L^\infty(\mu)$ and $\vec{p_i}\colon \mathbb{N}^d \to \mathbb{Z}^\ell$, $i = 1, \ldots, s$ be polynomial mappings of degree at most t. We can assume that $\|f_i\|_{L^\infty(\mu)} \le 1$ for $i = 1, \ldots, s$.

Let $\delta > 0$ and $k = k(d, \ell, s, t)$ be given by Proposition 6.1 and suppose that the bounded sequence $w\colon \mathbb{N}^d \to \mathbb{C}$ is k-regular. As remarked there, if all the polynomials are linear, then we can take $k = s$. By Theorem 3.9, the sequence w can be written as $w = \psi + u$, where ψ is a k-step nilsequence in d variables and $u \in \ell^\infty(\mathbb{N}^d)$ satisfies $\|u\|_{U^{k+1}(\mathbb{N}^d)} < \delta$. By Proposition 2.1, the limit

$$\lim \operatorname{Av} \psi(\mathbf{n}) \cdot T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s$$

exists in $L^2(\mu)$ and, by Proposition 6.1, for every Følner sequence $\mathbf{I} = (I_j)_{j\in\mathbb{N}}$,

$$\lim \|\operatorname{Av}_{\mathbf{I}}\, u(\mathbf{n}) \cdot T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s\|_{L^2(\mu)} \le 4\delta.$$

It follows that, for all sufficiently large $j, j' \in \mathbb{N}$, the difference between

$$\frac{1}{|I_j|} \sum_{\mathbf{n}\in I_j} u(\mathbf{n})\, T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s \tag{7.1}$$

and the similar average on $I_{j'}$ has a norm in $L^2(\mu)$ bounded by 8δ. Therefore, the averages (7.1) form a Cauchy sequence and thus converge in $L^2(\mu)$.

Furthermore, if we assume that the averages of $w(\mathbf{n})\psi(\mathbf{n})$ converge to zero for every k-step nilsequence ψ in d variables, then, by Corollary 3.8, $\|w\|_{U^k(\mathbb{N}^d)} = 0$ and the averages (7.1) converge to zero in $L^2(\mu)$, by Proposition 6.1. This completes the proof. □

7.2. *Proof of Theorem 2.4.* Let $k = k(d, \ell, s, t)$ be as in Theorem 2.5 and the sequence $a\colon \mathbb{N}^d \to \mathbb{C}$ be defined by

$$a(\mathbf{n}) := \int f_0 \cdot T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s \, d\mu, \quad \mathbf{n} \in \mathbb{N}^d.$$

By Theorem 2.5, a is an approximate k-step nilsequence in d variables and, by hypothesis (2.3), the averages (2.4) converge. This proves Part (i) of Theorem 2.4. Furthermore, we deduce that if the averages (2.3) converge to zero for every nilsequence ψ in d variables, then the averages (2.4) converge to zero.

Next, we prove Part (ii). Let $k = k(2d, \ell, 2s - 1, t)$ be as in Theorem 2.5 and let the sequence $b\colon \mathbb{N}^d \times \mathbb{N}^d \to \mathbb{C}$ be defined by

$$b(\mathbf{n}, \mathbf{n}') := \int T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s \cdot T_{\vec{p_1}(\mathbf{n}')} \overline{f}_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n}')} \overline{f}_s \, d\mu, \quad \mathbf{n}, \mathbf{n}' \in \mathbb{N}^d.$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

By Theorem 2.5, the sequence b is an approximate k-step nilsequence in $2d$ variables. Hence, by hypothesis (2.5), the averages

$$\frac{1}{N^d N'^d} \sum_{\mathbf{n}\in[N]^d} \sum_{\mathbf{n}'\in[N']^d} w(\mathbf{n})\, w(\mathbf{n}')\, b(\mathbf{n}, \mathbf{n}')$$

converge to some limit L when N and N' tend to $+\infty$. Let N_0 be such that the difference between this average and L is bounded in absolute value by ε for all $N, N' > N_0$. Expanding the square

$$\bigg\| \frac{1}{N^d} \sum_{\mathbf{n}\in[N]^d} w(\mathbf{n}) \cdot T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s \\ - \frac{1}{N'^d} \sum_{\mathbf{n}'\in[N']^d} w(\mathbf{n}') \cdot T_{\vec{p_1}(\mathbf{n}')} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n}')} f_s \bigg\|^2_{L^2(\mu)}$$

we obtain four terms of this form with alternate signs, and thus this square is bounded by 4ε. Hence, the averages in (2.6) form a Cauchy sequence in $L^2(\mu)$ and thus converge. This proves Part (ii) of Theorem 2.4. As above, if the limit (2.5) is zero for all k-step nilsequences ψ in $2d$ variables, then the limit (2.6) is equal to zero.

In order to get the last part of Theorem 2.4, we argue as before and use the last parts of Theorems 2.5 and 2.6. □

8. *Applications to arithmetic weights*

In this section we prove the main results stated in §2.6.

8.1. *Proof of Theorem 2.11.*

Theorem 2.11 follows immediately using the next result, which shows that hypothesis (ii) of Theorem 2.4 is satisfied.

THEOREM 8.1. *Let $d \in \mathbb{N}$ and let $\phi : \mathbb{N}^d \to \mathbb{C}$ be a good multiplicative function. Then, for every nilsequence $\psi : \mathbb{N}^d \to \mathbb{C}$, the limit*

$$\lim_{N_1,\ldots,N_d\to+\infty} \frac{1}{N_1 \cdots N_d} \sum_{\mathbf{n}\in[N_1]\times\cdots\times[N_d]} \phi(\mathbf{n})\, \psi(\mathbf{n}) \tag{8.1}$$

exists. Moreover, if the multiplicative function ϕ is aperiodic, then the limit is equal to zero for all nilsequences ψ in d variables.

Remarks.

- One can also deduce variants of the previous result that deal with polynomial sequences on nilmanifolds; indeed, any such sequence can be represented as a linear sequence on a different nilmanifold [**51**, §2.11].
- The case where $d = 1$ can be deduced from [**30**, Theorem 6.1] but, due to the finitary nature of the statement in [**30**], the argument used there is vastly more complicated.

We prove Theorem 8.1 in §8.3.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

8.2. *Proof of Theorems 2.12 and 2.13.* For $b \in \mathbb{N}$, we let ζ be a root of unity of order b and let f_b be the multiplicative function defined by $f_b(p^k) = \zeta$ for all primes p and all $k \in \mathbb{N}$†. Note that

$$\mathbf{1}_{S_{a,b}}(n) = \frac{1}{b} \sum_{j=0}^{b-1} \zeta^{-aj} (f_b(n))^j, \quad n \in \mathbb{N}. \tag{8.2}$$

It can be seen, using Theorem 2.10 (see [**31**, Proposition 2.10]), that, for $j = 1, \ldots, b-1$, the multiplicative function f^j is aperiodic.

Let

$$V_{\mathbf{n}} := T_{\vec{p_1}(\mathbf{n})} f_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} f_s, \quad \mathbf{n} \in \mathbb{N}^d.$$

Recall that $S := (S_{a_1,b_1} + c_1) \times \cdots \times (S_{a_d,b_d} + c_d)$. Using (8.2), we get that the averages

$$\frac{1}{N^d} \sum_{\mathbf{n} \in S \cap [N]^d} V_{\mathbf{n}}$$

are asymptotically equal (meaning that the relevant difference converges to zero in $L^2(\mu)$) to the averages

$$\frac{1}{N^d} \sum_{\mathbf{n} \in [N]^d} \prod_{i=1}^{d} \frac{1}{b_i} \sum_{j=0}^{b_i - 1} \zeta^{-a_i j} (f_{b_i}(n_i))^j \cdot V_{\mathbf{n}+\mathbf{c}}, \tag{8.3}$$

where $\mathbf{c} := (c_1, \ldots, c_d)$ and, as usual, $n_1, \ldots, n_d$ denote the coordinates of the vector $\mathbf{n} \in \mathbb{N}^d$. Since, for $i = 1, \ldots, d$ and $j = 1, \ldots, b_i - 1$, the multiplicative function $f_{b_i}^j$ is aperiodic, we get that, for $j_i \in \{0, \ldots, b_i - 1\}$, $i = 1, \ldots, d$, the multiplicative function $(n_1, \ldots, n_d) \mapsto \prod_{i=1}^d (f_{b_i}(n_i))^{j_i}$ is aperiodic unless $j_1 = \cdots = j_d = 0$. Keeping this in mind, expanding the product in (8.3) and using the second part of Theorem 2.11, we deduce that the averages (8.3) are asymptotically equal in $L^2(\mu)$ to the averages

$$\prod_{i=1}^{d} \frac{1}{b_i} \cdot \frac{1}{N^d} \sum_{\mathbf{n} \in [N]^d} V_{\mathbf{n}}.$$

Furthermore, the previous argument, applied for $V_{\mathbf{n}} := 1$, $\mathbf{n} \in [N]^d$, gives that

$$\lim_{N \to +\infty} \frac{|S \cap [N]^d|}{N^d} = \prod_{i=1}^{d} \frac{1}{b_i}.$$

Combining the above, we deduce that the difference

$$\frac{1}{|S \cap [N]^d|} \sum_{\mathbf{n} \in S \cap [N]^d} V_{\mathbf{n}} - \frac{1}{N^d} \sum_{\mathbf{n} \in [N]^d} V_{\mathbf{n}}$$

converges to zero in $L^2(\mu)$. Using this, in conjunction with Theorem 1.1, completes the proof of Theorem 2.12.

† In order to get results when the set $S_{a,b}$ is defined by counting prime factors with multiplicity, we define the completely multiplicative function f_b by $f_b(p^k) = \zeta^k$ for all primes p and all $k \in \mathbb{N}$.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

Using the previously established identity for $f_i := \mathbf{1}_A$, $i = 1, \ldots, d$, where $A \in \mathcal{X}$ is a set of positive measure, we deduce that

$$\lim_{N\to+\infty} \frac{1}{|S\cap[N]^d|} \sum_{\mathbf{n}\in S\cap[N]^d} \mu(A \cap T_{-\vec{p_1}(\mathbf{n})}A \cap \cdots \cap T_{-\vec{p_s}(\mathbf{n})}A)$$
$$= \lim_{N\to+\infty} \frac{1}{N^d} \sum_{\mathbf{n}\in[N]^d} \mu(A \cap T_{-\vec{p_1}(\mathbf{n})}A \cap \cdots \cap T_{-\vec{p_s}(\mathbf{n})}A) > 0,$$

where the positiveness of the last limit follows from the multiparameter polynomial Szemerédi theorem [**13**, Theorem 0.9]. This proves Theorem 2.13. □

8.3. *Correlations of multiplicative functions with nilsequences.* In this subsection, we prove Theorem 8.1 for $d = 2$; the general case is identical to this one modulo changes in notation. We begin with some preliminaries.

8.3.1. *Some classical facts about commutators.*

LEMMA 8.2. *Let G be a group and $i, j \in \mathbb{N}$. Then the commutator map $(g, h) \to [g, h]$ maps $G_i \times G_j$ to G_{i+j}. Moreover, it induces a bi-homomorphism from $(G_i/G_{i+1}) \times (G_j/G_{j+1})$ to G_{i+j}/G_{i+j+1}.*

Sketch of the proof. The first statement follows by induction from the classical three subgroups lemma.

Let $H, K, L \subset G$ and N be a normal subgroup of G. If $[[H, K], L] \subset N$ and $[[L, H], K] \subset N$, then $[[K, L], H] \subset N$.

The second statement follows from the identity

$$[xy, z] = x[y, z]x^{-1} \cdot [x, z] = [x, [y, z]] \cdot [y, z] \cdot [x, z] \quad \text{for all } x, y, z \in G. \tag{8.4}$$

□

LEMMA 8.3. *Let G be a group, $H^{(1)}, \ldots, H^{(k)}$ and $Q^{(1)}, \ldots, Q^{(\ell)}$ be normal subgroups of G, $H = H^{(1)} \cdot \cdots \cdot H^{(k)}$ and $Q = Q^{(1)} \cdot \cdots \cdot Q^{(\ell)}$. Then the commutator group $[H, Q]$ is the product of the groups $[H^{(i)}, Q^{(j)}]$ for $i = 1, \ldots, k$ and $j = 1, \ldots, \ell$.*

Proof. All the groups $[H^{(i)}, Q^{(j)}]$ are normal and included in $[H, Q]$, and thus it suffices to prove that $[H, Q]$ is contained in the product of these groups. If $h_i \in H^{(i)}$ for $i = 1, 2$ and $q \in Q^{(1)}$, it follows, from (8.4) and from the normality of $[H^{(1)}, Q^{(1)}]$, that $[h_1h_2, q] \in [H^{(1)}, Q^{(1)}] \cdot [H^{(2)}, Q^{(1)}]$. This proves the result in the case $k = 2$, $\ell = 1$. The result for k arbitrary and $\ell = 1$ follows by induction on k. The result for k, ℓ arbitrary follows by exchanging the roles of H and Q. □

8.3.2. *Some reductions.* We prove Theorem 8.1 for $d = 2$. Recall that ψ is an s-step nilsequence in two variables defined by

$$\psi(n_1, n_2) := \Psi(\tau_1^{n_1}\tau_2^{n_2} \cdot e_X), \quad n_1, n_2 \in \mathbb{N},$$

where $X = G/\Gamma$ is a nilmanifold, $\Psi \in C(X)$ and τ_1, τ_2 are two commuting elements of G. It is known [**51**, **56**] that the closure X' of $\{\tau_1^{n_1}\tau_2^{n_2} \cdot e_X : n_1, n_2 \in \mathbb{N}\}$ in X is a sub-nilmanifold of X. Substituting X' for X, we can assume that $\{\tau_1^{n_1}\tau_2^{n_2} \cdot e_X : n_1, n_2 \in \mathbb{N}\}$ is dense in X; it then follows, by [**51**, Theorem 1.4], that it is equidistributed in X.

We can assume, without loss of generality, that G is spanned by the connected component of the unit element and the elements τ_1 and τ_2. This implies that all commutator subgroups G_i, $i \geq 2$ are connected (see, for example [**11**, Theorem 4.1]).

Suppose that $s \geq 2$ and that G is s-step but not $(s-1)$-step nilpotent. Then the group $K_s := G_s/(G_s \cap \Gamma)$ is a finite dimensional torus, sometimes called the *vertical torus* [**38**], and acts freely on X. We denote this action by $(u, x) \mapsto u \cdot x$ for $u \in K_s$ and $x \in X$. Let $\widehat{K_s}$ be the dual group of K_s, that is, the group of continuous group homomorphisms from K_s to the circle group $\mathcal{S}^1$. If, for some $\chi \in \widehat{K_s}$, the function $\Psi \in C(X)$ satisfies $\Psi(u \cdot x) = \chi(u)\Psi(x)$ for every $u \in K_s$ and $x \in X$, then it is called a *nilcharacter* of X with *vertical frequency* χ. The linear span of nilcharacters is dense in $C(X)$ for the uniform norm. Therefore, it suffices to prove (8.1) when the function Ψ defining the nilsequence ψ is a nilcharacter.

If the vertical frequency χ of Ψ is the trivial character of K_s, then the function Ψ factorizes through the quotient of X under the action of this group. This quotient is the $(s-1)$-step nilmanifold $X' = G/(G_s\Gamma) = (G/G_s)/((\Gamma \cap G_s)/G_s)$. Writing τ_1', τ_2' for the images of τ_1, τ_2 in G/G_s, $\psi(n_1, n_2) = \Psi'(\tau_1'^{n_1}\tau_2'^{n_2} \cdot e_{X'})$ for some $\Psi' \in C(X')$.

Iterating this procedure, we reduce matters to considering the following two cases: (i) X is a 1-step nilmanifold, and (ii) X is an s-step nilmanifold for some $s \geq 2$ and Ψ is a nilcharacter with vertical frequency $\chi \neq 1$.

If (i) holds, then X is a compact Abelian group and we can further reduce matters to the case where Ψ is a character of X. Then the average in (8.1) can be rewritten as

$$\bigg(\frac{1}{N_1}\sum_{n_1=1}^{N_1} \phi_1(n_1)\,\mathrm{e}(n_1 t_1)\bigg) \cdot \bigg(\frac{1}{N_2}\sum_{n_2=1}^{N_2} \phi_2(n_2)\,\mathrm{e}(n_2 t_2)\bigg) \tag{8.5}$$

for some $t_1, t_2 \in \mathbb{R}$ and good multiplicative functions $\phi_1, \phi_2 : \mathbb{N} \to \mathbb{C}$. If t_1 or t_2 is irrational, then the limit is equal to zero, by a Theorem of Daboussi [**22**] (see also [**23**, **24**]). If both t_1 and t_2 are rational, then the limit exists since, by hypothesis, the multiplicative functions ϕ_1 and ϕ_2 are good. Furthermore, this limit is equal to zero if either ϕ_1 or ϕ_2 is aperiodic, that is, if ϕ is aperiodic.

Hence, it suffices to consider case (ii): that is, we can assume that X is an s-step nilmanifold for some $s \geq 2$ and Ψ is a nilcharacter with vertical frequency $\chi \neq 1$. Replacing X with its quotient by the kernel of χ, we are reduced to the case where $G_s = \mathcal{S}^1$ and Ψ has vertical frequency one: that is,

$$\Psi(u \cdot x) = u\,\Psi(x) \quad \text{for every } x \in X \text{ and } u \in G_s = \mathcal{S}^1. \tag{8.6}$$

Therefore, it suffices to prove the following proposition.

PROPOSITION 8.4. (Theorem 8.1-nilcharacter form) *Let $X = G/\Gamma$ be an s-step nilmanifold for some $s \geq 2$. Suppose that $G_s = \mathcal{S}^1$, $\Psi \in C(X)$ satisfies (8.6) and that τ_1, τ_2 are two commuting elements of G such that the sequence $(\tau_1^{n_1}\tau_2^{n_2} \cdot e_X)_{n_1,n_2\in\mathbb{N}}$*

is equidistributed in X. Then, for all multiplicative functions $\phi_1, \phi_2 \colon \mathbb{N} \to \mathbb{C}$ that are bounded by one,

$$\lim_{N_1, N_2 \to +\infty} \frac{1}{N_1 N_2} \sum_{n_1 \in [N_1],\, n_2 \in [N_2]} \phi_1(n_1)\, \phi_2(n_2)\, \Psi(\tau_1^{n_1} \tau_2^{n_2} \cdot e_X) = 0. \tag{8.7}$$

We proceed now to establish Proposition 8.4.

8.3.3. *A variant of Katái's lemma.* We use a two-dimensional variant of a result of Katái ([**48**], see also [**22**]); we omit its proof since it is identical to its one-dimensional version modulo changes in notation.

PROPOSITION 8.5. *Let $P_0 \in \mathbb{N}$, $\phi_1, \phi_2 \colon \mathbb{N} \to \mathbb{C}$ be multiplicative functions bounded by one and let $a \in \ell^\infty(\mathbb{N}^2)$. Suppose that*

$$\lim_{N_1, N_2 \to +\infty} \frac{1}{N_1 N_2} \sum_{n_1 \in [N_1],\, n_2 \in [N_2]} a(p_1 n_1, p_2 n_2)\, \overline{a(p'_1 n_1, p'_2 n_2)} = 0$$

for all distinct primes $p_1, p_2, p'_1, p'_2 \geq P_0$. Then

$$\lim_{N_1, N_2 \to +\infty} \sup_{\phi_1, \phi_2} \left| \frac{1}{N_1 N_2} \sum_{n_1 \in [N_1],\, n_2 \in [N_2]} \phi_1(n_1)\, \phi_2(n_2)\, a(n_1, n_2) \right| = 0,$$

where the sup is taken over all $\phi_1, \phi_2 \colon \mathbb{N} \to \mathbb{C}$ that are multiplicative and bounded by one.

8.3.4. *The nilmanifold Y.* In the subsequent work, p_1, p_2, p'_1, p'_2 are fixed distinct primes. By Proposition 8.5, in order to prove Proposition 8.4, it suffices to show that

$$\lim_{N_1, N_2 \to +\infty} \frac{1}{N_1 N_2} \sum_{n_1 \in [N_1], n_2 \in [N_2]} \Psi(\tau_1^{p_1 n_1} \tau_2^{p_2 n_2} \cdot e_X)\, \overline{\Psi(\tau_1^{p'_1 n_1} \tau_2^{p'_2 n_2} \cdot e_X)} = 0, \tag{8.8}$$

where Ψ satisfies (8.6). We let

$$Y := \overline{\{(\tau_1^{p_1 n_1} \tau_2^{p_2 n_2} \cdot e_X\,,\ \tau_1^{p'_1 n_1} \tau_2^{p'_2 n_2} \cdot e_X) \colon n_1, n_2 \in \mathbb{N}\}} \subset X \times X.$$

Then Y is a sub-nilmanifold of $X \times X$ and, by [**51**, Theorem 1.4], the 2-variable sequence above is equidistributed in Y. Writing m_Y for the Haar measure of Y, the limit in (8.8) is equal to

$$\int_Y \Psi(x)\, \overline{\Psi(x')}\, dm_Y(x, x'). \tag{8.9}$$

Hence, it remains to show that this integral is zero.

Let H be the smallest closed subgroup of G containing $\Gamma \times \Gamma$ and the shift elements $(\tau_1^{p_1}, \tau_1^{p'_1})$ and $(\tau_2^{p_2}, \tau_2^{p'_2})$. We claim that

$$Y = H/(\Gamma \times \Gamma).$$

Indeed, by the definition of H and Y, $H \cdot (e_X, e_X) \supset Y$. Furthermore, by the remark following [**50**, Theorem 2.21], $Y = H_1 \cdot (e_X, e_X)$ for some closed subgroup H_1 of $G \times G$ containing the shift elements $(\tau_1^{p_1}, \tau_1^{p'_1})$ and $(\tau_2^{p_2}, \tau_2^{p'_2})$. Since Y is compact, $H_1 \cap (\Gamma \times \Gamma)$ is cocompact in H_1 and thus $H_2 := H_1 \cdot (\Gamma \times \Gamma)$ is a closed subgroup of G. Since H_2 is a closed subgroup that contains the shift elements and $\Gamma \times \Gamma$, $H_2 \supset H$, and hence $H \cdot (e_X, e_X) \subset H_2 \cdot (e_X, e_X) = H_1 \cdot (e_X, e_X) = Y$. Therefore, $H \cdot (e_X, e_X) = Y$, which implies that $Y = H/(\Gamma \times \Gamma)$. This proves the claim.

8.3.5. *Projection on the Kronecker factor.* We denote by Z the compact Abelian group $G/(G_2\Gamma)$ and let $\pi\colon X\to Z$ and $p\colon G\to Z$ be the natural projections. For $i=1,2$, let $\alpha_i:=p(\tau_i)$ and let Z_i be the closed subgroup of Z spanned by α_i. Since $\{\tau_1^{n_1}\tau_2^{n_2}\cdot e_X\colon n_1,n_2\in\mathbb{N}\}$ is dense in X, $\{\alpha_1^{n_1}\alpha_2^{n_2}\cdot e_X\colon n_1,n_2\in\mathbb{N}\}$ is dense in Z and $Z=Z_1Z_2$. For $i=1,2$, we let

$$G^{(i)}:=p^{-1}(Z_i).$$

Then $G^{(i)}$ is a closed subgroup of G containing Γ and G_2, and hence is normal in G, and

$$G=G^{(1)}G^{(2)}.$$

Let

$$W:=(p\times p)(H)=(\pi\times\pi)(Y).$$

By the definition of Y, W is the closure in $Z\times Z$ of

$$\{(\alpha_1^{p_1n_1}\alpha_2^{p_2n_2}\cdot e_X,\alpha_1^{p'_1n_1}\alpha_2^{p'_2n_2}\cdot e_X)\colon n_1,n_2\in\mathbb{N}\}$$

and thus

$$W=\{(z_1^{p_1}z_2^{p_2},z_1^{p'_1}z_2^{p'_2})\colon z_1\in Z_1,z_2\in Z_2\}. \tag{8.10}$$

8.3.6. *Starting the induction.* For $i=1,2$, let $g_i\in G^{(i)}$. Then, by (8.10),

$$(p\times p)(g_1^{p_1},g_1^{p'_1})\in W\quad\text{and}\quad(p\times p)(g_2^{p_2},g_2^{p'_2})\in W.$$

We have $(p\times p)(H)=W$, the kernel of $p\times p$ is $(G_2\times G_2)(\Gamma\times\Gamma)$ and $\Gamma\times\Gamma\subset H$. Thus

$$(g_1^{p_1},g_1^{p'_1})\in H(G_2\times G_2)\quad\text{and}\quad(g_2^{p_2},g_2^{p'_2})\in H(G_2\times G_2).$$

For $i=1,2$, let $h_i\in G^{(i)}$. By (8.4),

$$(H(G_2\times G_2))_2\subset H_2(G_3\times G_3)$$

and thus

$$[(g_1^{p_1},g_1^{p'_1}),(h_1^{p_1},h_1^{p'_1})],\quad[(g_1^{p_1},g_1^{p'_1}),(h_2^{p_2},h_2^{p'_2})]\quad\text{and}\quad[(g_2^{p_2},g_2^{p'_2}),(h_2^{p_2},h_2^{p'_2})]$$

belong to $H_2(G_3\times G_3)$. By Lemma 8.2, these elements are equal modulo $G_3\times G_3$ to

$$([g_1,h_1]^{p_1^2},[g_1,h_1]^{p_1'^2}),\quad([g_1,h_2]^{p_1p_2},[g_1,h_2]^{p'_1p'_2})\quad\text{and}\quad([g_2,h_2]^{p_2^2},[g_2,h_2]^{p_2'^2}),$$

respectively.

Henceforth, for $i=1,2$, we denote $(G^{(i)})_2$ by $G_2^{(i)}$. For $u,v\in G_2$, $(uv)^{p_1^2}=u^{p_1^2}v^{p_1^2}$ mod G_3 and $(uv)^{p_1'^2}=u^{p_1'^2}v^{p_1'^2}$ mod G_3 and thus the set

$$L:=\{u\in G_2^{(1)}\colon(u^{p_1^2},u^{p_1'^2})\in H_2(G_3\times G_3)\}$$

is a subgroup of $G_2^{(1)}$. By the previous discussion, for $g_1,h_1\in G^{(1)}$, the set L contains $[g_1,h_1]$ and thus it is equal to $G_2^{(1)}$. Hence,

$$(u^{p_1^2},u^{p_1'^2})\in H_2(G_3\times G_3)\quad\text{for every }u\in G_2^{(1)}.$$

In the same way,

$$(u^{p_2^2},u^{p_2'^2})\in H_2(G_3\times G_3)\quad\text{for every }\ u\in G_2^{(2)},$$

$$(u^{p_1p_2},u^{p'_1p'_2})\in H_2(G_3\times G_3)\quad\text{for every }\ u\in[G^{(1)},G^{(2)}].$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

8.3.7. *The induction.* By induction on r, we show the following lemma.

LEMMA 8.6. *For* $2 \le r \le s$ *and* $i_1, \ldots, i_r \in \{1, 2\}$, *let*

$$s_i := |\{j : 1 \le j \le r, i_j = i\}| \quad \textit{for } i = 1, 2,$$
$$G^{(i_1,\ldots,i_r)} := [G^{(i_1)}, [G^{(i_2)}, \ldots, [G^{(i_{r-1})}, G^{(i_r)}] \ldots]] \subset G_r.$$

Then, for every $u \in G^{(i_1,\ldots,i_r)}$, $(u^{p_1^{s_1} p_2^{s_2}}, u^{p_1'^{s_1} p_2'^{s_2}}) \in H_r(G_{r+1} \times G_{r+1})$.

Proof. For $r = 2$, this was proved in the preceding subsection and the inductive step is proved by the same method. □

In the subsequent work, we only use this lemma for $r = s$.

8.3.8. *Conclusion of the proof.* We argue by contradiction. Suppose that

$$I := \int \Psi(x)\, \overline{\Psi}(x')\, dm_Y(x, x') \neq 0.$$

Recall that $G_s = \mathcal{S}^1$ and that Ψ satisfies (8.6).

Let $i_1, \ldots, i_s \in \{1, 2\}$ and $s_i = |\{j : 1 \le j \le s,\ i_j = i\}|$ for $i = 1, 2$. Since G is s-step nilpotent, the subgroup G_{s+1} is trivial. Hence, by Lemma 8.6, for every $u \in G^{(i_1,\ldots,i_s)}$, $(u^{p_1^{s_1} p_2^{s_2}}, u^{p_1'^{s_1} p_2'^{s_2}}) \in H_s$. Therefore, the measure m_Y is invariant under translation by $(u^{p_1^{s_1} p_2^{s_2}}, u^{p_1'^{s_1} p_2'^{s_2}})$. Hence,

$$I = \int \Psi(u^{p_1^{s_1} p_2^{s_2}} \cdot x)\, \overline{\Psi}(u^{p_1'^{s_1} p_2'^{s_2}} \cdot x')\, dm_Y(x, x') = u^{p_1^{s_1} p_2^{s_2} - p_1'^{s_1} p_2'^{s_2}} \cdot I.$$

Thus $u^{p_1^{s_1} p_2^{s_2} - p_1'^{s_1} p_2'^{s_2}} = 1$ for every $g \in G^{(i_1,\ldots,i_s)}$. Since $G^{(i_1,\ldots,i_s)}$ is a subgroup of the torus G_s and p_1, p_1', p_2, p_2' are distinct primes, it follows that the group $G^{(i_1,\ldots,i_s)}$ is finite.

Furthermore, since $G = G^{(1)} G^{(2)}$, using Lemma 8.3 and induction, we have that the group G_s is the product of the groups $G^{(i_1,\ldots,i_s)}$ for $i_1, \ldots, i_s \in \{1, 2\}$ and thus is finite, which is a contradiction since $G_s = \mathcal{S}^1$. This completes the proof of Proposition 8.4 and hence of Theorem 8.1. □

9. *Applications to Hardy field weights*

9.1. *Hardy field functions.* Let B be the collection of equivalence classes of real-valued functions defined on some half-line $[c, +\infty)$, where we identify two functions if they agree on some half-line. A *Hardy field* H is a subfield of the ring $(B, +, \cdot)$ that is closed under differentiation. A *Hardy field function* is a function that belongs to some Hardy field. An example of a Hardy field consists of all *logarithmic-exponential functions*, that is, all functions defined on some half-line $[c, +\infty)$ by a finite combination of the symbols $+, -, \times, :$, log, exp operating on the real variable t and on real constants. Examples include functions of the form $t^a (\log t)^b$ for every $a, b \in \mathbb{R}$. An important property of Hardy field functions is that we can relate their growth rates with the growth rates of their derivatives. The reader can find further discussion about Hardy fields in [**14–16**, **28**] and the references therein.

Let f be a Hardy field function. We say that it:

(i) has *at most polynomial growth* if $f(t)/t^m \to 0$ for some $m \in \mathbb{N}$;

(ii) *stays away from polynomials* if $|f(t) - p(t)|/\log t \to +\infty$ for every $p \in \mathbb{R}[t]$; and

(iii) is *asymptotically polynomial* if $f(t) - p(t) \to 0$ for some $p \in \mathbb{R}[t]$.

For our purposes, the key property of Hardy field functions that stay away from polynomials is that they satisfy Lemma 9.6, below. Examples of such functions include:

- t^a where a is a positive non-integer;
- $t^a(\log t)^b$, where $a > 0$ and $b \in \mathbb{R} \setminus \{0\}$; and
- $t^a + (\log t)^b$, where $a \in \mathbb{R}$ and $b > 1$.

9.2. *Convergence and recurrence results.* The main result of this section is the following theorem.

THEOREM 9.1. *Let $d \in \mathbb{N}$ and let $f_1, \ldots, f_d$ be Hardy field functions with at most polynomial growth that stay away from polynomials. We define the sequence $w\colon \mathbb{N}^d \to \mathbb{C}$ by*

$$w(\mathbf{n}) := e\Big(\sum_{i=1}^{d} f_i(n_i)\Big), \qquad \mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d.$$

Then for every system $(X, \mu, T_1, \ldots, T_\ell)$, functions $F_1, \ldots, F_s \in L^\infty(\mu)$ and polynomial mappings $\vec{p_i} : \mathbb{N}^d \to \mathbb{Z}^\ell$, $i = 1, \ldots, s$,

$$\lim_{N\to+\infty} \frac{1}{N^d} \sum_{\mathbf{n}\in[N]^d} w(\mathbf{n}) \cdot T_{\vec{p_1}(\mathbf{n})} F_1 \cdot \cdots \cdot T_{\vec{p_s}(\mathbf{n})} F_s = 0,$$

where the limit is taken in $L^2(\mu)$.

Remark. Related work for pointwise convergence when $d = \ell = s = 1$ appears in [**26**].

Using the $d = 1$ case of the previous result, we deduce, in §9.3, the following corollary.

COROLLARY 9.2. *Let f be a Hardy field function with at most polynomial growth. Then the sequence $w\colon \mathbb{N} \to \mathbb{C}$, defined by $w(n) := e(f(n))$, $n \in \mathbb{N}$, is a good universal weight for mean convergence of the averages (2.6) if and only if either f is asymptotically polynomial or f stays away from polynomials.*

Theorem 9.1 follows, immediately, from Part (ii) of Theorem 2.4 and the next result.

PROPOSITION 9.3. *Let $d \in \mathbb{N}$ and let $f_1, \ldots, f_d$ be Hardy field functions with at most polynomial growth that stay away from polynomials. Then, for every nilsequence $\psi : \mathbb{N}^d \to \mathbb{C}$,*

$$\lim_{N_1,\ldots,N_d\to+\infty} \frac{1}{N_1 \cdots N_d} \sum_{\mathbf{n}\in[N_1]\times\cdots\times[N_d]} e\Big(\sum_{i=1}^{d} f_i(n_i)\Big)\, \psi(\mathbf{n}) = 0.$$

We prove Proposition 9.3 in §9.3.

Next, we give some applications. Note that, for $0 < a < b < 1/2$ and $t \in [0, 1)$,

$$\mathbf{1}_{[a,b]}(\|t\|) = \mathbf{1}_{[a,b]}(t) + \mathbf{1}_{[1-b,1-a]}(t).$$

Since $\mathbf{1}_{[c,d]}(t)$ is Riemann integrable for all $c, d \in \mathbb{R}$ with $0 \leq c < d < 1$, for every $\varepsilon > 0$, there exist (1-periodic) trigonometric polynomials P_1, P_2 with zero constant terms such that

$$P_1(t) - \varepsilon \leq \mathbf{1}_{[c,d]}(t) - (d - c) \leq P_2(t) + \varepsilon, \quad t \in [0, 1]. \tag{9.1}$$

Using Proposition 9.3, with $2\pi k_i f_i$ in place of f_i for $k_i \in \mathbb{Z}$ not all of them zero, for $i = 1, \ldots, d$, we deduce the following corollary, using the estimate (9.1).

COROLLARY 9.4. *Let $d \in \mathbb{N}$ and let $f_1, \ldots, f_d$ be Hardy field functions with at most polynomial growth that stay away from polynomials. Let $a_i, b_i \in \mathbb{R}$ with $0 \leq a_i < b_i \leq 1/2$, $i = 1, \ldots, d$, and*

$$S := \{(n_1, \ldots, n_d) \in \mathbb{N}^d : \|f_1(n_1)\| \in [a_1, b_1], \ldots, \|f_d(n_d)\| \in [a_d, b_d]\}.$$

Then, for every nilsequence $\psi : \mathbb{N}^d \to \mathbb{C}$,

$$\lim_{N\to+\infty} \frac{1}{|S \cap [N]^d|} \sum_{n \in S\cap[N]^d} \psi(\mathbf{n}) = \lim_{N\to+\infty} \frac{1}{N^d} \sum_{\mathbf{n}\in[N]^d} \psi(\mathbf{n}).$$

We also deduce the following mean convergence and multiple recurrence result.

THEOREM 9.5. *Let $S \subset \mathbb{N}^d$ be as in Corollary 9.4. Then the density $d(S)$ of S is $(\prod_{i=1}^d 2(b_i - a_i))^{-1}$ and:*

(i) *the sequence $w := \mathbf{1}_S$ is a good universal weight for mean convergence of the averages (2.6) and the limit of these averages is equal to the limit obtained when $w := d(S)$; and*

(ii) *for every $d, \ell, s \in \mathbb{N}$, polynomial mappings $\vec{p_1}, \ldots, \vec{p_s} : \mathbb{N}^d \to \mathbb{Z}^\ell$ with zero constant term, system $(X, \mu, T_1, \ldots, T_\ell)$ and set $A \in \mathcal{X}$ with $\mu(A) > 0$,*

$$\lim_{N\to+\infty} \frac{1}{N^d} \sum_{\mathbf{n}\in[N]^d} \mathbf{1}_S(\mathbf{n})\, \mu(A \cap T_{-\vec{p_1}(\mathbf{n})} A \cap \cdots \cap T_{-\vec{p_s}(\mathbf{n})} A) > 0. \tag{9.2}$$

Proof. If f is a Hardy field function of at most polynomial growth that stays away from polynomials, then the sequence $(f(n))_{n\in\mathbb{N}}$ is uniformly distributed mod 1 (see [**15**]). The statement about the density of S follows from this fact.

Using Theorem 9.1, with $2\pi k_i f_i$ in place of f_i for $k_i \in \mathbb{Z}$ not all of them zero, for $i = 1, \ldots, d$, and the estimate (9.1), we deduce Part (i).

To prove Part (ii) we use Part (i) for $f_1 = \cdots = f_s = \mathbf{1}_A$, multiply by $\mathbf{1}_A$ and integrate with respect to μ. We deduce that the limit in (9.2) is the same as the one obtained when the constant sequence $d(S)$ takes the place of $\mathbf{1}_S$. The asserted positiveness then follows from the multiparameter polynomial Szemerédi theorem [**13**, Theorem 0.9]. □

9.3. *Proof of Proposition 9.3 and Corollary 9.2.* We start with some preliminary facts. Our assumptions on the functions $f_1, \ldots, f_d$ are used via the following lemma.

LEMMA 9.6. *Let f be a Hardy field function that stays away from polynomials and satisfies $f(t)/t^m \to 0$ as $t \to +\infty$ for some $m \in \mathbb{N}$. Let $k \geq m$ be an integer. Then there exist real numbers α_N with $\alpha_N \to 0$, polynomials $q_N \in \mathbb{R}[t]$ with $\deg(q_N) < k$, positive integers L_N with $L_N/N \to 0$ and $L_N^k|\alpha_N| \to +\infty$, such that*

$$f(N + n) = n^k \alpha_N + q_N(n) + o_{N\to+\infty}(1), \quad n \in [L_N].$$

Proof. This follows by noticing that the proof of [**28**, Lemma 3.5] applies to all $k \in \mathbb{N}$ with $k \geq m$ and then following the argument in the proof of [**28**, Lemma 3.4]. □

In the proof of Proposition 9.3, we use some quantitative equidistribution results from [**38**]. We record here some relevant notions.

- If G is a nilpotent group, then $g\colon \mathbb{N}^d \to G$ is a *polynomial sequence* if it has the form $g(\mathbf{n}) = \prod_{i=1}^{s} \tau_i^{p_i(\mathbf{n})}$, where, for $i = 1, \ldots, s$, $\tau_i \in G$ and $p_i\colon \mathbb{N}^d \to \mathbb{Z}$ are polynomials. The *degree* of the polynomial sequence (with a given representation) is the maximum of the degrees of the polynomials $p_1, \ldots, p_s$.
- For $N_1, \ldots, N_d \in \mathbb{N}$, we say that the finite sequence $(g(\mathbf{n}) \cdot e_X)_{\mathbf{n}\in[N_1]\times\cdots\times[N_d]}$ is δ*-equidistributed* in the nilmanifold X if, for every Lipschitz function $\Psi\colon X \to \mathbb{C}$ with $\|\Psi\|_{\mathrm{Lip}(X)} \leq 1$ and $\int_X \Psi\, dm_X = 0$,
$$\Bigl|\frac{1}{N_1\cdots N_d}\sum_{\mathbf{n}\in[N_1]\times\cdots\times[N_d]} \Psi(g(\mathbf{n})\cdot e_X)\Bigr| \leq \delta.$$
- An infinite sequence $(g(\mathbf{n}) \cdot e_X)_{\mathbf{n}\in\mathbb{N}^d}$ is *equidistributed in* X if, for all $\Psi \in C(X)$ with $\int_X \Psi\, dm_X = 0$, (note that the averages below are uniform)
$$\lim \mathrm{Av}\, \Psi(g(\mathbf{n}) \cdot e_X) = 0.$$
It is *totally equidistributed in* X if the sequence $(\mathbf{1}_{P_1\times\cdots\times P_d}(\mathbf{n}) \cdot g(\mathbf{n}) \cdot e_X)_{\mathbf{n}\in\mathbb{N}}$ is equidistributed in X for all infinite arithmetic progressions $P_1, \ldots, P_d \subset \mathbb{N}$.
- The *horizontal torus* of the nilmanifold $X = G/\Gamma$ is the compact Abelian group $Z := G/(G_2\Gamma)$. If G is connected, it is a finite dimensional torus. A *horizontal character* is a continuous group homomorphism $G \to \mathbb{T}$. It factors through the horizontal torus and induces a character $\eta\colon Z \to \mathbb{T}$; when G is connected, it is of the form $\mathbf{t} \mapsto \mathbf{k}\cdot\mathbf{t}$, where $\mathbf{k} \in \mathbb{Z}^s$, $\mathbf{t} \in \mathbb{T}^s$, $s := \dim(Z)$. In this case, we define $\|\eta\| := \|\mathbf{k}\|_1$: that is, the sum of the absolute values of the coordinates of $\mathbf{k}$.
- If $p\colon \mathbb{N}^d \to \mathbb{T}$ has the form $p(n_1, \ldots, n_d) = \sum_{j_1,\ldots,j_d} \alpha_{j_1,\ldots,j_d} n_1^{j_1}\cdots n_d^{j_d}$,, we define
$$\|p\|_{C^\infty[N_1]\times\cdots\times[N_d]} := \max_{(j_1,\ldots,j_d)\neq(0,\ldots,0)} N_1^{j_1}\cdots N_d^{j_d}\|\alpha_{j_1,\ldots,j_d}\|.$$
- If $X = G/\Gamma$ is a nilmanifold, then γ is a *rational element of* G if $\gamma^k \in \Gamma$ for some $k \in \mathbb{N}$.

We will use the following quantitative equidistribution result.

THEOREM 9.7. ([**38**, Theorem 8.6] and [**39**]) *Let $X := G/\Gamma$ be a nilmanifold with G connected and simply connected, $d, t \in \mathbb{N}$ and $\varepsilon > 0$. There exists $M := M(X, d, t, \varepsilon) > 0$ such that the following holds: for all $N_1, \ldots, N_d \in \mathbb{N}$ greater than M, if $g\colon \mathbb{N}^d \to G$ is a polynomial sequence of degree t and $(g(\mathbf{n}) \cdot e_X)_{\mathbf{n}\in[N_1]\times\cdots\times[N_d]}$ is not ε-equidistributed in X, then there exists a non-trivial horizontal character η such that*
$$0 < \|\eta\| \leq M \quad \text{and} \quad \|\eta\circ g\|_{C^\infty[N_1]\times\cdots\times[N_d]} \leq M.$$

Remark. For every horizontal character η, the sequence $\eta \circ g$ is a polynomial sequence in $\mathbb{T}$ of degree at most t.

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

We will use the following elementary result which is a two-dimensional variant of [**28**, Lemma 3.3].

LEMMA 9.8. *Let $a \in \ell^\infty(\mathbb{N}^2)$ be such that*

$$\lim_{N,N'\to+\infty} \frac{1}{L_N L'_{N'}} \sum_{\mathbf{n}\in(N+[L_N])\times(N'+[L'_{N'}])} a(\mathbf{n}) = 0 \tag{9.3}$$

for some sequences of positive integers $(L_N)_{N\in\mathbb{N}}$, $(L'_N)_{N\in\mathbb{N}}$ that satisfy $L_N/N \to 0$ and $L'_N/N \to 0$ as $N \to +\infty$. Then

$$\lim_{N,N'\to+\infty} \frac{1}{NN'} \sum_{\mathbf{n}\in[N]\times[N']} a(\mathbf{n}) = 0.$$

Proof. Let the sequence of positive integers $(k_i)_{i\in\mathbb{N}}$ be defined by $k_1 := 1$, $k_{i+1} := k_i + L_{k_i}$, $i \in \mathbb{N}$ and, similarly, let the sequence $(k'_i)_{i\in\mathbb{N}}$ be defined by $k'_1 := 1$, $k'_{i+1} := k'_i + L'_{k'_i}$, $i \in \mathbb{N}$. For $N \in \mathbb{N}$, let $i_N := \max\{i \in \mathbb{N} : k_i \le N\}$ and $i'_N := \max\{i \in \mathbb{N} : k'_i \le N\}$. Then the rectangles $(k_i, k_{i+1}] \times (k'_{i'}, k'_{i'+1}]$, where $i \in [i_N - 1]$ and $i' \in [i'_{N'} - 1]$, have the form $(k, k + L_k] \times (k', k' + L'_{k'}]$ and, together with a leftover set $E_{N,N'}$, form a partition of the rectangle $[N] \times [N']$. The set $E_{N,N'}$ is contained in the union of the rectangles $[N] \times (N' - L'_{k'_{i_{N'}}}, N']$ and $(N - L_{k_{i_N}}, N] \times [N']$ and, since $k_{i_N} \le N$ and $k'_{i'_N} \le N'$,,

$$|E_{N,N'}| \le N \max_{k\le N'}(L'_k) + N' \max_{k\le N}(L_k).$$

Since L_N/N, $L'_N/N \to 0$ as $N \to +\infty$, we get that $|E_{N,N'}|/(NN') \to 0$ as $N, N' \to +\infty$. Using this, the fact that $a : \mathbb{N}^2 \to \mathbb{C}$ is bounded and our assumption (9.3), we deduce that

$$\lim_{N,N'\to+\infty} \frac{1}{NN'} \sum_{\mathbf{n}\in[N]\times[N']} a(\mathbf{n}) = \lim_{k,k'\to+\infty} \frac{1}{L_k L'_{k'}} \sum_{\mathbf{n}\in(k,k+L_k]\times(k',k'+L'_{k'}]} a(\mathbf{n}) = 0.$$

This completes the proof. □

Proof of Proposition 9.3. We give the proof for $d = 2$; the proof in the general case is analogous.

Suppose that the nilsequence ψ has the form

$$\psi(n, n') = \Psi(\tau^n \tau'^{n'} e_X), \quad n, n' \in \mathbb{N},$$

for some nilmanifold $X = G/\Gamma$, commuting elements $\tau, \tau' \in G$ and function $\Psi \in C(X)$. By a remark made in §2.1.2, we can assume that the group G is connected and simply connected. Moreover, we can assume that Ψ is a Lipschitz function with $\|\Psi\|_{\mathrm{Lip}(X)} \le 1$.

By the infinitary factorization theorem [**38**, Corollary 1.12] (the same argument works for sequences in several variables), the sequence $g : \mathbb{N}^2 \to G$, defined by $g(n, n') := \tau^n \tau'^{n'}$, $n, n' \in \mathbb{N}$, can be factorized as

$$g(n, n') = g'(n, n')\, \gamma(n, n'), \quad n, n' \in \mathbb{N},$$

where:

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

- $g'\colon \mathbb{N}^2 \to G'$ is a polynomial sequence on a closed, connected and simply connected subgroup G' of G such that $X' := G'/(G' \cap \Gamma)$ is a nilmanifold;
- the sequence $(g'(n, n') \cdot e_{X'})_{n,n' \in \mathbb{N}}$ is totally equidistributed on X'; and
- the sequence $(\gamma(n, n') \cdot e_X)_{n,n' \in \mathbb{N}}$ is periodic and $\gamma(n, n')$ is a rational element of G for every $n, n' \in \mathbb{N}$.

Then, for some $r \in \mathbb{N}$ and all $(i_1, i_2) \in \{0, \ldots, r-1\}^2$, the sequence $(\gamma(n, n') \cdot e_X)_{n,n' \in \mathbb{N}}$ is constant in the set $r\mathbb{N}^2 + (i_1, i_2)$; say that it is equal to $\gamma_{i_1,i_2} \cdot e_X$ for some rational element γ_{i_1,i_2} in G. After partitioning $\mathbb{N}^2$ as a union of such sets, we are reduced to showing that

$$\lim_{N,N' \to +\infty} \frac{1}{NN'} \sum_{n \in [N], n' \in [N']} e(f(rn + i_1) + f'(rn' + i_2)) \times \Psi(g'(rn + i_1, rn' + i_2)\gamma_{i_1,i_2} \cdot e_X) = 0$$

for all $(i_1, i_2) \in \{0, \ldots, r-1\}^2$. Notice that, if h is a Hardy field function of at most polynomial growth that stays away from polynomials, then, also, $t \mapsto h(kt + l)$ has the same property for all $k \in \mathbb{N}$ and $l \in \mathbb{Z}$. Hence, it suffices to show that if f, f' satisfy the assumptions of Proposition 9.3, then

$$\lim_{N,N' \to +\infty} \frac{1}{NN'} \sum_{n \in [N], n' \in [N']} e(f(n) + f'(n'))\, \Psi(g'(n, n')\gamma \cdot e_X) = 0, \tag{9.4}$$

where $g'\colon \mathbb{N}^2 \to G'$ is such that the infinite polynomial sequence $(g'(n, n') \cdot e_X)_{n,n' \in \mathbb{N}}$ is equidistributed on X' and γ is a rational element of G.

Let $(L_N)_{N \in \mathbb{N}}$, $(L'_N)_{N \in \mathbb{N}}$ be sequences of positive integers that will be specified later; for the moment we only assume that $L_N, L'_N \to +\infty$ and $L_N/N, L'_N/N \to 0$ (we will also impose condition (9.8) later). Using Lemma 9.8, we see that it suffices to show that

$$\lim_{N,N' \to +\infty} \frac{1}{L_N L_{N'}} \sum_{n \in [L_N], n' \in [L_{N'}]} e(f(N + n) + f(N' + n')) \times \Psi_\gamma(g'_\gamma(N + n, N' + n') \cdot e_X) = 0, \tag{9.5}$$

where $g'_\gamma := \gamma^{-1} g' \gamma$ is a polynomial sequence on $G'_\gamma := \gamma^{-1} G \gamma$ and $\Psi_\gamma \in \mathrm{Lip}(X)$ is defined by $\Psi_\gamma(g \cdot e_X) := \Psi(\gamma g \cdot e_X)$ for $g \in G$.

In order to prove (9.5), we need to first gather some data. First, we claim that, for $N, N' \in \mathbb{N}$, the finite sequence

$$(g'(N + n, N' + n') \cdot e_X)_{n \in [L_N], n' \in [L_{N'}]}$$

is $\delta_{N,N'}$-equidistributed on X' for some $\delta_{N,N'} > 0$ that satisfy $\delta_{N,N'} \to 0$ as $N, N' \to +\infty$. Indeed, if this is not the case, then there exists $\delta > 0$, $N_m, N'_m \to +\infty$ and $\Psi_m \in \mathrm{Lip}(X')$ with $\|\Psi_m\|_{\mathrm{Lip}(X')} \le 1$ and $\int_{X'} \Psi_m \, dm_{X'} = 0$, such that

$$\left| \frac{1}{L_{N_m} L_{N'_m}} \sum_{(n,n') \in (N_m + [L_{N_m}]) \times (N'_m + [L_{N'_m}])} \Psi_m(g(n, n') \cdot e_{X'}) \right| \ge \delta \quad \text{for all } m \in \mathbb{N}. \tag{9.6}$$

By the Arzelá–Ascoli theorem, a subsequence of Ψ_m converges uniformly to some $\Psi_0 \in \mathrm{Lip}(X')$ with $\|\Psi_0\|_{\mathrm{Lip}(X')} \le 1$ and $\int_{X'} \Psi_0 \, dm_{X'} = 0$. Then (9.6) is satisfied for an infinite number of $m \in \mathbb{N}$ with Ψ_0 in place of Ψ_m and $\delta/2$ in place of δ. Since $L_{N_m}, L_{N'_m} \to +\infty$, $((N_m + [L_{N_m}]) \times (N'_m + [L_{N'_m}]))_{m\in\mathbb{N}}$ is a Følner sequence in $\mathbb{N}^2$ and we deduce that

$$\lim \mathrm{Av}\, \Psi_0(g'(n, n') \cdot e_X) \neq 0.$$

This contradicts our assumption that $(g'(n, n') \cdot e_X)_{n,n'\in\mathbb{N}}$ is equidistributed in X'.

From the aforementioned equidistribution property, we deduce, using [**30**, Corollary 5.5], that the finite sequence

$$(g'_\gamma(N + n, N' + n') \cdot e_X)_{n\in[L_N],n'\in[L_{N'}]}$$

is $\delta'_{N,N'}$-equidistributed, where $\delta'_{N,N'} \to 0$ as $N, N' \to +\infty$, on the nilmanifold $X'_\gamma := G'_\gamma/\Gamma'_\gamma$, where $\Gamma'_\gamma := \Gamma \cap G'_\gamma$.

We move now to the proof of (9.5). We apply Lemma 9.6 for the functions f, f' and we get sequences $(L_N)_{N\in\mathbb{N}}$, $(L'_N)_{N\in\mathbb{N}}$ satisfying the conditions in the lemma for some $k, k' \in \mathbb{N}$ such that $k, k' > \deg(g'_\gamma)$. After ignoring negligible errors, we deduce that, in (9.5), we can replace the finite sequences $(f(N + n))_{n\in[L_N]}$ and $(f'(N' + n'))_{n'\in[L'_{N'}]}$ by finite polynomial sequences $(p_N(n))_{n\in[L_N]}$ and $(p'_{N'}(n'))_{n'\in[L'_{N'}]}$, where

$$p_N(n) = n^k\alpha_N + q_N(n), \quad p'_{N'}(n') = n'^{k'}\alpha'_{N'} + q'_{N'}(n'), \quad n, n', N, N' \in \mathbb{N}, \tag{9.7}$$

for some real numbers α_N, α'_N satisfying $\alpha_N \to 0$ and $\alpha'_N \to 0$, and polynomials $q_N, q'_N \in \mathbb{R}[t]$, $N \in \mathbb{N}$, that satisfy

$$\deg(q_N) < k, \quad \deg(q'_N) < k', \quad \text{and} \quad L^k_N|\alpha_N|, L'^{k'}_N|\alpha'_N| \to +\infty. \tag{9.8}$$

We claim that the finite polynomial sequence

$$(p_N(n), p_{N'}(n'), g'_\gamma(N + n, N' + n') \cdot e_X)_{n\in[L_N],n'\in[L'_{N'}]} \tag{9.9}$$

is $\delta''_{N,N'}$-equidistributed in the nilmanifold $\mathbb{T}^2 \times X'_\gamma$, where $\delta''_{N,N'} \to 0$ as $N, N' \to +\infty$. Arguing by contradiction, suppose that this is not true. Then there exists $\delta > 0$ such that

$$\text{the sequence (9.9) is not } \delta\text{-equidistributed in } \mathbb{T}^2 \times X'_\gamma \text{ for some } N_m, N'_m \to +\infty. \tag{9.10}$$

The horizontal torus of the nilmanifold X'_γ has the form $\mathbb{T}^s$ for some $s \in \mathbb{N}$. Then the horizontal torus of the nilmanifold $\mathbb{T}^2 \times X'_\gamma$ is $\mathbb{T}^2 \times \mathbb{T}^s$. Let $\pi : G \to \mathbb{T}^s$ be the natural projection on the horizontal torus of X'_γ and let $r_{N,N'} : \mathbb{N}^2 \to \mathbb{T}^s$ be the polynomial sequence defined by $r_{N,N'}(n, n') := \pi(g'_\gamma(N + n, N' + n'))$. Then

$$\deg(r_{N,N'}) < \min(k, k') \quad \text{for every } N, N' \in \mathbb{N}. \tag{9.11}$$

By Theorem 9.7, we deduce that there exists $M > 0$ and $k_m, k'_m \in \mathbb{Z}$, $k''_m \in \mathbb{Z}^s$ such that, for those N_m, N'_m for which (9.10) holds and are greater than M,

$$0 < |k_m| + |k'_m| + \|k''_m\|_1 \leq M \tag{9.12}$$

and

$$\|k_m p_{N_m}(n) + k'_m p'_{N'_m}(n') + k''_m \cdot r_{N_m,N'_m}(n, n')\|_{C^\infty([L_{N_m}]\times[L'_{N'_m}])} \leq M. \tag{9.13}$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

If $k_m = k'_m = 0$ for an infinite number of $m \in \mathbb{N}$, then k''_m is non-zero for an infinite number of $m \in \mathbb{N}$ and, using [**30**, Lemma 5.3], we get a contradiction from the fact that the sequence $(g'_\gamma(N+n, N'+n') \cdot e_X)_{n\in[L_N],n'\in[L'_{N'}]}$ is $\delta'_{N,N'}$-equidistributed on the nilmanifold X'_γ, where $\delta'_{N,N'} \to 0$ as $N, N' \to +\infty$.

Suppose, next, that $k_m \neq 0$ for a infinite number of $m \in \mathbb{N}$. Using (9.13) (note that the polynomials p_N and $p_{N'}$ depend on different variables) in conjunction with (9.7), (9.8), (9.11), we obtain that

$$L^k_{N_m} \|k_m \alpha_{N_m}\| \leq M \quad \text{for an infinite number of } m \in \mathbb{N}.$$

Since $\alpha_N \to 0$ and $1 \leq |k_m| \leq M$, we get that $\|k_m \alpha_{N_m}\| = |k_m \alpha_{N_m}| \geq |\alpha_{N_m}|$ for an infinite number of $m \in \mathbb{N}$. We deduce that

$$L^k_{N_m} |\alpha_{N_m}| \leq M \quad \text{for infinitely many } m \in \mathbb{N}.$$

This contradicts (9.8). The argument is similar if $k'_m \neq 0$ for an infinite number of $m \in \mathbb{N}$.

Hence, the finite polynomial sequence (9.9) is $\delta''_{N,N'}$-equidistributed in the nilmanifold $\mathbb{T}^2 \times X'_\gamma$, where $\delta''_{N,N'} \to 0$ as $N, N' \to +\infty$. We deduce that the limit in the left-hand side of (9.5) is equal to

$$\int e(t) \cdot e(t') \cdot F_\gamma(x) \, dm_{\mathbb{T}^2 \times X'_\gamma} = 0,$$

where the last identity follows because $m_{\mathbb{T}^2 \times X'_\gamma} = m_{\mathbb{T}^2} \times m_{X'_\gamma}$. This completes the proof. □

Proof of Corollary 9.2. Let f be a Hardy field function of polynomial growth. We consider the following three cases.

If f stays away from polynomials, then the conclusion follows from the $d = 1$ case of Theorem 9.1 and the corresponding averages converge to zero in $L^2(\mu)$.

Suppose, next, that f is asymptotically polynomial, that is, $f(t) - p(t) \to 0$ for some $p \in \mathbb{R}[t]$. In this case, the mean convergence of the averages (2.6) follows from Proposition 2.1 and the well-known fact that sequences of the form $n \mapsto e(p(n))$ are nilsequences.

Lastly, suppose that $f = p + g$ for some polynomial $p \in \mathbb{R}[t]$ and Hardy field function g that satisfies $|g(t)| \to +\infty$ and $|g(t)| \leq C \log t$ for some $C > 0$ and all sufficiently large $t \in \mathbb{R}_+$. Let $p(t) = \sum_{i=0}^{\ell} \alpha_i t^i$, $t \in \mathbb{R}$, for some $\ell \in \mathbb{N}$ and $\alpha_1, \ldots, \alpha_\ell \in \mathbb{R}$. For $i = 1, \ldots, \ell$, we consider the commuting transformations $T_i t := t + \alpha_i$, $t \in \mathbb{T}$, acting on $\mathbb{T}$ with the Haar measure $m_\mathbb{T}$, and the function $h \in L^\infty(m_\mathbb{T})$ defined by $h(t) := e(-t)$, $t \in \mathbb{T}$. Then

$$\frac{1}{N}\sum_{n=1}^{N} e(f(n))\, h\Big(\prod_{i=1}^{\ell} T_i^{n^i} t\Big) = e(-t + \alpha_0)\, \frac{1}{N}\sum_{n=1}^{N} e(g(n)) \quad \text{for every } N \in \mathbb{N},\ t \in \mathbb{T}.$$

By [**28**, Proof of Theorem 3.1] (see, also, [**16**, Proof of Theorem 3.3]), we get that the last averages do not converge as $N \to +\infty$. Hence, the sequence $n \mapsto e(f(n))$ is not a good universal weight for weak convergence of averages of the form (2.6), even when $s = 1$.

If f is any Hardy field function and $p \in \mathbb{R}[t]$ is any polynomial, then it is known that the limit $\lim_{t\to+\infty}(f(t) - p(t))/\log t$ either exists or else is $\pm\infty$. Hence, every Hardy field function with at most polynomial growth is covered in one of the previous three cases and the proof is complete. □

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

Acknowledgement. The second author was partially supported by Centro de Modelamiento Matemático, Universitad de Chile.

A. *Appendix. Seminorms on $L^\infty(\mu)$ and related factors*

Let $(X, \mu, T_1, \ldots, T_\ell)$ be a system. We recall, here, the definition and some properties of the seminorms $|||\cdot|||_k$ on $L^\infty(\mu)$ and of the factors $\mathcal{Z}_k$ defined in [**44**] for the ergodic case and in [**19**] for the general case. These two papers deal only with the case of a single transformation, the generalization to the case of several commuting transformations is analogous and is given below.

A.1. *The seminorms $|||\cdot|||_k$.* We write $\mathcal{I}(\overrightarrow{T})$ for the σ-algebra of sets invariant under all transformations $T_1, \ldots, T_\ell$. For $f \in L^\infty(\mu)$, we define

$$|||f|||_1 := \|\mathbb{E}_\mu(f \mid \mathcal{I}(\overrightarrow{T}))\|_{L^2(\mu)} \tag{A.1}$$

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

$$|||f|||_{k+1} := (\text{Lim Av}_{\overrightarrow{n}}\ |||f \cdot T_{\overrightarrow{n}}\overline{f}\,|||_k^{2^k})^{1/2^{k+1}}, \tag{A.2}$$

where, as usual, we use the notation $T_{\overrightarrow{n}} = \prod_{i=1}^{\ell} T_i^{n_i}$ for $\overrightarrow{n} = (n_1, \ldots, n_\ell)$. By induction,

$$|||f|||_k \le \|f\|_{L^\infty(\mu)} \quad \text{for every } k \in \mathbb{N}.$$

In case of ambiguity, we write $|||f|||_{\mu,k}$ or $|||f|||_{\overrightarrow{T},\mu,k}$. If $\mu = \int \mu_x \, d\mu(x)$ is the ergodic decomposition of μ under $\overrightarrow{T}$, then, for every $f \in L^\infty(\mu)$ and every $k \in \mathbb{N}$,

$$|||f|||_{\mu,k}^{2^k} = \int |||f|||_{\mu_x,k}^{2^k} \, d\mu(x). \tag{A.3}$$

For $f \in L^\infty(\mu)$, by (A.1),

$$\left|\int f \, d\mu\right| \le |||f|||_1. \tag{A.4}$$

Writing $\overrightarrow{T} \times \overrightarrow{T}$ for the $\mathbb{Z}^\ell$-action on $X \times X$ induced by $T_1 \times T_1, \ldots, T_\ell \times T_\ell$,

$$\begin{aligned}
|||f \otimes \overline{f}\,|||^2_{\overrightarrow{T}\times\overrightarrow{T},\mu\times\mu,1} &= \|\mathbb{E}_{\mu\times\mu}(f \otimes \overline{f} \mid \mathcal{I}(\overrightarrow{T}\times\overrightarrow{T}))\|^2_{L^2(\mu\times\mu)} \\
&= \lim \text{Av}_{\overrightarrow{n}} \left|\int f \cdot T_{\overrightarrow{n}}\overline{f} \, d\mu\right|^2 \text{ by the ergodic theorem} \\
&\le \text{limsup Av}_{\overrightarrow{n}} \|\mathbb{E}_\mu(f \cdot T_{\overrightarrow{n}}\overline{f} \mid \mathcal{I}(\overrightarrow{T}))\|^2_{L^2(\mu)} \\
&= \text{limsup Av}_{\overrightarrow{n}}\ |||f \cdot T_{\overrightarrow{n}}\overline{f}\,|||^2_{\overrightarrow{T},\mu,1} \\
&= |||f|||^4_{\overrightarrow{T},\mu,2} \quad \text{by (A.1) and (A.2).}
\end{aligned}$$

By induction, using the relation (A.2) for the measures $\mu \times \mu$ and μ, we deduce that, for every $k \in \mathbb{N}$,

$$|||f \otimes \overline{f}\,|||_{\overrightarrow{T}\times\overrightarrow{T},\mu\times\mu,k} \le |||f|||^2_{\overrightarrow{T},\mu,k+1}. \tag{A.5}$$

https://doi.org/10.1017/etds.2016.19 Published online by Cambridge University Press

A.2. *The factors $\mathcal{Z}_k$.* For $k \in \mathbb{Z}_+$, the factor $\mathcal{Z}_k$ of X is characterized by the property

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

Equivalently,

$$L^\infty(\mathcal{Z}_k, \mu) = \left\{ f \in L^\infty(\mu) : \int f \cdot g \, d\mu = 0 \text{ for every } g \in L^\infty(\mu) \text{ with } |||g|||_{k+1} = 0 \right\}.$$

In case of ambiguity, we write $\mathcal{Z}_k(X, \mu, \vec{T})$.

We say that $(X, \mu, \vec{T})$ is a *system of order k* if the σ-algebra $\mathcal{Z}_k$ coincides with the σ-algebra $\mathcal{X}$, or, equivalently, if $|||\cdot|||_{k+1}$ is a norm on $L^\infty(\mu)$.

PROPOSITION A.1. *Let $(X, \mu, \vec{T})$ be a system of order one and let $\mu = \int \mu_x \, d\mu(x)$ be the ergodic decomposition of μ under $\vec{T}$. Then for μ-almost every $x \in X$, the system $(X, \mu_x, \vec{T})$ is isomorphic to an ergodic rotation on a compact Abelian group.*

Remark. It is not hard to show that the converse also holds.

Proof. Since $(X, \mathcal{X}, \mu)$ is a Lebesgue space, there exists a countable sequence $(f_n)_{n\in\mathbb{N}}$ of bounded Borel functions (defined everywhere) that is dense in $L^1(\mu)$ and in $L^1(\mu_x)$ for every $x \in X$. By [**19**, Corollary 3.3], there exists a Borel set $X_1 \subset X$ with $\mu(X_1) = 1$, such that, for every $x \in X_1$ and every $n \in \mathbb{N}$, the function f_n belongs to $L^\infty(\mathcal{Z}_1(X, \mu_x, \vec{T}))$. For $x \in X_1$, it follows, by density, that every $f \in L^1(\mu_x)$ belongs to $L^1(X, \mathcal{Z}_1(X, \mu_x, \vec{T}))$. The σ-algebras $\mathcal{X}$ and $\mathcal{Z}_1(X, \mu_x, \vec{T})$ coincide up to μ_x-null sets and $(X, \mu_x, \vec{T})$ is a system of order one for μ-almost every $x \in X$. It is well known that an ergodic system of order one is isomorphic to an ergodic rotation on a compact Abelian group, and the proof is complete. □

REFERENCES

[**1**] I. Assani, D. Duncan and R. Moore. Pointwise characteristic factors for Wiener–Wintner double recurrence theorem. *Ergod. Th. & Dynam. Sys.* to appear. *Preprint*, 2014, arXiv:1402.7094.

[**2**] I. Assani and R. Moore. Extension of Wiener–Wintner double recurrence theorem to polynomials. *J. Anal. Math.* to appear. *Preprint*, 2014, arXiv:1409.0463.

[**3**] I. Assani and R. Moore. A good universal weight for nonconventional ergodic averages in norm. *Ergod. Th. & Dynam. Sys.* to appear. *Preprint*, 2015, arXiv:1503.08863.

[**4**] I. Assani and R. Moore. A good universal weight for multiple recurrence averages with commuting transformations in norm. *Preprint*, 2015, arXiv:1506.06730.

[**5**] J. Auslander. *Minimal Flows and Their Extensions*. North-Holland, Amsterdam, 1988.

[**6**] L. Auslander, L. Green and F. Hahn. *Flows on homogeneous spaces. With the assistance of L. Markus and W. Massey, and an appendix by L. Greenberg. (Annals of Mathematics Studies, 53)*. Princeton University Press, Princeton, NJ, 1963.

[**7**] T. Austin. On the norm convergence of nonconventional ergodic averages. *Ergod. Th. & Dynam. Sys.* **30** (2010), 321–338.

[**8**] T. Austin. Pleasant extensions retaining algebraic structure, I. *J. Anal. Math.* **125** (2015), 1–36.

[**9**] T. Austin. Pleasant extensions retaining algebraic structure, II. *J. Anal. Math.* **126** (2015), 1–111.

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

[**11**] V. Bergelson, B. Host and B. Kra. Multiple recurrence and nilsequences. *Invent. Math.* **160**(2) (2005), 261–303, With an appendix by I. Ruzsa.

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

[**13**] V. Bergelson and R. McCutcheon. An ergodic IP polynomial Szemerédi theorem. *Mem. Amer. Math. Soc.* **146** (2000), 695.

[**14**] M. Boshernitzan. An extension of Hardy's class L of 'Orders of Infinity'. *J. Anal. Math.* **39** (1981), 235–255.

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

[**16**] M. Boshernitzan, G. Kolesnik, A. Quas and M. Wierdl. Ergodic averaging sequences. *J. Anal. Math.* **95** (2005), 63–103.

[**17**] A. Camarena and B. Szegedy. Nilspaces, nilmanifolds and their morphisms. *Preprint*, 2012, arXiv:1009.3825v3.

[**18**] Q. Chu. Convergence of weighted polynomial multiple ergodic averages. *Proc. Amer. Math. Soc.* **137**(4) (2009), 1363–1369.

[**19**] Q. Chu, N. Frantzikinakis and B. Host. Commuting averages with polynomial iterates of distinct degrees. *Proc. Lond. Math. Soc.* (3) **102**(5) (2011), 801–842.

[**20**] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France* **112**(2) (1984), 143–175.

[**21**] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *Probabilités, Publ. Inst. Rech. Math. Rennes, 1987-1*. Université Rennes I, Rennes, 1988, pp. 1–31.

[**22**] H. Daboussi. Fonctions multiplicatives presque périodiques B. D'après un travail commun avec Hubert Delange. *Journées Arithmétiques de Bordeaux (Conference, Université Bordeaux, Bordeaux, 1974) (Astérisque, 24–25)*. Société Mathématique de France, Paris, 1975, pp. 321–324.

[**23**] H. Daboussi and H. Delange. Quelques proprietes des functions multiplicatives de module au plus egal 1. *C. R. Acad. Sci. Paris Ser. A* **278** (1974), 657–660.

[**24**] H. Daboussi and H. Delange. On multiplicative arithmetical functions whose modulus does not exceed one. *J. Lond. Math. Soc.* (2) **26**(2) (1982), 245–264.

[**25**] T. Eisner. Linear sequences and weighted ergodic theorems. *Abstr. Appl. Anal.* (2013), Art. ID 815726.

[**26**] T. Eisner and B. Krause. (Uniform) convergence of twisted ergodic averages. *Ergod. Th. & Dynam. Sys.* to appear. *Preprint*, 2014, arXiv:1407.4736.

[**27**] P. Elliott. *Probabilistic Number Theory I*. Springer, New York, 1979.

[**28**] N. Frantzikinakis. Equidistribution of sparse sequences on nilmanifolds. *J. Anal. Math.* **109** (2009), 353–395.

[**29**] N. Frantzikinakis. Multiple correlation sequences and nilsequences. *Invent. Math.* **202**(2) (2015), 875–892.

[**30**] N. Frantzikinakis and B. Host. Higher order Fourier analysis of multiplicative functions and applications. *J. Amer. Math. Soc.* to appear. *Preprint*, 2014, arXiv:1403.0945.

[**31**] N. Frantzikinakis and B. Host. Multiple ergodic theorems for arithmetic sets. *Trans. Amer. Math. Soc.* to appear. *Preprint*, 2015, arXiv:1503.07154.

[**32**] N. Frantzikinakis, B. Host and B. Kra. The multidimensional Szemerédi theorem along shifted primes. *Israel J. Math.* **194**(1) (2013), 331–348.

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

[**34**] H. Furstenberg and B. Weiss. A mean ergodic theorem for $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2} x)$. *Convergence in ergodic theory and probability (Columbus, OH, 1993) (Ohio State University Mathematical Research Institute Publications, 5)*. de Gruyter, Berlin, 1996, pp. 193–227.

[**35**] E. Glasner. *Ergodic Theory via Joinings (Mathematical Surveys Monographs, vol. 101)*. American Mathematical Society, Providence, RI, 2003.

[**36**] A. Granville and K. Soundararajan. *Multiplicative Number Theory: The Pretentious Approach*. Book manuscript in preparation.

[**37**] B. Green and T. Tao. An inverse theorem for the Gowers $U^3(G)$-norm. *Proc. Edinb. Math. Soc.* **51**(1) (2008), 73–153.

[**38**] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)* **175**(2) (2012), 465–540.

[**39**] B. Green and T. Tao. On the quantitative distribution of polynomial nilsequences-erratum. *Ann. of Math. (2)* **179**(3) (2014), 1175–1183. arXiv:1311.6170v3.

[**40**] B. Green, T. Tao and T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Ann. of Math. (2)* **176**(2) (2012), 1231–1372.

[**41**] J. T. Griesmer. Ergodic averages, correlation sequences, and sumsets. *PhD Thesis*, Ohio State University, 2009. Available at: http://rave.ohiolink.edu/etdc/view?acc_num=osu1243973834.

[**42**] G. Halász. Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen. *Acta Math. Acad. Sci. Hung.* **19** (1968), 365–403.

[**43**] B. Host. Ergodic seminorms for commuting transformations and applications. *Studia Math.* **195** (2009), 31–49.
[**44**] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)* **161** (2005), 397–488.
[**45**] B. Host and B. Kra. Convergence of polynomial ergodic averages. *Israel J. Math.* **149** (2005), 1–19.
[**46**] B. Host and B. Kra. Uniformity seminorms on ℓ^∞ and applications. *J. Anal. Math.* **108** (2009), 219–276.
[**47**] B. Host and B. Kra. A point of view on Gowers uniformity norms. *New York J. Math.* **18** (2012), 213–248.
[**48**] I. Kátai. A remark on a theorem of H. Daboussi. *Acta Math. Hungar.* **47** (1986), 223–225.
[**49**] A. Kechris. *Classical descriptive set theory (Graduate Texts in Mathematics, 156)*. Springer, New York, 1995.
[**50**] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Ergod. Th. & Dynam. Sys.* **25**(1) (2005), 201–213.
[**51**] A. Leibman. Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold. *Ergod. Th. & Dynam. Sys.* **25**(1) (2005), 215–225.
[**52**] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Israel J. Math.* **146** (2005), 303–316.
[**53**] A. Leibman. Multiple polynomial correlation sequences and nilsequences. *Ergod. Th. & Dynam. Sys.* **30**(3) (2010), 841–854.
[**54**] A. Leibman. Nilsequences, null-sequences, and multiple correlation sequences. *Ergod. Th. & Dynam. Sys.* **35**(1) (2015), 176–191.
[**55**] A. Leibman. Correction to the paper 'Nilsequences, null-sequences, and multiple correlation sequences'. *Preprint*, 2012, arXiv:1205.4004.
[**56**] E. Lesigne.. Sur une nil-variété, les parties minimales associées à une translation sont uniquement ergodiques. *Ergod. Th. & Dynam. Sys.* **11**(2) (1991), 379–391.
[**57**] D. Meiri. Generalized correlation sequences, *Master's thesis*, Tel Aviv University, 1990. Available at: http://taalul.com/David/Math/ma.pdf.
[**58**] W. Parry. Ergodic properties of affine transformations and flows on nilmanifolds. *Amer. J. Math.* **91** (1969), 757–771.
[**59**] W. Parry. Dynamical systems on nilmanifolds. *Bull. Lond. Math. Soc.* **2** (1970), 37–40.
[**60**] D. Rudolph. Eigenfunctions of $T \times S$ and the Conze–Lesigne algebra. *Ergodic Theory and its Connections with Harmonic Analysis (Alexandria, 1993). (London Mathematical Society Lecture Note Series, 205)*. Cambridge University Press, Cambridge, 1995, pp. 369–432.
[**61**] B. Szegedy. On higher order Fourier analysis. *Preprint*, 2012, arXiv:1203.2260v1.
[**62**] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergod. Th. & Dynam. Sys.* **28**(2) (2008), 657–688.
[**63**] T. Tao. Deducing the inverse theorem for the multidimensional Gowers norms from the one-dimensional version, *Blog entry*, https://terrytao.wordpress.com/2015/07/24.
[**64**] M. Walsh. Norm convergence of nilpotent ergodic averages. *Ann. of Math. (2)* **175**(3) (2012), 1667–1688.
[**65**] E. Wirsing. Das asymptotische Verhalten von Summen uber multiplikative Funktionen, II. *Acta Math. Acad. Sci. Hungar* **18** (1967), 411–467.
[**66**] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.* **20** (2007), 53–97.
[**67**] P. Zorin-Kranich. Norm convergence of multiple ergodic averages on amenable groups. *J. Anal. Math.* to appear. *Preprint*, 2011, arXiv:1111.7292.
[**68**] P. Zorin-Kranich. A uniform nilsequence Wiener–Wintner theorem for bilinear ergodic averages. *Preprint*, 2015, arXiv:1504.04647.