Annals of Mathematics **175** (2012), 1667–1688
http://dx.doi.org/10.4007/annals.2012.175.3.15

Norm convergence of nilpotent ergodic averages

By Miguel N. Walsh

Abstract

We show that multiple polynomial ergodic averages arising from nilpotent groups of measure preserving transformations of a probability space always converge in the L^2 norm.

1. Introduction

The purpose of this paper is to prove the following result.

Theorem 1.1. *Let G be a nilpotent group of measure preserving transformations of a probability space $(X, \mathcal{X}, \mu)$. Then, for every $T_1, \dots, T_l \in G$, the averages*

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{j=1}^{d} \left(T_1^{p_{1,j}(n)} \cdots T_l^{p_{l,j}(n)} \right) f_j \tag{1.1}$$

always converge in $L^2(X, \mathcal{X}, \mu)$ for every $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$ and every set of integer valued polynomials $p_{i,j}$.

This result was conjectured in the present form by Bergelson and Leibman, who also showed that even $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f S^n g$ need not exist if T and S only generate a solvable group [6].

1.1. *Historical background.* Partial results towards Theorem 1.1 have a rich history. Notice that when $d = l = 1$ and the polynomial is linear it reduces to the classical mean ergodic theorem. The only case of Theorem 1.1 that was fully settled is that in which $T_1 = \cdots = T_l$, that is, when G is a cyclic group. The study of this case originated in the seminal work of Furstenberg [10] on Szemerédi's theorem, while a general solution when the polynomials are linear was later provided by Host and Kra [15] following the work of several authors (with a different proof subsequently found by Ziegler [22]). Convergence for general polynomials was established by Bergelson [5] under the assumption of

The author was partially supported by a CONICET doctoral fellowship.

weakly mixing, while the first unconditional nonlinear result was obtained by Furstenberg and Weiss [11]. The general result for cyclic groups and arbitrary polynomials was finally settled by Host and Kra [14] and Leibman [18].

Another case of Theorem 1.1 that is known is that in which G is abelian and every polynomial is linear. Here, the case $d = 2$ was proven by Conze and Lesigne [8], and assuming extra ergodicity hypothesis on the transformations, Zhang [21] gave a proof for $d = 3$ and Frantzikinakis and Kra [9] for general d. Without these assumptions, this result was established by Tao [19] and by now possesses several different proofs [1], [13], [20]. However, when G is abelian but the polynomials are arbitrary, very little was known. It was shown by Chu, Frantzikinakis and Host [7] that

$$\frac{1}{N}\sum_{n=1}^{N} T_1^{p_1(n)} f_1 \cdots T_l^{p_l(n)} f_l \tag{1.2}$$

converges whenever the polynomials p_i have distinct degrees, but the convergence of (1.2) has remained open for arbitrary polynomials. Notice that (1.2) corresponds to taking $p_{i,j} = 0$ whenever $i \neq j$ in Theorem 1.1. More generally, very little was known until now for convergence of $\mathbb{Z}^d$ actions along polynomials. A particular result in this direction is the convergence of the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^{n^2} f\left(T^{n^2} S^n\right) g,$$

which was established by Austin [2], [3].

Finally, when G is only assumed to be nilpotent the results are much scarcer. Prior to this paper, it was known by the work of Bergelson and Leibman [6] that the averages

$$\frac{1}{N}\sum_{n=1}^{N} T^n f S^n g$$

always converge in L^2, but even in the linear case no convergence result has been previously established for more than two transformations.

1.2. *Overview of the proof.* Our proof of Theorem 1.1 does not make use of the aforementioned results and therefore provides an alternative proof of these statements, which in many cases is substantially simpler than the original ones. In particular, we do not make use of the machinery of characteristic factors that is heavily used in previous literature. The price we pay in doing so is that we do not obtain any explicit description of the limits. In this sense, our approach is similar to that of Tao [19] in that we use a weak inverse theorem (see Lemma 3.4) to decompose our functions into the sum of a random component, which is easily treatable, and a structured one, which can be handled by an adequate induction. Interestingly, we find that our decomposition

is best carried out by adapting ideas of Gowers related to the Hahn-Banach separation theorem [12], and this is done in Section 2. This is arguably the first time that these ideas are used in a purely ergodic theoretical context.

The main new ingredient of the proof is the concept of an L-reducible function (Definition 3.3), which will play the role of the structured component of our decompositions. We refer to Section 3 for precise definitions, but for now let us discuss what these are in the linear abelian case. Here, an L-reducible function σ with respect to a set of transformations $T_1, \dots, T_j$, is a function for which the behavior of $T_j^n \sigma$ can be somewhat recovered from that of the set $T_1^n b_1, \dots, T_{j-1}^n b_{j-1}$ for some prescribed set of functions b_i. This way, the problem of convergence for the set of transformations $T_1, \dots, T_j$ is reduced to the analogous question for the smaller set $T_1, \dots, T_{j-1}$, and one may then proceed inductively. The details of these reductions are carried out in Section 3.

When either G is not abelian or the polynomials are not linear, the system of transformations to which L-reducible functions allow us to pass does not admit such a simple expression. In general, it will consist of twice as many transformations as the original one and the degree of the polynomials involved may not necessarily decrease, so that it may seem that we have not gained much with this procedure. As it turns out however, one can define a suitable notion of complexity for every set of transformations and show that the above process does indeed lead us to a set of lower complexity. The proof that every system of transformations of the type studied in Theorem 1.1 reduces in finitely many steps to one consisting only of the identity transformation $Ix = x$ is performed in Section 4, and this completes the proof of Theorem 1.1.

The methods of this paper immediately shield some further convergence results, and these are discussed in Section 5. We also include an appendix with several examples of how the induction process mentioned in the previous paragraph works in some concrete cases.

Acknowledgments. I would like to thank my advisor Román Sasyk for a careful reading of the manuscript and several helpful discussions. I would also like to thank Tim Austin, Vitaly Bergelson and Nikos Frantzikinakis for their useful comments.

2. **Decompositions through the Hahn-Banach theorem**

In this section we will review some of the tools developed by Gowers and adapt them to the context of our problem. For a much better discussion of these topics the reader is referred to Gowers's paper [12]. We begin by stating the Hahn-Banach separation theorem in the form it will be needed.

THEOREM 2.1 (Geometric Hahn-Banach). *Let A be an open convex subset containing 0 of a real topological vector space V, and suppose $v \in V$ does not*

lie in A. Then there exists some continuous linear functional $\phi : V \to \mathbb{R}$ such that $\phi(v) \geq 1$ and $\phi(w) < 1$ for every $w \in A$.

The idea of Gowers to obtain decompositions can roughly be described as follows. While it may be difficult to check directly whether an arbitrary function can be described by the sum of a structured and a random component, if such a decomposition fails to exist, then an application of the Hahn-Banach theorem would allow us to find, since these sets tend to be convex, some large functional that does not correlate with random functions (therefore having a kind of structure itself) nor with structured functions (therefore also having some randomness). This way, we are only left with proving that no object can be random and structured at the same time, which generally tends to be an easier task.

We will concentrate on the study of a real Hilbert space $\mathcal{H}$ with norm $\|\cdot\|$. In order to apply the above scheme, we will need the following corollary.

Corollary 2.2 (cf. [12, Cor. 3.2]). *Let $A_1, \ldots, A_n$ be open convex subsets containing 0 of some real Hilbert space $\mathcal{H}$. Let $c_1, \ldots, c_n > 0$ be positive real numbers, and suppose $f \in \mathcal{H}$ cannot be written as $\sum_{j=1}^n c_j f_j$ with $f_j \in A_j$. Then there exists some $\phi \in \mathcal{H}$ such that $\langle \phi, f \rangle \geq 1$ and $\langle \phi, g_i \rangle < c_i^{-1}$ for every $g_i \in A_i$.*

Proof. Since the set $A := \sum_{i=1}^n c_i A_i$ will be an open convex set in $\mathcal{H}$ containing 0 but not f, it follows by the Hahn-Banach theorem that there exists some $\phi \in \mathcal{H}$ satisfying $\langle \phi, f \rangle \geq 1$ and $\langle \phi, g \rangle < 1$ for every $g \in A$. The result follows immediately, since $c_i g_i \in A$ for every $g \in A_i$. $\square$

Given a positive real number δ and some decreasing function $\eta : \mathbb{R}^+ \to \mathbb{R}^+$, we will consider the sequence $C_1^{\delta,\eta}, \ldots, C_{\lceil 2\delta^{-2} \rceil}^{\delta,\eta}$ defined recursively by

$$C_{\lceil 2\delta^{-2} \rceil}^{\delta,\eta} := 1, C_{n-1}^{\delta,\eta} := \max \left\{ C_n^{\delta,\eta}, 2\eta(C_n^{\delta,\eta})^{-1} \right\}. \tag{2.1}$$

We shall also write $C^{\delta,\eta} := C_1^{\delta,\eta}$. These constants will provide the parameters for the decomposition obtained below, and the fact that they are independent of the specific decomposition will allow us to do *a priori* modifications on our set of structured functions so that they are better suited to the resulting bounds.

Given some norm $\|\cdot\|_N$ on $\mathcal{H}$ equivalent to $\|\cdot\|$, we define its dual norm by

$$\|f\|_N^* := \sup_{\|g\|_N \leq 1} |\langle f, g \rangle|.$$

Notice that $\|\cdot\|_N^*$ is then also equivalent to $\|\cdot\|$. We will be concerned with the study of an infinite family of norms $(\|\cdot\|_N)_{N \in \mathbb{N}}$ measuring increasing rates of structure and for which their dual norms $(\|\cdot\|_N^*)_{N \in \mathbb{N}}$ measure decreasing rates

of randomness. As it turns out, we will need to work with this large family of norms *simultaneously*, so that if we know one of the components is random at a level A (that is, $\|\|_A^*$ is small), we need the other component to be structured at a much higher level B (that is, $\|\|_B$ must be small for some B much larger than A). This is accomplished by the following result.

PROPOSITION 2.3. *Let $(\|\cdot\|_N)_{N\in\mathbb{N}}$ be a family of norms on $\mathcal{H}$ equivalent to $\|\cdot\|$ and satisfying $\|\cdot\|_{N+1}^* \leq \|\cdot\|_N^*$ for every N. Let $0 < \delta, c < 1$ be positive real numbers, $\eta : \mathbb{R}^+ \to \mathbb{R}^+$ some decreasing function and $\psi : \mathbb{N} \to \mathbb{N}$ some function satisfying $\psi(N) \geq N$ for all N. Then, for every integer $M_\bullet > 0$, there exists a sequence*

$$M_\bullet \leq M_1 \leq \cdots \leq M_{\lceil 2\delta^{-2}\rceil} \leq M^\bullet = O_{M_\bullet,\delta,c,\psi}(1),$$

which does not depend on the specific family of norms, with the property that for any $f \in \mathcal{H}$ with $\|f\| \leq 1$, we can find some $1 \leq i \leq \lceil 2\delta^{-2}\rceil$ and integers A, B with $M_\bullet \leq A < cM_i < \psi(M_i) \leq B$, such that we have the decomposition $f = f_1 + f_2 + f_3$ with

$$\|f_1\|_B < C_i^{\delta,\eta}, \|f_2\|_A^* < \eta(C_i^{\delta,\eta}), \|f_3\| < \delta.$$

Proof. Our proof is modeled on the proof of Proposition 3.5 of [12]. Set $A_1 := M_\bullet$, $M_1 := \lceil c^{-1}A_1 + 1\rceil$ and $B_1 := \psi(M_1)$. If there is no decomposition of the desired form with these parameters and $C_1 := C_1^{\delta,\eta}$, we may apply Corollary 2.2 to obtain some $\phi_1 \in \mathcal{H}$ such that $\langle \phi_1, f\rangle \geq 1$, $\|\phi_1\|_{B_1}^* \leq C_1^{-1}$, $\|\phi_1\|_{A_1}^{**} \leq \eta(C_1)^{-1}$ and $\|\phi_1\| \leq \delta^{-1}$, where we are using the fact that if $\|\cdot\|_N$ is some norm equivalent to $\|\cdot\|$, then $\{f \in \mathcal{H} : \|f\|_N < 1\}$ is an open convex set in $\mathcal{H}$ containing 0.

Recursively, if we cannot find a decomposition with parameters A_{j-1}, M_{j-1}, B_{j-1}, C_{j-1}, then we set $A_j := B_{j-1}$, $M_j := \lceil c^{-1}A_j + 1\rceil$, $B_j := \psi(M_j)$ and $C_j := C_j^{\delta,\eta}$. If no such decomposition exists with these parameters, we can then use Corollary 2.2 to find some $\phi_j \in \mathcal{H}$ with properties analogous to the ones above. This way we construct a sequence of elements obeying the orthogonality relationships

$$\begin{aligned}|\langle \phi_j, \phi_i\rangle| &\leq \|\phi_j\|_{A_j}^{**}\|\phi_i\|_{A_j}^* \leq \|\phi_j\|_{A_j}^{**}\|\phi_i\|_{B_i}^* \\ &\leq \eta(C_j)^{-1}C_i^{-1} \leq 1/2,\end{aligned}$$

whenever $i < j$, by construction of C_k. But then, by the bounds on $\|\phi_i\|$, we obtain upon expanding the inner product

$$\|\phi_1 + \cdots + \phi_r\|^2 \leq \delta^{-2}r + \frac{r^2 - r}{2} \tag{2.2}$$

for each $r \leq \lceil 2\delta^{-2} \rceil$. On the other hand, the condition $\langle \phi_i, f \rangle \geq 1$ for all i implies that the left-hand side of (2.2) is at least r^2. Since this is absurd for $r = \lceil 2\delta^{-2} \rceil$, the result follows. □

Finally, we also prove the following lemma that will be needed later.

LEMMA 2.4 (cf. [12, Cor. 3.5]). *Let $\Sigma \subseteq \mathcal{H}$ be a bounded set and suppose the norm*

$$\|f\|_\Sigma := \inf \left\{ \sum_{j=0}^{k-1} |\lambda_j| : f = \sum_{j=0}^{k-1} \lambda_j \sigma_j, \sigma_j \in \Sigma \right\} \tag{2.3}$$

is well defined and equivalent to $\|\cdot\|$. Then its dual norm is given by $\|f\|_\Sigma^* = \sup_{\sigma \in \Sigma} |\langle f, \sigma \rangle|$.

Proof. Given some $f \in \mathcal{H}$, it is clear on one hand that

$$\sup_{\sigma \in \Sigma} |\langle f, \sigma \rangle| \leq \sup_{\|g\|_\Sigma \leq 1} |\langle f, g \rangle|.$$

On the other hand, for every $\epsilon > 0$, if $g = \sum_{j=0}^{k-1} \lambda_j \sigma_j$ with $\sum_{j=0}^{k-1} |\lambda_j| < 1 + \epsilon$, then $|\langle f, g \rangle| \leq (1+\epsilon) \sup_{\sigma \in \Sigma} |\langle f, \sigma \rangle|$. The result follows. □

3. Norm convergence for systems of finite complexity

From now on fix a nilpotent group G and a probability space X as in the statement of Theorem 1.1. By a G-sequence we shall mean a sequence $\{g(n)\}_{n \in \mathbb{Z}}$ taking values in G. An ordered tuple $\mathbf{g} = (g_1, \ldots, g_j)$ of G-sequences will be called a system, and for each system one can ask whether the corresponding ergodic averages

$$\mathcal{A}_N^{\mathbf{g}}[f_1, \ldots, f_j] := \mathbb{E}_{n \in [N]} \prod_{i=1}^{j} g_i(n) f_i \tag{3.1}$$

converge in $L^2(X)$ for every $f_1, \ldots, f_j \in L^\infty(X)$. Here, for a finite set A we write $\mathbb{E}_{x \in A} f(x) := \frac{1}{|A|} \sum_{x \in A} f(x)$, and for every positive integer N it is $[N] := \{1, \ldots, N\}$. We say two systems are equivalent if they consist of the same G-sequences, so for example if g, h are G-sequences, then the system (h, g) is equivalent to the system (g, h), and so is (g, h, h). Clearly, the convergence of the averages of the form (3.1) for some system implies the convergence of the averages associated to every equivalent system, since $T(f_1)T(f_2) = T(f_1 f_2)$ for every $T \in G$ and $f_1, f_2 \in L^\infty(X)$.

To each pair of G-sequences g, h we will associate, for each positive integer m, the G-sequence

$$\langle g | h \rangle_m(n) := g(n) g(n+m)^{-1} h(n+m),$$

and we define the m-reduction of a system $\mathbf{g} = (g_1, \dots, g_j)$ to be the system

$$\mathbf{g}_m^* = (g_1, \dots, g_{j-1}, \langle g_j | 1_G \rangle_m, \langle g_j | g_1 \rangle_m, \dots, \langle g_j | g_{j-1} \rangle_m),$$

where by a slight abuse of notation we write 1_G for the G-sequence $1_G(n) := 1_G$, where 1_G is the identity of G. The main purpose of this section will be to show that one can deduce the convergence of the averages (3.1) for some system $\mathbf{g}$ from knowing this (actually, the slightly stronger Theorem 3.2 below) for every reduction $\mathbf{g}_m^*$ of $\mathbf{g}$. This leads us to define the complexity of a system.

Definition 3.1 (Complexity of a system). We say a system $\mathbf{g}$ has complexity 0 if it is equivalent to the trivial system (1_G) (that is, the system consisting only on the sequence 1_G). Recursively, we say a system $\mathbf{g}$ has complexity d, for some positive integer $d \geq 1$, if it is not of complexity d' for any $0 \leq d' < d$ and it is equivalent to some system $\mathbf{h}$ for which every reduction $\mathbf{h}_m^*$ has complexity $\leq d-1$. We say a system has finite complexity if it has complexity d for some integer $d \geq 0$.

Given a system $\mathbf{g} = (g_1, \dots, g_j)$, some set of functions $f_1, \dots, f_j \in L^\infty(X)$ and a pair of integers N, N', write

$$\mathcal{A}^{\mathbf{g}}_{N,N'}[f_1, \dots, f_j] := \mathcal{A}^{\mathbf{g}}_{N'}[f_1, \dots, f_j] - \mathcal{A}^{\mathbf{g}}_{N}[f_1, \dots, f_j].$$

We have the following result.

THEOREM 3.2. *Let G and X be as above, and let $d \geq 0$. Let $F : \mathbb{N} \to \mathbb{N}$ be some nondecreasing function with $F(N) \geq N$ for all N, and let $\varepsilon > 0$ be some positive real number. Then, for every integer $M > 0$, there exists a sequence of integers*

$$M \leq M_1^{\varepsilon,F,d} \leq \dots \leq M_{K_{\varepsilon,d}}^{\varepsilon,F,d} \leq M^{\varepsilon,F,d} = O_{d,F,\varepsilon,M}(1), \tag{3.2}$$

for some $K_{\varepsilon,d} = O_{\varepsilon,d}(1)$, such that for every system $\mathbf{g} = (g_1, \dots, g_j)$ of complexity at most d and every choice of functions $f_1, \dots, f_j \in L^\infty(X)$ with $\|f_i\|_\infty \leq 1$, there exists some $1 \leq i \leq K_{\varepsilon,d}$ such that

$$\left\| \mathcal{A}^{\mathbf{g}}_{N,N'}[f_1, \dots, f_j] \right\|_{L^2(X)} \leq \varepsilon \tag{3.3}$$

for every $M_i^{\varepsilon,F,d} \leq N, N' \leq F(M_i^{\varepsilon,F,d})$.

This type of statement already appears in the works of Tao [19] and of Avigad, Gerhardy and Towsner [4]. Clearly, Theorem 3.2 implies that the averages (3.1) converge in $L^2(X)$ for every system $\mathbf{g}$ of finite complexity, since otherwise one could find some $\varepsilon > 0$ and some increasing function $F : \mathbb{N} \to \mathbb{N}$ such that $\left\| \mathcal{A}^{\mathbf{g}}_{N,F(N)}[f_1, \dots, f_d] \right\|_{L^2(X)} > \varepsilon$ for every integer N. The usefulness of Theorem 3.2 lies on its uniformity over all systems of a fixed complexity, which plays an important role in the inductive argument. In fact, the ergodic averages (3.1) associated to a system $\mathbf{g}$ for which the reductions $\mathbf{g}_m^*$ do not satisfy

stability bounds that are uniform on m may not necessarily converge, even if the ergodic averages associated to each individual reduction $\mathbf{g}_m^*$ do converge.

The rest of this section is devoted to the proof of Theorem 3.2. In Section 4 we will show that every system of the form given in Theorem 1.1 has finite complexity, thereby completing the proof of that theorem.

3.1. *L-reducible functions.* Since Theorem 3.2 is trivially true when $d = 0$, we may proceed by induction. Thus, let $d > 0$ be some positive integer, and assume the result holds for every $d' < d$. Let F and $0 < \varepsilon < 1$ be as in the statement of the theorem, and let $\mathbf{g} = (g_1, \dots, g_j)$ be some system of complexity at most d. Since it clearly sufficies to prove the result for any system equivalent to $\mathbf{g}$, by definition of the complexity we may assume without lost of generality that $\mathbf{g}_m^*$ has complexity $\leq d-1$ for every positive integer m.

Let C^* denote the quantity $C^{\delta,\eta}$ defined in (2.1) associated to $\delta := \varepsilon/(2^5 3)$ and $\eta(x) := \varepsilon^2/(2^3 3^3 x)$ so that, in particular, C^* depends only on ε. We will sometimes use the shorthands $\|\cdot\|_\infty$ for $\|\cdot\|_{L^\infty(X)}$, $\|\cdot\|_2$ for $\|\cdot\|_{L^2(X)}$ and $\langle\cdot,\cdot\rangle$ for $\langle\cdot,\cdot\rangle_{L^2(X)}$. The following definition will be crucial.

Definition 3.3 (reducible functions). Given a positive integer L, we say $\sigma \in L^\infty(X)$, $\|\sigma\|_\infty \leq 1$, is an L-reducible function (with respect to $\mathbf{g}$) if there exists some integer $M > 0$ and a family $b_0, b_1, \dots, b_{j-1} \in L^\infty(X)$ with $\|b_i\|_\infty \leq 1$, such that for every positive integer $l \leq L$,

$$\left\| g_j(l)\sigma - \mathbb{E}_{m\in[M]} \left(\langle g_j | 1_G\rangle_m(l)\right) b_0 \prod_{i=1}^{j-1} \left(\langle g_j | g_i\rangle_m(l)\right) b_i \right\|_{L^\infty(X)} < \frac{\varepsilon}{16C^*}.$$

Reducible functions will play a role similar to the one played by basic anti-uniform functions in [19]. We stress that we do not care for the value of M in Definition 3.3. We will show in Lemma 3.4 below that every function giving rise to a large average must resemble a reducible function. The main feature of these objects is that the role of the G-sequence g_j on the averages (3.1) can essentially be recovered by means of the set of G-sequences $\langle g_j|1_G\rangle_m, \langle g_j|g_1\rangle_m, \dots, \langle g_j|g_{j-1}\rangle_m$.

Lemma 3.4 (Weak inverse result for ergodic averages). *Assume the inequality* $\|\mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_{j-1}, u]\|_2 > \varepsilon/6$ *holds for some* $\|u\|_{L^\infty(X)} \leq 3C$, *some* $1 \leq C \leq C^*$ *and some* $f_1, \dots, f_{j-1} \in L^\infty(X)$ *with* $\|f_i\|_\infty \leq 1$. *Then, there exists some constant* $0 < c_1 < 1$, *depending only on* ε, *such that for every positive integer* $L < c_1 N$, *there is an* L*-reducible function* σ *with* $\langle u, \sigma\rangle > 2\eta(C)$.

Proof. We begin by noticing that $\|\mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_{j-1}, u]\|_2^2 = \langle u, h\rangle$, where

$$h := \mathbb{E}_{n\in[N]} g_j(n)^{-1} \mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_{j-1}, u] \prod_{i=1}^{j-1} g_j(n)^{-1} g_i(n) f_i. \tag{3.4}$$

We claim $\sigma := h/3C$ is an L-reducible function for every $L < c_1 N$ and some $0 < c_1 < 1$ depending only on ε, from where the result immediately follows since by the observation above it is $\langle u, \sigma \rangle > 2\eta(C)$.

It remains to prove this claim. Write $c_1 := \frac{\varepsilon}{96(C^*)^2}$ and assume that $0 < l < c_1 N$. Then, if we shift $[N]$ to $[N] + l$, we see that the right-hand side of (3.4) changes by a magnitude of at most $6lC^*/N < \varepsilon/(16C^*)$ (since $\|\mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_{j-1}, u]\|_\infty \leq 3C \leq 3C^*$), and thus

$$\left\| h - \mathbb{E}_{n\in[N]} g_j(l+n)^{-1} \mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_{j-1}, u] \prod_{i=1}^{j-1} g_j(l+n)^{-1} g_i(l+n) f_i \right\|_{L^\infty(X)} < \frac{\varepsilon}{16C^*}.$$

Applying $g_j(l)$ we get

$$\left\| g_j(l)h - \mathbb{E}_{n\in[N]} \left(\langle g_j | 1_G \rangle_n(l)\right) \mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_{j-1}, u] \prod_{i=1}^{j-1} \left(\langle g_j | g_i \rangle_n(l)\right) f_i \right\|_{L^\infty(X)} < \frac{\varepsilon}{16C^*}.$$

The claim then follows with $M := N$, $b_0 := \frac{1}{3C} \mathcal{A}_N[f_1, \dots, f_{j-1}, u]$ and $b_i := f_i$. □

As mentioned previously, the advantage of L-reducible functions is that they allow us to reduce the study of the ergodic averages of $\mathbf{g}$ to the study of averages arising from the reductions $\mathbf{g}_m^*$, which we already know to satisfy uniform stability bounds by the induction hypothesis. This idea is carried out in the next proposition.

Proposition 3.5. *For every positive integer M_*, there exists a sequence*

$$M_* \leq M_1 \leq \dots \leq M_{\tilde{K}} \leq M^* = O_{M_*, \varepsilon, d, F}(1) \tag{3.5}$$

depending only on M_*, ε, d and F, and with $\tilde{K}$ depending only on ε and d, such that if $f_1, \dots, f_{j-1} \in L^\infty(X)$ with $\|f_i\|_\infty \leq 1$ and $f = \sum_{t=0}^{k-1} \lambda_t \sigma_t$, where $\sum_{t=0}^{k-1} |\lambda_t| \leq C^*$ and each σ_t is an L-reducible function for some $L \geq F(M^*)$, then there exists some $1 \leq i \leq \tilde{K}$ such that

$$\left\| \mathcal{A}_{N,N'}^{\mathbf{g}}[f_1, \dots, f_{j-1}, f] \right\|_{L^2(X)} \leq \varepsilon/4$$

for every pair $M_i \leq N, N' \leq F(M_i)$.

Proof. For every σ_t, let $M^{(t)}$ be the integer coming from the definition of an L-reducible function and let $b_i^{(t)} \in L^\infty(X)$ be the corresponding family of functions. It follows from the definition of an L-reducible function that for every $N \leq L$ and every $0 \leq t \leq k-1$, we may replace $\mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_{j-1}, \sigma_t]$ by

$$\mathbb{E}_{m\in[M^{(t)}]} \left[\mathbb{E}_{n\in[N]} \left(\prod_{i=1}^{j-1} g_i(n) f_i \right) \left(\left(\langle g_j | 1_G \rangle_m(n)\right) b_0^{(t)} \right) \left(\prod_{i=1}^{j-1} \left(\langle g_j | g_i \rangle_m(n)\right) b_i^{(t)} \right) \right],$$

at the cost of an L^∞ error of at most $\varepsilon/(16C^*)$. Thus, we get by Minkowski's inequality that for $N, N' \leq L$, $\left\|\mathcal{A}^{\mathbf{g}}_{N,N'}[f_1,\dots,f_{j-1},f]\right\|_2$ is bounded by

(3.6)
$$\left(\sum_{t=0}^{k-1} |\lambda_t| \mathbb{E}_{m\in[M^{(t)}]} \left\|\mathcal{A}^{\mathbf{g}^*_m}_{N,N'}\left[f_1,\dots,f_{j-1},b_0^{(t)},b_1^{(t)},\dots,b_{j-1}^{(t)}\right]\right\|_{L^2(X)}\right) + \varepsilon/8.$$

We are thus given a large family of averages coming from the lower complexity systems $\mathbf{g}^*_m$. Write $\gamma := \frac{\varepsilon}{16C^*}$. Clearly, it would suffice to find a suitable interval on which each of these lower dimensional averages is bounded by γ. Although this will not be possible, we will indeed show by repeated applications of the induction hypothesis that we can get such a bound for all but a negligible subset of these averages. In order to do this, consider nondecreasing functions $F_1,\dots,F_r : \mathbb{N}\to\mathbb{N}$, for some $r = O_{\varepsilon,d}(1)$, to be specified, defined recursively by $F_r := F$ and $F_{i-1}(N) := \max_{1\leq M\leq N} F_i(M^{\gamma,F_i,d-1})$, where we are using the notation in the statement of Theorem 3.2. Also, let $K := K_{\gamma,d-1}$ be as in that theorem. For each tuple $1\leq i_1,\dots,i_s\leq K$, $s\leq r$ and integer M, we define recursively

$$M^{(i_1,\dots,i_s)} := \left(\left(\left(M_{i_1}^{\gamma,F_1,d-1}\right)_{i_2}^{\gamma,F_2,d-1}\right)\cdots\right)_{i_s}^{\gamma,F_s,d-1}.$$

Thus, $M^{(i_1)}$ is the integer $M_{i_1}^{\gamma,F_1,d-1}$ obtained in (3.2) by starting at M, $M^{(i_1,i_2)}$ is the integer $M_{i_2}^{\gamma,F_2,d-1}$ obtained by starting the sequence (3.2) at $M = M^{(i_1)}$, etc. In particular, notice that this sequence depends only on ε, F, d and M. Observe also that since each of the averages in (3.6) satisfies

$$\left\|\mathcal{A}^{\mathbf{g}^*_m}_{N,N'}\left[f_1,\dots,f_{j-1},b_0^{(t)},b_1^{(t)},\dots,b_{j-1}^{(t)}\right]\right\|_{L^\infty(X)} \leq 1, \tag{3.7}$$

the sum on (3.6) is bounded by $\sum_{t=0}^{k-1}|\lambda_t| \leq C^*$.

We now proceed as follows. By the induction hypothesis we know that each of the reduced averages in (3.6) is bounded by γ for every pair $N, N' \in [M_*^{(i)}, F_1(M_*^{(i)})]$ and some $1\leq i\leq K$, which depends on the particular average. By the pigeonhole principle and (3.7), this implies that we may find some $1\leq i_1\leq K$ such that the contribution to (3.6) of those averages which are not bounded by γ for every pair $N,N'\in[M_*^{(i_1)}, F_1(M_*^{(i_1)})]$ is at most $\left(\frac{K-1}{K}\right)C^*$. We now apply the induction hypothesis to these remaining averages with the function F_2, the parameter γ and the starting point $M_*^{(i_1)}$. This way, for each of these remaining averages, we know that there exists some $1\leq i\leq K$ such that the average is bounded by γ for every pair $N,N'\in[M_*^{(i_1,i)}, F_2(M_*^{(i_1,i)})]$. Since by construction of F_1 it is $[M_*^{(i_1,i)}, F_2(M_*^{(i_1,i)})] \subseteq [M_*^{(i_1)}, F_1(M_*^{(i_1)})]$, we see that those averages that we bounded in the previous step remain bounded by γ on each of these new intervals. Thus, we may apply the pigeonhole principle as before to find some $1\leq i_2\leq K$ such that the contribution to

(3.6) of those averages that are not bounded by γ for every pair $N, N' \in [M_*^{(i_1,i_2)}, F_2(M_*^{(i_1,i_2)})]$ is at most $\left(\frac{K-1}{K}\right)^2 C^*$.

Iterating the above process r times, we find a tuple $1 \leq i_1, \ldots, i_r \leq K$ such that the set of reduced averages that are not bounded by γ for every pair $N, N' \in [M_*^{(i_1,\ldots,i_r)}, F_r(M_*^{(i_1,\ldots,i_r)})] = [M_*^{(i_1,\ldots,i_r)}, F(M_*^{(i_1,\ldots,i_r)})]$, contributes at most

$$\left(\frac{K-1}{K}\right)^r C^* < \varepsilon/16$$

to (3.6), upon choosing r sufficiently large in terms of ε and d. Since the sum over the remaining terms will be bounded by $\sum_{t=0}^{k-1} |\lambda_t|\gamma < \varepsilon/16$, we conclude that (3.6), and therefore $\left\|\mathcal{A}^{\mathbf{g}}_{N,N'}[f_1, \ldots, f_{j-1}, f]\right\|_2$, is bounded by $\varepsilon/4$, for every $N, N' \in [M_*^{(i_1,\ldots,i_r)}, F(M_*^{(i_1,\ldots,i_r)})]$. Notice that while the specific integer $M_*^{(i_1,\ldots,i_r)}$ we have obtained depends on the set of functions $f_1, \ldots, f_{j-1}, f$ and the system $\mathbf{g}$, this integer belongs to the sequence $M_*^{(j_1,\ldots,j_r)}$, $1 \leq j_1, \ldots, j_r \leq K$, which depends only on F, ε, d and M_*. The result follows from this observation with the sequence (3.5) given by the integers $M_*^{(j_1,\ldots,j_r)}$, $1 \leq j_1, \ldots, j_r \leq K$. □

3.2. *Proof of Theorem* 3.2. We can now conclude the proof of Theorem 3.2. As it was done before, we fix X, G, F, ε, d and $\mathbf{g}$ as in the statement of that theorem and assume without lost of generality that each reduction $\mathbf{g}^*_m$ of $\mathbf{g}$ is of complexity at most $d-1$ and that the result is already proven for every $d' < d$. We will also write M_0 for the integer M to be chosen as the starting point of the sequence (3.2) in Theorem 3.2. Let δ, η be as specified at the beginning of Section 3.1, and write $C_i := C_i^{\delta,\eta}$ for the constants defined in (2.1). Given some positive integer L, write Σ_L for the set of L-reducible functions and set

$$\Sigma_L^+ := \Sigma_L \cup B_2(\delta/C^*),$$

where we write $B_2(\delta/C^*)$ for the set of $f \in L^2(X)$ with $\|f\|_2 \leq \delta/C^*$. Consider on $L^2(X)$ the norms $\|\cdot\|_L := \|\cdot\|_{\Sigma_L^+}$ defined as in (2.3). It is easy to see that these norms are well defined and equivalent to $\|\cdot\|_{L^2(X)}$ (by the presence of the small L^2 ball and the fact that reducible functions are bounded by 1). Also, notice that $\Sigma_{L+1}^+ \subseteq \Sigma_L^+$ for every L, which in turn implies (by Lemma 2.4) that $\|\cdot\|_{L+1}^* \leq \|\cdot\|_L^*$.

Given any integer M write $\psi(M) := F(M^*)$, where M^* is the integer obtained in Proposition 3.5 with $M_* = M$. Let $f_1, \ldots, f_j \in L^\infty(X)$, $\|f_i\|_\infty \leq 1$ be given, and consider for f_j a decomposition of the form provided in Proposition 2.3 with $(\|\cdot\|_L)_{L\in\mathbb{N}}$, ψ, δ, η as above and c equal to the constant c_1 in Lemma 3.4. This allows us to find a constant $1 \leq C_i \leq C^*$ and some integer

M with $M_0 \leq M = O_{M_0,\varepsilon,F,d}(1)$ such that

$$f_j = \sum_{t=0}^{k-1} \lambda_t \sigma_t + u + v, \tag{3.8}$$

where $\sum_{t=0}^{k-1} |\lambda_t| \leq C_i$, each σ_t belongs to Σ_B^+ for some $B \geq \psi(M)$ (and therefore to $\Sigma_{\psi(M)}^+$), $\|u\|_A^* < \eta(C_i)$ for some $A < c_1 M$ and $\|v\|_2 < \delta$. We remark that this constant C_i is the one defined in (2.1) and that the integer M obtained belongs to the sequence given in Proposition 2.3, which does not depend on the family of norms $(\|\cdot\|_L)_{L\in\mathbb{N}}$ and in the present case is therefore independent of the particular system $\mathbf{g}$ we have fixed (although it certainly depends on its complexity d, as well as on ε, F and M_0). Since $\|\sum_t^* \lambda_t \sigma_t\|_2 \leq \delta$, where the sum is restricted to those $\sigma_t \in B_2(\delta/C^*)$, we may assume that each σ_t in (3.8) actually belongs to $\Sigma_{\psi(M)}$, at the cost of softening our bound on v to $\|v\|_2 < 2\delta$.

We would like to use Lemma 3.4 to study the function u, but first we need to gain some control on its L^∞ norm. In order to do this, denote by $S \in \mathcal{X}$ the set of points on which the inequality $|v(s)| \leq C_i$ holds (in particular one has $\mu(S^c) < (2\delta/C_i)^2$), and write $v' := u\mathbf{1}_{S^c} + v$. From the fact that $\|\sigma_j\|_{L^\infty(X)} \leq 1$ for every $\sigma_j \in \Sigma_{\psi(M)}$, (3.8) and the definition of S, one easily checks that $|u\mathbf{1}_{S^c}(x)| \leq 3|v(x)|$ almost everywhere and therefore $\|u\mathbf{1}_{S^c}\|_2 \leq 3\|v\|_2$. Hence, it follows that for every pair of integers N, N',

$$\begin{aligned} \|\mathcal{A}_{N,N'}[f_1,\dots,f_{j-1},v']\|_2 &\leq \|\mathcal{A}_{N'}[f_1,\dots,f_{j-1},v']\|_2 + \|\mathcal{A}_N[f_1,\dots,f_{j-1},v']\|_2 \\ &\leq 2(4\|v\|_2) \\ &< \varepsilon/3, \end{aligned} \tag{3.9}$$

where we are using Minkowski's inequality and the fact that $\|f_i\|_\infty \leq 1$ for every $1 \leq i \leq j-1$. Consider now $u\mathbf{1}_S$. Similarly as above, one sees that $\|u\mathbf{1}_S\|_{L^\infty(X)} \leq 3C_i$. Also, it follows from Lemma 2.4 that for every $\sigma \in \Sigma_A$, it is

$$\begin{aligned} |\langle u_{\mathbf{1}S}, \sigma\rangle| &\leq |\langle u, \sigma\rangle| + |\langle u_{\mathbf{1}S^c}, \sigma_{\mathbf{1}S^c}\rangle| \\ &\leq \|u\|_A^* + \|u\mathbf{1}_{S^c}\|_2 \|\sigma\mathbf{1}_{S^c}\|_2 \\ &< \eta(C_i) + 12\delta^2/C_i \\ &< 2\eta(C_i). \end{aligned}$$

We are now in a position to apply Lemma 3.4, which implies that for every pair $N, N' \geq M$,

$$\begin{aligned} \|\mathcal{A}_{N,N'}[f_1,\dots,f_{j-1},u\mathbf{1}_S]\|_2 &\leq \|\mathcal{A}_{N'}[f_1,\dots,f_{j-1},u\mathbf{1}_S]\|_2 \\ &\quad + \|\mathcal{A}_N[f_1,\dots,f_{j-1},u\mathbf{1}_S]\|_2 \\ &\leq \varepsilon/3. \end{aligned} \tag{3.10}$$

It only remains to analyze $\sum_{t=0}^{k-1} \lambda_t \sigma_t$. But we may now invoke Proposition 3.5 to conclude from our choice of ψ that

$$\left\| \mathcal{A}_{N,N'}[f_1, \ldots, f_{j-1}, \sum_{t=0}^{k-1} \lambda_t \sigma_t] \right\|_{L^2(X)} < \varepsilon/3 \tag{3.11}$$

for every pair $M_i \leq N, N' \leq F(M_i)$ and some $M_i \in [M, \psi(M)]$ that belongs to the corresponding sequence (3.5). Theorem 3.2 then follows from (3.8), (3.9), (3.10), (3.11) and Minkowski's inequality.

4. **The complexity of polynomial systems**

In this section we will prove that every system of the form given in Theorem 1.1 has finite complexity, thereby finishing the proof of that theorem. In order to do this, we begin by reviewing some facts about polynomial sequences in nilpotent groups. For a detailed treatment of this topic, the reader is referred to the work of Leibman [16], [17].

For a G-sequence $g = (g(n))_{n\in\mathbb{Z}}$ taking values in a nilpotent group G and some integer m, we define the operator D_m that takes g to the G-sequence $(D_m g)(n) := g(n)g(n+m)^{-1}$. In particular, $\langle g|h\rangle_m(n) = (D_m g)(n)h(n+m)$ for every pair of G-sequences g, h and every positive integer m. We say that a G-sequence is *polynomial* if there exists some positive integer d such that for every choice of integers $m_1, \ldots, m_d$, we have $D_{m_1} \cdots D_{m_d} g = 1_G$, where we recall that 1_G stands for the constant sequence that equals the identity of G. It is known that if $(g(n))_{n\in\mathbb{Z}}$ is a sequence in a nilpotent group G that is of the form

$$g(n) = T_1^{p_1(n)} \cdots T_k^{p_k(n)} \tag{4.1}$$

for some $T_1, \ldots, T_k \in G$ and some set of integer valued polynomials $p_1, \ldots, p_k$, then g is a polynomial sequence. Indeed, each $T_i^{p_i(n)}$ is clearly a polynomial sequence and the product of polynomial sequences is polynomial by Lemma 4.1 below. (The converse also holds; see, for example, [16].)

By a polynomial system we shall mean a system $\mathbf{g} = (g_1, \ldots, g_j)$, where each g_i, $1 \leq i \leq j$, is a polynomial sequence. We define the size of such a system to be $|\mathbf{g}| = j$. To prove Theorem 1.1 it will suffice, by Theorem 3.2 and the fact that sequences of the form (4.1) are polynomial, to prove that every polynomial system has finite complexity.

In order to proceed, we will need to define the degree of a polynomial sequence. Unfortunately, the natural choice of taking the least positive integer d for which every d successive application of the above operators returns the identity is not appropriate for our purposes, since with this definition the set of polynomial sequences of degree $\leq d$ need not form a group. In order to amend this, we need to introduce some notation. Write $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and

$\mathbb{N}_* = \mathbb{N}_0 \cup \{-\infty\}$. We say a vector $\overline{d} = (d_1, \dots, d_k) \in \mathbb{N}_*^k$ is *superadditive* if $d_i \leq d_{i+1}$ for every $1 \leq i < k$ and $d_i + d_j \leq d_{i+j}$ for every pair i, j, where we are using the conventions $-\infty + t = -\infty$ for every $t \in \mathbb{N}_*$ and $-\infty < r$ for every $r \in \mathbb{N}_0$. Also, given a superadditive vector $\overline{d} = (d_1, \dots, d_k)$ and some nonnegative integer t, we write $\overline{d} - t := (d_1', \dots, d_k')$, where $d_i' = d_i - t$ if $t \leq d_i$ and $d_i' = -\infty$ otherwise. Notice that $\overline{d} - t$ so defined is also a superadditive vector.

Fix a nilpotent group G of nilpotency class s, and let $G = G_1 \supset G_2 \supset \cdots \supset G_s \supset G_{s+1} = \{1_G\}$ be its lower central series. As in [16], [17], we say a sequence $g = (g(n))_{n \in \mathbb{Z}}$ taking values in G is a polynomial sequence of (vector) degree $\leq (d_1, \dots, d_s)$ if $(D_{m_1} \cdots D_{m_{d_k+1}} g)(n) \in G_{k+1}$ for every n, every $1 \leq k \leq s$ and every choice of $m_1, \dots, m_{d_k+1} \in \mathbb{Z}$. If $d_k = -\infty$, we take this to mean that g itself takes values in G_{k+1}. We will make use of the following results of Leibman.

Lemma 4.1 ([17, §3]). *Let $\overline{d} = (d_1, \dots, d_s)$ be a superadditive vector, and let $t, t_1, t_2 \geq 0$ be nonnegative integers. Then we have the following properties*:

(1) *If g is a polynomial sequence of degree $\leq \overline{d} - t$, then $D_m g$ is a polynomial sequence of degree $\leq \overline{d} - (t+1)$ for every $m \in \mathbb{Z}$.*
(2) *The set of polynomial sequences of degree $\leq \overline{d} - t$ forms a group.*
(3) *If g is a polynomial sequence of degree $\leq \overline{d} - t_1$ and h is a polynomial sequence of degree $\leq \overline{d} - t_2$, then $[g, h]$ is a polynomial sequence of degree $\leq \overline{d} - (t_1 + t_2)$, where $[g, h](n) := g^{-1}(n) h^{-1}(n) g(n) h(n)$.*

Remark. The results of [17] concern the operators

$$(\widetilde{D}_m g)(n) := g(n)^{-1} g(n+m) = (D_m g^{-1})(n).$$

Nevertheless, using Lemma 4.1 for these operators and a straightforward descending induction on t, one can easily check that a G-sequence g has degree $\leq \overline{d} - t$ with respect to the operators $\widetilde{D}_m$ if and only if it has degree $\leq \overline{d} - t$ with respect to the operators D_m, from where we recover Lemma 4.1 as stated.

We say a polynomial system $\mathbf{g} = (g_1, \dots, g_j)$ has degree $\leq d$ if the degree of g_i is $\leq d$ for every $1 \leq i \leq j$. We will show that any system of degree $\leq \overline{d}$, for some superadditive vector $\overline{d} = (d_1, \dots, d_s)$, has finite complexity. Notice that this is enough to prove Theorem 1.1, since if a polynomial sequence g has degree $\leq (d_1, \dots, d_s)$, then it also has degree $\leq (d, 2d, \dots, sd)$, with $d = \max\{d_i : 1 \leq i \leq s\}$, and this last vector is clearly superadditive.

Given a polynomial system $\mathbf{g}$, we are concerned with the process that consists of passing from $\mathbf{g}$ to an equivalent system $\mathbf{g}'$, then taking the m-reduction $(\mathbf{g}')_m^*$ of $\mathbf{g}'$ for some m, passing to an equivalent system $((\mathbf{g}')_m^*)'$ and then taking the m'-reduction of this for some m', etc. What we are free to choose in the above process is to which equivalent system we apply

the reductions (but not the integer on which we subsequently reduce), and our objective is to show that there exists some constant C, depending on $\mathbf{g}$, such that for every sequence of positive integers $m, m', m'', \ldots$, we can go to the trivial system (1_G) by means of at most C repetitions of the above transformations. This clearly implies that the complexity of $\mathbf{g}$ is at most C.

In order to simplify notation we will omit the reference to the specific sequence of integers on which we reduce. So, for instance, we will generically refer to the reduction of a system $\mathbf{g} = (g_1, \ldots, g_j)$ to be the system

$$\mathbf{g}^* = (g_1, \ldots, g_{j-1}, \langle g_j | 1_G \rangle, \langle g_j | g_1 \rangle, \ldots, \langle g_j | g_{j-1} \rangle).$$

Similarly, we have the identity

$$\langle g | h \rangle(n) = Dg(n)(Dh(n))^{-1} h(n)$$

provided, of course, that the omitted subindices are the same. We define a *step* to be the process of passing from a system $\mathbf{g}$ to the reduction $(\mathbf{g}')^*$ of some system $\mathbf{g}'$ equivalent to $\mathbf{g}$. We will show that one can pass from a polynomial system $\mathbf{g}$ to the trivial system in a number of steps that is bounded in terms of the size and degree of $\mathbf{g}$.

We define the complete reduction of a system $\mathbf{g}$ to be the system

$$\mathbf{g}^{**} = (g_1, \ldots, g_{j-1}, \langle g_j | g_1 \rangle, \ldots, \langle g_j | g_{j-1} \rangle).$$

Thus, $\mathbf{g}^{**} = \mathbf{g}^* \setminus \{\langle g_j | 1_G \rangle\}$. We define a *complete step* in the same way as a step, but with the reduction replaced by the complete reduction. Complete steps are needed for a technical reason related to the inductive process to be applied. Precisely, in order to handle steps involving systems of degree $\leq \overline{d}$, we will need to assume some control on both steps and complete steps over systems of degree $\leq \overline{d} - 1$.

Theorem 1.1 follows from Theorem 3.2 and the following result.

THEOREM 4.2. *Let* $\mathbf{g}$ *be a polynomial system of size* $|\mathbf{g}| \leq C_1$ *and degree* $\leq \overline{d}$ *for some superadditive vector* $\overline{d} = (d_1, \ldots, d_s)$. *Then*

- *one can go from* $\mathbf{g}$ *to the trivial system* (1_G) *in* $O_{C_1, \overline{d}}(1)$ *steps,*
- *one can go from* $\mathbf{g}$ *to a system consisting of a single sequence of degree* $\leq \overline{d}$ *in* $O_{C_1, \overline{d}}(1)$ *complete steps*

for every sequence of positive integers $m, m', m'', \ldots$. *In particular,* $\mathbf{g}$ *has complexity* $O_{C_1, \overline{d}}(1)$.

Proof. Let $\overline{d}$ be as in the statement. We begin by noticing that the result is trivially true for systems of degree $\leq \overline{d} - (d_s + 1) = (-\infty, \ldots, -\infty)$, since 1_G is the only sequence lying in G_{k+1} for every $1 \leq k \leq s$. We will proceed by induction. Since $\overline{d} - t$ is superadditive for every $t \geq 0$, it will suffice to prove

that if Theorem 4.2 holds for systems of degree $\leq \overline{d}-1$, then it also holds for systems of degree $\leq \overline{d}$.

Thus, let $\mathbf{g}$ be as in the statement. We will first prove that we can go from $\mathbf{g}$ to the trivial system in $O_{C_1,\overline{d}}(1)$ steps (and therefore, that $\mathbf{g}$ has complexity $O_{C_1,\overline{d}}(1)$). In order to do this, observe that $\mathbf{g}$ can be rewritten in the form

$$\mathbf{g} = \mathbf{h}_0 \cup \bigcup_{i=1}^{l} s_i \mathbf{h}_i \tag{4.2}$$

for some polynomial sequences $s_1, \ldots, s_l$ of degree $\leq \overline{d}$, $l \leq C_1$, and some polynomial systems $\mathbf{h}_i$ of degree $\leq \overline{d}-1$ and size $\leq C_1$, with $\mathbf{h}_0$ possibly empty. (For example, one could simply take $s_i = g_i$ and $h_i = (1_G)$ for every $1 \leq i \leq l$.) Here, if $\mathbf{h} = (h_1, \ldots, h_k)$, then $s\mathbf{h}$ is the system $(sh_1, \ldots, sh_k)$ and the union of two systems $(h_1, \ldots, h_k)$, $(h'_1, \ldots, h'_r)$ is understood to be the system $(h_1, \ldots, h_k, h'_1, \ldots, h'_r)$.

The idea will be to show that for systems of the form (4.2) one can perform steps in such a way that the resulting systems are also of the form (4.2) for the same set of sequences $s_1, \ldots, s_l$. Furthermore, we will show that in finitely many steps we may actually discard the sequence s_l, therefore arriving at a system like (4.2) in which only the sequences $s_1, \ldots, s_{l-1}$ are present. Iterating this l times we shall then end up with a system of degree $\leq \overline{d}-1$, from where one can proceed by induction.

In order to carry out this plan, we begin by observing that if s_i, s_j are sequences of degree $\leq \overline{d}$ and h_i, h_j are sequences of degree $\leq \overline{d}-1$, we have

$$\begin{aligned} \langle s_j h_j | s_i h_i \rangle &= D(s_j h_j)(D(s_i h_i))^{-1} s_i h_i \\ &= s_i D(s_j h_j)(D(s_i h_i))^{-1} \left[D(s_j h_j)(D(s_i h_i))^{-1}, s_i \right] h_i \\ &= s_i h^{j,i} \end{aligned} \tag{4.3}$$

for some polynomial sequence $h^{j,i}$ that is seen to have degree $\leq \overline{d}-1$ by Lemma 4.1. Furthermore, if $s_i = s_j = s$, it is easy to check that

$$\langle s h_j | s h_i \rangle = s \langle h_j | h_i \rangle.$$

It follows from these formulas that, provided $|\mathbf{h}_l| > 1$, the reduction $\mathbf{g}^*$ of $\mathbf{g}$ is equivalent to a system of the form

$$\mathbf{h}_0^{(1)} \cup \left(\bigcup_{i=1}^{l-1} s_i \mathbf{h}_i^{(1)} \right) \cup s_l \mathbf{h}_l^{**}, \tag{4.4}$$

for some systems $\mathbf{h}_0^{(1)}, \mathbf{h}_1^{(1)}, \ldots, \mathbf{h}_{l-1}^{(1)}$ of degree $\leq \overline{d}-1$ and size $|\mathbf{h}_0^{(1)}| \leq 2|\mathbf{h}_0|+1$ and $|\mathbf{h}_i^{(1)}| \leq 2|\mathbf{h}_i|$ for every other i, and where we recall that $\mathbf{h}_l^{**}$ refers to the complete reduction of $\mathbf{h}_l$. Explicitly, if $\mathbf{h}_i = (h_{i,1}, \ldots, h_{i,j_i})$ for every $0 \leq i \leq l$,

then

$$\mathbf{h}_0^{(1)} = (\langle s_l h_{l,j_l} | 1_G \rangle, h_{0,1}, \ldots, h_{0,j_0}, \langle s_l h_{l,j_l} | h_{0,1} \rangle, \ldots, \langle s_l h_{l,j_l} | h_{0,j_0} \rangle),$$

while $s_i \mathbf{h}_i^{(1)}$ equals

$$(s_i h_{i,1}, \ldots, s_i h_{i,j_i}, \langle s_l h_{l,j_l} | s_i h_{i,1} \rangle, \ldots, \langle s_l h_{l,j_l} | s_i h_{i,j_i} \rangle)$$

for every $1 \leq i \leq l-1$. We see by (4.3) that this is of the desired form.

Observe now that if $\mathbf{h}$ is equivalent to $\mathbf{h}'$, then the system $s\mathbf{h}$ is also equivalent to $s\mathbf{h}'$. Since by the induction hypothesis we know that one can pass from $\mathbf{h}_l$ to a system $\mathbf{h}$ consisting of a single sequence of degree $\leq \overline{d}-1$ in $O_{C_1,\overline{d}-1}(1)$ complete steps, it follows from the above observation and (4.4) that we may pass from $\mathbf{g}$ to a system of the form

$$\mathbf{h}_0^{(2)} \cup \left(\bigcup_{i=1}^{l-1} s_i \mathbf{h}_i^{(2)} \right) \cup s_l \mathbf{h}, \tag{4.5}$$

in $O_{C_1,\overline{d}-1}(1)$ steps, where each system $\mathbf{h}_i^{(2)}$ has degree $\leq \overline{d}-1$ and size $O_{C_1,\overline{d}}(1)$, and $\mathbf{h}$ is a system consisting of a single sequence of degree $\leq \overline{d}-1$. But then we see from (4.3) that the reduction of (4.5) will be of the form

$$\mathbf{h}_0^{(3)} \cup \left(\bigcup_{i=1}^{l-1} s_i \mathbf{h}_i^{(3)} \right),$$

with each $\mathbf{h}_i^{(3)}$ having degree $\leq \overline{d}-1$ and size $O_{C_1,\overline{d}}(1)$. We have therefore succeeded in discarding the sequence s_l from our system. We can now repeat the same process as before with s_{l-1} in place of s_l. Since the size of $\mathbf{h}_{l-1}^{(3)}$ is $O_{C_1,\overline{d}}(1)$, we see that this new process finishes in $O_{C_1,\overline{d}}(1)$ steps, leaving us with a system of the form

$$\mathbf{h}_0^{(4)} \cup \left(\bigcup_{i=1}^{l-2} s_i \mathbf{h}_i^{(4)} \right).$$

Therefore, iterating the above process l times, we are finally left in $O_{C_1,\overline{d}}(1)$ steps with a system of degree $\leq \overline{d}-1$ from where we may apply the induction hypothesis to obtain the trivial system in $O_{C_1,\overline{d}}(1)$ further steps, thereby completing the proof of the finite complexity of $\mathbf{g}$.

Now it only remains to show that one can pass from $\mathbf{g}$ to a system consisting of a single sequence of degree $\leq \overline{d}$ in $O_{C_1,\overline{d}}(1)$ complete steps. But it is clear that the above reasoning to pass from $\mathbf{g}$ to a system of degree $\leq \overline{d}-1$ works in exactly the same way for complete steps, since the only things that may change are the systems $\mathbf{h}_0^{(1)}, \mathbf{h}_0^{(2)}, \mathbf{h}_0^{(3)}, \ldots$, which nevertheless will always have degree $\leq \overline{d}-1$ and whose size may only be smaller than in the previous case. Thus, the above reasoning allows us to pass to a system of degree $\leq \overline{d}-1$ from where we may apply induction, as long as after any of the complete steps we are not left with a system that can be written in its entirety as $s_i\mathbf{h}$ for

some s_i as above and $\mathbf{h}$ of degree $\leq \overline{d}-1$ and size $|\mathbf{h}|=1$ (because if the whole system has size 1, the complete reduction is not defined). But since in such a case we are already done, this completes the proof of Theorem 4.2 and therefore of Theorem 1.1. $\square$

5. **Further results**

The next result is easily seen to follow from the methods of this paper.

THEOREM 5.1. *Let G be a nilpotent group of measure preserving transformations of a probability space $(X,\mathcal{X},\mu)$. Then, for every $T_1,\dots,T_l\in G$, every $f_1,\dots,f_r\in L^\infty(X)$, every set of polynomials $p_{i,j}:\mathbb{Z}^d\to\mathbb{Z}$ and every Følner sequence $\{\Phi_N\}_{N=1}^\infty$ in $\mathbb{Z}^d$, the averages*

$$\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\prod_{j=1}^{r}\left(T_1^{p_{1,j}(u)}\dots T_l^{p_{l,j}(u)}\right)f_j \tag{5.1}$$

converge in $L^2(X,\mathcal{X},\mu)$.

During the proof of Theorem 1.1 we used crucially the fact that the L^∞ norm is an algebra norm ($\|fg\|_\infty\leq\|f\|_\infty\|g\|_\infty$). While this is not true for the L^2 norm, if we are concerned with the study of a single function $f\in L^2(X)$, this issue is no longer present. Furthermore, in this case our polynomial systems will always have size 1, a fact that allows us to drop the hypothesis of nilpotency on our group G (because we no longer need the product of polynomial sequences to be polynomial). More generally, it is easy to see from these observations that our methods produce the following result, which was also conjectured by Bergelson and Leibman in [6].

THEOREM 5.2. *Let G be a group of unitary operators on a Hilbert space $\mathcal{H}$. If $(g(n))_{n\in\mathbb{Z}}$ is a polynomial sequence in G, then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}g(n)u$$

exists for every $u\in\mathcal{H}$.

This was established by Bergelson and Leibman [6] for nilpotent G. While our result drops this hypothesis, it should be noted that it is not presently known if there are polynomial sequences whose study does not essentially reduce to the nilpotent case.

Appendix A. **Some examples of reductions**

We now provide some concrete examples of how the process studied in Section 4 returns the trivial system for some polynomial systems. Given systems $\mathbf{g}$ and $\mathbf{h}$, we write $\mathbf{g}\sim\mathbf{h}$ to mean that both systems are equivalent and we

write $\mathbf{g} \xrightarrow{m} \mathbf{h}$ to mean that $\mathbf{h}$ is the m-reduction of $\mathbf{g}$. Given measure preserving transformations T, S, C of a probability space, we will use the convention of writing $T^{n^2}S^nC$ for the G-sequence g given by $g(n) = T^{n^2}S^nC$.

Example A.1. As a trivial example, suppose our system is constant i.e., of the form $(C_1, \dots, C_j)$ for some constant G-sequences $C_1, \dots, C_j$. Then its m-reduction is equivalent to $(1_G, C_1, \dots, C_{j-1})$ for every m so, in particular, we get the trivial system after at most j steps.

Example A.2. If we are given a system of the form

$$(L_1^nC_{1,1}, \dots, L_1^nC_{1,i_1}, \dots, L_k^nC_{k,1}, \dots, L_k^nC_{k,i_k}),$$

with the transformations generating an abelian group, its m-reduction will be equivalent to

$$\begin{aligned}(L_k^{-m}, L_1^nC_{1,1}, \dots, L_1^nC_{1,i_1}, L_1^nC_{1,1}L_k^{-m}L_1^m, \dots, L_1^nC_{1,i_1}L_k^{-m}L_1^m, \dots,\\ L_{k-1}^nC_{k-1,1}, \dots, L_{k-1}^nC_{k-1,i_{k-1}}, L_{k-1}^nC_{k-1,1}L_k^{-m}L_{k-1}^m, \dots,\\ L_{k-1}^nC_{k-1,i_{k-1}}L_k^{-m}L_{k-1}^m, L_k^nC_{k,1}, \dots, L_k^nC_{k,i_k-1}).\end{aligned}$$

In particular, L_i^n appears twice as many times as before for every $1 \le i < k$, while L_k^n appears one time less, therefore disappearing after i_k steps. Notice also that at each step we get twice plus one as many constant sequences as before. Applying these observations we see that the system $(L_1^nC_1, \dots, L_j^nC_j)$ reduces to one consisting only of constant sequences after at most $a(j)$ steps, with $a : \mathbb{N} \to \mathbb{N}$ the function recursively defined by $a(1) = 1$ and $a(n+1) = a(n) + 2^{a(n)}$. We may then proceed as in Example A.1.

Example A.3. Consider now a system of the form

$$(S_1^nC_1, \dots, S_{j-1}^nC_{j-1}, T^{n^2}S_j^nC_j)$$

for commuting T, S_i, C_i. The m-reduction is given by

$$\begin{aligned}(S_1^nC_1, \dots, S_{j-1}^nC_{j-1}, T^{-2mn-m^2}S_j^{-m}, T^{-2mn-m^2}S_j^{-m}S_1^{n+m}C_1, \dots,\\ T^{-2mn-m^2}S_j^{-m}S_{j-1}^{n+m}C_{j-1}),\end{aligned}$$

which is a linear system and therefore reduces to the trivial system by the procedure discussed in Example A.2.

Example A.4. Let the system be $(T^{n^2}, T^{n^2}S^n)$ for commuting T and S. The m-reduction is given by

$$\begin{aligned}&(T^{n^2}, T^{-2mn-m^2}S^{-m}, T^{n^2}S^{-m})\\ &\quad \xrightarrow{l} (T^{n^2}, T^{-2mn-m^2}S^{-m}, T^{-2ln-l^2}, T^{n^2}, T^{-2ln-l^2-2m(n+l)-m^2}S^{-m}),\end{aligned}$$

and this is equivalent to a system of the form studied in Example A.3.

Example A.5. Consider (T^n, S^n) with T and S generating a nilpotent group. The m_1-reduction is given by

$$(T^n, S^{-m_1}, S^{-m_1}T^{m_1}T^n).$$

Write $C := S^{-m_1}T^{m_1}$, $C^{(1)} := S^{-m_1}$. Then the m_2-reduction of this is given by

$$\begin{aligned}(T^n, C^{(1)}, CT^{-m_2}C^{-1}, CT^{-m_2}C^{-1}T^{m_2}T^n, CT^{-m_2}C^{-1}C^{(1)})\\ \sim (C^{(1)}, C^{(2)}, C^{(3)}, T^n, [C^{-1}, T^{m_2}]T^n)\end{aligned}$$

for some constant G-sequences $C^{(2)}, C^{(3)}$ that depend on m_1, m_2. By the same reasoning we see that after reducing at $m_3, \dots, m_l$ (and passing to equivalent systems), we get the system

$$(C^{(1)}, C^{(2)}, \dots, C^{(c(l))}, T^n, [[[[C^{-1}, T^{m_2}]^{-1}, T^{m_3}]^{-1} \dots]^{-1}, T^{m_l}]T^n),$$

for some constant G-sequences $C^{(i)}$, $1 \le i \le c(l)$, with $c : \mathbb{N} \to \mathbb{N}$ the increasing function defined recursively by $c(1) = 1$ and $c(n+1) = 2c(n)+1$. Clearly, this is equivalent to $(C^{(0)}, C^{(1)}, \dots, C^{(c(l))}, T^n)$ for some l of size at most $s+1$, with s the nilpotency class of the group. Since any reduction of this last system will be a constant system of size $c(l+1)$, it follows that our original system (S^n, T^n) reduces to the trivial one in at most $s + 2 + c(s+2)$ steps.

Example A.6. Our last example is the system (T^{n^2}, S^{n^2}) for commuting T and S. We have

$$\begin{aligned}(T^{n^2}, S^{n^2}) &\overset{m_1}{\to} (T^{n^2}, S^{-2nm_1-m_1^2}, S^{-2nm_1-m_1^2}T^{n^2}T^{2nm_1+m_1^2})\\ &\sim (S^{-2nm_1-m_1^2}, T^{n^2}, S^{-2nm_1-m_1^2}T^{n^2}T^{2nm_1+m_1^2})\\ &\overset{m_2}{\to} (S^{-2nm_1-m_1^2}, T^{n^2}, T^{-2nm_2-m_2^2-2m_1m_2}S^{2m_1m_2},\\ &\qquad T^{-2nm_2-m_2^2-2m_1m_2}S^{-2nm_1-m_1^2}, T^{n^2}S^{2m_1m_2}T^{-2m_1m_2})\\ &\overset{m_3}{\to} (S^{-2nm_1-m_1^2}, T^{n^2}, T^{-2nm_2-m_2^2-2m_1m_2}S^{2m_1m_2},\\ &\qquad T^{-2nm_2-m_2^2-2m_1m_2}S^{-2nm_1-m_1^2}, T^{-2nm_3-m_3^2},\\ &\qquad T^{-2nm_3-m_3^2}S^{-2nm_1-2m_1m_3-m_1^2}, T^{n^2},\\ &\qquad T^{-2nm_3-m_3^2-2nm_2-2m_2m_3-m_2^2-2m_1m_2}S^{2m_1m_2},\\ &\qquad T^{-2nm_3-m_3^2-2nm_2-2m_2m_3-m_2^2-2m_1m_2}S^{-2nm_1-2m_1m_3-m_1^2}),\end{aligned}$$

and this last system is equivalent to one of the form studied in Example A.3.

References

[1] T. Austin, On the norm convergence of non-conventional ergodic averages, *Ergodic Theory Dynam. Systems* **30** (2010), 321–338. MR 2599882. Zbl 1206.37003. http://dx.doi.org/10.1017/S014338570900011X.

[2] ———, Pleasant extensions retaining algebraic structure, I, preprint. arXiv 0905.0518.

[3] ———, Pleasant extensions retaining algebraic structure, II, preprint. arXiv 0910.0907.

[4] J. Avigad, P. Gerhardy, and H. Towsner, Local stability of ergodic averages, *Trans. Amer. Math. Soc.* **362** (2010), 261–288. MR 2550151. Zbl 1187.37010. http://dx.doi.org/10.1090/S0002-9947-09-04814-4.

[5] V. Bergelson, Weakly mixing PET, *Ergodic Theory Dynam. Systems* **7** (1987), 337–349. MR 0912373. Zbl 0645.28012. http://dx.doi.org/10.1017/S0143385700004090.

[6] V. Bergelson and A. Leibman, A nilpotent Roth theorem, *Invent. Math.* **147** (2002), 429–470. MR 1881925. Zbl 1042.37001. http://dx.doi.org/10.1007/s002220100179.

[7] Q. Chu, N. Frantzikinakis, and B. Host, Ergodic averages of commuting transformations with distinct degree polynomial iterates, *Proc. Lond. Math. Soc.* **102** (2011), 801–842. MR 2795725. Zbl 1218.37009. http://dx.doi.org/10.1112/plms/pdq037.

[8] J.-P. Conze and E. Lesigne, Théorèmes ergodiques pour des mesures diagonales, *Bull. Soc. Math. France* **112** (1984), 143–175. MR 0788966. Zbl 0595.28018. Available at http://www.numdam.org/item?id=BSMF_1984__112__143_0.

[9] N. Frantzikinakis and B. Kra, Convergence of multiple ergodic averages for some commuting transformations, *Ergodic Theory Dynam. Systems* **25** (2005), 799–809. MR 2142946. Zbl 1140.37300. http://dx.doi.org/10.1017/S0143385704000616.

[10] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204–256. MR 0498471. Zbl 0347.28016.

[11] H. Furstenberg and B. Weiss, A mean ergodic theorem for $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, in *Convergence in Ergodic Theory and Probability* (Columbus, OH, 1993), *Ohio State Univ. Math. Res. Inst. Publ.* **5**, de Gruyter, Berlin, 1996, pp. 193–227. MR 1412607. Zbl 0869.28010.

[12] W. T. Gowers, Decompositions, approximate structure, transference, and the Hahn-Banach theorem, *Bull. Lond. Math. Soc.* **42** (2010), 573–606. MR 2669681. Zbl 05770950. http://dx.doi.org/10.1112/blms/bdq018.

[13] B. Host, Ergodic seminorms for commuting transformations and applications, *Studia Math.* **195** (2009), 31–49. MR 2539560. Zbl 1230.37006. http://dx.doi.org/10.4064/sm195-1-3.

[14] B. Host and B. Kra, Convergence of polynomial ergodic averages, *Israel J. Math.* **149** (2005), 1–19, Probability in mathematics. MR 2191208. Zbl 1085.28009. http://dx.doi.org/10.1007/BF02772534.

[15] ———, Nonconventional ergodic averages and nilmanifolds, *Ann. of Math.* **161** (2005), 397–488. MR 2150389. Zbl 1077.37002. http://dx.doi.org/10.4007/annals.2005.161.397.

[16] A. Leibman, Polynomial sequences in groups, *J. Algebra* **201** (1998), 189–206. MR 1608723. Zbl 0908.20029. http://dx.doi.org/10.1006/jabr.1997.7269.

[17] ———, Polynomial mappings of groups, *Israel J. Math.* **129** (2002), 29–60. MR 1910931. Zbl 1007.20035. http://dx.doi.org/10.1007/BF02773152.

[18] ———, Convergence of multiple ergodic averages along polynomials of several variables, *Israel J. Math.* **146** (2005), 303–315. MR 2151605. Zbl 1080.37002. http://dx.doi.org/10.1007/BF02773538.

[19] T. Tao, Norm convergence of multiple ergodic averages for commuting transformations, *Ergodic Theory Dynam. Systems* **28** (2008), 657–688. MR 2408398. Zbl 1181.37004. http://dx.doi.org/10.1017/S0143385708000011.

[20] H. Towsner, Convergence of diagonal ergodic averages, *Ergodic Theory Dynam. Systems* **29** (2009), 1309–1326. MR 2529651. Zbl 1179.37006. http://dx.doi.org/10.1017/S0143385708000722.

[21] Q. Zhang, On convergence of the averages $(1/N)\sum_{n=1}^{N} f_1(R^n x) f_2(S^n x) f_3(T^n x)$, *Monatsh. Math.* **122** (1996), 275–300. MR 1414862. Zbl 0911.28012. http://dx.doi.org/10.1007/BF01320190.

[22] T. Ziegler, Universal characteristic factors and Furstenberg averages, *J. Amer. Math. Soc.* **20** (2007), 53–97. MR 2257397. Zbl 1198.37014. http://dx.doi.org/10.1090/S0894-0347-06-00532-7.

(Received: December 6, 2011)

Universidad de Buenos Aires, Buenos Aires, Argentina
E-mail: mwalsh@dm.uba.ar