Multiple recurrence and nilsequences

Vitaly Bergelson[1], **Bernard Host**[2], **Bryna Kra**[3]
with an Appendix by Imre Ruzsa[4]

[1] Department of Mathematics, The Ohio State University, 100 Math Tower, 231 West 18th Avenue, Columbus, OH 43210-1174, USA. e-mail: `vitaly@math.ohio-state.edu`

[2] Équipe d'analyse et de mathématiques appliquées, Université de Marne la Vallée, 77454 Marne la Vallée Cedex, France. e-mail: `host@math.univ-mlv.fr`

[3] Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208, USA. e-mail: `kra@math.northwestern.edu`

[4] Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, POB 127, H-1364 Budapest, Hungary. e-mail: `ruzsa@renyi.hu`

May 4, 2004

Abstract. Aiming at a simultaneous extension of Khintchine's and Furstenberg's Recurrence theorems, we address the question if for a measure preserving system $(X,\mathcal{X},\mu,T)$ and a set $A\in\mathcal{X}$ of positive measure, the set of integers n such that $\mu(A\cap T^nA\cap T^{2n}A\cap\ldots\cap T^{kn}A)>\mu(A)^{k+1}-\varepsilon$ is syndetic. The size of this set, surprisingly enough, depends on the length $(k+1)$ of the arithmetic progression under consideration. In an ergodic system, for $k=2$ and $k=3$, this set is syndetic, while for $k\geq 4$ it is not. The main tool is a decomposition result for the multicorrelation sequence $\int f(x)f(T^nx)f(T^{2n}x)\ldots f(T^{kn}x)\,d\mu(x)$, where k and n are positive integers and f is a bounded measurable function. We also derive combinatorial consequences of these results, for example showing that for a set of integers E with upper Banach density $d^*(E)>0$ and for all $\varepsilon>0$, the set

$$\{n\in\mathbb{Z}\colon d^*\big(E\cap(E+n)\cap(E+2n)\cap(E+3n)\big)>d^*(E)^4-\varepsilon\}$$

is syndetic.

1. Introduction

1.1. Ergodic theory results

We begin by recalling two classical results of early ergodic theory. Let $(X,\mathcal{X},\mu,T)$ be a measure preserving probability system with an invertible

* The first author was partially supported by NSF grant DMS-0245350 and the third author was partially supported by NSF grant DMS-0244994.

measure preserving transformation T. (For brevity, we call $(X,\mathcal{X},\mu,T)$ a *system*.) Let $A\in\mathcal{X}$ be a set of positive measure. The Poincaré Recurrence Theorem states that

$$\mu(A\cap T^nA)>0 \textit{ for infinitely many values of } n.$$

The Khintchine Recurrence Theorem [K] states that the measure $\mu(A\cap T^nA)$ is 'large' for many values of n. Before stating the result precisely, we need a definition:

Definition 1.1. *A subset E of the integers $\mathbb{Z}$ is* syndetic *if $\mathbb{Z}$ can be covered by finitely many translates of E.*

In other words, E has *bounded gaps*, meaning that there exists an integer $L>0$ such that every interval of length L contains at least one element of E.

In [K], Khintchine strengthened the Poincaré Recurrence Theorem, showing that under the same assumptions:

$$\textit{For every } \varepsilon>0,\ \big\{n\in\mathbb{Z}\colon \mu(A\cap T^nA)>\mu(A)^2-\varepsilon\big\} \textit{ is syndetic.}$$

More recently, Furstenberg proved a Multiple Recurrence Theorem:

Theorem (Furstenberg [F1]). *Let $(X,\mathcal{X},\mu,T)$ be a system, let $A\in\mathcal{X}$ be a set with $\mu(A)>0$ and let $k\geq 1$ an integer. Then*

$$\liminf_{N-M\to+\infty}\frac{1}{N-M}\sum_{n=M}^{N-1}\mu(A\cap T^nA\cap T^{2n}A\cap\cdots\cap T^{kn}A)\ >0\ .$$

The $\liminf$ is actually a limit; see [HK]. (See also [Z2].)

In particular, there exist infinitely many integers n such that $\mu(A\cap T^nA\cap T^{2n}A\cap\cdots\cap T^{kn}A)>0$. Furstenberg's Theorem can thus be considered as a far reaching generalization of the Poincaré Recurrence Theorem, which corresponds to $k=1$. Our interest is in the existence of a theorem that has the same relation to the Khintchine Recurrence Theorem as Furstenberg's Theorem has to the Poincaré Recurrence Theorem. More precisely, under the same assumptions we ask if the set

$$\big\{n\in\mathbb{Z}\colon \mu\big(A\cap T^nA\cap\cdots\cap T^{kn}A\big)>\mu(A)^{k+1}-\varepsilon\big\} \tag{1.1}$$

is syndetic for every positive integer k and every $\varepsilon>0$. While it follows from Furstenberg's Theorem that for some constant $c=c(A)>0$, the set $\{n\in\mathbb{Z}\colon \mu(A\cap T^nA\cap\cdots\cap T^{kn}A)>c\}$ is syndetic, we are asking if this can be strengthened to make the set in (1.1) syndetic for every positive integer k and $c=\mu(A)^{k+1}-\varepsilon$ for every $\varepsilon>0$.

Under the hypothesis of ergodicity, the answer is positive for $k=2$ and $k=3$ and surprisingly enough, is negative for all $k\geq 4$. Under the assumption of ergodicity, we generalize the Khintchine Recurrence Theorem:

Theorem 1.2. *Let $(X,\mathcal{X},\mu,T)$ be an ergodic system and let $A\in\mathcal{X}$ be a set of positive measure. Then for every $\varepsilon>0$, the subsets*

$$\big\{n\in\mathbb{Z}:\mu(A\cap T^nA\cap T^{2n}A)>\mu(A)^3-\varepsilon\big\} \tag{1.2}$$

and

$$\big\{n\in\mathbb{Z}:\mu(A\cap T^nA\cap T^{2n}A\cap T^{3n}A)>\mu(A)^4-\varepsilon\big\} \tag{1.3}$$

of $\mathbb{Z}$ are syndetic.

While ergodicity is not needed for Khintchine's Theorem, it is essential in Theorem 1.2. Theorem 2.1 provides a counterexample in the general (nonergodic) case.

For arithmetic progressions of length ≥ 5, the result analogous to Theorem 1.2 does not hold. Using the result of Ruzsa contained in the Appendix, in Section 2.3 we show

Theorem 1.3. *There exists an ergodic system $(X,\mathcal{X},\mu,T)$ such that for all integers $\ell\geq 1$, there exists a set $A=A(\ell)\in\mathcal{X}$ with $\mu(A)>0$ and*

$$\mu\big(A\cap T^nA\cap T^{2n}A\cap T^{3n}A\cap T^{4n}A\big)\leq\mu(A)^\ell/2 \tag{1.4}$$

for every integer $n\neq 0$.

In fact, we find the slightly better upper bound $\mu(A)^{-c\log\mu(A)}$ for some set A of positive arbitrarily small measure and some positive constant c.

1.2. Combinatorial results

We recall a definition:

Definition 1.4. *The* upper Banach density *of a subset E of $\mathbb{Z}$ is:*

$$d^*(E)=\lim_{N\to+\infty}\sup_{M\in\mathbb{Z}}\frac{1}{N}\big|E\cap[M,M+N-1]\big|\ .$$

Note that the limit exists by subadditivity and is the infimum of the sequence. Furstenberg used his Multiple Recurrence Theorem to make the beautiful connection between ergodic theory and combinatorics and prove Szemerédi's Theorem:

Theorem (Szemerédi [S]). *A subset of integers with positive upper Banach density contains arithmetic progressions of arbitrary finite length.*

Using a variation of Furstenberg's Correspondence Principle (Proposition 3.1) and Theorem 1.2, we immediately deduce:

Corollary 1.5. *Let E be a set of integers with positive upper Banach density. Then for every $\varepsilon > 0$, the sets*

$$\left\{n \in \mathbb{Z} : d^*\big(E \cap (E+n) \cap (E+2n)\big) > d^*(E)^3 - \varepsilon\right\}$$

and

$$\left\{n \in \mathbb{Z} : d^*\big(E \cap (E+n) \cap (E+2n) \cap (E+3n)\big) > d^*(E)^4 - \varepsilon\right\}$$

are syndetic.

Roughly speaking, this means that given a set E with positive upper Banach density, for 'many' values of n, E contains 'many' arithmetic progressions of length 3 (or of length 4) with difference n. (The *difference* of the arithmetic progression $\{a, a+n, \dots, a+kn\}$ is the integer $n > 0$.)

The following result follows from the proof of Theorem 1.3 and shows that the analogous result does not hold for longer progressions:

Corollary 1.6. *For every positive integer ℓ, there exists a set of integers $E = E(\ell)$ with positive upper Banach density such that*

$$d^*\big(E \cap (E+n) \cap (E+2n) \cap (E+3n) \cap (E+4n)\big) \le d^*(E)^\ell / 2$$

for every nonzero integer n.

1.3. Nilsequences

We now explain the main ideas behind the ergodic theoretic results of Section 1.1.

Fix an integer $k \geq 1$. Given an ergodic system $(X, \mathcal{X}, \mu, T)$ and a set $A \in \mathcal{X}$ of positive measure, the key ingredient for our ergodic results is the analysis of the sequence

$$\mu\big(A \cap T^n A \cap T^{2n} A \cap \cdots \cap T^{kn} A\big) .$$

More generally, for a real valued function $f \in L^\infty(\mu)$, we consider the *multicorrelation* sequence

$$I_f(k,n) := \int f(x) \cdot f(T^n x) \cdot \ldots \cdot f(T^{kn} x)\, d\mu(x) . \tag{1.5}$$

When $k = 1$, Herglotz's Theorem states that the correlation sequence $I_f(1,n)$ is the Fourier transform of some positive finite measure $\sigma = \sigma_f$ on the torus $\mathbb{T}$:

$$I_f(1,n) := \int f \cdot f \circ T^n \, d\mu = \widehat{\sigma}(n) := \int_{\mathbb{T}} e^{2\pi i n t}\, d\sigma(t) .$$

By decomposing the measure σ into its continuous part σ^c and its discrete part σ^d, we can write the sequence $I_f(1,n)$ as the sum of two sequences

$$I_f(1,n) = \widehat{\sigma^c}(n) + \widehat{\sigma^d}(n) .$$

The sequence $\{\widehat{\sigma^c}(n)\}$ tends to 0 in uniform density:

Definition 1.7. *Let $\{a_n : n \in \mathbb{Z}\}$ be a bounded sequence. We say that a_n* tends to zero in uniform density, *and we write* UD-Lim $a_n = 0$, *if*

$$\lim_{N\to+\infty} \sup_{M\in\mathbb{Z}} \frac{1}{M} \sum_{n=M}^{M+N-1} |a_n| = 0 .$$

Equivalently, UD-Lim $a_n = 0$ if and only if for any $\varepsilon > 0$, the set $\{n \in \mathbb{Z} : |a_n| > \varepsilon\}$ has upper Banach density zero (cf. Bergelson [Ber], definition 3.5).

The sequence $\{\widehat{\sigma^d}(n)\}$ is *almost periodic*, and hence there exists a compact abelian group G, a continuous real valued function ϕ on G, and $a \in G$ such that $\widehat{\sigma^d}(n) = \phi(a^n)$ for all n.

The compact abelian group G is an inverse limit of compact abelian Lie groups. Thus any almost periodic sequence can be uniformly approximated by an almost periodic sequence arising from a compact abelian Lie group.

We find a similar decomposition for the multicorrelation sequences $I_f(k,n)$ for $k \geq 2$. The notion of an almost periodic sequence is replaced by that of a *nilsequence*, which we now define:

Definition 1.8. *Let $k \geq 1$ be an integer and let $X = G/\Lambda$ be a k-step nilmanifold. Let ϕ be a continuous real (or complex) valued function on G and let $a \in G$ and $e \in X$. The sequence $\{\phi(a^n \cdot e)\}$ is called a* basic k-step nilsequence. *A* k-step nilsequence *is a uniform limit of basic k-step nilsequences.*

(For the precise definition of a nilmanifold, see section 4.1.) Note that a 1-step nilsequence is the same as an almost periodic sequence. Examples of 2-step nilsequences are given in Section 4.3.

While an inverse limit of compact abelian Lie groups is a compact group, an inverse limit of k-step nilmanifolds is not, in general, the homogeneous space of some locally compact group (see Rudolph [R]). This explains why the definition of a nilsequence is not a direct generalization of the definition of an almost periodic sequence.

The general decomposition result is:

Theorem 1.9. *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system, let $f \in L^\infty(\mu)$ and let $k \geq 1$ be an integer. The sequence $\{I_f(k,n)\}$ is the sum of a sequence tending to zero in uniform density and a k-step nilsequence.*

We explain how Theorems 1.9 and 1.2 are related.

Definition 1.10. *Let $\{a_n : n \in \mathbb{Z}\}$ be a bounded sequence of real numbers. The* syndetic supremum *of this sequence is*

$$\text{synd-sup}\, a_n := \sup\Big\{ c \in \mathbb{R} : \{n \in \mathbb{Z} : a_n > c\} \textit{ is syndetic } \Big\} .$$

In Section 4.3, we show that the syndetic supremum of a nilsequence is equal to its supremum.

We use the following simple lemma several times:

Lemma 1.11. *Let $\{a_n\}$ and $\{b_n\}$ be two bounded sequences of real numbers. If* $\text{UD-Lim}(a_n - b_n) = 0$*, then* $\text{synd-sup}\, a_n = \text{synd-sup}\, b_n$.

Therefore, the syndetic supremums of the sequences $\{\mu(A \cap T^n A \cap T^{2n} A)\}$ and $\{\mu(A \cap T^n A \cap T^{2n} A \cap T^{3n} A)\}$ are equal to the supremums of the associated nilsequences and we are reduced to showing that they are greater than or equal to $\mu(A)^3$ and $\mu(A)^4$, respectively. This is carried out in Section 8, completing the proof of Theorem 1.2.

1.4. Conventions and notation

Given a system $(X, \mathcal{X}, \mu, T)$, in general we omit the σ-algebra from our notation and write (X, μ, T).

For a system (X, μ, T), a *factor* is used with two meanings: it is a T-invariant sub-σ-algebra $\mathcal{Y}$ or a system (Y, ν, S) and a measurable map $\pi : X \to Y$ such that $\pi\mu = \nu$ and $S \circ \pi = \pi \circ T$. These two definitions coincide under the identification of the σ-algebra $\mathcal{Y}$ of Y with the invariant sub-σ-algebra $\pi^{-1}(\mathcal{Y})$ of $\mathcal{X}$.

In a slight abuse of vocabulary, we say that Y is a factor of X. If f is an integrable function on X, we denote the conditional expectation of f on the factor $\mathcal{Y}$ by $\mathbb{E}(f \mid \mathcal{Y})$. We write $\mathbb{E}(f \mid Y)$ for the function on Y defined by $\mathbb{E}(f \mid \mathcal{Y}) = \mathbb{E}(f \mid Y) \circ \pi$. This expectation is characterized by

$$\text{for all } g \in L^\infty(\nu), \qquad \int_X f \cdot g \circ \pi \, d\mu = \int_Y \mathbb{E}(f \mid Y) \cdot g \, d\nu \, .$$

Throughout the article, we implicitly assume that the term “bounded function” means bounded and measurable.

1.5. Outline of the paper

In Section 2, we construct two examples, the first showing that ergodicity is a necessary assumption for Theorem 1.2 and the second, a counterexample for progressions of length ≥ 5 (Theorem 1.3). In Section 3, we use a variant of Furstenberg’s Correspondence Principle to derive combinatorial consequences of the ergodic theoretic statements. The bulk of the paper is devoted to describing the decomposition of multicorrelation sequences and proving Theorem 1.2. We start by reviewing nilsystems and construction of certain factors in Section 4, and then in Section 5 explicitly describe the limit of an average along arithmetic progressions in a nilsystem. In Section 6, we introduce technical notions needed for the decomposition and in

Section 7 we complete the proof of the decomposition. Section 8 combines these results and proves Theorem 1.2.

Acknowledgments: We thank I. Ruzsa for the combinatorial construction (contained in the Appendix) which is the starting point for the construction of the counterexample of Theorem 1.3. We also thank E. Lesigne for pointing us to the version of the Correspondence Principle we use and A. Leibman for useful explanations about nilsystems.

2. Combinatorial and ergodic counterexamples

In this section, c denotes a universal constant, with the understanding that its value may change from one use to the next. Let $m_{\mathbb{T}}$ denote the Haar measure on the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$.

2.1. A counterexample for a nonergodic system

In order to show that ergodicity is necessary in Theorem 1.2, we use the following result of Behrend:

Theorem (Behrend [Beh]). *For every integer $L > 0$, there exists a subset Λ of $\{0,1,\dots,L-1\}$ having more than $L\exp(-c\sqrt{\log L})$ elements that does not contain any nontrivial arithmetic progression of length 3.*

Theorem 2.1. *There exists a (nonergodic) system (X,μ,T) and, for every integer $\ell \geq 1$, there exists a subset A of X of positive measure so that*

$$\mu\big(A\cap T^nA\cap T^{2n}A\big) \leq \frac{1}{2}\mu(A)^{\ell}\ . \tag{2.1}$$

for every nonzero integer n.

We actually construct a set A of arbitrarily small positive measure with $\mu\big(A\cap T^nA\cap T^{2n}A\big) \leq \mu(A)^{-c\log(\mu(A))}$ for every integer $n \neq 0$ and a positive universal constant c.

Proof. Let $X = \mathbb{T}\times\mathbb{T}$, endowed with its Haar measure $\mu = m_{\mathbb{T}}\times m_{\mathbb{T}}$ and with the transformation T given by $T(x,y) = (x,y+x)$. Let Λ be a subset of $\{0,1,\dots,L-1\}$, not containing any nontrivial arithmetic progression of length 3. Define

$$B = \bigcup_{j\in\Lambda}\Big[\frac{j}{2L},\frac{j}{2L}+\frac{1}{4L}\Big)\ , \tag{2.2}$$

which we consider as a subset of the torus and set $A = \mathbb{T}\times B$.

For every integer $n \neq 0$ we have $T^n(x,y) = (x, y+nx)$ and

$$\mu\big(A\cap T^nA\cap T^{2n}A\big) = \iint_{\mathbb{T}\times\mathbb{T}} 1_B(y)1_B(y+nx)1_B(y+2nx)\,dm_{\mathbb{T}}(y)\,dm_{\mathbb{T}}(x) = \iint_{\mathbb{T}\times\mathbb{T}} 1_B(y)1_B(y+x)1_B(y+2x)\,dm_{\mathbb{T}}(y)\,dm_{\mathbb{T}}(x)\ .$$

We now bound this last integral. Let $x,y \in \mathbb{T}$ be such that the expression in this integral is nonzero. The three points $y, y+x$ and $y+2x$ belong to B and we can write

$$y = \frac{i}{2L} + a\ ;\ y+x = \frac{j}{2L} + b\ ;\ y+2x = \frac{k}{2L} + c$$

for integers i,j,k belonging to Λ and $a,b,c \in [0,1/4L) \pmod 1$. Then

$$\frac{i-2j+k}{2L} = -a+2b-c \in \big(\frac{-1}{2L}, \frac{1}{2L}\big)$$

and thus $i-2j+k=0$. The integers i,j,k form an arithmetic progression in Λ and so the only possibility is that they are all equal, giving that the three points $y, y+x, y+2x$ belong to the same subinterval of B. Therefore, $x \in (-1/(8L), 1/(8L)) \pmod 1$ and, for every $n \neq 0$,

$$\mu\big(A\cap T^nA\cap T^{2n}A\big) = \iint_{\mathbb{T}\times\mathbb{T}} 1_B(y)1_B(y+x)1_B(y+2x)\,dm_{\mathbb{T}}(y)\,dm_{\mathbb{T}}(x) \leq \frac{m_{\mathbb{T}}(B)}{4L}\ .$$

We have $m_{\mathbb{T}}(B) = |\Lambda|/(4L)$. By Behrend's Theorem, we can choose Λ of cardinality on the order of $L\exp(-c\sqrt{\log L})$. By taking L sufficiently large, an easy computation gives the bound (2.1). $\square$

2.2. *Quadratic configurations*

Out next goal is to show that the results of Theorem 1.2 do not hold for arithmetic progression of length ≥ 5. We start with a definition designed to describe certain patterns that do not occur.

Definition 2.2. *An* integer polynomial *is a polynomial taking integer values on the integers. When P is a nonconstant integer polynomial of degree ≤ 2, the subset*

$$\big\{P(0), P(1), P(2), P(3), P(4)\big\}$$

of $\mathbb{Z}$ is called a quadratic configuration of 5 terms, *written QC5 for short.*

Note that any QC5 contains at least 3 distinct elements. An arithmetic progression of length 5 is a QC5, corresponding to a polynomial of degree 1.

We mimic the construction in Theorem 2.1 for this setup:

Lemma 2.3. *Let Λ be a subset of $\{0,1,\dots,L-1\}$ not containing any QC5. For $j\in\Lambda$, let $I_j\subset\mathbb{T}$ be the interval*

$$I_j=[\frac{j}{4L},\frac{j}{4L}+\frac{1}{16L})\ .$$

and let B be the union of the intervals I_j for $j\in\Lambda$. Let $x,y,z\in\mathbb{T}$ be such that the five points

$$a_i=x+iy+i^2z \pmod 1,\ i=0,1,\dots,4$$

belong to B. Then $2y$ belongs $\pmod 1$ to the interval $\big(\frac{-1}{4L},\frac{1}{4L}\big)$.

Proof. We consider $a_0,\dots,a_4$ as real numbers belonging to the interval $[0,1)$; by the definition of B, they actually all belong to $[0,1/4)$. For $i=0,\dots,4$, let $j_i\in E$ be the integer such that $a_i\in I_{j_i}$.

The five elements $a_0,\dots,a_4$ of $\mathbb{T}$ satisfy the relations

$$a_3=a_0-3a_1+3a_2 \pmod 1 \text{ and } a_4=a_1-3a_2+3a_3 \pmod 1\ .$$

The real number $a_0-3a_1+3a_2$ belongs to the interval

$$J=\Big(\frac{j_0-3j_1+3j_2}{4L}-\frac{3}{16L},\frac{j_0-3j_1+3j_2}{4L}+\frac{1}{4L}\Big)$$

and this interval is contained in $(-3/4,1)$. As $a_3=a_0-3a_1+3a_2 \pmod 1$ and $a_3\in[0,1/4)$, the equality $a_3=a_0-3a_1+3a_2$ holds in $\mathbb{R}$ and thus $a_3\in J$.

Moreover $a_3\in I_{j_3}$ and, for every $j\neq j_0-3j_1+3j_2$, the interval I_j has empty intersection with the interval J. We get that $j_3=j_0-3j_1+3j_2$. By the same computation we have that $j_4=j_1-3j_2+3j_3$.

From these two relations it follows that there exists an integer polynomial Q with $j_i=Q(i)$ for $i=0,\dots,4$. This polynomial must be constant, for otherwise $\{j_0,\dots,j_4\}$ would be a QC5 in Λ. Therefore the five points a_i, $i=0,\dots,4$ belong to the same subinterval I_j of B. Since $2y=-3a_0+4a_1-a_2 \pmod 1$, we have that $2y\in(\frac{-1}{4L},\frac{1}{4L})$. □

The next counterexample relies on a combinatorial construction communicated to us by Imre Ruzsa; his construction is reproduced in the Appendix.

Theorem 2.4 (I. Ruzsa). *For every integer $L>0$ there exists a subset Λ of $\{0,1,\dots,L-1\}$ having more than $L\exp(-c\sqrt{\log L})$ elements that does not contain any QC5.*

2.3. Counterexample for longer progressions in ergodic theory.

Proof (Proof of Theorem 1.3). We first define a particular system. Recall that $\mathbb{T}$ denotes the torus and $m_{\mathbb{T}}$ its Haar measure. Define $X = \mathbb{T} \times \mathbb{T}$ and $\mu = m_{\mathbb{T}} \times m_{\mathbb{T}}$.

Let $\alpha \in \mathbb{T}$ be an irrational and let X be endowed with the transformation T given by

$$T(x,y) = (x+\alpha, y+2x+\alpha) .$$

It is classical that this system is ergodic. This also follows from the discussion in Section 4.2.

Let Λ be a subset of $\{0,\dots,L-1\}$ not containing any QC5, B the subset of $\mathbb{T}$ defined in Lemma 2.3 and $A = \mathbb{T} \times B$.

For every integer n and every $(x,y) \in X$ we have

$$T^n(x,y) = (x+n\alpha, y+2nx+n^2\alpha) .$$

Thus for $n \neq 0$ we have

$$\begin{aligned}
&\mu(A\cap T^nA\cap T^{2n}A\cap T^{3n}A\cap T^{4n}A)\\
&\quad = \iint_{\mathbb{T}\times\mathbb{T}} 1_B(y)1_B(y+2nx+n^2\alpha)1_B(y+4nx+4n^2\alpha)\\
&\qquad\qquad 1_B(y+6nx+9n^2\alpha)1_B(y+8nx+16n^2\alpha)\,dm_{\mathbb{T}}(x)\,dm_{\mathbb{T}}(y) .
\end{aligned}$$

Let $x,y \in \mathbb{T}$ and $n \neq 0$ be such that the expression in the last integral is not zero. By Lemma 2.3, $4nx$ belongs $\pmod 1$ to the interval $(\frac{-1}{4L}, \frac{1}{4L})$. Since $y \in B$, we have

$$\mu(A\cap T^nA\cap T^{2n}A\cap T^{3n}A\cap T^{4n}A) \leq \frac{m_{\mathbb{T}}(B)}{2L} = \frac{|\Lambda|}{32L^2} .$$

By Ruzsa's Theorem, we can choose Λ of cardinality on the order of $L\exp(-c\sqrt{\log L})$. By choosing L sufficiently large, an easy computation gives the bound (1.4). □

2.4. Counterexample for longer progressions for sets of integers with positive density.

Proof (Proof of Corollary 1.6). Let (X,μ,T) and A be the system and the subset of X defined in Section 2.3. Fix some $x \in X$ and define $E = \{m \in \mathbb{Z} \colon T^m x \in A\}$.

It is classical that the topological dynamical system (X,T) is uniquely ergodic; this also follows from the discussion in Section 4.2 and Theorem 4.1. Since the boundary of A has zero measure, we have $d^*(E) = \mu(A)$. By the same argument $d^*\big(E\cap(E+n)\cap(E+2n)\cap(E+3n)\cap(E+4n)\big) = \mu(A\cap T^nA\cap T^{2n}A\cap T^{3n}A\cap T^{4n}A)$. The same argument as in the proof of Theorem 1.3 gives the announced result. □

3. Translation between combinatorics and ergodic theory

3.1. Correspondence principle

In order to obtain combinatorial corollaries of our ergodic theoretic results, we need the following version of Furstenberg's Correspondence Principle:

Proposition 3.1. *Let E be a set of integers with positive upper Banach density. There exist an ergodic system $(X,\mathcal{X},\mu,T)$ and a set $A\in\mathcal{X}$ with $\mu(A)=d^*(E)$ such that $\mu(T^{m_1}A\cap\cdots\cap T^{m_k}A)\le d^*\big((E+m_1)\cap\cdots\cap(E+m_k)\big)$ for all integers $k\ge 1$ and all $m_1,\dots,m_k\in\mathbb{Z}$.*

The observation that it suffices to prove the ergodic theoretic results for ergodic systems was transmitted to us by Lesigne (personal communication); the proof we give is almost entirely contained in Furstenberg [F2].

Proof. We proceed as in the proof of Lemma 3.7 of [F2].

Let $\{0,1\}^{\mathbb{Z}}$ be endowed with the product topology and the shift map T given by $(Tx)_n=x_{n+1}$ for all $n\in\mathbb{Z}$. Define $e\in\{0,1\}^{\mathbb{Z}}$ by setting $e_n=1$ if $n\in E$ and $e_n=0$ otherwise. Let X be the closure of the orbit of e under T, meaning the closure of $\big\{T^m e\colon m\in\mathbb{Z}\big\}$. Set $A=\{x\in X\colon x_0=1\}$. It is a clopen (closed and open) subset of X. It follows from the definition that for every integer n, we have $T^n e\in A$ if and only if $n\in E$.

By definition of $d^*(E)$, there exist two sequences $\{M_i\}$ and $\{N_i\}$ of integers, with $N_i\to+\infty$, such that

$$\lim_{i\to\infty}\frac{1}{N_i}\big|E\cap[M_i,M_i+N_i-1]\big|\to d^*(E)\ .$$

In the proof of Lemma 3.7 in [F2], it is shown that there exists an invariant probability measure ν on X such that $\nu(A)$ is equal to the above limit and thus is equal to $d^*(E)$. By using the ergodic decomposition of ν under T, we have that there exists an ergodic invariant probability measure μ on X with $\mu(A)\ge d^*(E)$.

Let $m_1,\dots,m_k$ be integers. The set $T^{m_1}A\cap\cdots\cap T^{m_k}A$ is a clopen set. Its indicator function is continuous and by Proposition 3.9 of [F2], there exist two sequences $\{K_i\}$ and $\{L_i\}$ of integers, with $L_i\to+\infty$, such that

$$\begin{aligned}\mu(T^{m_1}A\cap\cdots\cap T^{m_k}A)&=\lim_i\frac{1}{L_i}\sum_{n=K_i}^{K_i+L_i-1}1_{T^{m_1}A\cap\cdots\cap T^{m_k}A}(T^n e)\\&=\lim_i\frac{1}{L_i}\big|(E+m_1)\cap\cdots\cap(E+m_k)\cap\{K_i,\dots,K_i+L_i-1\}\big|\\&\qquad\le d^*\big((E+m_1)\cap\cdots\cap(E+m_k)\big)\ .\end{aligned}$$

By using this bound with $k=1$ and $m_1=0$, we have that $\mu(A)\le d^*(E)$ and thus $\mu(A)=d^*(E)$. □

Using the modified correspondence principle and Theorem 1.2, we immediately deduce Corollary 1.5.

3.2. Combinatorial consequence for progressions of length 3 *and* 4

Szemerédi's Theorem can be formulated in a finite version:

Theorem (Szemerédi [S]). *For every integer $k \geq 1$ and every $\delta > 0$, there exists $N(k,\delta)$ such that for all $N \geq N(k,\delta)$, any subset E of $\{1,\dots,N\}$ with at least δN elements contains an arithmetic progression of length k.*

Similar to the finite version of Szemerédi's Theorem, we can derive a finite version, albeit a somewhat weaker one, of Corollary 1.5. We begin with a remark.

Write $\lfloor x \rfloor$ for the integer part of the real number x. From the finite version of Szemerédi's Theorem, it is not difficult to deduce that every subset E of $\{1,\dots,N\}$ with at least δN elements contains at least $\lfloor c(k,\delta)N^2 \rfloor$ arithmetic progressions of length k, where $c(k,\delta)$ is some constant. Therefore the set E contains at least $\lfloor c(k,\delta)N \rfloor$ progressions of length k with the same difference.

In view of Corollary 1.5, it is natural to ask the following question:

Question. Is it true that for every $\delta > 0$ and every $\varepsilon > 0$, there exists $N(\varepsilon,\delta)$ such that for every $N > N(\varepsilon,\delta)$, every subset E of $\{1,\dots,N\}$ with $|E| \geq \delta N$ elements contains at least $(1-\varepsilon)\delta^3 N$ arithmetic progressions of length 3 with the same difference and at least $(1-\varepsilon)\delta^4 N$ arithmetic progressions of length 4 with the same difference?[1]

We are not able answer this question but can prove a weaker result with a relatively intricate formulation:

Corollary 3.2. *For all real numbers $\delta > 0$ and $\varepsilon > 0$ and every integer $K > 0$, there exists an integer $M(\delta,\varepsilon,K) > 0$ such that for all $N > M(\delta,\varepsilon,K)$ and every subset $E \subset \{1,\dots,N\}$ with $|E| \geq \delta N$ there exist:*

- *a subinterval J of $\{1,\dots,N\}$ with length K and an integer $s > 0$ such that*
$$\big|E \cap (E-s) \cap (E-2s) \cap J\big| \geq (1-\varepsilon)\delta^3 K\,. \tag{3.1}$$
- *a subinterval J' of $\{1,\dots,N\}$ with length K and an integer $s' > 0$ such that*
$$\big|E \cap (E-s') \cap (E-2s') \cap (E-3s') \cap J\big| \geq (1-\varepsilon)\delta^4 K\,. \tag{3.2}$$

Statement (3.1) means that $E \cap J$ contains at least $(1-\varepsilon)\delta^3 K$ starting points of arithmetic progressions of length 3 included in E, all with the same difference. Statement (3.2) has the analogous meaning for progressions of length 4.

[1] Recently Ben Green gave a positive answer to this question for progressions of length 3 (preprint available at http://www.arxiv.org, math.CO/0310476).

Proof. We only prove the result for progressions of length 3, as the proof for progressions of length 4 is identical.

Assume that the result does not hold. Then there exist $\delta > 0$, $\varepsilon > 0$, $K > 0$, a sequence $\{N_i\}$ tending to $+\infty$ and for every i a subset E_i of $\{1,\dots,N_i\}$ with $|E_i| \geq \delta N_i$ such that for every subinterval J of length K of $\{1,\dots,N_i\}$, relation (3.1) (with E_i substituted for E) is false.

We can assume that $N_i > K$ for every i. By induction, we build a sequence $\{M_i\}$ of positive integers with $M_{i+1} > 2M_i + 2N_i$ and $M_{i+1} > M_i + N_i + N_{i+1}$ for every i. Define the set E to be the union of the sets $M_i + E_i$. We have $d^*(E) = \limsup_i |E_i|/N_i \geq \delta$.

Fix an integer $s > 0$. By construction, for every integer M there exist i and an interval J of length K, included in $[M_i+1, M_i+N_i]$, such that $E \cap [M, M+K-1] \subset E \cap J$. We deduce that

$$\sup_M \big|E \cap (E-s) \cap (E-2s) \cap [M, M+K-1]\big| = \sup_i \sup_{\substack{J \subset [M_i+1, M_i+N_i] \\ |J|=K}} \big|E \cap (E-s) \cap (E-2s) \cap J\big| . \quad (3.3)$$

By construction, if for some integer $m \in E$ we have $m+s \in E$, $m+2s \in E$ and $m \in [M_i+1, M_i+N_i]$, then the integers $m+s$ and $m+2s$ also belong to the same interval. Therefore, for $J \subset [M_i+1, M_i+N_i]$,

$$E \cap (E-s) \cap (E-2s) \cap J = \big(E_i \cap (E_i - s) \cap (E_i - 2s) \cap (J - M_i)\big) + M_i .$$

Putting this into Equation (3.3), we have that

$$\sup_M \big|E \cap (E-s) \cap (E-2s) \cap [M, M+K-1]\big| = \sup_i \sup_{\substack{I \subset \{1,\dots,N_i\} \\ |I|=K}} \big|E_i \cap (E_i - s) \cap (E_i - 2s) \cap I\big| \leq (1-\varepsilon)\delta^3 K$$

by definition of the sets E_i.

We deduce that for every $s > 0$, we have $d^*\big(E \cap (E-s) \cap (E-2s)\big) \leq (1-\varepsilon)\delta^3$ and Corollary 1.5 provides a contradiction. □

The answer to the similar question for longer arithmetic progressions is negative: there exist significant subsets of $\{1,\dots,N\}$ that contain very few arithmetic progressions of length ≥ 5 with the same difference.

Proposition 3.3. *For all integers $\ell > 0$, there exists $\delta > 0$ such that for infinitely many values of N, there exists a subset E of $\{1,\dots,N\}$ with $|E| \geq \delta N$ that contains no more than $\frac{1}{2}\delta^\ell N$ arithmetic progressions of length 5 with the same difference.*

We give only the main ideas of the proof, as it lies a bit far from the main focus of this paper.

Let L and B be as in the proof of Theorem 1.3 (Sections 2.2 and 2.3). Let α be an irrational which is well approximated by a sequence $\{p_j/q_j\}$ of rationals with q_j prime and define:

$$F = \{n \in \mathbb{N}\colon\ n^2\alpha \in B \pmod 1\}\ .$$

Let N be one of the primes q_j and $E = F \cap \{1,\dots,N\}$.

Then $|E|/N$ is close to $m(B)$. Let s be a positive integer.

Let the integer n be such that the arithmetic progression $\{n, n+s, \dots, n+4s\}$ is included in E. By Lemma 2.3, $2ns\alpha$ belongs modulo 1 to the interval $(\frac{-1}{4L}, \frac{1}{4L})$. Note that n and s are smaller than $N = q_j$. By approximating α by p_j/q_j and using the primality of q_j we see that the number of possible values of n for a given s is bounded by cN/L for some positive constant c.

Therefore E contains fewer than cN/L progressions of length 5 with the same difference. For L sufficiently large we get the announced bound. Once again, the bound we actually find is $\delta^{-c\log(\delta)}N$ for some constant $c > 0$. □

4. Preliminaries

4.1. Nilsystems

We review some definitions and properties needed in the sequel. The notation introduced here is used throughout the rest of this paper.

Let G be a group. For $g, h \in G$ we write $[g,h] = g^{-1}h^{-1}gh$. When A and B are two subsets of G we write $[A,B]$ for the subgroup of G spanned by $\{[a,b] : a \in A,\ b \in B\}$. The lower central series

$$G = G_1 \supset G_2 \supset \cdots \supset G_j \supset G_{j+1} \supset \dots$$

of G is defined by

$$G_1 = G \text{ and } G_{j+1} = [G, G_j] \text{ for } j \geq 1\ .$$

Let $k \geq 1$ be an integer. We say that *G is k-step nilpotent* if $G_{k+1} = \{1\}$.

Let G be a k-step nilpotent Lie group and let Λ be a discrete cocompact subgroup. The compact manifold $X = G/\Lambda$ is called a *k-step nilmanifold.* The group G acts naturally on X be left translation and we write $(g,x) \mapsto g \cdot x$ for this action. There exists a unique Borel probability measure μ on X invariant under this action, called the *Haar measure* of X.

The fundamental properties of nilmanifolds were established by Malcev [M]. We make use of the following property several times, which appears in [M] for connected groups and is proved in Leibman [Lei2] in a similar way for the general case:

– *For every integer $j \geq 1$, the subgroups G_j and ΛG_j are closed in G. It follows that the group $\Lambda_j = \Lambda \cap G_j$ is cocompact in G_j.*

Let t be a fixed element of G and let $T : X \to X$ be the transformation $x \mapsto t \cdot x$. Then (X,T) is called a *k-step topological nilsystem* and (X,μ,T) is called a *k-step nilsystem*.

Fundamental properties of nilsystems were established by Auslander, Green and Hahn [AGH] and by Parry [Pa1]. Further ergodic properties were proven by Parry [Pa2] and Lesigne [Les] when the group G is connected, and generalized by Leibman [Lei2].

We summarize various properties of nilsystems that we need:

Theorem 4.1. *Let $(X = G/\Lambda,\mu,T)$ be a k-step nilsystem with T the translation by the element $t \in G$. Then:*

1. *(X,T) is uniquely ergodic if and only if (X,μ,T) is ergodic if and only if (X,T) is minimal if and only if (X,T) is transitive (that is, if there exists a point x whose orbit $\{T^n x \colon n \in \mathbb{Z}\}$ is dense).*
2. *Let Y be the closed orbit of some point $x \in X$. Then Y can be given the structure of a nilmanifold, $Y = H/\Gamma$, where H is a closed subgroup of G containing t and Γ is a closed cocompact subgroup of H.*
3. *For any continuous function f on X and any sequences of integers $\{M_i\}$ and $\{N_i\}$ with $N_i \to +\infty$ the averages*

$$\frac{1}{N_i}\sum_{n=M_i}^{M_i+N_i-1} f(T^n x)$$

converge for all $x \in X$.

Assume furthermore that

(H) G is spanned by the connected component of the identity and the element t.

Then:

4. *The groups G_j, $j \geq 2$, are connected.*
5. *The nilsystem (X,μ,T) is ergodic if and only if the rotation induced by t on the compact abelian group $G/G_2\Lambda$ is ergodic.*
6. *If the nilsystem (X,μ,T) is ergodic then its Kronecker factor is $Z = G/G_2\Lambda$ with the rotation induced by t and with the natural factor map $X = G/\Lambda \to G/G_2\Lambda = Z$.*

For connected groups, parts 1, 2 and 3 of this theorem can be deduced from [AGH], [F3] and [Pa1], while parts 4 and 6 are proved in [Pa1]. When G is connected and simply connected and, more generally, when G can be imbedded as a closed subgroup of a connected simply connected k-step nilpotent Lie group, all parts of this theorem were proved in [Les]. In the case that the group is simply connected, the result follows from Lesigne's proof. The general case for parts 1, 2, 3, 4 and 6 follow from [Lei2]. The proof of part 4 was transmitted to us by Leibman (personal communication) and we outline it here.

Proof. (Part 4) Assume that property (H) holds and let $G_{(0)}$ be the connected component of the identity 1 of G. The second commutator G_2 is spanned by the commutators of the generators of G. Since $[t,t]=1$, therefore we have $G_2=[G_{(0)},G]$. For $h\in G$, the map $g\mapsto[g,h]$ is continuous from $G_{(0)}$ to G_2 and maps 1 to 1. Thus its range is included in the connected component of 1 in G_2. We get that G_2 is connected.

We proceed by induction for the commutator subgroups of higher order. Assume that the n-th commutator subgroup G_n is connected. Proceeding as above, using the connectedness of G_n, we have that for $g\in G_n$ and $h\in G$, $[g,h]$ belongs to the connected component of 1 in G_{n+1}. Thus G_{n+1} is connected. □

We also use (in Section 5.2) a generalization of part 5 of Theorem 4.1 for two commuting translations on a nilmanifold, but it is just as simple to state it for ℓ commutating translations:

Theorem 4.2 (Leibman [Lei2], Theorem 2.17). *Let $X=G/\Lambda$ be a k-step nilmanifold endowed with its Haar measure μ. Let $t_1,\dots,t_\ell$ be commuting elements of G and let $T_1,\dots,T_\ell$ be the associated translations on X. Assume that*

(H') G is spanned by the connected component of the identity and the elements $t_1,\dots,t_\ell$.

Then the action of $\mathbb{Z}^\ell$ on X spanned by $T_1,\dots,T_\ell$ is ergodic if and only if the induced action on $G/G_2\Lambda$ is ergodic.

We return to the case of a single transformation. Let (X,μ,T) be an *ergodic* k-step nilsystem. There are several ways to represent X as a nilmanifold G/Λ. For our purposes, we reduce to a particular choice of the representation.

Assume that $X=G/\Lambda$ and let $t\in G$ be the element defining T. The connected component $G_{(0)}$ of the identity in G projects to an open subset of X. By ergodicity, the subgroup $\langle G_{(0)},t\rangle$ of G spanned by $G_{(0)}$ and t projects onto X. Substituting this group for G and $\Lambda\cap\langle G_{(0)},t\rangle$ for Λ, we have reduced to the case that hypothesis (H) is satisfied.

Let Λ' be the largest normal subgroup of G included in Λ. Substituting G/Λ' for G and Λ/Λ' for Λ, hypothesis (H) remains valid and we have reduced to the case that

(L) *Λ does not contain any nontrivial normal subgroup of G.*

4.2. Two examples

We start by reviewing the simplest examples of 2-step nilsystems.

4.2.1. Let $G = \mathbb{Z} \times \mathbb{T} \times \mathbb{T}$, with multiplication given by

$$(k,x,y) * (k',x',y') = (k+k',x+x',y+y'+2kx') \; .$$

Then G is a Lie group. Its commutator subgroup is $\{0\} \times \{0\} \times \mathbb{T}$ and G is 2-step nilpotent. The subgroup $\Lambda = \mathbb{Z} \times \{0\} \times \{0\}$ is discrete and cocompact. Let X denote the nilmanifold G/Λ and we maintain the notation of the preceding Section, with one small modification. Here $Z = \mathbb{T}$, $m = m_{\mathbb{T}}$ is the Haar measure of $\mathbb{T}$, and the factor map $\pi : X \to Z$ is given by $(k,x,y) \mapsto x$; it is thus more natural to use additive notation for Z.

Let α be an irrational point in $\mathbb{T}$, $a = (1,\alpha,\alpha)$ and $T : X \to X$ the translation by a. Then (X,μ,T) is a 2-step nilsystem. Note that hypotheses (H) and (L) are satisfied. Since α is irrational the rotation (Z,m,T) is ergodic and (X,μ,T) is ergodic by part 6 of Theorem 4.1.

We give an alternate description of this system. The map $(k,x,y) \mapsto (x,y)$ from G to $\mathbb{T}^2$ induces a homeomorphism of X onto $\mathbb{T}^2$. Identifying X with $\mathbb{T}^2$ via this homeomorphism, the measure μ becomes equal to $m_{\mathbb{T}} \times m_{\mathbb{T}}$, and the transformation T of X is given for $(x,y) \in \mathbb{T}^2 = X$ by

$$T(x,y) = (x+\alpha, y+2x+\alpha) \; .$$

This is exactly the system used in the construction of the counterexample in Section (2.3).

4.2.2. We also review another standard example of a 2-step ergodic nilsystem. Let G be the Heisenberg group $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$, with multiplication given by

$$(x,y,z) * (x',y',z') = (x+x',y+y',z+z'+xy') \; . \tag{4.1}$$

Then G is a 2-step nilpotent Lie group. The subgroup $\Lambda = \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ is discrete and cocompact. Let $X = G/\Lambda$ and let T be the translation by $t = (t_1,t_2,t_3) \in G$ with t_1,t_2 independent over $\mathbb{Q}$ and $t_3 \in \mathbb{R}$. We have that $(G/\Lambda, T)$ is a nilsystem. Hypothesis (H) is obviously satisfied since G is connected. Here the compact abelian group $G/G_2\Lambda$ is isomorphic to $\mathbb{T}^2$ and the rotation on $\mathbb{T}^2$ by (t_1,t_2) is ergodic. Therefore the system $(G/\Lambda, T)$ is uniquely ergodic.

Note that hypothesis (L) is not satisfied by G and Λ. The reduction explained above consists here in taking the quotient of G and Λ by the subgroup $\Lambda' = \{0\} \times \{0\} \times \mathbb{Z}$. We get that X is the quotient of G/Λ' by Λ/Λ' where $G/\Lambda' = \mathbb{R} \times \mathbb{R} \times \mathbb{T}$ with multiplication given by (4.1) and $\Lambda/\Lambda' = \mathbb{Z} \times \mathbb{Z} \times \{0\}$.

4.3. Nilsequences

For clarity, we repeat some of the definitions given in the introduction. Let $X = G/\Lambda$ be a k-step nilmanifold, ϕ be a continuous function on X, $e \in X$

and $t \in G$. The sequence $\{a_n\}$ given by $a_n = \phi(t^n \cdot e)$ is called a k-step *basic nilsequence.* We say that a bounded sequence is a k-*step nilsequence* if it is a uniform limit of k-step basic nilsequences.

Let X, e, t and ϕ be as above and let Y be the closed orbit of e. By part 2 of Theorem 4.1, (Y,T) can be given the structure of a nilsystem. Since this system is transitive, it is minimal by part 1 of the same theorem. Let $S = \sup_{n\in\mathbb{Z}} a_n$ and $\varepsilon > 0$. The set $U = \{x \in Y : \phi(x) > S - \varepsilon\}$ is a nonempty open subset of Y. By minimality of (Y,T), the set $\{n \in \mathbb{Z} : t^n \cdot e \in U\}$ is syndetic and thus $\text{synd-sup}\, a_n \geq S - \varepsilon$. Therefore for every basic nilsequence $\{a_n\}$, we have

$$\text{synd-sup}\, a_n = \sup_{n\in\mathbb{Z}} a_n \ .$$

This property passes to uniform limits and is therefore valid for every nilsequence.

The Cartesian product of two k-step nilsystems is again a k-step nilsystem and so the family of basic k-step nilsequences is a subalgebra of ℓ^∞. Therefore the family of k-step nilsequences is a closed subalgebra of ℓ^∞. This algebra is clearly invariant under translation and invariant under complex conjugation.

We give two examples of 2-step nilsequences, arising from the two examples of 2-step nilsystems given above.

4.3.1. Let (X,T) be the nilsystem defined in Section 4.2.1. We identify X with $\mathbb{T}\times\mathbb{T}$. Let $e = (0,0)$. For every integer n, we have $T^n e = (n\alpha, n^2\alpha)$. Let k and ℓ be two integers and let ϕ be the function on X given by $\phi(x,y) = \exp(2\pi i(kx+\ell y))$. The sequence

$$\big\{\exp(2\pi i(kn+\ell n^2)\alpha)\big\}$$

is a 2-step nilsequence.

4.3.2. Let (X,T) denote the system defined in Section 4.2.2. We use the first representation of this system.

We first define a continuous function on X. Let f be a continuous function on $\mathbb{R}$, tending sufficiently fast to 0 at infinity. For $(x,y,z) \in \mathbb{R}^3$, define

$$\psi(x,y,z) := \exp(2\pi iz) \sum_{k\in\mathbb{Z}} \exp(2\pi ikx)\, f(y+k) \ .$$

Then ψ is a continuous function on $\mathbb{R}^3$ and an immediate computation gives that for all $(x,y,z) \in \mathbb{R}^3$ and for all $(p,q,r) \in \mathbb{Z}^3$,

$$\psi\big((x,y,z) * (p,q,r)\big) = \psi(x,y,z) \ .$$

Therefore the function ψ on $G = \mathbb{R}^3$ induces a continuous function ϕ on the quotient X of G by $\Lambda = \mathbb{Z}^3$. Let e be the image of $(0,0,0)$ in X. For every integer n we have $\phi(t^n \cdot e) = \psi(t^n)$ and $t^n = (nt_1, nt_2, nt_3 + \frac{n(n-1)}{2} t_1 t_2)$. Therefore the sequence $\{a_n\}$ given by

$$a_n = \exp\big(2\pi i n t_3\big) \exp(2\pi i \frac{n(n-1)}{2} t_1 t_2) \sum_{k \in \mathbb{Z}} \exp(2\pi i k n t_1) f(nt_2 + k)$$

is a 2-step nilsequence.

4.4. Construction of certain factors

In this Section, (X, μ, T) is an ergodic system.

We review the construction of some factors in Host and Kra [HK]. These are the factors that control the limiting behavior of the multiple ergodic averages associated to the expressions $I_f(k,n)$. We begin recalling some well known facts about the Kronecker factor.

4.4.1. The Kronecker factor and the ergodic decomposition of $\mu \times \mu$ Let $(Z(X), m, T)$ denote the Kronecker factor of (X, μ, T) and let $\pi : X \to Z(X)$ be the factor map.

When there is no ambiguity, we write Z instead of $Z(X)$. We recall that Z is a compact abelian group, endowed with a Borel σ-algebra $\mathcal{Z}$ and Haar measure m. The transformation $T : Z \to Z$ has the form $z \mapsto \alpha z$ for some fixed element α of Z.

For $s \in Z$, we define a measure μ_s on $X \times X$ by

$$\int_{X \times X} f(x) f'(x') \, d\mu_s(x, x') = \int_Z \mathbb{E}(f \mid Z)(z) \cdot \mathbb{E}(f' \mid Z)(sz) \, d\mu(z) \ . \tag{4.2}$$

For every $s \in Z$ the measure μ_s is invariant under $T \times T$ and is ergodic for m-almost every s. The ergodic decomposition of $\mu \times \mu$ under $T \times T$ is

$$\mu \times \mu = \int_Z \mu_s \, dm(s) \ .$$

4.4.2. The factors Z_k. We recall some constructions of Sections 3 and 4 in [HK]. For an integer $k \geq 0$, we write $X^{[k]} = X^{2^k}$ and $T^{[k]} : X^{[k]} \to X^{[k]}$ for the map $T \times T \times \ldots \times T$, taken 2^k times.

We define a probability measure $\mu^{[k]}$ on $X^{[k]}$, invariant under $T^{[k]}$ by induction. Set $\mu^{[0]} = \mu$. For $k \geq 0$, let $\mathscr{I}^{[k]}$ be the σ-algebra of $T^{[k]}$-invariant subsets of $X^{[k]}$. Then $\mu^{[k+1]}$ is the relatively independent square of $\mu^{[k]}$ over $\mathscr{I}^{[k]}$. This means that if F', F'' are bounded functions on $X^{[k]}$,

$$\int_{X^{[k+1]}} F'(x') F''(x'') \, d\mu^{[k+1]}(x', x'') := \int_{X^{[k]}} \mathbb{E}(F' \mid \mathscr{I}^{[k]}) \, \mathbb{E}(F'' \mid \mathscr{I}^{[k]}) \, d\mu^{[k]} \ , \tag{4.3}$$

where $x = (x', x'')$ is an element of $X^{[k+1]}$, considered under the natural identification of $X^{[k+1]}$ with $X^{[k]} \times X^{[k]}$.

For a bounded function f on X we can define

$$|||f|||_k := \left(\int_{X^{[k]}} \prod_{j=0}^{2^k-1} f(x_j)\, d\mu^{[k]}(x) \right)^{1/2^k} \tag{4.4}$$

because this last integral is nonnegative. It is shown in [HK] that for every integer $k \geq 1$, $|||\cdot|||_k$ is a seminorm on $L^\infty(\mu)$.

The seminorms define factors of $\mathcal{X}$. Namely, the sub-σ-algebra $\mathcal{Z}_{k-1}(X)$ of $\mathcal{X}$ is characterized by

$$\text{for } f \in L^\infty(\mu),\ \mathbb{E}(f|\mathcal{Z}_{k-1}(X)) = 0 \text{ if and only if } |||f|||_k = 0\ . \tag{4.5}$$

The factor $Z_k(X)$ is the factor of X associated to $\mathcal{Z}_k(X)$. This gives that $Z_0(X)$ is the trivial factor, $Z_1(X)$ is the Kronecker factor. When there is no ambiguity, we write Z_k and $\mathcal{Z}_k$ instead of $Z_k(X)$ and $\mathcal{Z}_k(X)$.

4.4.3. The factors associated to the measures μ_s For each $s \in Z$ such that $(X \times X, \mu_s, T \times T)$ is ergodic, for each integer $k \geq 1$, a measure $(\mu_s)^{[k]}$ on $(X \times X)^{[k]}$ can be defined in the way that $\mu^{[k]}$ was defined from μ. Furthermore, a seminorm $\|\cdot\|_{s,k}$ on $L^\infty(\mu_s)$ can be associated to this measure, in the same way that the seminorm $\|\cdot\|_k$ is associated to $\mu^{[k]}$. In Section 3 of [HK], it is shown that under the natural identification of $(X \times X)^{[k]}$ with $X^{[k+1]}$, we have

$$\mu^{[k+1]} = \int_Z (\mu_s)^{[k]}\, dm(s)\ .$$

It follows from definition (4.4) that for every $f \in L^\infty(\mu)$,

$$|||f|||_{k+1}^{2^{k+1}} = \int_Z |||f \otimes f|||_{s,k}^{2^k}\, dm(s)\ .$$

From this we immediately deduce:

Proposition 4.3. *Let $k \geq 2$ be an integer and let f be a bounded function on X. If f has zero conditional expectation on $\mathcal{Z}_k$, then for m-almost every $s \in Z$ the function $f \otimes f$, considered as a function on $(X \times X, \mu_s)$, has zero conditional expectation on $\mathcal{Z}_{k-1}(X \times X, \mu_s, T \times T)$.*

4.4.4. Inverse limits of nilsystems. We say that the system (X, T) is an *inverse limit* of a sequence of factors $\{(X_j, T)\}$ if $\{\mathcal{X}_j\}_{j\in\mathbb{N}}$ is an increasing sequence of sub-σ-algebras invariant under the transformation T such that $\bigvee_{j\in\mathbb{N}} \mathcal{X}_j = \mathcal{X}$ up to null sets. If each system (X_j, T) is isomorphic to a k-step nilsystem, then we say that (X, T) is an *inverse limit of k-step nilsystems.*

Theorem 10.1 of [HK] states that for every ergodic system (X, μ, T) and every integer $k \geq 1$ the system $Z_k(X)$ is an inverse limit of k-step nilsystems.

4.5. Arithmetic progressions

We continue assuming that (X,μ,T) is an ergodic system. From Theorem 12.1 of [HK] and the characterization (4.5) of the factor Z_{k-1} we have:

Theorem 4.4. *Let $k \geq 2$ be an integer and $f_0, f_1, \dots, f_k$ bounded functions on X. If at least one of these functions has zero conditional expectation on $\mathcal{Z}_{k-1}$ then for all sequences $\{M_i\}$ and $\{N_i\}$ of integers with $N_i \to +\infty$,*

$$\lim_{i\to\infty} \frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \int f_0(x)\, f_1(T^n x)\, f_2(T^{2n}x) \dots f_k(T^{kn}x)\, d\mu(x) = 0\ .$$

Corollary 4.5. *Let $k \geq 2$ be an integer and let $f_0, f_1, \dots, f_k$ be bounded functions on X. If at least one of these functions has zero conditional expectation on $\mathcal{Z}_k$ then*

$$\int f_0(x)\, f_1(T^n x)\, f_2(T^{2n}x) \dots f_k(T^{kn}x)\, d\mu(x)$$

converges to zero in uniform density.

Proof. Let $\{M_i\}$ and $\{N_i\}$ be two sequences of integers, with $N_i \to +\infty$. For m-almost every $s \in Z$, one of the functions $f_0 \otimes f_0, \dots, f_k \otimes f_k$ has zero conditional expectation on $\mathcal{Z}_k(X \times X, \mu_s, T \times T)$ by Proposition 4.3 and thus the averages on $\{M_i, \dots, M_i + N_i - 1\}$ of

$$\int f_0(x)\, f_0(x')\, f_1(T^n x)\, f_1(T^n x') \dots f_k(T^{kn}x)\, f_k(T^{kn}x')\, d\mu_s(x,x')$$

converge to zero by Theorem 4.4 applied to the system $(X \times X, \mu_s, T \times T)$ and to the functions $f_0 \otimes f_0, \dots, f_k \otimes f_k$. Integrating with respect to s we get

$$\frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \Big(\int f_0(x)\, f_1(T^n x) \dots f_k(T^{kn}x)\, d\mu(x) \Big)^2 \to 0$$

and the result follows. □

Recall that for a bounded measurable function f on X and an integer $k \geq 1$, we defined

$$I_f(k,n) = \int f(x)\, f(T^n x) \dots f(T^{kn}x)\, d\mu(x)\ .$$

Even more generally, one can consider the same expression with $k+1$ distinct bounded functions $f_0, \dots, f_k$. However, this gives no added information for the problems we are studying and so we restrict to the above case.

Corollary 4.6. *Let $k \geq 1$ be an integer, let f be a bounded function on X and let $g = \mathbb{E}(f \mid \mathcal{Z}_k)$. Then $I_f(k,n) - I_g(k,n)$ converges to zero in uniform density.*

Let k, f and g be as in this Corollary. We consider g as a function defined on Z_k. Note that the functions f and g have the same integral.

Since the system Z_k is an inverse limit of a sequence of k-step nilsystems, the function g can be approximated arbitrarily well in L^1-norm by its conditional expectation on one of these nilsystems. We use this remark in the proof of Theorem 1.2 in Section 8.

5. The limit of the averages

In this section $k \geq 1$ is an integer, $(X = G/\Lambda, \mu, T)$ is an ergodic k-step nilsystem and the transformation T is translation by the element $t \in G$. We keep the notation of Section 4.1 and assume that hypotheses (H) and (L) are satisfied.

Recall that we let G_j denote the j-th commutator of G and that $\Lambda_j = \Lambda \cap G_j$. We have that $G = G_1$, but sometimes it is convenient to use both notations in the same formula.

For $f \in L^\infty(\mu)$, we first study the averages of the sequence $I_f(k,n)$. This establishes a short proof of a recent result by Ziegler [Z1]. We use some algebraic constructions based on ideas of Petresco [Pe] and Leibman [Lei1].

We explain the idea behind this construction. It is natural to define an arithmetic progression of length $k+1$ in G as an element of G^{k+1} of the form $(g, hg, h^2g, \dots, h^kg)$ for some $g, h \in G$. Unfortunately, these elements do not form a subgroup of G^{k+1}. However such elements do span the subgroup $\tilde{G}$ (defined in the next section), which could thus be called *the group of arithmetic progressions of length* $k+1$ *in* G.

Similarly, one is tempted to define an arithmetic progression of length $k+1$ in X as a point in X^{k+1} of the form $(x, h\cdot x, h^2\cdot x, \dots, h^k\cdot x)$ for some $x \in X$ and $h \in G$. Once again, it is more fruitful to take a broader definition, calling an arithmetic progression in X a point from the set $\tilde{X}$ (again defined below), which is the image of the group $\tilde{G}$ under the natural projection on X^{k+1}.

5.1. Some algebraic constructions

Define the map $j : G \times G_1 \times G_2 \times \cdots \times G_k \to G^{k+1}$ by

$$j(g, g_1, g_2, \dots, g_k) = \big(g, gg_1, gg_1^2g_2, \dots, gg_1^{\binom{k}{1}}g_2^{\binom{k}{2}}\cdots g_k^{\binom{k}{k}}\big)$$

and let $\tilde{G}$ denote the range of the map j:

$$\tilde{G} = j(G \times G_1 \times G_2 \times \cdots \times G_k)\ .$$

Similarly, we define a map $j^* : G_1 \times G_2 \times \cdots \times G_k \to G^k$ by

$$j^*(g_1, g_2, \dots, g_k) = \big(g_1, g_1^2g_2, \dots, g_1^{\binom{k}{1}}g_2^{\binom{k}{2}}\cdots g_k^{\binom{k}{k}}\big)\ .$$

Finally we define

$$\begin{aligned}\tilde{G}^* &= j^*(G_1 \times G_2 \times \cdots \times G_k) \\ &= \left\{(h_1,h_2,\ldots,h_k) \in G^k \colon (1,h_1,h_2,\ldots,h_k) \in \tilde{G}\right\} .\end{aligned}$$

The following results are found in Leibman [Lei1]:

Theorem 5.1.

1. *$\tilde{G}$ is a subgroup of G^{k+1}.*
2. *The commutator group $(\tilde{G})_2$ of $\tilde{G}$ is*

$$(\tilde{G})_2 = \tilde{G} \cap G_2^{k+1} = j(G_2 \times G_2 \times G_2 \times G_3 \times \cdots \times G_k) .$$

It follows from part 1 that

3. $\tilde{G}^*$ is a subgroup of G^k.

Moreover, for $g \in G$, $(h_1,h_2,\ldots,h_k) \in G_1 \times G_2 \times \cdots \times G_k$ and

$$(g_1,g_2,\ldots,g_k) = j^*(h_1,h_2,\ldots,h_k) ,$$

we have

$$(g^{-1}g_1g, g^{-1}g_2g, \ldots, g^{-1}g_kg) = j^*(g^{-1}h_1g, g^{-1}h_2g, \ldots, g^{-1}h_kg) \tag{5.1}$$

It follows that

4. For $(g_1,g_2,\ldots,g_k) \in \tilde{G}^*$ and $g \in G$, we have

$$(g^{-1}g_1g, g^{-1}g_2g, \ldots, g^{-1}g_kg) \in \tilde{G}^* .$$

The maps j and j^* are injective, continuous and proper (the inverse image of a compact set is compact). It follows that $\tilde{G}$ and $\tilde{G}^*$ are closed subgroups of G^{k+1} and G^k, respectively, and thus are Lie groups.

We also define the two elements

$$\tilde{t} = (1,t,t^2,\ldots,t^k) \text{ and } t^\Delta = (t,t,\ldots,t)$$

of $\tilde{G}$ and the element

$$\tilde{t}^* = (t,t^2,\ldots,t^k)$$

of $\tilde{G}^*$. Write $\tilde{T}$ and T^Δ for the translations by $\tilde{t}$ and t^Δ on X^{k+1}, respectively, and $\tilde{T}^*$ for the translation by $\tilde{t}^*$ on X^k.

5.2. The nilmanifold $\tilde{X}$.

Define $\tilde{\Lambda} = \tilde{G} \cap \Lambda^{k+1}$.

Then $\tilde{\Lambda}$ is a discrete subgroup of $\tilde{G}$ and it is easy to check that $\tilde{\Lambda} = j(\Lambda \times \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_k)$. Therefore $\tilde{\Lambda}$ is cocompact in $\tilde{G}$. We write

$$\tilde{X} = \tilde{G}/\tilde{\Lambda}$$

and let $\tilde{\mu}$ denote the Haar measure of this nilmanifold. Note that $\tilde{X}$ is imbedded in X^{k+1} in a natural way. Since $\tilde{t}$ and t^Δ belong to $\tilde{G}$, this nilmanifold is invariant under the transformations $\tilde{T}$ and T^Δ.

Lemma 5.2. *The nilmanifold $\tilde{X}$ is ergodic (and thus uniquely ergodic) for the action spanned by $\tilde{T}$ and T^Δ.*

Proof. Since the groups G_j, $j > 1$, are connected and G satisfies condition (H), it follows that the group $\tilde{G}$ is spanned by $\tilde{t}$, t^Δ and the connected component of the identity. By Theorem 4.2, it suffices to show that the action induced by $\tilde{T}$ and T^Δ on $\tilde{G}/(\tilde{G})_2\tilde{\Lambda}$ is ergodic.

By part 2 of Theorem 5.1, the map

$$(g_1, g_2, \dots, g_k) \mapsto (g_1 \bmod G_2, g_2 \bmod G_2)$$

induces an isomorphism from $\tilde{G}/(\tilde{G})_2$ onto $G/G_2 \times G/G_2$. Thus the compact abelian group $\tilde{G}/(\tilde{G})_2\tilde{\Lambda}$ can be identified with $G/G_2\Lambda \times G/G_2\Lambda$, and the transformations induced by $\tilde{T}$ and T^Δ are $\mathrm{Id} \times T$ and $T \times T$, respectively. The action spanned by these transformations is obviously ergodic. □

5.3. The nilmanifolds $\tilde{X}_x$.

For $x \in X$ we define

$$\tilde{X}_x = \left\{ (x_1, x_2, \dots, x_k) \in X^k \colon (x, x_1, x_2, \dots, x_k) \in \tilde{X} \right\} .$$

Clearly, for every $x \in X$ the compact set $\tilde{X}_x$ is invariant under translations by elements of $\tilde{G}^*$. We give to each of these sets the structure of a nilmanifold, quotient of this group.

Fix $x \in X$ and let a be a lift of x in G. The point $(x, x, \dots, x)$ (k times) clearly belongs to $\tilde{X}_x$. Let $(x_1, x_2, \dots, x_k) \in \tilde{X}_x$. The point $(x, x_1, x_2, \dots, x_k)$ belongs to $\tilde{X}$ and we can lift it to an element of $\tilde{G}$ that we can write $(g, gg_1, gg_2, \dots, gg_k)$ with $g \in G$ and $(g_1, g_2, \dots, g_k) \in \tilde{G}^*$. Writing $\lambda = a^{-1}g$ and $h_i = a\lambda g_i \lambda^{-1} a^{-1}$ for $1 \le i \le k$, we have $\lambda \in \Lambda$, $(1, h_1, h_2, \dots, h_k)$ belongs to $\tilde{G}$ by Remark 4 above and

$$(g, gg_1, gg_2, \dots, gg_k) = (1, h_1, h_2, \dots, h_k) \cdot (a, a, a, \dots, a) \cdot (\lambda, \lambda, \lambda, \dots, \lambda) .$$

This gives that $(x_1, x_2, \dots, x_k)$ is the image of $(x, x, \dots x)$ under translation by $(h_1, h_2, \dots, h_k)$, which belongs to $\tilde{G}^*$.

Therefore the action of $\tilde{G}^*$ on $\tilde{X}_x$ is transitive. The stabilizer of $(x,x,\dots,x)$ for this action is the group

$$\tilde{\Lambda}_x := \left\{(a\lambda_1 a^{-1}, a\lambda_2 a^{-1}, \dots, a\lambda_k a^{-1}) \colon (\lambda_1, \lambda_2, \dots, \lambda_k) \in \Lambda^k \cap \tilde{G}^*\right\} .$$

$\Lambda^k \cap \tilde{G}^*$ is a discrete subgroup of $\tilde{G}^*$ and it is easy to check that it is equal to $j^*(\Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_k)$ and thus is cocompact in $\tilde{G}^*$. It follows that $\tilde{\Lambda}_x$ is a discrete and cocompact subgroup of $\tilde{G}^*$.

We can thus identify $\tilde{X}_x$ with the nilmanifold $\tilde{G}^*/\tilde{\Lambda}_x$. Let $\tilde{\mu}_x$ denote the Haar measure of $\tilde{X}_x$.

Lemma 5.3. $\tilde{\mu} = \displaystyle\int_X \delta_x \otimes \tilde{\mu}_x \, d\mu(x)$.

Proof. Let $\tilde{\mu}'$ be the measure defined by this integral. This measure is concentrated on $\tilde{X}$. By Lemma 5.2 it suffices to show that it is invariant under $\tilde{T}$ and T^Δ.

Recall that $\tilde{T}^*$ is the translation by $\tilde{t}^* = (t, t^2, \dots, t^k)$, which belongs to $\tilde{G}^*$ and thus this transformation preserves $\tilde{X}_x$ and $\tilde{\mu}_x$ for every x. Therefore, for every $x \in X$, the measure $\delta_x \otimes \tilde{\mu}_x$ is invariant under $\tilde{T} = \mathrm{Id} \times \tilde{T}^*$ and so $\tilde{\mu}'$ is invariant under this transformation.

Let $x \in X$. Consider the image of $\tilde{\mu}_x$ under $T \times \cdots \times T$ (k times). This measure is concentrated on X_{Tx} and by remark 4 in Section 5.1, it is easy to check that it is invariant under $\tilde{G}^*$. Thus it is equal to the Haar measure $\tilde{\mu}_{Tx}$. Therefore the image of $\delta_x \otimes \tilde{\mu}_x$ under T^Δ is $\delta_{Tx} \otimes \tilde{\mu}_{Tx}$. It follows that $\tilde{\mu}'$ is invariant under T^Δ. $\square$

5.4. The limit of the averages.

Given this background, we give a short proof of Ziegler's result [Z1]:

Theorem 5.4 (Ziegler [Z1]). *Let $f_1, f_2, \dots, f_k$ be continuous functions on X and let $\{M_i\}$ and $\{N_i\}$ be two sequences of integers such that $N_i \to +\infty$. For μ-almost every $x \in X$,*

$$\frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \to \int f_1(x_1) f_2(x_2) \dots f_k(x_k) \, d\tilde{\mu}_x(x_1, x_2, \dots, x_k) \quad (5.2)$$

as $i \to \infty$

Proof. For $x \in X$, the point $(x,x,\dots,x)$ belongs to the nilmanifold $\tilde{X}_x$. By part 3 of Theorem 4.1 applied to the nilsystem $(\tilde{X}_x, \tilde{T}^*)$ and this point, the averages in Equation (5.2) converge everywhere to some function ϕ. Therefore we are left with computing this function.

Let f be a continuous function on X. We have

$$\begin{aligned}
\int f(x)\phi(x)\,d\mu(x)& \\
&= \lim_{i\to\infty}\int \frac{1}{N_i}\sum_{n=M_i}^{M_i+N_i-1} f(x)\prod_{j=1}^{k} f_j(T^{jn}x)\,d\mu(x)\\
&= \lim_{i\to\infty}\int \frac{1}{N_i^2}\sum_{m=M_i}^{M_i+N_i-1}\sum_{n=M_i}^{M_i+N_i-1} f(T^m x)\prod_{j=1}^{k} f_j(T^{jn+m}x)\,d\mu(x)\ .
\end{aligned}$$

The point $(x,x,\dots,x)$ belongs to $\tilde{X}$ and for all n and m, its image under $\tilde{T}^n(T^\Delta)^m$ is the point $(T^m x, T^{n+m}x,\dots,T^{kn+m}x)$. By Lemma 5.2, $\tilde{X}$ is uniquely ergodic for the action spanned by $\tilde{T}$ and T^Δ and the average in the last integral converges everywhere to

$$\int_{\tilde{X}} f(x_0)\,f_1(x_1)\dots f_k(x_k)\,d\tilde{\mu}(x_0,x_1,\dots,x_k)\ .$$

Using Lemma 5.3, we have

$$\begin{aligned}
\int_X f(x)\phi(x)\,d\mu(x) &= \int_{\tilde{X}} f(x_0)\,f_1(x_1)\dots f_k(x_k)\,d\tilde{\mu}(x_0,x_1,\dots,x_k)\\
&= \int_X f(x)\Big(\int_{\tilde{X}_x} f_1(x_1)\dots f_k(x_k)\,d\tilde{\mu}_x(x_1,\dots,x_k)\Big)\,d\mu(x)
\end{aligned}$$

and the result follows. □

Corollary 5.5. *For μ-almost every $x\in X$, the nilsystem $(\tilde{X}_x,\tilde{\mu}_x,\tilde{T}^*)$ is ergodic.*

Proof. Recall that $\tilde{T}^*$ preserves $\tilde{\mu}_x$ for every x. Let $\mathscr{F}$ be a countable family of continuous functions on X that is dense in $\mathscr{C}(X)$ in the uniform norm. By Theorem 5.4, there exists a subset X_0 of X, with $\mu(X_0)=1$, such that

$$\frac{1}{N}\sum_{n=0}^{N-1}\prod_{j=1}^{k} f_j(T^{jn}x) \to \int\prod_{j=1}^{k} f_j(x_j)\,d\tilde{\mu}_x(x_1,x_2,\dots,x_k)$$

as $N\to+\infty$ for every $x\in X_0$ and for all functions $f_1,f_2,\dots,f_k\in\mathscr{F}$. Since $\mathscr{F}$ is dense, the same result holds for arbitrary continuous functions. It follows that for $x\in X_0$, the orbit of $(x,x,\dots,x)$ under $\tilde{T}^*$ is dense in the support of the measure $\tilde{\mu}_x$. Since the support of this measure is $\tilde{X}_x$, we have that the action of $\tilde{T}^*$ on $\tilde{X}_x$ is transitive. By Theorem 4.1, $(\tilde{X}_x,\tilde{\mu}_x,\tilde{T}^*)$ is ergodic. □

Corollary 5.6. *Let $f_1, f_2, \dots, f_k$ be continuous functions on X and let $\{M_i\}$ and $\{N_i\}$ be two sequences of integers such that $N_i \to +\infty$. For μ-almost every $x \in X$ and for every $(g_1, g_2, \dots, g_k) \in \tilde{G}^*$,*

$$\frac{1}{N_i}\sum_{n=M_i}^{M_i+N_i-1} f_1(T^n g_1 \cdot x)\, f_2(T^{2n} g_2 \cdot x) \dots f_k(T^{kn} g_k \cdot x) \\ \to \int f_1(x_1)\, f_2(x_2) \dots f_k(x_k)\, d\tilde{\mu}_x(x_1, x_2, \dots, x_k) \quad (5.3)$$

as $i \to \infty$

Proof. Let $x \in X$ be such that the nilsystem $(\tilde{X}_x, \tilde{\mu}_x, \tilde{T}^*)$ is ergodic. For every $(g_1, g_2, \dots, g_k) \in \tilde{G}^*$, the point $(g_1 \cdot x, g_2 \cdot x, \dots, g_k \cdot x)$ belongs to $\tilde{X}_x$ and the convergence in Formula (5.3) follows from the unique ergodicity of $(\tilde{X}_x, \tilde{T}^*)$. □

6. Using the Cartesian square

In this section, we begin the proof of Theorem 1.9. We first construct a nilsystem in order to replace the sequence $I_f(k,n)$ (defined in 1.5) by another sequence $J_f(k,n)$ so that the difference between the two sequences tends to 0 in uniform density. In the next section, we complete the proof of Theorem 1.9 by showing that the sequence $J_f(k,n)$ is a nilsequence.

To pass from the convergence results obtained in the preceding section to a more precise description of the sequence $\{I_f(k,n)\}$, we consider the Cartesian square of the groups, manifolds, etc. studied in the previous section. This enables passage from the uniform Cesaro convergence results to uniform density convergence results.

6.1. The group H.

Define

$$H = \left\{(g,h) \in G \times G : hg^{-1} \in G_2\right\} .$$

H is a closed subgroup of $G \times G$ and is a k-step nilpotent Lie group. By induction, its commutator subgroups H_j, $j \geq 1$, are given by

$$H_j = \left\{(g,h) \in G_j \times G_j : hg^{-1} \in G_{j+1}\right\} .$$

We build the groups $\tilde{H}$ and $\tilde{H}^*$ from H in the same way that the groups $\tilde{G}$ and $\tilde{G}^*$ were built from G (in Section 5.1), using the maps

$$i\colon H \times H_1 \times H_2 \times \cdots \times H_k \to H^{k+1}$$

and

$$i^*\colon H_1 \times H_2 \times \cdots \times H_k \to H^k ,$$

defined analogously to the maps j and j^*. By part 1 of Theorem 5.1, $\tilde{H}$ is a subgroup of H^{k+1} and thus $\tilde{H}^*$ is a subgroup of H^k. These subgroups are closed and thus $\tilde{H}$ and $\tilde{H}^*$ are Lie groups.

The group $\tilde{H}^*$ is included in $(G\times G)^k$. We identify this last group with $G^k\times G^k$ in the obvious way and consider $\tilde{H}^*$ as a subset of $G^k\times G^k$. For $\big((g_1,h_1),(g_2,h_2),\dots,(g_k,h_k)\big)\in H_1\times H_2\times\cdots\times H_k$, we have

$$i^*\big((g_1,h_1),(g_2,h_2),\dots,(g_k,h_k)\big)=\big(j^*(g_1,g_2,\dots,g_k),j^*(h_1,h_2,\dots,h_k)\big)\ . \tag{6.1}$$

In a similar way, we consider $\tilde{H}$ as a subset of $G^{k+1}\times G^{k+1}$.

6.2. The nilmanifolds X_s, $\widetilde{X}_s$ and $\widetilde{X}_{s(x,y)}$.

Recall the ergodic decomposition

$$\mu\times\mu=\int_Z \mu_s\,dm(s)$$

of $\mu\times\mu$ under $T\times T$, where m is the Haar measure of the Kronecker factor Z of X.

By part 6 of Theorem 4.1, Z is equal to $G/G_2\Lambda$ and the factor map $\pi\colon X\to Z$ is the natural projection $G/\Lambda\to G/G_2\Lambda$. When f is a bounded function on X and $g\in G$, we write $f\circ g$ for the function $x\mapsto f(g\cdot x)$ on X. We have

$$\mathbb{E}(f\circ g\mid Z)(z)=\mathbb{E}(f\mid Z)(\pi(g)z)\ .$$

Therefore it follows from definition (4.2) of μ_s that for every $s\in Z$, this measure is concentrated on the closed subset

$$X_s=\big\{(x,y)\in X\times X\colon \pi(y)\pi(x)^{-1}=s\big\} \tag{6.2}$$

of $X\times X$. It also follows that for all $s\in Z$, all bounded functions f,f' on X and all $(g,h)\in H$,

$$\int f\circ g(x)\,f'\circ h(x)\,d\mu_s(x,x')=\int f(x)\,f'(x')\,d\mu_s(x,x')\ .$$

This means that the measure μ_s is invariant under translation by elements of H.

Let $s\in Z$. By its definition (6.2), the set X_s is invariant under the action of H by translation and this action is transitive. We give X_s the structure of a nilmanifold, quotient of this group. Write

$$\Theta=H\cap(\Lambda\times\Lambda)=\big\{(\lambda,\lambda')\in\Lambda\times\Lambda\colon \lambda'\lambda^{-1}\in\Lambda_2\big\}\ .$$

This group is discrete and cocompact in H. Let $a\in G$ be a lift of s in G and let e_X be the base point of X (that is, the image in X of the unit element 1 of G). Then the stabilizer in H of the point $(e_X,a\cdot e_X)$ of X_s is

$$\Theta_a=\big\{(\lambda,a\lambda'a^{-1})\colon (\lambda,\lambda')\in\Theta\big\}$$

and this group is a discrete cocompact subgroup of H. Thus we can identify X_s with the nilmanifold H/Θ_a. Since the measure μ_s is concentrated on X_s and is invariant under the action of H, it is equal to the Haar measure of this nilmanifold.

Let $s \in Z$ be such that μ_s is ergodic for $T \times T$. Then the nilsystem $(X_s, \mu_s, T \times T)$ is ergodic (note that $T \times T$ is the translation by the element (t,t) of H).

The k-step nilpotent Lie group H, its subgroup Θ_a and its element (t,t) satisfy properties (H) and (L) (see Section 4.1) that were used for G, Λ and t in the preceding Section. Therefore all the constructions of this Section can be carried out with H, Θ_a and (t,t) substituted for G, Λ and t. In particular, we can define the nilmanifold $\widetilde{X}_s$, its Haar measure $\widetilde{\mu}_s$ and, for $(x,y) \in X_s$, the nilmanifold $\widetilde{X}_{s(x,y)}$ and its Haar measure $\widetilde{\mu}_{s(x,y)}$. Note that $\widetilde{X}_s$ is included in $(X \times X)^{k+1}$. We identify this set with $X^{k+1} \times X^{k+1}$ in the natural way and consider $\widetilde{X}_s$ as contained in $X^{k+1} \times X^{k+1}$. Similarly, $\widetilde{X}_{s(x,y)}$ is included in $X^k \times X^k$.

We rewrite Corollary 5.6 for this situation. We consider only the case that all the functions on X_s are equal to $f \otimes f$ for some function f on X.

Corollary 6.1. *Let f be a continuous function on X and let $\{M_i\}$ and $\{N_i\}$ be two sequences of integers with $N_i \to +\infty$. For m-almost every $s \in Z$, for μ_s-almost every $(x,y) \in X_s$ and for every $\big((g_1, g_2, \dots, g_k), (h_1, h_2, \dots, h_k)\big) \in \tilde{H}^*$,*

$$\frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \prod_{j=1}^{k} f(T^{jn} g_j \cdot x) f(T^{jn} h_j \cdot y)$$

converges as $i \to \infty$ to

$$\int \prod_{j=1}^{k} f(x_j) f(y_j) \, d\widetilde{\mu}_{s(x,y)}\big((x_1, x_2, \dots, x_k), (y_1, y_2, \dots, y_k)\big) \ .$$

In order to use this result, we need a more precise description of the measures $\widetilde{\mu}_{s(x,y)}$ and thus of the groups $\tilde{H}$ and $\tilde{H}^*$.

6.3. The groups $\vec{G}^*$ and $\vec{G}$

Clearly, $\tilde{H}^* \subset \tilde{G}^* \times \tilde{G}^*$. Define

$$\vec{G}^* = \{ g = (g_1, g_2, \dots, g_k) \in \tilde{G}^* \colon \big((1,1,\dots,1), (g_1, g_2, \dots, g_k)\big) \in \tilde{H}^* \} \ .$$

Then $\vec{G}^*$ is a closed subgroup of G^k and thus is a Lie group. By Equation (6.1), the injectivity of j^*, and the above description of the groups H_j we have

$$\vec{G}^* = j^*(G_2 \times G_3 \times \cdots \times G_{k+1}) \ .$$

Moreover, when $g = (g_1, g_2, \dots, g_k) \in \tilde{G}^*$ we have $(g, g) \in \tilde{H}^*$. It follows that

$$\tilde{H}^* = \{(g,h) \in \tilde{G}^* \times \tilde{G}^* \colon hg^{-1} \in \vec{G}^*\} \tag{6.3}$$

and that $\vec{G}^*$ is a normal subgroup of $\tilde{G}^*$.

The following Lemma is taken from [Lei1]:

Lemma 6.2. *For $g \in G$ and $(g_1, g_2, \dots, g_k) \in \tilde{G}^*$, we have*

$$\big([g_1, g], [g_2, g], \dots, [g_k, g]\big) \in \vec{G}^* \ .$$

Proof. Let $(h_1, h_2, \dots, h_k) \in G_1 \times G_2 \times \cdots \times G_k$ be the inverse image of $(g_1, g_2, \dots, g_k)$ under j^*. For $1 \le \ell \le k$, we have $[g, h_\ell] \in G_{\ell+1}$ and thus $(h_\ell, g^{-1}h_\ell g) \in H_\ell$. We get that

$$\begin{aligned}
&\big((g_1, g_2, \dots, g_k), (g^{-1}g_1g, g^{-1}g_2g, \dots, g^{-1}g_kg)\big)\\
&\qquad = \big(j^*(h_1, h_2, \dots, h_k), j^*(g^{-1}h_1g, g^{-1}h_2g, \dots, g^{-1}h_kg)\big)\\
&\qquad\qquad = i^*\big((h_1, g^{-1}h_1g), (h_2, g^{-1}h_2g), \dots, (h_k, g^{-1}h_kg)\big) \in H^*
\end{aligned}$$

and the result follows from characterization (6.3) of $\tilde{H}^*$. □

In particular, it follows that

$$\begin{aligned}
&\text{if } (g_1, g_2, \dots, g_k) \in \vec{G}^* \text{ and } g \in G,\\
&\qquad\qquad \text{then } (gg_1g^{-1}, gg_2g^{-1}, \dots, gg_kg^{-1}) \in \vec{G}^* \ .
\end{aligned} \tag{6.4}$$

We also define

$$\begin{aligned}
\vec{G} &= \{(g, gh_1, gh_2, \dots, gh_k) \colon g \in G,\ (h_1, h_2, \dots, h_k) \in \vec{G}^*\}\\
&= j(G \times G_2 \times G_3 \times \cdots \times G_{k+1}) \ .
\end{aligned}$$

It follows from Remark (6.4) that $\vec{G}$ is a subgroup of G^{k+1}. It is clearly included in $\tilde{G}$. By using the normality of $\vec{G}^*$ in $\tilde{G}^*$ and Lemma 6.2, we have that $\vec{G}$ is a normal subgroup of $\tilde{G}$. As j is a proper map, $\vec{G}$ is closed in G^{k+1} and is a Lie group.

6.4. The nilmanifold $\vec{X}$ and the nilmanifolds $\vec{X}_x$.

Define $\vec{\Lambda} := \vec{G} \cap \Lambda^{k+1}$. It is a discrete subgroup of $\vec{G}$ and it is easy to check that $\vec{\Lambda} = j(\Lambda \times \Lambda_2 \times \Lambda_3 \times \cdots \times \Lambda_{k+1})$. Thus $\vec{\Lambda}$ is cocompact in $\vec{G}$. We write

$$\vec{X} = \vec{G}/\vec{\Lambda}$$

and let $\vec{\mu}$ denote the Haar measure of this nilmanifold. $\vec{X}$ is imbedded in X^{k+1} in a natural way and $\vec{X} \subset \tilde{X}$ since $\vec{G} \subset \tilde{G}$.

Lemma 6.3. *The nilmanifold $\vec{X}$ and its Haar measure $\vec{\mu}$ are invariant under translation by any element of $\tilde{\Lambda} = \Lambda^{k+1} \cap \tilde{G}$.*

Proof. Let $x = (x, x_1, x_2, \dots, x_k) \in \vec{X}$ and let $\lambda \in \tilde{\Lambda}$. We show that $\lambda \cdot x$ (with the obvious interpretation) belongs to $\vec{X}$.

The point x is the image in $\vec{X}$ of an element g of $\vec{G}$. Thus $\lambda \cdot x$ is the image in $\tilde{X}$ of the element $\lambda g = (\lambda g \lambda^{-1})\lambda$ of $\tilde{G}$ and thus also of the element $\lambda g \lambda^{-1}$. Since $\vec{G}$ is a normal subgroup of $\tilde{G}$, this last element belongs to $\vec{G}$ and $\lambda \cdot x \in \vec{G}$.

Let $\lambda \in \tilde{\Lambda}$. The measure $\vec{\mu}$ is invariant under the action of $\vec{G}$ and thus its image under translation by λ is invariant under the action of $\lambda \vec{G} \lambda^{-1} = \vec{G}$. Since this measure is concentrated on $\vec{X}$, it is equal to $\vec{\mu}$. □

For $x \in X$ we write

$$\vec{X}_x = \left\{(x_1, x_2, \dots, x_k) \in X^k : (x, x_1, x_2, \dots, x_k) \in \vec{X}\right\} .$$

Let $x \in X$. Proceeding as in Subsection 5.3 we note that $\vec{\Lambda}^* := \vec{G}^* \cap \Lambda^k = j^*(\Lambda_2 \times \Lambda_3 \times \cdots \times \Lambda_{k+1})$ is a discrete cocompact subgroup of $\vec{G}^*$ and that for every $x \in X$, the compact set $\vec{X}_x$ can be identified with a nilmanifold, quotient of the group $\vec{G}^*$ by some conjugate $\vec{\Lambda}_x$ of the group $\vec{\Lambda}^*$.

Let $\vec{\mu}_x$ be the Haar measure of $\vec{X}_x$.

Lemma 6.4. $\vec{\mu} = \displaystyle\int_X \delta_x \otimes \vec{\mu}_x \, d\mu(x)$.

Proof. The proof is similar to the proof of Lemma 5.3. The measure defined by the integral above is concentrated on $\vec{X}$ and thus it suffices to prove that $\vec{X}$ is invariant under $\vec{G}$. It is clearly invariant under translation by elements of the form $(1, g_1, g_2, \dots, g_k)$ with $(g_1, g_2, \dots, g_k) \in \vec{G}^*$ and so we are reduced to showing that $\vec{X}$ is invariant under translation by $(g, g, g, \dots, g)$ ($k+1$ times) for every $g \in G$.

Let $g \in G$ and let $x \in X$. Since $\vec{X}$ is invariant under translation by $(g, g, g, \dots, g)$, we have that the image of $\vec{X}_x$ under $g = (g, g, \dots, g)$ (k times) is $\vec{X}_{g \cdot x}$. The image of $\vec{\mu}_x$ under translation by g is thus concentrated on $\vec{X}_x$. It is invariant under translation by $g\vec{G}^* g^{-1}$ and this group is equal to $\vec{G}^*$ by (6.4). Thus this measure is equal to $\vec{\mu}_{g \cdot x}$.

Taking the integral over $x \in X$, we have that the measure given by the integral in the Lemma is invariant under $(g, g, g, \dots, g)$. □

6.5. Approximating the sequence $\{I_f(k,n)\}$ up to density zero.

For a bounded function f on X, an integer $k \geq 1$ and an integer n, we define:

$$J_f(k,n) = \int_{\vec{X}} f(x_0) f(T^n x_1) \dots f(T^{kn} x_k) \, d\vec{\mu}(x_0, x_1, \dots, x_k) . \qquad (6.5)$$

Proposition 6.5. *Let f be a bounded function on X and let $k \geq 1$ be an integer. Then the sequence $\{I_f(k,n) - J_f(k,n)\}$ converges to zero in uniform density.*

Proof. Given a sequence $\{(M_i, M_i + N_i)\}$ of intervals with $N_i \to +\infty$ we show that for any $f \in L^\infty(\mu)$,

$$\frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \big(I_f(k,n) - J_f(k,n)\big)^2 \to 0\,. \tag{6.6}$$

Since the continuous functions are dense, we can restrict to the case that the function f is continuous.

Let $g = (g_1, g_2, \dots, g_k)$ and $h = (h_1, h_2, \dots, h_k)$ be two elements of $\vec{G}^*$.
The four elements

$$\big((1,\dots,1),(1,\dots 1)\big)\,,\quad \big((g_1,\dots,g_k),(1,\dots,1)\big)\,,$$
$$\big((1,\dots,1),(h_1,\dots,h_k)\big) \text{ and } \big((g_1,\dots,g_k),(h_1,\dots,h_k)\big)$$

of $G^k \times G^k$ belong to $\tilde{H}^*$ by formula (6.3).

We use Corollary 6.1 with these four elements. The four limits given by this Corollary are the same. Taking differences, we have that for m-almost every $s \in Z$, for μ_s-almost every (x,y), for every g and $h \in \vec{G}^*$, the averages on $[M_i, M_i + N_i - 1]$ of the product

$$\Big(f(x)\prod_{j=1}^{k} f(T^{jn} g_j \cdot x) - f(x)\prod_{j=1}^{k} f(T^{jn}x)\Big)\cdot$$
$$\Big(f(y)\prod_{j=1}^{k} f(T^{jn} h_j \cdot y) - f(y)\prod_{j=1}^{k} f(T^{jn}y)\Big)$$

converge to zero.

Let $\vec{m}^*$ be the Haar measure of $\vec{G}^*$. Fix $s \in Z$ and $(x,y) \in X_s$. Recall that $\vec{\mu}_x$ is the Haar measure of the nilmanifold $\vec{X}_x = \vec{G}^*/\vec{\Lambda}_x$. Let $K \subset \vec{G}^*$ be a fundamental domain of the projection $\vec{G}^* \to \vec{X}_x$. Then the image of the measure $1_K \cdot \vec{m}^*$ under this projection is equal to a constant multiple of $\vec{\mu}_x$. Similarly, when L is a fundamental domain for the projection $\vec{G}^* \to \vec{X}_y$, the image of $1_L \cdot \vec{m}^*$ under this projection is a constant multiple of $\vec{\mu}_y$. Taking the integral for $g \in K$ and $h \in L$ with respect to the measure $\vec{m}^*$ in the last convergence we have:

For m-almost every $s \in Z$ and for μ_s-almost every (x,y), the averages on $[M_i, M_i + N_i - 1]$ of

$$\Big(\int f(x)\prod_{j=1}^{k} f(T^{jn}x_j)\,d\vec{\mu}_x(x_1,x_2,\dots,x_k) - f(x)\prod_{j=1}^{k} f(T^{jn}x)\Big)\cdot$$
$$\Big(\int f(y)\prod_{j=1}^{k} f(T^{jn}y_j)\,d\vec{\mu}_y(y_1,y_2,\dots,y_k) - f(y)\prod_{j=1}^{k} f(T^{jn}y)\Big)$$

converge to zero. Since this holds for m-almost every $s \in Z$ and for μ_s-almost every (x,y), it holds for $\mu\times\mu$-almost every $(x,y)\in X\times X$. Taking the integral with respect to $\mu\times\mu$ and using Lemma 6.4 we have the convergence (6.6). □

7. $J_f(k,n)$ is a nilsequence

In this Section we show that the sequence $\{J_f(k,n)\}$ introduced in Section 6.5 is a nilsequence.

We first explain the idea behind the construction. Two arithmetic progressions in X (see the discussion in the beginning of Section 5) are equivalent if one can pass from one to the other using translation by some element of $\vec{G}$. The strategy of the proof is the following: $I_f(k,n)$ is the average of the function

$$(x_0,x_1,\dots,x_k)\mapsto f(x_0)f(x_1)\dots f(x_k)$$

on the set of progressions of the form $(x,t^nx,\dots,t^{kn}x)$. In Proposition 6.5, we have shown that up to a small error, one can replace this average by the average on the set of arithmetic progressions that are equivalent to these. In Proposition 7.2, we define a continuous function $\phi(y)$, where $y\in Y$ is an equivalence class of arithmetic progressions, that is exactly this average. The transformation on Y can be viewed as multiplying the difference of a progression by t, meaning that

$$(x_0,x_1,x_2,\dots,x_k)\mapsto(x_0,t\cdot x_1,t^2x_2,\dots,t^kx_k)$$

induces the transformation S on Y.

7.1. The nilsystem (Y,ν,S).

We first build an ergodic nilsystem. Let K denote the group $\tilde{G}/\vec{G}$, let $p\colon \tilde{G}\to K$ be the natural projection and let $\Gamma = p(\tilde{\Lambda})$.

Since $\vec{G}/\vec{\Lambda} = \vec{G}/(\tilde{\Lambda}\cap\vec{G})$ is compact, it follows that $\vec{G}\tilde{\Lambda}$ is closed in $\tilde{G}$. Thus Γ is a closed subgroup of K. It is discrete because it is countable and it is cocompact because $G\tilde{\Gamma}$ is cocompact in $\tilde{G}$.

Let Y denote the nilmanifold K/Γ, ν be its Haar measure, $s = p(\tilde{t})\in K$ and S be the translation by s on Y.

Lemma 7.1. *The nilsystem (Y,ν,S) is ergodic (and thus is uniquely ergodic and minimal).*

Proof. We know that $\tilde{G}$ is spanned by its connected component of the identity and the elements t^{Δ} and $\tilde{t}$. Since $t^{\Delta}\in\vec{G}$, it follows that K is spanned by the connected component of the identity and s. Therefore by part 6 of Theorem 4.1 we only have to show that the rotation induced by S on the compact abelian group $K/(K_2\Gamma)$ is ergodic. We identify $K/(K_2\Gamma)$ with $G/((\tilde{G})_2\vec{G}\tilde{\Lambda})$.

We have already noted that the map

$$q:(g,g_1,\dots,g_k)\mapsto(g \bmod G_2, g_1 \bmod G_2)$$

induces an isomorphism from $\tilde{G}/(\tilde{G})_2$ onto $(G/G_2)\times(G/G_2)$. We have $q(\vec{G})=\{(u,u)\colon u\in G/G_2\}$ and

$$q(\tilde{\Lambda})=\{(\lambda \bmod G_2,\lambda' \bmod G_2)\colon \lambda,\lambda'\in\Lambda\}\ .$$

Therefore the map $(g,g_1,\dots,g_k)\mapsto g_1g^{-1} \bmod G_2\Lambda$ induces an isomorphism

$$K/(K_2\Gamma)=\tilde{G}/\big((\tilde{G})_2\vec{G}\tilde{\Lambda}\big)\to G/(G_2\Lambda)\ .$$

The image of s under this map is equal to the image of t under the natural projection $G\mapsto G/G_2\Lambda$. As X is ergodic, the rotation by this element of the compact abelian group $G/G_2\Lambda$ is ergodic. Therefore, the rotation induced by S on $K/(K_2\Lambda)$ is ergodic. □

7.2. Two examples

We give a description of the nilsystem (Y,ν,S) when X is each of the two systems described in Subsection 4.2.

We first study the general case of an ergodic 2-step nilsystem $(X=G/\Lambda,\mu,T)$, assuming that hypotheses (H) and (L) are satisfied. The commutator map $(g,h)\mapsto[g,h]$ is an antisymmetric bilinear map from $G\times G$ to G_2 and it is trivial on $G_2\times G$ and on $G\times G_2$. Therefore it induces a bilinear map $B\colon G/G_2\times G/G_2\to G_2$.

We have:

$$\tilde{G}=\big\{(g,gg_1,gg_1^2g_2)\colon g,g_1\in G,\ g_2\in G_2\big\}$$
$$\vec{G}=\big\{(h,hh_2,hh_2^2)\colon h\in G,\ h_2\in G_2\big\}\ .$$

Let $K'=(G/G_2)\times G_2$ with multiplication given by

$$(v,w)*(v',w')=\big(vv',ww'B(v,v')\big)\ .$$

Then it is easy to check that K' is a group and that the map

$$(g,gg_1,gg_1^2g_2)\mapsto(g_1 \bmod G_2,g_2)$$

is a group homomorphism from $\tilde{G}$ onto K'. The kernel of this homomorphism is $\vec{G}$. Therefore we can identify the groups $K = \tilde{G}/\vec{G}$ and K'. Under this identification, Γ is equal to $M \times \{1\}$, where M is the image of Λ in G/G_2 under the natural projection $G \to G/G_2$. The element s of K is $(\beta, 1)$, where β is the image of t in G/G_2 under the natural projection.

7.2.1. The example of Section 4.2.1 Here $G/G_2 = \mathbb{Z} \times \mathbb{T}$ and $G_2 = \{0\} \times \{0\} \times \mathbb{T}$ is identified with $\mathbb{T}$. We have $K = \mathbb{Z} \times \mathbb{T} \times \mathbb{T}$, $\Gamma = \mathbb{Z} \times \{0\} \times \{0\}$ and $s = (1, \alpha, 0)$. The bilinear map $B\colon (\mathbb{Z} \times \mathbb{T}) \times (\mathbb{Z} \times \mathbb{T}) \to \mathbb{T}$ is given by

$$B\big((k,x),(k',x')\big) = 2(kx' - k'x)$$

and multiplication on K is given by

$$(k,x,y) * (k',x',y') = \big(k+k', x+x', y+y'+2(kx'-k'x)\big)\ .$$

The map $(k,x,y) \mapsto (x, y+2kx)$ induces a homeomorphism of $Y = K/\Gamma$ onto $\mathbb{T}^2$, mapping the Haar measure of Y to the Haar measure $m_{\mathbb{T}} \times m_{\mathbb{T}}$ of $\mathbb{T}^2$. Under this identification of Y with $\mathbb{T}^2$, the transformation S takes the form

$$S(x,z) = (x+\alpha, z+2\alpha+4x)\ .$$

Thus Y is a factor of X, with factor map $(x,y) \mapsto (x, 2y)$.

7.2.2. The example of Section 4.2.2 We use the reduced representation of this system. Here $G/G_2 = \mathbb{R} \times \mathbb{R}$, $K = \mathbb{R} \times \mathbb{R} \times \mathbb{T}$, $\Gamma = \mathbb{Z} \times \mathbb{Z} \times \{0\}$ and $s = (t_1, t_2)$. For (x,y) and $(x',y') \in G/G_2 = \mathbb{R} \times \mathbb{R}$, $B\big((x,y),(x',y')\big) = xy' - x'y$. The multiplication in K is given by

$$(x,y,z) * (x',y',z') = (x+x', y+y', z+z'+xy'-x'y)\ .$$

7.3. A nilsequence

Proposition 7.2. *Let (Y, ν, S) be the ergodic nilsystem of Lemma 7.1. Let f be a bounded function on X and let $k \geq 1$ be an integer. Then there exists a continuous function ϕ on Y such that $J_f(k,n) = \phi(S^n e_Y)$ for every integer n, where e_Y denotes the base point in Y. In particular, the sequence $\{J_f(k,n)\}$ is a basic nilsequence.*

Proof. Define the function ψ on $\tilde{G}$ by

$$\psi(g_0, g_1, \ldots, g_k) = \int_{\vec{X}} \prod_{j=0}^{k} f(g_j \cdot x_j)\, d\vec{\mu}(x_0, x_1, \ldots, x_k)\ . \tag{7.1}$$

The function ψ is clearly continuous and satisfies $\psi(\tilde{t}^n) = J_f(k,n)$ for every integer n.

The measure $\vec{\mu}$ is invariant under (left) translation by elements of $\vec{G}$ by definition, and by left translation by elements of $\tilde{\Lambda}$ by Lemma 6.3. Thus the function ψ is invariant under (right) translations by elements of $\vec{G}\tilde{\Lambda}$.

Writing r for the natural projection $\tilde{G} \to Y = \tilde{G}/\vec{G}\tilde{\Lambda}$, we get that there exists a continuous function ϕ on Y with $\psi = \phi \circ r$. For every integer n we have $r(\tilde{t}^n) = s^n e_Y$ and thus $J_f(k,n) = \phi(S^n e_Y)$. $\square$

7.4. A decomposition

We are ready to prove Theorem 1.9.

Proof. Assume that $k \geq 1$ is an integer and let $f \in L^\infty(\mu)$. Without loss of generality, we can assume that $\|f\|_\infty \leq 1$.

Let $\tilde{f} = \mathbb{E}(f \mid \mathcal{Z}_k(X))$. Then by Corollary 4.6 the sequence $\{I_f(k,n) - I_{\tilde{f}}(k,n)\}$ converges to zero in uniform density. Thus it suffices to prove the theorem for the function $\tilde{f}$ substituted for f, meaning that we can assume that f is measurable with respect to $Z_k(X)$.

$Z_k(X)$ is the inverse limit of a sequence of ergodic k-step nilsystems (see Section 4.4). Let r be a positive integer. There exists a factor X' of $Z_k(X)$, which is a k-step nilsystem, such that

$$\|f - \mathbb{E}(f \mid \mathcal{X}')\|_1 \leq 1/(k+1)r \ .$$

Let $f' = \mathbb{E}(f \mid X')$. For every n, $|I_f(k,n) - I_{f'}(k,n)| \leq 1/r$. By Proposition 6.5 and Proposition 7.2 the sequence $\{I_{f'}(k,n)\}$ can be decomposed as a sum of a k-step nilsequence and a sequence tending to zero in uniform density. We thus have

$$I_f(k,n) = a_r(n) + b_r(n) + c_r(n)$$

where $|a_r(n)| \leq 1/r$ for every n, $\text{UD-Lim}\, b_r(n) = 0$ and $c_r(n)$ is an elementary k-step nilsequence. For $s \neq r$ we have

$$c_r(n) - c_s(n) = \big(a_r(n) - a_s(n)\big) + \big(b_r(n) - b_s(n)\big) \ .$$

We have $\text{UD-Lim}\big(b_r(n) - b_s(n)\big) = 0$ and $\sup_n |a_r(n) - a_s(n)| \leq 1/r + 1/s$. Thus by Lemma 1.11, $\text{synd-sup}|c_r(n) - c_s(n)| \leq 1/r + 1/s$. Since the sequence $\{c_r(n) - c_s(n)\}$ is a nilsequence, $\sup_n |c_r(n) - c_s(n)| \leq 1/r + 1/s$. Therefore $\{c_r(n)\}$ is a Cauchy sequence in ℓ^∞ for uniform convergence, and it converges uniformly to some sequence $\{c(n)\}$. This sequence is a k-step nilsequence and one can immediately check that the sequence $\{I_f(k,n) - c(n)\}$ converges to zero in uniform density. $\square$

8. Proof of Theorem 1.2

Theorem 1.2 follows immediately from the next one, with $f = 1_A$.

Theorem 8.1. *Let (X,μ,T) be an ergodic system and f a nonnegative bounded function on X. Then*

$$\text{synd-sup}\, I_f(2,n) \geq \Big(\int f\,d\mu\Big)^3 . \tag{8.1}$$

$$\text{synd-sup}\, I_f(3,n) \geq \Big(\int f\,d\mu\Big)^4 . \tag{8.2}$$

Let us summarize the steps already proved. Let k be equal to either 2 or to 3. We proceed as in the proof of Theorem 1.9, first replacing f by its conditional expectation on $Z_k(X)$ and then by its conditional expectation on a k-step nilsystem factor of $Z_k(X)$. We note that the operators of conditional expectation preserve the integral. By Corollary 4.5 and Lemma 1.11 the first conditional expectation does not change the synd-sup of the sequence $\{I_f(k,n)\}$; if the k-step nilsystem is well chosen the second expectation changes the synd-sup of this sequence by less than any given positive number. Therefore we are left with showing the theorem under the additional hypothesis that (X,μ,T) is an ergodic k-step nilsystem. We use the notation of Sections 5, 6 and 7.

By Proposition 6.5, the difference between the sequences $\{I_f(k,n)\}$ and $\{J_f(k,n)\}$ converges to zero in uniform density and thus they have the same synd-sup by Lemma 1.11. Let ϕ be the function on Y defined as in Proposition 7.2 and let ψ be the function on $\tilde{G}$ defined by Equation (7.1) in the proof of the same proposition. Since (Y,S) is minimal, we have

$$\begin{aligned}\text{synd-sup}\, J_f(k,n) = \text{synd-sup}\, \phi(S^n e_Y) &= \sup_n \phi(S^n e_Y) \\ &= \sup_{y\in Y} \phi(y) = \sup_{g\in\tilde{G}} \psi(g) \ .\end{aligned}$$

and we are reduced to showing that

$$\sup_{g\in\tilde{G}} \psi(g) \geq \Big(\int f\,d\mu\Big)^{k+1} \tag{8.3}$$

for $k = 2$ and $k = 3$.

8.1. One more reduction.

Recall that the group G_k is connected and closed. Since G is k-step nilpotent, G_k is included in the center of G and so is abelian. By hypothesis (L), $G_k \cap \Lambda = \{1\}$ and thus G_k is compact. More precisely, G_k is a torus. We write m_k for its Haar measure.

Let $G' = G/(G_k\Lambda_{k-1})$, $p : G \to G'$ be the natural projection, $t' = p(t)$ and $\Lambda' = \Lambda/\Lambda_{k-1}$.

Then G' is a $(k-1)$-step nilpotent Lie group but we actually consider it as a k-step nilpotent group. Λ' is a discrete cocompact subgroup of G'. We write $X' = G'/\Lambda'$, μ' for its Haar measure and T' for the translation by t' on X'. (X',μ',T') is a $(k-1)$-step nilsystem but we actually consider it as a k-step nilsystem. This system is a factor of (X,μ,T) in a natural way. Let $q : X \to X'$ be the factor map. For any bounded function f on X we have

$$\mathbb{E}(f \mid X')(q(x)) = \int_{G_k} f(u\cdot x)\,dm_k(x)\ . \tag{8.4}$$

The hypotheses (H) and (L) are satisfied and thus we can build the groups $\widetilde{G'}$, $\overrightarrow{G'}$,… and the nilmanifolds $\overrightarrow{X'}$,… associated to G' and X', with the same properties.

We note that $(G_k)^{k+1} \subset \tilde{G}$ and thus that $\widetilde{G'} = \tilde{G}/(G_k)^{k+1}$. Also, $\overrightarrow{X'}$ is the image of $\vec{X}$ under the natural projection $X^{k+1} \to {X'}^{k+1}$ and the Haar measure $\vec{\mu}'$ is the image of $\vec{\mu}$ under the same map.

Let f be a bounded function on X, ψ the function on $\tilde{G}$ associated to f as above, $f' = \mathbb{E}(f \mid X')$ and ψ' the function on $\widetilde{X'}$ associated to f'. By Equation (8.4), we have

$$\begin{aligned}\psi'\big(p(g_0), p(g_1), \dots, p(g_k)\big)\\ = \int_{(G_k)^{k+1}} \psi(u_0g_0, u_1g_1, \dots, u_kg_k)\,dm_k(u_0)\,dm_k(u_1)\,\dots\,dm_k(u_k)\ .\end{aligned}$$

In particular, $\sup_{g\in\tilde{G}} \psi(g) \geq \sup_{g'\in\widetilde{G'}} \psi'(g')$. Since f' is nonnegative and has the same integral as f, we are left with showing inequality (8.3) with G' substituted for G, f' substituted for f and ψ' substituted for ψ. In other words we can assume without loss that G is $(k-1)$-step nilpotent and that $\Lambda_{k-1} = G_{k-1} \cap \Lambda$ is trivial.

Note that G_{k-1}^{k+1} is not included in $\tilde{G}$. For this reason, the same method cannot be used to reduce the level of the nilmanifold once again.

8.2. The case $k = 2$.

In this case we can assume that G is a compact abelian group and that Λ is trivial. We have $X = G$ and μ is its Haar measure. The nilmanifold $\vec{X}$ is the diagonal $\big\{(x,x,x) : x \in X\big\}$ and its Haar measure $\vec{\mu}$ is the image of μ under the map $x \mapsto (x,x,x)$.

Let f be a bounded function on X and let ψ be the associated function on $\tilde{G}$. For $(g_0,g_1,g_2) \in \tilde{G}$ we have

$$\psi(g_0,g_1,g_2) = \int_X f(g_0x)\,f(g_1x)\,f(g_2x)\,d\mu(x)$$

and thus

$$\sup_{g\in\tilde{G}} \psi(g) \geq \psi(1,1,1) = \int_X f(x)^3\,d\mu(x) \geq \Big(\int_X f\,d\mu\Big)^3$$

and the proof is complete.

8.3. The case $k = 3$.

In this case we can assume that G_3 is trivial and that $G_2 \cap \Lambda$ is trivial. G_2 is connected, compact and included in the center of G and thus is abelian. Therefore it is a torus. We write m_2 for its Haar measure. We have

$$\tilde{G} = \big\{(g, gh, gh^2u, gh^3u^3) : g, h \in G, u \in G_2\big\}\ ;$$
$$\vec{X} = \big\{(x, v\cdot x, v^2\cdot x, v^3\cdot x) : x \in X,\ v \in G_2\big\}$$

and $\vec{\mu}$ is the image of $\mu \times m_2$ under the map $(x,v) \mapsto (x, v\cdot x, v^2\cdot x, v^3\cdot x)$.

Let f be a bounded function on X and let ψ be the associated function on $\tilde{G}$. For $(g, gh, gh^2u, gh^3u^3) \in \tilde{G}$,

$$\begin{aligned}&\psi(g, gh, gh^2u, gh^3u^3)\\ &= \int\Big(\int f(g\cdot x)\,f(ghv\cdot x)\,f(gh^2uv^2\cdot x)\,f(gh^3u^3v^3\cdot x)\,dm_2(v)\Big)\,d\mu(x)\ .\end{aligned}$$

We have

$$\begin{aligned}\sup_{g\in\tilde{G}} \psi(g) &\geq \int_{G_2\times G_2\times G_2} \psi(g, gh, gh^2u, gh^3u^3)\,dm_2(g)\,dm_2(h)\,dm_2(u)\\
&= \int_X\Big(\int_{G_2\times G_2\times G_2} f(g\cdot x)\,f(ghv\cdot x)\,f(gh^2uv^2\cdot x)\,f(gh^3u^3v^3\cdot x)\\
&\qquad\qquad dm_2(v)\,dm_2(g)\,dm_2(h)\Big)\,d\mu(x)\\
&= \int_X\Big(\int_{G_2\times G_2\times G_2} f(g\cdot x)\,f(h\cdot x)\,f(hw\cdot x)\,f(gw^3\cdot x)\\
&\qquad\qquad dm_2(g)\,dm_2(h)\,dm_2(w)\Big)\,d\mu(x)\ .\end{aligned} \tag{8.5}$$

Let $\widehat{G_2}$ be the dual group of the compact abelian group G_2, that is the group of continuous group homomorphisms from G_2 to the circle. For $x \in X$ we write $\widehat{f_x}$ for the Fourier transform of the function f_x defined on G_2 by $f_x(u) = f(u\cdot x)$:

$$\text{for } \gamma \in \widehat{G_2},\ \ \widehat{f_x}(\gamma) = \int_{G_2} f(u\cdot x)\overline{\gamma(u)}\,dm_2(u)\ .$$

The inner integral in the double integral (8.5) is equal to

$$\sum_{\gamma\in\widehat{G_2}} |\widehat{f_x}(\gamma)\,\widehat{f_x}(\gamma^3)|^2 \geq \Big|\widehat{f_x}(1)\Big|^4 = \Big|\int f(u\cdot x)\,dm_2(u)\Big|^4$$

We obtain

$$\sup_{g\in\tilde{G}} \psi(g) \geq \int \Big| \int f(u\cdot x)\, dm_2(u) \Big|^4 d\mu(x)$$
$$\geq \Big(\int f(u\cdot x) dm_2(u)\, d\mu(x) \Big)^4 = \Big(\int f\, d\mu \Big)^4$$

and the proof is complete.

Appendix: a combinatorial example by Imre Ruzsa (proof of Theorem 2.4)

We use the definitions of Section 2.2.

Lemma 8.2. *Let $d > 0$ be an integer and $a_0, a_1, \dots, a_4$ five points in $\mathbb{R}^d$, all having the same Euclidean norm and satisfying the relations*

$$a_0 - 3a_1 + 3a_2 - a_3 = 0\ , \tag{8.6}$$
$$a_1 - 3a_2 + 3a_3 - a_4 = 0\ . \tag{8.7}$$

Then these points are equal.

Proof. By adding relations (8.6) and (8.7), we have that

$$a_0 + 2a_3 = a_4 + 2a_1\ .$$

Setting $s = (a_0 + 2a_3)/3$, $a = a_1 - s$ and $b = a_3 - s$, we have

$$a_0 = s - 2b\ ;\ a_1 = s + a\ ;\ a_2 = s + a + b\ ;\ a_3 = s + b\ ;\ a_4 = s - 2a\ . \tag{8.8}$$

Taking the square of the norm of these vectors and subtracting $\|s\|^2$, we find that the five following numbers are equal:

$$\|a\|^2 + 2\langle a, s\rangle\ ; \tag{8.9}$$
$$4\|a\|^2 - 4\langle a, s\rangle\ ; \tag{8.10}$$
$$\|b\|^2 + 2\langle b, s\rangle\ ; \tag{8.11}$$
$$4\|b\|^2 - 4\langle b, s\rangle\ ; \tag{8.12}$$
$$\|a\|^2 + \|b\|^2 + 2\langle a, b\rangle + 2\langle a, s\rangle + 2\langle b, s\rangle\ . \tag{8.13}$$

Equality between (8.9) and (8.10) yields $\langle a, s\rangle = \|a\|^2/2$, and the equality (8.11) = (8.12) yields $\langle b, s\rangle = \|b\|^2/2$. From equality (8.9) = (8.11), we have that $\|a\| = \|b\|$. The common value of the four first numbers is then $2\|a\|^2$, and (8.13) is equal to $4\|a\|^2 + 2\langle a, b\rangle$. As these values are equal, $\langle a, b\rangle = -\|a\|^2$ and this is possible only if $b = -a$. Now $\langle a, s\rangle = -\langle a, s\rangle = \|a\|^2$ and so indeed $a = b = 0$. We conclude that $a_0 = a_1 = a_2 = a_3 = a_4$. □

From this point the proof follows the line of Behrend's argument [Beh] for three-term arithmetic progressions. Let m,d,r be positive integers, and let Λ be defined to be

$$\Big\{x_0+x_1m+\cdots+x_{d-1}m^{d-1} : x_i\in\mathbb{Z},\ 0\le x_i<m/4,\ \sum_{i=0}^{d-1}x_i=r\Big\}\ . \quad (8.14)$$

We claim that Λ does not contain any QC5. Let P be a quadratic integer polynomial such that $P(0),P(1),\dots,P(4)$ belong to Λ. For $j=0,1,\dots,4$ we write

$$P(i)=\sum_{j=0}^{d-1}x_{i,j}m^j \text{ with } x_{i,j}\in\mathbb{Z},\ 0\le x_{i,j}<m/4,\ \sum_{i=0}^{d-1}x_{i,j}=r\ .$$

The integers $P(0),P(1),\dots,P(4)$ are related by the equations:

$$P(0)-3P(1)+3P(2)-P(3)=0 \text{ and } P(1)-3P(2)+3P(3)-P(4)=0$$

and we have that

$$\sum_{j=0}^{d-1}\big(x_{0,j}-3x_{1,j}+3x_{2,j}-x_{3,j}\big)m^j=0\ ;$$
$$\sum_{j=0}^{d-1}\big(x_{1,j}-3x_{2,j}+3x_{3,j}-x_{4,j}\big)m^j=0\ .$$

The left hand side of each of these equations is the value at m of some polynomial whose coefficients are integers belonging to the interval $(-m,m)$. As m is a root of this polynomial, it is identically zero and

$$x_{0,j}-3x_{1,j}+3x_{2,j}-x_{3,j}=0 \text{ and } x_{1,j}-3x_{2,j}+3x_{3,j}-x_{4,j}=0$$

for $j=0,1,\dots,d-1$. The five points $a_0,a_1,\dots,a_4\in\mathbb{R}^d$ given by

$$a_i=(x_{i,0},x_{i,1},\dots,x_{i,d-1})$$

satisfy relations (8.6) and (8.7) and all have the same Euclidean norm. By Lemma 8.2 they are equal and thus $P(0)=P(1)=\cdots=P(4)$; P is constant and our claim is proven.

For d,m given, let F be the set of integers of the form $x_0+x_1m+\cdots+x_{d-1}m^{d-1}$ where $x_i\in\mathbb{Z}$ and $0\le x_i<m/4$ for $0\le i\le d-1$. If two vectors $(x_0,x_1,\dots,x_{d-1})$ and $(x'_0,x'_1,\dots,x_1d-1)$ of this form give the same element of F, then

$$\sum_{i=0}^{d-1}(x_i-x'_i)m^i=0$$

and by the same argument as above the vectors $(x_0,x_1,\dots,x_{d-1})$ and $(x'_0,x'_1,\dots,x'_{d-1})$ are equal. Therefore, F has at least $(m/4)^d$ elements and there exists r, $0\le r<d(m/4)^2$, such that the set Λ defined by (8.14) has at least $(m/4)^{d-2}d^{-1}$ elements. Note that $\Lambda\subset\{0,\dots,L-1\}$ for $L=m^d$. Choosing $m=2^d$, we have a set Λ of the announced order of magnitude. $\square$

References

[AGH] L. Auslander, L. Green and F. Hahn. Flows on homogeneous spaces. *Ann. Math. Studies* **53**, Princeton Univ. Press (1963).
[Beh] F. A. Behrend. On sets of integers which contain no three in arithmetic progression. *Proc. Nat. Acad. Sci.* **23** (1946), 331–332.
[Ber] V. Bergelson. Weakly mixing PET. *Ergod. Th. & Dynam. Sys.* **7** (1987), 337–349.
[F3] H. Furstenberg. Strict ergodicity and transformation of the torus. *Amer. J. of Math.*, **83** (1961), 573–601.
[F1] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, **31** (1977), 204–256.
[F2] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton Univ. Press (1981).
[HK] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. To appear, *Ann. of Math.* Available at http://www.math.psu.edu/kra (2002).
[K] A. Y. Khintchine. Eine Verschärfung des Poincaréschen "Wiederkehr-satzes". *Comp. Math.*, **1** (1934), 177–179.
[Lei1] A. Leibman. Polynomial sequences in groups. *J. of Algebra*, **201** (1998), 189–206.
[Lei2] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold. To appear, *Ergod. Th. & Dynam. Sys.* Available at http://www.math.ohio-state.edu/ leibman/preprints (2002).
[Les] E. Lesigne. Sur une nil-variété, les parties minimales associées à une translation sont uniquement ergodiques. *Ergod. Th. & Dynam. Sys.* **11** (1991), 379–391.
[M] A. Malcev. On a class of homogeneous spaces. *Amer. Math. Soc. Transl.* **39** (1951)
[Pa1] W. Parry. Ergodic properties of affine transformations and flows on nilmanifolds. *Amer. J. Math. 91* (1969), 757–771.
[Pa2] W. Parry. Dynamical systems on nilmanifolds. *Bull. London Math. Soc.* **2** (1970), 37–40.
[Pe] J. Petresco. Sur les commutateurs. *Math. Z.* **61** (1954), 348–356.
[R] D. J. Rudolph. Eigenfunctions of $T \times S$ and the Conze-Lesigne algebra. *Ergodic Theory and its Connections with Harmonic Analysis*, Eds. K. Petersen and I. Salama. Cambridge University Press, New York (1995), 369–432.
[S] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** (1975), 199-245.
[Z1] T. Ziegler. A non-conventional ergodic theorem for a nilsystem. To appear, *Ergod. Th. & Dynam. Sys.* Available at http://www.arxiv.org, math.DS/0204058 v1 (2002).
[Z2] T. Ziegler. Universal characteristic factors and Furstenberg averages. Preprint. Available at http://www.arxiv.org/abs/math.DS/0403212 (2004).