A nilpotent Roth theorem

V. Bergelson (vitaly@math.ohio-state.edu),
A. Leibman (leibman@math.ohio-state.edu)

Department of Mathematics
The Ohio State University
Columbus, OH 43210, USA

Abstract

Let T and S be invertible measure preserving transformations of a probability measure space $(X,\mathcal{B},\mu)$. We prove that if the group generated by T and S is nilpotent, then $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}u(T^nx)v(S^nx)$ exists in L^2-norm for any $u,v\in L^\infty(X,\mathcal{B},\mu)$. We also show that for $A\in\mathcal{B}$ with $\mu(A)>0$ one has $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mu(A\cap T^{-n}A\cap S^{-n}A)>0$. By the way of contrast, we bring examples showing that if measure preserving transformations T, S generate a solvable group, then (i) the above limits do not have to exist; (ii) the double recurrence property fails, that is, for some $A\in\mathcal{B}$, $\mu(A)>0$, one may have $\mu(A\cap T^{-n}A\cap S^{-n}A)=0$ for all $n\in\mathbb{N}$. Finally, we show that when T and S generate a nilpotent group of class $\leq c$, $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}u(T^nx)v(S^nx)=\int ud\mu\int vd\mu$ in $L^2(X)$ for all $u,v\in L^\infty(X)$ if and only if $T\times S$ is ergodic on $X\times X$ and the group generated by $T^{-1}S,T^{-2}S^2,\ldots,T^{-c}S^c$ acts ergodically on X.

0. Introduction

Let $(X,\mathcal{B},\mu)$ be a probability measure space and let T be a measure preserving transformation of X. The Furstenberg's *ergodic Roth theorem* (see [F1], [F3]) asserts that for any $u,v\in L^\infty(X,\mathcal{B},\mu)$ the limit

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}u(T^nx)v(T^{2n}x) \tag{0.1}$$

exists in L^2-norm. If $A\in\mathcal{B}$ is a set of positive measure and $u=v=1_A$, then one can show that $f(x)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}u(T^nx)v(T^{2n}x)$ satisfies $\int_A fd\mu>0$. This, via Furstenberg's correspondence principle ([F3]), implies Roth's theorem, namely the fact

---

1991 *Mathematical Subject Classification.* Primary 28D15; Secondary 20F18, 11L15.

The authors gratefully acknowledge support received from the National Science Foundation (USA) via grant DMS-9706057.

that any set $E \subseteq \mathbb{N}$ with positive upper density $\bar{d}(E) = \limsup_{N\to\infty} \frac{|E\cap\{1,\ldots,N\}|}{N}$ contains arithmetic progressions of length 3.

Roth's theorem, published in 1952 ([R]), provided verification of the first nontrivial case of the Erdös–Turán conjecture ([ET]) that any set $E \subseteq \mathbb{N}$ with $\bar{d}(E) > 0$ contains arbitrarily long arithmetic progressions. The conjecture was settled in affirmative by Szemerédi in [Sz]. The whole topic was given new life in 1977 when Furstenberg ([F1]) gave an ergodic proof of Szemerédi's theorem. Furstenberg's seminal approach gave rise to a new discipline, Ergodic Ramsey Theory, which placed Szemerédi's theorem into the right perspective and has led to new discoveries (see for example [FK1], [FK2], [FK3], [BL1], [BL2], [L2]), which benefited both the ergodic theory and combinatorics (see [F2] for the introduction to Ergodic Ramsey Theory and [B2] for an update on some of more recent developments).

Some of the natural problems in Ergodic Ramsey Theory as well as the problems in the ergodic theory of multiple recurrence concentrate around the study of the behavior of ergodic averages of the form

$$F_N^{(k)} = \frac{1}{N}\sum_{n=1}^{N} u_1(T_1^n x)\ldots u_k(T_k^n x), \tag{0.2}$$

where T_i are measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$ and $u_i \in L^\infty(X, \mathcal{B}, \mu)$, $i = 1, \ldots, k$. Two questions related to the limiting behavior of $F_N^{(k)}$ are of major importance. First, one wants to know whether the $\lim_{N\to\infty} F_N^{(k)}$ exists (weakly, in norm, or even almost everywhere). Second, if all the u_i are equal to 1_A, where $A \in \mathcal{B}$, $\mu(A) > 0$, one wants to know if

$$\liminf_{N\to\infty} \int_X F_N^{(k)} d\mu = \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(T_1^{-n}A \cap \ldots \cap T_k^{-n}A) > 0. \tag{0.3}$$

While the first question has an intrinsic value for purely ergodic reasons, the second one is also related to applications of ergodic theory to combinatorics.

It turns out that the answers to these two questions depend heavily on the nature of the (semi)group $G(T_1, \ldots, T_k)$ generated by the transformations T_i. We describe first the state of current knowledge with respect to the second question. Let us assume that the transformations T_i are invertible. If the group $G(T_1, \ldots, T_k)$ is nilpotent, it was shown in [L2] that the lim inf in (0.3) is always positive. On the other hand, some examples discussed in Section 4 below show that if $G(T_1, T_2)$ is a solvable, and even polycyclic, non-nilpotent group, it may happen that for some $A \in \mathcal{B}$ with $\mu(A) > 0$, $T_1^{-n}A \cap T_2^{-n}A = \emptyset$ for all $n > 0$. (We strongly believe that *any* solvable non-virtually nilpotent group G possesses a measure preserving action on a space X such that for some $T_1, T_2 \in G$ and $A \subset X$ with $\mu(A) > 0$ one has $T_1^{-n}A \cap T_2^{-n}A = \emptyset$ for all $n > 0$; see Conjecture 5.4 below.)

Much less is known about the existence of the $\lim_{N\to\infty} F_N^{(k)}$. Even in the case of powers of the same transformation, $T_i = T^{a_i}$, the existence of the limit in norm is known only for $k \leq 3$ (proved for totally ergodic T in [CL2] and in full generality in [FW] and [HK]). In the more general case of commuting transformations the best results to date are due to Conze and Lesigne ([CL1]), who established the existence of the $\lim_{N\to\infty} F_N^{(2)}$ in norm, and to Zhang ([Zh]), who proved the existence of $\lim_{N\to\infty} F_N^{(3)}$ under the assumption that T_i and $T_i^{-1}T_j$, $i, j = 1, 2, 3$, $i \neq j$, are ergodic.

Our main result is the following extension of the Conze-Lesigne theorem from [CL1]:

Theorem A. *Let T and S be two invertible measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$. Assume that the group of transformations generated by T and S is nilpotent. Then the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} u(T^n x) v(S^n x) \tag{0.4}$$

exists in L^2-norm for all $u, v \in L^\infty(X)$.

Note that it immediately follows from Theorem A that the limit (0.4) exists also in the L^1-norm. Moreover, it is not hard to show that the L^1-convergence holds for any $u \in L^p(X)$ and $v \in L^q(X)$ with $\frac{1}{p} + \frac{1}{q} = 1$. The problem of the almost everywhere convergence is however much more delicate: even in the case of commuting T, S this is an open question, though it is generally believed to have a positive answer. The best result in this direction is due to Bourgain, who established in [Bo] the almost everywhere convergence of the averages of the form $\frac{1}{N}\sum_{n=1}^{N} u(T^{an}x)v(T^{bn}x)$, $a, b \in \mathbb{Z}$.

While we believe that Theorem A extends to expressions (0.2) with $k > 2$ (see Conjecture 5.5 below), such an extension seems to require a significant progress in our understanding of the characteristic factors of measure preserving systems. (See [F4] and [FW], where the difficulties arising already in the case of powers of a single transformation are discussed.)

The proof of Theorem A follows in general the scheme of the proof of its commutative analogue in [CL1]. However, the fact that T and S are not necessarily commuting poses additional hindrances. In the case when T and S commute, one analyses the behavior of T and S with respect to a factor $(X, \mathcal{D}, \mu)$ of $(X, \mathcal{B}, \mu)$ (*a factor* is the measure space determined by a sub-σ-algebra $\mathcal{D}$ of $\mathcal{B}$) with the property that $L^2(X, \mathcal{D}, \mu)$ is spanned by the limits of the ergodic averages $\frac{1}{N}\sum_{n=1}^{N} T^{-n}S^n u(x) = \frac{1}{N}\sum_{n=1}^{N} u(T^{-n}S^n x)$ for $u \in L^2(X, \mathcal{B}, \mu)$. If T and S commute, then $T^{-n}S^n = (T^{-1}S)^n$, and $\mathcal{D}$ is the σ-algebra generated by $T^{-1}S$-invariant sets. Then the actions of T and S on $L^2(X,\mathcal{D},\mu)$ coincide, which plays an essential role in the proof.

Now, in the case where T and S generate a nilpotent non-commutative group G, one does not have the relation $T^{-n}S^n = (T^{-1}S)^n$, so that even the existence of

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} T^{-n}S^n u(x) \tag{0.5}$$

is no longer obvious. What comes to help is the fact that $T^{-n}S^n$ is a polynomial sequence. A sequence $g(n)$ taking values in a nilpotent group G is called *a polynomial sequence*, or *a G-polynomial*, if the derivative $Dg(n)$ defined by $Dg(n) = g(n)^{-1}g(n+1)$ trivializes after finitely many applications (that is, there exists $d \in \mathbb{N}$ such that $D^d g \equiv \mathbf{1}_G$). We prove the following theorem, which is a nilpotent generalization of a von Neumann-type ergodic polynomial theorem for commuting operators (see for example [B2], section 2):

Theorem C. *Let G be a nilpotent group of unitary operators on a Hilbert space $\mathcal{H}$. Then for any G-polynomial $g(n)$ and any $u \in \mathcal{H}$, $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} g(n)u$ exists.*

Remark. Theorem C should be seen as yet another theorem belonging to the variety of results which extend recurrence and equidistribution from the group actions framework to polynomial setup. See, for example, [S], where N. Shah establishes an analogue of Weyl equidistribution theorem for polynomial trajectories on homogeneous spaces.

Applying Theorem C to the G-polynomial $g(n) = T^{-n}S^n$ and the Hilbert space $\mathcal{H} = L^2(X,\mathcal{B},\mu)$, one obtains the existence of the limit (0.5). However, unlike the commutative case, T and S do not have to coincide on the subspace of $L^2(X,\mathcal{B},\mu)$ which is spanned by the limits of the form (0.5). This complicates the situation and makes necessary introducing a technique which is based on a nilpotent analogue of the following well known fact from analytic number theory:

Proposition. (See [Hua], Chapter 1) *For any $\varepsilon > 0$ and $r \in \mathbb{N}$ there is $L \in \mathbb{N}$ such that if λ is a primitive root of unity of degree $k > L$, and if a polynomial $p(n) = a_r n^r + \ldots + a_1 n \in \mathbb{Z}[n]$ satisfies* g.c.d.$(k, a_1, \ldots, a_r) = 1$, *then $\left|\frac{1}{k}\sum_{n=1}^{k} \lambda^{p(n)}\right| < \varepsilon$.*

To formulate the nilpotent version of the Proposition, let us introduce some notation. Given a nilpotent group G of unitary operators on a Hilbert space $\mathcal{H}$ and a G-polynomial $g(n)$ satisfying $g(0) = \mathbf{1}_G$, let $\mathcal{H}^{\mathrm{rat}}(g) = \{u \in \mathcal{H} \mid \text{the sequence } g(n)u,\ n \in \mathbb{Z}, \text{ is periodic}\}$. Let H be the subgroup of G generated by the elements of g and let $\mathcal{H}^{(l)}(g) = \{u \in \mathcal{H} \mid P^l u = u \text{ for all } P \in H\}$. One can show that $\mathcal{H}^{\mathrm{rat}}(g) = \bigcup_{l\in\mathbb{N}} \mathcal{H}^{(l)}(g)$, and that for every $l \in \mathbb{N}$, g is periodic on $\mathcal{H}^{(l)}(g)$: there is $K \in \mathbb{N}$ such that $g(n+K)u = g(n)u$ for all $u \in \mathcal{H}^{(l)}(g)$ and all $n \in \mathbb{Z}$. It follows from the proof of Theorem C (see Theorem 2.17

below) that for any $u \in \mathcal{H}$, $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} g(n)u$ lies in the closure of $\mathcal{H}^{\mathrm{rat}}(g)$. The following theorem may be viewed as a nilpotent analogue of the Proposition above, and is also an enhancement of Theorem 2.17:

Theorem D. *For every $r \in \mathbb{N}$ and $\varepsilon > 0$ there is $L \in \mathbb{N}$ such that if g is a G-polynomial of degree $\leq r$ with $g(0) = \mathbf{1}_G$, and $u \in \mathcal{H}$ is such that $u \perp \mathcal{H}^{(l)}(g)$ for all $l \leq L$, then $\left\|\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} g(n)u\right\| \leq \varepsilon\|u\|$.*

It follows from Theorem D, applied to $\mathcal{H} = L^2(X,\mathcal{B},\mu)$ and $g(n) = T^{-n}S^n$, that for any $u \in L^2(X,\mathcal{B},\mu)$ the "major" portion of the limit $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} T^{-n}S^n u$ belongs to the space $\mathcal{H}^{(L!)}(g)$, where L is large enough. Now, since $g(n)$ is periodic on $\mathcal{H}^{(L!)}(g)$, one has $T^{-K}S^K|_{\mathcal{H}^{(L!)}(g)} = \mathbf{1}_{\mathcal{H}^{(L!)}(g)}$ for some $K \in \mathbb{N}$. In other words, T^K and S^K coincide on $\mathcal{H}^{(L!)}(g)$. One is able then to conclude the proof of Theorem A by analyzing the behavior of T and S with respect to the T,S-invariant factor of $(X,\mathcal{B},\mu)$ which is determined by the subspace $\mathcal{H}^{(L!)}(g)$.

As a corollary of Theorem A we obtain a strong form of double recurrence:

Theorem E. *Let T, S be measure preserving transformations of $\mathbf{X}$ generating a nilpotent group. Then for any $A \in \mathcal{B}$ with $\mu(A) > 0$, $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu\big(T^{-n}A \cap S^{-n}A \cap A\big) > 0$.*

We remark that the fact that $\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu\big(T^{-n}A \cap S^{-n}A \cap A\big) > 0$ is a special case of a general nilpotent Szemerédi theorem proved in [L2]; the novelty of Theorem E is that the lim inf is replaced by lim, and that the proof of this theorem is direct and relatively simple.

We also bring the following theorem, which establishes the necessary and sufficient conditions for the limit (0.4) to be the "right" one:

Theorem B. *Under the assumptions of Theorem A, let the group G generated by T and S have nilpotency class $\leq c$ (that is, if $G_1 = G$ and $G_{k+1} = [G, G_k]$, $k = 1, 2, \ldots$, the subgroup G_{c+1} is trivial). Then one has*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} u(T^n x)v(S^n x) = \int u d\mu \int v d\mu$$

in L^2-norm for all $u, v \in L^\infty(X)$ if and only if the transformation $T \times S$ is ergodic on $X \times X$, and the group generated by $\{T^{-n}S^n,\ 1 \leq n \leq c\}$ acts ergodically on X.

Theorem B naturally generalizes (for the case of 2 transformations) the "commutative" result from [BB]:

Theorem BB. ([BB]) *Let $T_1, \ldots, T_k$ be commuting invertible measure preserving trans-*

formations of a probability measure space $(X, \mathcal{B}, \mu)$. *Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} u_1(T_1^n x) \dots u_k(T_k^n x) = \int u_1 d\mu \dots \int u_k d\mu$$

in L^2*-norm for all* $u_1, \dots, u_k \in L^\infty(X)$ *if and only if* $T_1 \times \dots \times T_k$ *is ergodic on* $X \times \dots \times X$ *and all the transformations* $T_i^{-1} T_j$, $i \neq j$, *are ergodic on* X.

It is worth mentioning that in complete analogy with the commutative situation, Theorems A – E also admit "uniform" versions which are obtained by replacing in the formulations Cesàro limits $\lim\limits_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N}$ by the uniform Cesàro limits $\lim\limits_{N-K\to\infty} \frac{1}{N-K} \sum_{n=K+1}^{N}$. (See Remark 2.2 in Section 2.)

One would like to know whether Theorems A and B hold in more general situations where the group $G = G(T, S)$ is not nilpotent. While Theorems A and B trivially extend to the case when G is virtually nilpotent, namely, contains a nilpotent subgroup of finite index, we show in Section 4 that if G is a solvable group of exponential growth, then the limit (0.4) does not have to exist even weakly, and even under the assumptions of Theorem B. (Actually we show that for suitably chosen L^∞-functions u, v, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int u(T^n x) v(S^n x) d\mu$ fails to exist.)

The structure of the paper is as follows. Section 1 is devoted to algebraic preliminaries about polynomial sequences in nilpotent groups. In Section 2 we consider a (finitely generated) nilpotent group G of unitary operators on a Hilbert space $\mathcal{H}$, build a structure theory of its action with respect to subspaces of rational spectrum of its elements, and prove Theorems C and D. The proofs of Theorems A, B and E, based on the results obtained in the Sections 1 and 2, are brought in Section 3. Section 4 is dedicated to various counterexamples, which show that the class of nilpotent groups is the natural domain where the ergodic Szemerédi and Roth theorems hold. Finally, in Section 5 we discuss possible refinements of our results and some open problems.

1. Polynomial sequences in nilpotent groups

1.1. Let G be a nilpotent group of class c and let $G = G_1 \supset G_2 \supset \dots \supset G_c \supset G_{c+1} = \{\mathbf{1}_G\}$ be its lower central series: $G_{k+1} = [G, G_k]$, $k = 1, \dots, c$. For a sequence $g\colon \mathbb{Z} \longrightarrow G$ in G, we define *the derivative of* g as the sequence $Dg\colon \mathbb{Z} \longrightarrow G$ with $Dg(n) = g(n)^{-1} g(n+1)$. g is *a polynomial sequence*, or *a* G*-polynomial*, if $D^d g \equiv \mathbf{1}_G$ for some $d \in \mathbb{N}$.

In particular, sequences $g\colon \mathbb{Z} \longrightarrow G$ satisfying $Dg \equiv \mathbf{1}_G$ are *the constant* G*-polynomials* $g \equiv T_0 \in G$, sequences satisfying $D^2 g \equiv \mathbf{1}_G$ are *the linear* G*-polynomials* $g(n) = T_1 T_0^n$,

$T_0, T_1 \in G$, *the quadratic G-polynomials* are of the form

$$\begin{cases} T_2(T_1)(T_1T_0)(T_1T_0^2)\dots(T_1T_0^{n-1}), \ n>0 \\ T_2(T_0T_1^{-1})(T_0^2T_1^{-1})\dots(T_0^{-n}T_1^{-1}), \ n\le 0, \end{cases}$$

etc.

1.2. *An integral polynomial* is a polynomial $p \in \mathbb{Q}[n]$ taking on integer values on the integers. It can be easily shown that any integral polynomial is a linear combination over $\mathbb{Z}$ of binomial coefficients $b_k(n) = \frac{n(n-1)\dots(n-k+1)}{k!}$, $k \ge 0$. Thus, if p is an integral polynomial of degree $\le r$, the least common multiple of the denominators of its coefficients divides $r!$, and hence the polynomials $r!p(n)$ and $p(r!n)$ have integer coefficients.

It can be shown (see, for example, [L1]) that for a nilpotent group G, a sequence g in G is a G-polynomial if and only if it is representable in the form

$$g(n) = T_1^{p_1(n)}\dots T_t^{p_t(n)}, \tag{1.1}$$

where $T_1,\dots,T_t \in G$ and $p_1,\dots,p_t$ are integral polynomials. If $g(0) = \mathbf{1}_G$, one can additionally assume that in (1.1), $p_1(0) = \ldots = p_t(0) = 0$. Moreover, if B is any subset of G with the property that $B \cap G_k$ generates the group G_k for each $k = 1,\dots,c$, then $g(n)$ has the representation (1.1) with $T_1,\dots,T_t \in B$.

1.3. G-polynomials form a group with respect to element-wise multiplication: $(gh)(n) = g(n)h(n)$; we will denote this group by $\wp G$.

It easily follows from the definition of a G-polynomial that if H is a subgroup of G and a G-polynomial g takes values in H: $g(n) \in H$ for all $n \in \mathbb{Z}$, then g is an H-polynomial. So, $\wp H$ is a subgroup of $\wp G$. It is also clear that if H is a normal subgroup in G and g is a G-polynomial, then the sequence $g(n)H$, $n \in \mathbb{Z}$, is a (G/H)-polynomial.

1.4. Following the analogy with ordinary polynomials, one has a temptation to define the degree of a G-polynomial g as the smallest nonnegative integer d for which $D^{d+1}g = \text{const}$. Regrettably, the degree so defined fails to have an important property: G-polynomials of degree $\le d$ do not, generally speaking, form a subgroup of $\wp G$. We do not have any "canonical" way of defining the degree; the following one seems to us to be best suited for our purposes (it is an adaptation for nilpotent groups of a more general definition in [L1]; cf. 1.4 and Theorem 1.12 in [L1]): Let for $1 \le k \le c$, d_k be the smallest nonnegative integer with the property $D^{d_k+1}g \in \wp G_{k+1}$ (that is, $\big(D^{d_k+1}g\big)(n) \in G_{k+1}$ for all $n \in \mathbb{Z}$); if $g \in \wp G_{k+1}$ we put $d_k = -\infty$. For $k = 1,\dots,c$, let $s_k = \max\limits_{l\le k,\ k_1+\ldots+k_l\le k} \sum_{j=1}^{l} d_{k_j}$. Then *the vector degree* of g is $\bar{s} = (s_1,\dots,s_c)$. Note the the vector degree $\bar{s}$ is *superadditive*: for any k,l with $k+l \le c$, one has $s_k + s_l \le s_{k+l}$. It can be shown that, given a superadditive vector $\bar{r} = (r_1, \ldots, r_c)$, G-polynomials of degree $\leq \bar{r}$ (that is, with $s_k \leq r_k$, $k = 1, \ldots, c$) form a group (see [L1]). One can also show that $g \in \wp G$ has degree $\leq \bar{r}$ if and only if g is representable in the form (1.1), $g(n) = T_1^{p_1(n)} \ldots T_t^{p_t(n)}$, so that, for $i = 1, \ldots, t$, if $T_i \in G_{k_i} \setminus G_{k_i+1}$, then $\deg p_i \leq r_{k_i}$.

On most occasions, however, one can work with a scalar "pseudodegree" which is given by the last component s_c of the vector degree $\bar{s}$, and will be denoted by $\deg g$. Under this definition, $\deg g \leq r$ implies that in the representation (1.1), $p_1, \ldots, p_t$ can be assumed to have degree $\leq r$. Warning: G-polynomials of degree $\leq r$ do not necessarily form a group.

1.5. It is clear that $\deg Dg \leq \deg g - 1$. Moreover, for every $m \in \mathbb{Z}$, the degree of the G-polynomial $g_m(n) = g(n)^{-1} g(n+m)$ is smaller than the degree of g. Indeed, for $m > 0$

$$\begin{aligned} g_m(n) = g(n)^{-1}g(n+1)g(n+1)^{-1}g(n+2)\ldots g(n+m-1)^{-1}g(n+m) \\ = Dg(n)Dg(n+1)\ldots Dg(n+m-1), \end{aligned}$$

and for $m < 0$

$$\begin{aligned} g_m(n) = g(n)^{-1}g(n-1)g(n-1)^{-1}g(n-2)\ldots g(n+m-1)^{-1}g(n+m) \\ = (Dg)^{-1}(n-1)(Dg)^{-1}(n-2)\ldots(Dg)^{-1}(n+m). \end{aligned}$$

Thus g_m lies in the subgroup of $\wp G$ generated by $Dg(n)$ and $Dg(n+k)$, $k \in \mathbb{Z}$. Let g have vector degree $(s_1, \ldots, s_c)$; then the vector degrees of Dg and of all of its shifts $Dg(n+k)$, $k \in \mathbb{Z}$, are majorized by $(s_1 - 1, \ldots, s_c - 1)$. Hence, the vector degree of g_m is also majorized by $(s_1 - 1, \ldots, s_c - 1)$ and so, $\deg g_m \leq s_c - 1 = \deg g - 1$.

1.6. If $\deg g \leq r$, one has $D^{r+1} g \equiv \mathbf{1}_G$. It follows (by induction on $\deg g$) that g is uniquely determined by its values at $0, 1, \ldots, r$. In particular, for every $n \in \mathbb{Z}$, $g(n)$ lies in the subgroup of G generated by $g(0), g(1), \ldots, g(r)$.

1.7. Now we bring in a unified form some general facts about nilpotent groups and polynomial sequences which will be needed in the sequel. We omit the proofs of the facts which are either well known or routine (see [H] for details).

1.8. Lemma. *Any subgroup H of a finitely generated nilpotent group G is finitely generated.*

1.9. Lemma. *Let G be a finitely generated nilpotent group. Then any nondecreasing sequence $H_1 \subseteq H_2 \subseteq \ldots$ of subgroups of G stabilizes.*

1.10. Lemma. *Any finitely generated torsion-free nilpotent group G possesses a finite central series with infinite cyclic factors: $\{\mathbf{1}_G\} = H_0 \subset H_1 \subset \ldots \subset H_s = G$ such that, for each $i = 1, \ldots, s$, H_i/H_{i-1} is in the center of G/H_{i-1} and $H_i/H_{i-1} \simeq \mathbb{Z}$.*

1.11. Given a subset B of a group G, we will denote by $\langle B\rangle$ the subgroup generated by B. For $T_1,\dots,T_t\in G$, let $\langle T_1,\dots,T_t\rangle=\big\langle\{T_1,\dots,T_t\}\big\rangle$.

1.12. Lemma. *Assume that a nilpotent group G is generated by its elements $T_1,\dots,T_t$, and let $d_1,\dots,d_t\in\mathbb{N}$. Then the subgroup $H=\langle T_1^{d_1},\dots,T_t^{d_t}\rangle$ has finite index in G.*

1.13. For a subgroup H of a group G we denote by $N(H)$ the normalizer of H in G: $N(H)=\{T\in G \mid T^{-1}HT=H\}$. We also define $N^0(H)=H$, $N^k(H)=N\big(N^{k-1}(H)\big)$, $k=1,2,\dots$.

1.14. Lemma. *If H is a subgroup of a nilpotent group G of class c, then $N^c(H)=G$.*

1.15. Let H be a subgroup of a group G. We will call the set of elements of G which have finite order modulo H *the closure of H* (in G) and denote it by $\overline{H}$:

$$\overline{H}=\big\{T\in G \bigm| T^n\in H \text{ for some } n\in\mathbb{N}\big\}.$$

We will say that H *is closed* if $H=\overline{H}$.

1.16. Lemma. *If G is a nilpotent group and H is its subgroup, then $\overline{H}$ is a closed subgroup of G.*

Proof. It follows immediately from definition that $\overline{H}$ is closed. To see that $\overline{H}$ is a subgroup, let $P_1,P_2\in\overline{H}$, that is, $P_1^{n_1},P_2^{n_2}\in H$ for some $n_1,n_2\in\mathbb{N}$. Since by Lemma 1.12 the subgroup $E=\langle P_1^{n_1},P_2^{n_2}\rangle$ has finite index in the group $\langle P_1,P_2\rangle$, we have $(P_1P_2)^n\in E\subseteq H$ for some $n\in\mathbb{N}$, which implies $P_1P_2\in\overline{H}$. ∎

1.17. Lemma. *If H is a subgroup of a finitely generated nilpotent group G, then H has finite index in $\overline{H}$.*

Proof. This follows from Lemmas 1.8 and 1.12. ∎

1.18. Proposition. *If H is a closed subgroup of a nilpotent group G, then $N(H)$ is closed in G as well.*

Proof. Let $T\in\overline{N(H)}$, that is, $T^d\in N(H)$ for some $d\in\mathbb{N}$. We may assume that G is generated by H and T; let $G=G_1\supset\ldots\supset G_c\supset G_{c+1}=\{\mathbf{1}_G\}$ be the lower central series.

We will prove that for any $k\ge 2$ and any G_k-polynomial g with $g(0)=\mathbf{1}_G$ there is $m\in\mathbb{N}$ such that the G_k-polynomial $g(mn)$ can be written as a product $g(mn)=h(n)\tilde{g}(n)$ with $h\in\wp H$ and $\tilde{g}\in\wp G_{k+1}$. The group G_k is generated by G_{k+1} and elements of the form

$$S=\big[R_1,\big[R_2,\dots,[R_{k-1},R_k]\dots\big]\big] \tag{1.2}$$

where either $R_i = T$ or $R_i \in H$, $i = 1, \dots, k$. For such S,

$$S^{d^k} \equiv \big[R_1^d, \big[R_2^d, \dots, [R_{k-1}^d, R_k^d] \dots\big]\big] \bmod G_{k+1}.$$

Since $T^d \in N(H)$, it is easy to see that $R = \big[R_1^d, \big[R_2^d, \dots, [R_{k-1}^d, R_k^d] \dots\big]\big] \in H$. Now, if $p(n)$ is an integral polynomial with $p(0) = 0$, then for some $r \in \mathbb{N}$, $p(rn) \in \mathbb{Z}[n]$, and then $p(rn) = nq(n)$ where $q(n) \in \mathbb{Z}[n]$. Thus, for $m = rd^k$ we have

$$S^{p(mn)} = S^{p(rd^k n)} = S^{d^k nq(d^k n)} \equiv R^{nq(d^k n)} \bmod G_{k+1} \equiv f(n) \bmod G_{k+1}$$

with $f(n) = R^{nq(d^k n)} \in \wp H$.

Now, if g is a G_k-polynomial with $g(0) = \mathbf{1}_G$, then one can find $S_1, \dots, S_s$ of the form (1.2) such that $g(n) \equiv S_1^{p_1(n)} \dots S_s^{p_s(n)} \bmod G_{k+1}$ where p_i, $i = 1, \dots, t$, are integral polynomials with $p_i(0) = 0$. Thus for a suitable $m \in \mathbb{N}$, we have $g(mn) \equiv f_1(n) \dots f_t(n) \bmod G_{k+1}$, with $f_1, \dots, f_t \in \wp H$. For $h = f_1 \dots f_t \in \wp H$ this gives $g(mn) = h(n)\tilde{g}(n)$ with $\tilde{g} \in \wp G_{k+1}$.

Now let g be a G_2-polynomial with $g(0) = \mathbf{1}_G$. Then for some $m_2 \in \mathbb{N}$, $g(m_2 n) = h_2(n) g_3(n)$ where $h_2 \in \wp H$ and $g_3 \in \wp G_3$. In its turn, for some $m_3 \in \mathbb{N}$ we have $g_3(m_3 n) = h_3(n) g_4(n)$ where $h_3 \in \wp H$ and $g_4 \in \wp G_4$, and etc. As a result, for $m = m_2 \dots m_c$ we get $g(mn) \in \wp H$.

Take any $P \in H$ and define a G_2-polynomial g by $g(n) = [P, T]^n$. Then for some $m \in \mathbb{N}$ we have $g(mn) \in \wp H$, which implies $[P, T]^m \in H$. Since H is closed, $[P, T] = P^{-1} T^{-1} P T \in H$, and so $T^{-1} P T \in H$. This proves that $T \in N(H)$. ∎

1.19. Proposition. *Let H be a subgroup of a finitely generated nilpotent group G. Then H is closed if and only if there is a subnormal series $H = H_0 \subset H_1 \subset \dots \subset H_t = G$ with infinite cyclic factors; that is, for each $j = 1, \dots, t$, the subgroup H_{j-1} is normal in H_j and $H_j / H_{j-1} \simeq \mathbb{Z}$.*

Proof. First, assume that $H = H_0 \subset H_1 \subset \dots \subset H_t = G$ is a subnormal series with infinite cyclic factors. Let $T \in G$, $T \notin H$. Let $1 \le j \le t$ be such that $T \in H_j \setminus H_{j-1}$. Since $H_j / H_{j-1} \simeq \mathbb{Z}$, $T^n \notin H_{j-1}$ for all $n \ne 0$, and so, $T^n \notin H$ for all $n \ne 0$. Hence, H is closed in G.

Conversely, assume that H is closed in G. Then $N(H)/H$ is a finitely generated torsion-free nilpotent group. By Lemma 1.10, $N(H)/H$ possesses a central series $\{\mathbf{1}_{N(H)/H}\} = \tilde{H}_0 \subset \tilde{H}_1 \subset \dots \subset \tilde{H}_s = N(H)/H$ with infinite cyclic factors. Denote by H_i the preimage of $\tilde{H}_i$ in $N(H)$, $i = 0, \dots, s$. Then $H = H_0 \subset H_1 \subset \dots \subset H_s = N(H)$ is a subnormal series with infinite cyclic factors.

By Proposition 1.18, the subgroup $N(H)$ is also closed in G. So, we can continue our series by a subnormal series $N(H) = H_s \subset H_{s+1} \subset \dots \subset H_{s_1} = N^2(H)$ with infinite cyclic factors. Since $N^2(H)$ is closed, there is a series $N^2(H) = H_{s_1} \subset H_{s_1+1} \subset \ldots \subset H_{s_2} = N^3(H)$, etc. Since, by Lemma 1.14, $N^c(H) = G$, the proposition follows. ∎

1.20. Proposition. *Let H be a closed subgroup of a nilpotent group G, and let $g \in \wp G \setminus \wp H$. Then the set $\{n \in \mathbb{Z} \mid g(n) \in H\}$ is finite.*

Proof. Since $\{g(n),\ n \in \mathbb{Z}\}$ lies in a finitely generated subgroup of G, we may assume that G is finitely generated. By Proposition 1.19, there exists a subnormal series $H = H_0 \subset H_1 \subset \ldots \subset H_t = G$ with $H_j/H_{j-1} \simeq \mathbb{Z}$ for all $j = 1, \ldots, t$. Let $1 \leq j \leq t$ be such that $g \in \wp H_j \setminus \wp H_{j-1}$. Let T be a generator of H_j/H_{j-1}. The image $\tilde{g}$ of g in H_j/H_{j-1} is a (H_j/H_{j-1})-polynomial, and hence is of the form $\tilde{g}(n) = T^{p(n)}$, where p is an integral polynomial. Since $p \not\equiv 0$, p has only finitely many roots, and so, $\tilde{g}(n) = \mathbf{1}_{H_j/H_{j-1}}$ for at most finitely many $n \in \mathbb{Z}$. It follows that $g(n) \in H_{j-1}$ for at most finitely many $n \in \mathbb{Z}$. ∎

1.21. Proposition. *Let H be a subgroup of a nilpotent group G, let $g \in \wp G \setminus \wp H$. Then either $g(n) \in H$ for only finitely many $n \in \mathbb{Z}$, or the sequence $g(n)H$ is periodic.*

Proof. By Lemma 1.16, $\overline{H}$ is a closed subgroup in G. Thus if $g \notin \wp\overline{H}$, then by Proposition 1.20, $g(n) \in \overline{H}$ for at most finitely many $n \in \mathbb{Z}$.

Assume that $g \in \wp\overline{H}$. Let $\deg g \leq r$. Since $\{g(n),\ n \in \mathbb{Z}\}$ lies in a finitely generated subgroup of G, we may assume that G is finitely generated. By Lemma 1.17, H has finite index in $\overline{H}$, so $g(n)H$ runs through a finite set of left cosets of H. Let $g(m)H$ be one of them. We can write the $\overline{H}$-polynomial $g'(n) = g(m)^{-1}g(n+m)$ as $g'(n) = S_1^{p_1(n)} \ldots S_s^{p_s(n)}$, where $S_1, \ldots, S_s \in \overline{H}$ and $p_1, \ldots, p_s$ are integral polynomials of degree $\leq r$ satisfying $p_1(0) = \ldots = p_s(0) = 0$. Let $d \in \mathbb{N}$ be such that $S_j^d \in H$ for all $j = 1, \ldots, s$. For every $j = 1, \ldots, s$, since $\deg p_j \leq r$ and $p_j(0) = 0$, one has $p_j(r!n) \in \mathbb{Z}[n]$ and so, $p_j(r!n) = nq_j(n)$ with $q_j \in \mathbb{Z}[n]$. Thus $p_j(r!dn) \equiv 0 \bmod d$, $j = 1, \ldots, s$, for all $n \in \mathbb{Z}$. Hence, $g'(r!dn) \in H$ and so, $g(r!dn + m) \in g(m)H$ for all $n \in \mathbb{Z}$. Since this is true for all $m \in \mathbb{Z}$, the sequence $g(n)H$ is periodic (and its period divides $r!d$). ∎

1.22. Let H be a subgroup of a group G. We define *the period of G relative to H*, $\mathrm{per}_H(G)$, as the minimal $d \in \mathbb{N}$ such that $T^d \in H$ for all $T \in G$; if such d does not exist, we say that $\mathrm{per}_H(G)$ is infinite. We also define $\mathrm{per}(G) = \mathrm{per}_{\{\mathbf{1}_G\}}(G)$. If H is a normal subgroup of G, then, clearly, $\mathrm{per}_H(G) = \mathrm{per}(G/H)$.

1.23. Lemma. *Let H be a subgroup of a group G. For any $S \in G$, $\mathrm{per}_{SHS^{-1}}(G) = \mathrm{per}_H(G)$. Furthermore, if φ is an automorphism of G, then $\mathrm{per}_{\varphi(H)}(G) = \mathrm{per}_H(G)$.*

1.24. Corollary. *Let H be a subgroup of a group G, let Φ be a set of automorphisms of G and let $\tilde{H} = \bigcap_{\varphi \in \Phi} \varphi(H)$. Then $\mathrm{per}_{\tilde{H}}(G) = \mathrm{per}_H(G)$. In particular, for $\hat{H} =$*

$\bigcap_{S\in G} SHS^{-1}$, $\mathrm{per}_{\hat{H}}(G)=\mathrm{per}_H(G)$.

1.25. The following lemma refines Lemma 1.12:

Lemma. *Let G be a nilpotent group of class c, let $G_2=[G,G]$, and let H be a subgroup of G. Then* $\mathrm{per}(G/(HG_2))\le \mathrm{per}_H(G)\le \big(\mathrm{per}(G/(HG_2))\big)^c$.

Proof. The first inequality is obvious. By Corollary 1.24, we may replace H by the subgroup $\bigcap_{S\in G} S^{-1}HS$ and hence may assume that H is normal in G. After factorizing G by H, we have only to check that $\mathrm{per}(G)\le \big(\mathrm{per}(G/G_2)\big)^c$.

Let $K=\mathrm{per}(G/G_2)$; we may assume that K is finite. Let $G=G_1\supset\ldots\supset G_c\supset G_{c+1}=\{\mathbf{1}_G\}$ be the lower central series of G. For $1\le k\le c$, the group G_k is generated by elements of the form $S=\big[S_1,[S_2,\ldots,[S_{k-1},S_k]\ldots]\big]$. Since $S_k^K\in G_2$, for any such S we have

$$S^K\equiv\big[S_1,[S_2,\ldots,[S_{k-1},S_k^K]\ldots]\big]\equiv \mathbf{1}_G \bmod G_{k+1},$$

that is, $S^K\in G_{k+1}$. So, $\mathrm{per}_{G_{k+1}}(G_k)\le K$ and hence, $\mathrm{per}(G)\le K^c$. ■

1.26. Let H be a subgroup of a nilpotent group G and g be a G-polynomial. We define *the period of g with respect to H*, $\mathrm{per}_H(g)$, as the minimal $k\in\mathbb{N}$ for which $g(n+k)\equiv g(n)\bmod H$ for all $n\in\mathbb{Z}$. That is, $\mathrm{per}_H(g)$ is the length of the period of the sequence $\{g(n)H,\ n\in\mathbb{Z}\}$ (or is infinite, if this sequence is non-periodic). We also define $\mathrm{per}(g)=\mathrm{per}_{\{\mathbf{1}_G\}}(g)$.

1.27. The rest of this section is devoted to the presentation of some rather technical facts which will be needed in Section 2. We start with formulating an important lemma from [Hua] and its easily derivable corollaries:

Lemma. ([Hua], Lemma 2.3) *Let q be a prime number, let $t,r\in\mathbb{N}$ and let $p\in\mathbb{Z}[n]$ with $\deg p\le r$. Assume that not all coefficients of p are divisible by q. Then the number of solutions of the congruence $p(n)\equiv 0\bmod q^t$ on the interval $\{1,\ldots,q^t\}$ does not exceed $Cq^{t-\frac{t}{r}}$, where C depends on r only.*

1.28. Corollary. *Let q be a prime number, let $t,r\in\mathbb{N}$ and let $p\in\mathbb{Z}[n]$ with $\deg p\le r$ and $p(0)=0$. Let $t_1\in\mathbb{N}$ be such that the coefficients of p are divisible by q^{t_1} but not by q^{t_1+1}, and let $t_2=t-t_1$. Then on the interval $\{1,\ldots,q^t\}$, $p(n)$ takes on any value $\bmod q^t$ at most $Cq^{t_1}q^{t_2-\frac{t_2}{r}}=Cq^{t-\frac{t_2}{r}}$ times, where C depends on r only. As a consequence, $p(n)\bmod q^t$ takes on at least $\frac{1}{C}q^{\frac{t_2}{r}}$ distinct values in $\mathbb{Z}_{q^t}$.*

1.29. Corollary. *Let q be a prime number, let $t, r \in \mathbb{N}$ and let $p \in \mathbb{Q}[n]$ be an integral polynomial of degree $\leq r$ satisfying $p(0) = 0$. Assume that the set $\{p(n) \bmod q^t \mid n \in \mathbb{Z}\}$ generates a subring $q^{t_1}\mathbb{Z}_{q^t}$ of $\mathbb{Z}_{q^t}$, and let $t_2 = t - t_1$. Then on any interval of length q^t in $\mathbb{Z}$, $p(n) \bmod q^t$ takes on any value at most $C' q^{t-\frac{t_2}{r}}$ times, and hence takes on at least $\frac{1}{C'} q^{\frac{t_2}{r}}$ distinct values in $\mathbb{Z}_{q^t}$, with C' depending on r only.*

1.30. Let us remind that, given a set $A \subseteq \mathbb{Z}$, *the density of A* is defined as $\lim_{n\to\infty} \frac{|A\cap\{-n,\ldots,n\}|}{2n+1}$ (if it exists).

Let H be a subgroup of a nilpotent group G and let g be a G-polynomial. It follows from Proposition 1.21 that for any $Q \in G$, the set $\{n \in \mathbb{Z} \mid g(n) \in QH\}$ has density (which can be zero).

1.31. The following proposition demonstrates that if g is a G-polynomial whose elements generate G, then, loosely speaking, $\mathrm{per}_H(g)$ is large if and only if $\mathrm{per}_H(G)$ is large, and, in this case, the values of the sequence $g(n)H$ are distributed with some uniformity.

Proposition. *Let G be a nilpotent group.*
(a) *Let H be a subgroup of G and let g be a G-polynomial of degree $\leq r$. Then $\mathrm{per}_H(g) \leq r!\,\mathrm{per}_H(G)$.*
(b) *For any $r, L \in \mathbb{N}$ and $\delta > 0$ there is $K \in \mathbb{N}$ with the following property: if H is a subgroup of G with $\mathrm{per}_H(G) > K$ and g is a G-polynomial of degree $\leq r$ such that $g(0) = \mathbf{1}_G$ and the elements of g generate G, then $\mathrm{per}_H(g) > L$ and for any left coset QH of H the set $\{n \in \mathbb{Z} \mid g(n) \in QH\}$ has density $< \delta$.*

Proof. In fact, part (a) of the proposition was already established in the proof of Proposition 1.21: if $\mathrm{per}_H(G) < \infty$, then $G = \overline{H}$, and so, the sequence $g(n)H$ is periodic with period dividing $r!\,\mathrm{per}_H(G)$.
(b) Choose large enough $K \in \mathbb{N}$ (it will be clear from the proof how large K has to be chosen) and assume that $\mathrm{per}_H(G) > K$. The factor-group $G' = G/(HG_2)$ is abelian and finitely generated, so G' is representable as a finite product of cyclic subgroups: $G' = B_1 \times \ldots \times B_k$, where each B_j is a cyclic group of order d_j, and d_j is either infinite or of the form $d_j = q_j^{t_j}$ with prime q_j. If at least one of d_j is infinite, then $\mathrm{per}_{HG_2}(G) = \mathrm{per}_H(G) = \infty$. Otherwise, $\mathrm{per}(G') = \mathrm{per}_{HG_2}(G)$ is the least common multiple of $d_1, \ldots, d_k$. Let $d_1 = \max\{d_1, \ldots, d_k\}$. Since by Lemma 1.25 $\mathrm{per}_{HG_2}(G) > K^c$, we have $d_1! > K^c$, that is, d_1 is large if K is large.

Now let g be a G-polynomial of degree $\leq r$ satisfying $g(0) = \mathbf{1}_G$. Let P be a generator of B_1. The group B_1 is isomorphic to a factor-group of G, namely, $B_1 \simeq \big(G/(HG_2)\big)/(B_2 \times \ldots \times B_k)$. Hence, we can consider the image of g in B_1, which is a B_1-polynomial and as such has form $P^{p(n)}$, where p is a polynomial $\mathbb{Z} \longrightarrow \mathbb{Z}_{d_1}$ of degree $\leq r$ satisfying $p(0) = 0$.

(If $d_1 = \infty$ we put $\mathbb{Z}_\infty = \mathbb{Z}$.) Since $\{g(n),\ n \in \mathbb{Z}\}$ generates G, $\{p(n),\ n \in \mathbb{Z}\}$ generates $\mathbb{Z}_{d_1}$. By Corollary 1.29, on any interval of length d_1 in $\mathbb{Z}$ the polynomial $p(n)$ takes on at least $\frac{d_1^{1/r}}{C'}$ distinct values in $\mathbb{Z}_{d_1}$, each value at most $C'{d_1}^{1-\frac{1}{r}}$ times. (If $d_1 = \infty$, $p(n)$ takes on infinitely many distinct values in $\mathbb{Z}_{d_1} = \mathbb{Z}$, each value finitely many times.) It follows that $g(n)H$ runs through $\geq \frac{d_1^{1/r}}{C'}$ distinct left cosets of H, and for every $Q \in G$ the set $\bigl\{n \in \mathbb{Z} \bigm| g(n) \in QH\bigr\}$ has density $\leq \frac{C'}{d_1^{1/r}}$. Choosing K so large that $d_1 > (C'/\delta)^r$ and $d_1 > (C'L)^r$, we get the result. ■

1.32. Corollary. *Let G be a nilpotent group, let H be a subgroup of G, let $r,k,b \in \mathbb{N}$ and let g be a G-polynomial of degree $\leq r$. Define $\tilde{g}(n) = g(b+kn)$. If $\mathrm{per}_H(g)$ is large, then $\mathrm{per}_H(\tilde{g})$ is large. (That is: for any r,k and N there is L such that if $\mathrm{per}_H(g) > L$, then $\mathrm{per}_H(\tilde{g}) > N$.)*

Proof. We may assume that $g(0) = \mathbf{1}_G$ and that G is generated by the elements of g. If $\mathrm{per}_H(g)$ is large, then for any $Q \in G$ the density of the set $\bigl\{n \in \mathbb{Z} \bigm| g(n) \in QH\bigr\}$ is small. This implies that the density of the set $\bigl\{n \in \mathbb{Z} \bigm| \tilde{g}(n) \in QH\bigr\}$ is small, and therefore $\mathrm{per}_H(\tilde{g})$ must be large. ■

1.33. Proposition. *Let $r \in \mathbb{N}$, let G be a nilpotent group and let $f\colon \mathbb{Z}^2 \longrightarrow G$ be such that for every $m \in \mathbb{Z}$, $g_m(n) = f(m,n)$ is a G-polynomial of degree $\leq r$, and for every $n \in \mathbb{Z}$, $\hat{g}_n(m) = f(m,n)$ is a G-polynomial of degree $\leq r$. ($f(m,n)$ is "a G-polynomial in two variables".) Also assume that the values of f generate G and that $f(m,0) = f(0,n) = \mathbf{1}_G$ for all $m,n \in \mathbb{Z}$. For any L and $\delta > 0$ there is K such that if H is a subgroup of G with $\mathrm{per}_H(G) > K$, then the set $\bigl\{m \in \mathbb{Z} \bigm| \mathrm{per}_H(g_m) < L\bigr\}$ has density $< \delta$. (K depends on r, L, δ and the nilpotency class of G, but not on f.)*

Proof. Let $G_2 = [G,G]$. By Lemma 1.25, if $\mathrm{per}_H(G)$ is large, then $\mathrm{per}(G/(HG_2))$ is also large. Hence, the finite abelian group $G' = G/(HG_2)$ contains a cyclic subgroup $B = \langle P \rangle$ whose order is either infinite or equals q^t, where q is a prime number and q^t is large. We will consider the second case only. (The analysis of the first case is analogous and, actually, simpler.)

The $\mathbb{Z}^2$-sequence $f(m,n)$ is represented in B by a sequence $P^{p(m,n)}$, where p is a polynomial $\mathbb{Z}^2 \longrightarrow \mathbb{Z}_{q^t}$ of degree $\leq r$ satisfying $p(m,0) = p(0,n) = 0$. Replacing, if needed, p by $(r!)^2 p$, we may assume that p has integer coefficients. Since the values of f generate G, the values of p generate $\mathbb{Z}_{q^t}$. It follows that at least one of the coefficients of p is not divisible by q.

Write $p(m,n) = p_r(m)n^r + \ldots + p_1(m)n$, and assume that it is the polynomial p_j which has a coefficient not divisible by q. Since q^t is a large number, we have two possibilities:
a) q is large: $q > rL$ and $q > r/\delta$. In this case, after factorizing B by the subgroup $\langle P^{q^{t-1}} \rangle$,

we may assume that $t=1$ and $B\simeq\mathbb{Z}_q$. Now, p_j is a nonzero polynomial $\mathbb{Z}\longrightarrow\mathbb{Z}_q$, and hence the set $U=\{m\in\mathbb{Z}\mid p^{(j)}(m)\equiv 0 \bmod q\}$ has density $\leq r/q<\delta$. On the other hand, if $m\notin U$ then $p(m,n)$, as a polynomial in n with values in $\mathbb{Z}_q$, is not identically zero and hence, takes on $\geq q/r>L$ distinct values. Therefore, the sequence $g_m(n)H$ takes on more than L distinct values, that is, $\mathrm{per}_H(g_m)>L$ for all $m\notin U$.
b) t is large: let t_1 be such that $Cq^{-t_1/r}<\delta$, t_2 be such that $\frac{1}{C}q^{t_2/r}>L$, and $t>t_1+t_2$. If m is such that $p_j(m)$ is not divisible by q^{t_1}, then, by Corollary 1.28, $p(m,n)$ as a polynomial in n takes on $>\frac{1}{C}q^{t_2/r}>L$ distinct values. Hence, $\mathrm{per}_H(g_m)>L$ for such m. On the other hand, by the same Corollary 1.28, the set $\{m\in\mathbb{Z}\mid p_j(m)\equiv 0 \bmod q^{t_1}\}$ has density $\leq Cq^{t_1(1-\frac{1}{r})}/q^{t_1}=Cq^{-t_1/r}<\delta$. ∎

1.34. As before, let G be a nilpotent group and g be a G-polynomial of degree $\leq r$ with $g(0)=\mathbf{1}_G$. Consider the mapping $f\colon\mathbb{Z}^2\longrightarrow G$ given by $f(m,n)=g(m)^{-1}g(n)^{-1}g(n+m)$. f is a G-polynomial in two variables, satisfying $f(m,0)=f(0,n)=\mathbf{1}_G$ for any $m,n\in\mathbb{Z}$. Let E be the group generated by the values of f, and let $G_2=[G,G]$.

Lemma. *If the elements of g generate G, then $G_2\subseteq E$. In particular, E is normal in G.*

Proof. For any $m,n\in\mathbb{Z}$ we have

$$\begin{aligned}\big[g(m),g(n)\big]=g(m)^{-1}g(n)^{-1}g(m)g(n)&=g(m)^{-1}g(n)^{-1}g(n+m)g(m+n)^{-1}g(m)g(n)\\&=f(m,n)f(n,m)^{-1}.\end{aligned}$$

∎

1.35. Lemma. *Under the assumptions of* 1.34*, let $K=\mathrm{per}(E)$ and for $b\in\mathbb{Z}$ let $T_b=g(b)^{-1}g(b+K)$. Then for any $n,b\in\mathbb{Z}$, $g(b+Kn)=g(b)T_b^n$.*

Proof. We have $T_b=g(b)^{-1}g(b+K)=Dg(b)Dg(b+1)\dots Dg(b+K-1)$. Thus for any $n\in\mathbb{Z}$,

$$\begin{aligned}g(b+nK)=g(b)\big(Dg(b)&Dg(b+1)\dots Dg(b+K-1)\big)\dots\\&\big(Dg(b+Kn-K)Dg(b+Kn-K+1)\dots Dg(b+Kn-1)\big)=g(b)T_b^n.\end{aligned}$$

∎

1.36. Clearly, $\mathrm{per}_{G_2}(E)\leq\mathrm{per}(E)$. The following proposition demonstrates that, under the notation of 1.34, $\mathrm{per}(E)$ is large if and only if $\mathrm{per}_{G_2}(E)$ is large.

Proposition. *Let G be a nilpotent group. For any K and r there is N such that if g is a G-polynomial of degree $\leq r$ with $g(0)=\mathbf{1}_G$, whose elements generate G and such that the subgroup E generated by $f(m,n)=g(m)^{-1}g(n)^{-1}g(n+m)$, $m,n\in\mathbb{Z}$, satisfies $\mathrm{per}(E)>N$, then $\mathrm{per}_{G_2}(E)>K$.*

Proof. Let $r \in \mathbb{N}$, let g be a G-polynomial of degree $\leq r$ with $g(0) = \mathbf{1}_G$, let the elements of g generate G, let E be the group generated by $g(m)^{-1}g(n)^{-1}g(n+m)$, $m, n \in \mathbb{Z}$, and let $K = \mathrm{per}_{G_2}(E)$. We will show that $\mathrm{per}(E)$ does not exceed some constant which depends on K and r only. Let $L = \mathrm{per}_{G_2}(Dg)$; since $Dg(n) = g(1)f(1,n)$ and $f(1,n)$ is an E-polynomial, Proposition 1.31(a) gives $L < r!K$. Define

$$T = g(L) = Dg(0)Dg(1)\dots Dg(L-1).$$

Take any $b \in \mathbb{Z}$, let $b = a + nL$ with $0 \leq a \leq L-1$. Then

$$\begin{aligned} &g\big(b+L\big) \equiv g(b)Dg(b)\dots Dg(b+L-1) \equiv g(b)Dg(a+nL)\dots Dg(a+nL+L-1) \\ &\equiv g(b)Dg(a)Dg(a+1)\dots Dg(L-1)Dg(0)Dg(1)\dots Dg(a-1) \equiv g(b)Dg(0)\dots Dg(L-1) \\ &\equiv g(b)T \bmod G_2. \end{aligned} \tag{1.3}$$

Denote by H the group generated by G_2 and T. Then (1.3) says that $\mathrm{per}_H(g) \leq L$. Since g generates G, by Proposition 1.31(b) $M = \mathrm{per}_H(G)$ does not exceed some constant depending on L and r only.

Now, let $G = G_1 \subset G_2 \subset \dots \subset G_c \subset G_{c+1} = \{\mathbf{1}_G\}$ be the lower central series of G. Consider an element $S = \big[S_1, [S_2, \dots, [S_{k-1}, S_k]\dots]\big]$ of G_k for $k \geq 2$. Let $S_{k-1}^M = T^{a_1}P_1$ and $S_k^M = T^{a_2}P_2$ for $a_1, a_2 \in \mathbb{Z}$ and $P_1, P_2 \in G_2$. Then

$$S^{M^2} \equiv \big[S_1, [S_2, \dots, [S_{k-1}^M, S_M^K]\dots]\big] \equiv \big[S_1, [S_2, \dots, [T^{a_1}, T^{a_2}]\dots]\big] \equiv \mathbf{1}_G \bmod G_{k+1},$$

that is, $S^{M^2} \in G_{k+1}$. It follows that $\mathrm{per}_{G_{k+1}}(G_k) \leq M^2$ for all $k = 2, \dots, c$ and so, $\mathrm{per}(G_2) \leq M^{2(c-1)}$. Hence, $\mathrm{per}(E) \leq KM^{2(c-1)}$. ■

2. Rational spectrum and rationally-primitive actions

Throughout this section G stands for a finitely generated nilpotent group of unitary operators on a Hilbert space $\mathcal{H}$.

2.1. For a sequence $\{u_n\}_{n\in\mathbb{N}}$ of elements of a Hilbert space *the Cesàro limit* $\mathrm{C}\text{-}\lim_n u_n$ is $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^N u_n$ (if it exists, of course). For a sequence $\{a_n\}_{n\in\mathbb{N}}$ of real numbers we also denote $\limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^N a_n$ by $\mathrm{C}\text{-}\limsup_n a_n$.

Our main tool in this section is the following modification of the *van der Corput lemma* (cf. [B1]):

Lemma. *Let $\{u_n\}_{n\in\mathbb{N}}$ be a bounded sequence of elements of a Hilbert space: $\|u_n\| < b$ for all $n \in \mathbb{N}$. Then $\limsup_{N\to\infty}\big\|\frac{1}{N}\sum_{n=1}^N u_n\big\|^2 \leq \mathrm{C}\text{-}\limsup_m \mathrm{C}\text{-}\limsup_n \mathrm{Re}\langle u_n, u_{n+m}\rangle$. In particular, if $\mathrm{C}\text{-}\lim_n\langle u_n, u_{n+m}\rangle$ exists for every $m \in \mathbb{N}$ and $\mathrm{C}\text{-}\lim_m \mathrm{C}\text{-}\lim_n\langle u_n, u_{n+m}\rangle = 0$, then $\mathrm{C}\text{-}\lim_n u_n = 0$.*

Proof. The proof is standard. Let $a_m = \text{C-}\limsup_n \operatorname{Re}\langle u_n, u_{n+m}\rangle$ for $m \in \mathbb{N}$ and $a = \text{C-}\limsup_m a_m$. Fix $M \in \mathbb{N}$. For any $N \in \mathbb{N}$ we have

$$
\begin{aligned}
\Big\|\frac{1}{N}\sum_{n=1}^{N} u_n\Big\|^2 &= \Big\|\frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{n=1}^{N} u_n\Big\|^2 = \frac{1}{M^2}\Big\|\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} u_n\Big\|^2 \\
&\leq \frac{1}{M^2}\Big\|\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} u_{n+m-1}\Big\|^2 + \frac{b^2}{N^2} \leq \frac{1}{M^2}\frac{1}{N}\sum_{n=1}^{N}\Big\|\sum_{m=1}^{M} u_{n+m-1}\Big\|^2 + \frac{b^2}{N^2} \\
&= \frac{1}{M^2}\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}\|u_{n+m-1}\|^2 + \frac{1}{M^2}\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M-1}\sum_{k=m+1}^{M} 2\operatorname{Re}\langle u_{n+m-1}, u_{n+k-1}\rangle + \frac{b^2}{N^2} \\
&\qquad\leq \frac{b^2}{M} + \frac{2}{M^2}\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M-1}(M-m)\operatorname{Re}\langle u_{n+m-1}, u_{n+k-1}\rangle + \frac{b^2}{N^2}.
\end{aligned}
$$

If N is large, the last expression is smaller than $\frac{b^2}{M} + \frac{2}{M^2}\sum_{m=1}^{M}(M-m)a_m + \delta$, where $\delta > 0$ is arbitrarily small.

Now let $M_0 \in \mathbb{N}$ be such that $\frac{1}{K}\sum_{m=1}^{K} a_m < a + \delta$ for all $K > M_0$. Since $a_m \leq b^2$ for all m, for any $M > M_0$ we can write

$$
\begin{aligned}
\frac{2}{M^2}\sum_{m=1}^{M}(M-m)a_m &= \frac{2}{M^2}\sum_{K=1}^{M}\sum_{m=1}^{K} a_m = \frac{2}{M^2}\Big(\sum_{K=1}^{M_0}\sum_{m=1}^{K} a_m + \sum_{K=M_0+1}^{M}\sum_{m=1}^{K} a_m\Big) \\
&\leq \frac{2}{M^2}\Big(\frac{M_0^2 b^2}{2} + \sum_{K=M_0+1}^{M} K(a+\delta)\Big) \leq \frac{2}{M^2}\Big(\frac{M_0^2 b^2}{2} + \frac{M(M+1)}{2}(a+\delta)\Big).
\end{aligned}
$$

When M is large enough, it is smaller than $a + 2\delta$. ■

2.2. Remark. One can prove in a completely analogous fashion the following uniform version of the van der Corput lemma:

Lemma. *If $\{u_n\}_{n\in\mathbb{N}}$ is a bounded sequence in a Hilbert space, then*

$$
\limsup_{N-K\to\infty}\Big\|\frac{1}{N-K}\sum_{n=K+1}^{N} u_n\Big\|^2 \leq \limsup_{M\to\infty}\frac{1}{M}\sum_{m=1}^{M}\Big(\limsup_{N-K\to\infty}\frac{1}{N-K}\sum_{n=K+1}^{N}\operatorname{Re}\langle u_n, u_{n+m}\rangle\Big).
$$

It is this uniform version of the van der Corput lemma which allows one to obtain (practically without altering the proofs otherwise) the uniform versions of Theorems A – E alluded to in the Introduction.

2.3. A unitary operator T on a Hilbert space $\mathcal{H}$ is called *ergodic* if $Tu \neq u$ for all nonzero $u \in \mathcal{H}$. A unitary operator T is called *totally ergodic* if T^n is ergodic for all nonzero $n \in \mathbb{Z}$. We will say that G is *totally ergodic* if every $T \in G \setminus \{\mathbf{1}_G\}$ is totally ergodic. Notice that if G is totally ergodic it has no torsion: $T^n \neq \mathbf{1}_G$ for any $T \neq \mathbf{1}_G$ and $n \neq 0$.

2.4. Proposition. *If G is totally ergodic, then for any nonconstant G-polynomial g and any $u \in \mathcal{H}$, $\text{C-}\lim_n g(n)u = 0$.*

Proof. The statement is true for linear G-polynomials, that is, for g of the form $g(n) = PT^n$, $T \neq \mathbf{1}_G$: since T is ergodic on $\mathcal{H}$, $\text{C-}\lim_n PT^n u = 0$ for all $u \in \mathcal{H}$. If g is not linear, then for every $m \in \mathbb{Z}$, $m \neq 0$, $g(n)^{-1}g(n+m)$ is a nonconstant G-polynomial of degree $< \deg g$. (Indeed, $Dg(n) = g(n)^{-1}g(n+1)$ is nonconstant in this case. Using the fact that G is torsion-free, one can deduce that the G-polynomial $g(n)^{-1}g(n+m) = Dg(n)Dg(n+1)\ldots Dg(n+m-1)$ is also nonconstant. We omit the proof.) So, we may use induction on the degree of g and write for $u \in \mathcal{H}$, $\text{C-}\lim_n \langle g(n)u, g(n+m)u\rangle = \text{C-}\lim_n \langle u, g(n)^{-1}g(n+m)u\rangle = 0$. By the van der Corput lemma, applied to the sequence $g(n)u$, $\text{C-}\lim_n g(n)u = 0$. ■

2.5. For $u \in \mathcal{H}$, the set $\{T \in G \mid T^m u = u \text{ for some } m \in \mathbb{N}\}$ is a subgroup of G. Indeed, it is the closure of the stabilizer $\{T \in G \mid Tu = u\}$ of u in G.

For a subgroup H of G, we define $\mathcal{H}^{\text{inv}}(H) = \{u \in \mathcal{H} \mid Pu = u \text{ for all } P \in H\}$ and $\mathcal{H}^{\text{rat}}(H) = \{u \in \mathcal{H} \mid Hu \text{ is finite}\}$. Both the subspace of invariant vectors, $\mathcal{H}^{\text{inv}}(H)$, and the rational spectrum subspace, $\mathcal{H}^{\text{rat}}(H)$, are H-invariant. Note that while $\mathcal{H}^{\text{inv}}(H)$ is a closed subspace, $\mathcal{H}^{\text{rat}}(H)$ may not be closed.

Specializing to the subgroup generated by $T \in G$, we will denote $\mathcal{H}^{\text{inv}}(T) = \{u \in \mathcal{H} \mid Tu = u\}$ and $\mathcal{H}^{\text{rat}}(T) = \{u \in \mathcal{H} \mid T^n u = u \text{ for some } n \in \mathbb{N}\}$.

2.6. Lemma. (a) *Let H be a subgroup of G. Then $\mathcal{H}^{\text{inv}}(H) = \bigcap_{P\in H} \mathcal{H}^{\text{inv}}(P)$ and $\mathcal{H}^{\text{rat}}(H) = \bigcap_{P\in H} \mathcal{H}^{\text{rat}}(P)$.*
(b) *Let $T \in G$ and $F = THT^{-1}$. Then $\mathcal{H}^{\text{inv}}(F) = T\big(\mathcal{H}^{\text{inv}}(H)\big)$ and $\mathcal{H}^{\text{rat}}(F) = T\big(\mathcal{H}^{\text{rat}}(H)\big)$.*

Proof. (a) The first statement is obvious, the second follows from Lemma 1.12.

(b) For any $P \in H$ and $u \in \mathcal{H}^{\text{inv}}(H)$ we have $TPT^{-1}(Tu) = Tu$. For $P \in H$ and $v \in \mathcal{H}^{\text{rat}}(H)$, the set $\{(TPT^{-1})^n Tv,\ n \in \mathbb{Z}\} = \{TP^n v,\ n \in \mathbb{Z}\}$ is finite. ■

2.7. Now we assume that G is not totally ergodic on $\mathcal{H}$. We fix a maximal subgroup H of G such that $\mathcal{H}^{\text{rat}}(H)$ is nontrivial; such a subgroup exists by Lemma 1.9.

2.8. Proposition. (a) *H is a closed subgroup of G.*
(b) *For any $T \notin H$ and any nonzero $u \in \mathcal{H}^{\text{rat}}(H)$, the orbit $\{T^n u,\ n \in \mathbb{Z}\}$ is infinite.*

(c) *Every element T of the normalizer $N(H)$ of H preserves $\mathcal{H}^{\mathrm{rat}}(H)$: $T\big(\mathcal{H}^{\mathrm{rat}}(H)\big) = \mathcal{H}^{\mathrm{rat}}(H)$, and for any $T \in N(H) \setminus H$, $\text{C-}\lim_n T^{mn}u = 0$ for all $u \in \mathcal{H}^{\mathrm{rat}}(H)$ and $m \in \mathbb{N}$.*
(d) *For any $T \notin N(H)$, $T\big(\mathcal{H}^{\mathrm{rat}}(H)\big) \perp \mathcal{H}^{\mathrm{rat}}(H)$.*

Proof. (a) and (b) are clear from the definition of H: if $T \in G$ were such that either $T^n \in H$ for some $n \in \mathbb{N}$, or $T^m u = u$ for some $m \in \mathbb{N}$ and nonzero $u \in \mathcal{H}^{\mathrm{rat}}(H)$, then we could add T to H.

If $T \in N(H)$, then for every $P \in H$, $Q = T^{-1}PT \in H$. It follows that for every $u \in \mathcal{H}^{\mathrm{rat}}(H)$ the set $\big\{P^n Tu,\ n \in \mathbb{Z}\big\} = \big\{TQ^n u,\ n \in \mathbb{Z}\big\}$ is finite. Hence $Tu \in \mathcal{H}^{\mathrm{rat}}(H)$, which implies that T preserves $\mathcal{H}^{\mathrm{rat}}(H)$. If $T \in N(H)\backslash H$, then T^m has no nonzero invariant vectors in $\mathcal{H}^{\mathrm{rat}}(H)$ for any $m \in \mathbb{N}$, thus we have $\text{C-}\lim_n T^{mn}u = 0$ for all $u \in \mathcal{H}^{\mathrm{rat}}(H)$. This gives (c).

If $T \notin N(H)$, then for some $P \in H$ one has $Q = T^{-1}PT \notin H$. So, for any $u \in \mathcal{H}^{\mathrm{rat}}(H)$, $u \neq 0$, the set $\{Q^n u,\ n \in \mathbb{Z}\} = \big\{T^{-1}P^n Tu,\ n \in \mathbb{Z}\big\}$ is infinite, and hence $Tu \notin \mathcal{H}^{\mathrm{rat}}(H)$. By Lemma 1.14, $T \in N^k(H) \setminus N^{k-1}(H)$ for some $2 \leq k \leq c$. We will use induction on k to prove that, in fact, $Tu \perp \mathcal{H}^{\mathrm{rat}}(H)$.

First, let $T \in N^2(H) \setminus N(H)$, and let $P \in H$ be such that $Q = T^{-1}PT \notin H$. Since $T \in N\big(N(H)\big)$, $Q \in N(H)$. Thus, for any $v \in \mathcal{H}^{\mathrm{rat}}(H)$ we have

$$\langle Tu, v\rangle = \text{C-}\lim_n \langle P^n Tu, P^n v\rangle = \text{C-}\lim_n \langle TQ^n u, P^n v\rangle = 0,$$

since $P^m v = v$ for some $m \in \mathbb{N}$ and $\text{C-}\lim_n Q^{mn}u = 0$ by (c).

Now assume that for some k, $2 \leq k \leq c-1$, for all $Q \in N^k(H) \setminus N(H)$ one has $Q\mathcal{H}^{\mathrm{rat}}(H) \perp \mathcal{H}^{\mathrm{rat}}(H)$. Since, by Proposition 1.18, $N(H)$ is closed in G, $Q^n \notin N(H)$ for all $n \in \mathbb{Z}$, $n \neq 0$. So, $Q^n u$, $n \in \mathbb{Z}$, are all pairwise orthogonal for distinct n. In particular, $\text{C-}\lim_n Q^{mn}u = 0$ for any $m \in \mathbb{Z}$, $m \neq 0$. Let $T \in N^{k+1}(H) \setminus N(H)$, let $P \in H$ be such that $Q = T^{-1}PT \notin H$. Then $Q \in N^k(H)$ and for any $v \in \mathcal{H}^{\mathrm{rat}}(H)$ we again have

$$\langle Tu, v\rangle = \text{C-}\lim_n \langle P^n Tu, P^n v\rangle = \text{C-}\lim_n \langle TQ^n u, P^n v\rangle = 0.$$

∎

2.9. Proposition. *Let g be a G-polynomial with $g(0) = \mathbf{1}_G$. If $g \in \wp H$ then for any $u \in \mathcal{H}^{\mathrm{rat}}(H)$, the sequence $g(n)u$ is periodic; in particular, $\text{C-}\lim_n g(n)u$ exists. If $g \notin \wp H$, then for any $u \in \mathcal{H}^{\mathrm{rat}}(H)$ one has $\text{C-}\lim_n g(n)u = 0$.*

Proof. The first statement of the proposition is clear from the representation $g(n) = P_1^{p_1(n)} \dots P_r^{p_r(n)}$, $P_1, \dots, P_r \in H$.

Let $g \notin \wp H$. Let $k \in \mathbb{N}$ be such that $g \in \wp N^k(H)$ and $g \notin \wp N^{k-1}(H)$. Consider the mapping $f \colon \mathbb{Z}^2 \longrightarrow G$, $f(m,n) = g(m)^{-1}g(n)^{-1}g(n+m)$. For every $m \in \mathbb{Z}$, $g_m(n) = f(m,n)$ is a G-polynomial. Assume that $g_m \in \wp N^{k-1}(H)$ for infinitely many $m \in \mathbb{Z}$; then, for every $n \in \mathbb{Z}$, the G-polynomial $\hat{g}_n(m) = f(m,n)$ meets $N^{k-1}(H)$ infinitely many times, and since $N^{k-1}(H)$ is closed, $\hat{g}_n(m) \in \wp N^{k-1}(H)$ for all $m \in \mathbb{Z}$. Thus, we have two possibilities: either $g_m(n) \in \wp N^{k-1}(H)$ for only finitely many $m \neq 0$, or $f(m,n) \in N^{k-1}(H)$ for all $m, n \in \mathbb{Z}$.

In the first case, since $\deg\bigl(g(n)^{-1}g(n+m)\bigr) < \deg g$, we may apply induction on the degree of g and assume that $\operatorname{C-lim}_n g(n)^{-1}g(n+m)u = 0$ for all but finitely many $m \in \mathbb{Z}$. Then

$$\operatorname{C-lim}_m \operatorname{C-lim}_n \bigl\langle g(n)u, g(n+m)u \bigr\rangle = \operatorname{C-lim}_m \bigl\langle u, \operatorname{C-lim}_n g(n)^{-1}g(n+m)u \bigr\rangle = 0,$$

and by the van der Corput lemma, $\operatorname{C-lim}_n g(n)u = 0$.

Now let us assume that $f(m,n) = g(m)^{-1}g(n)^{-1}g(n+m) \in N^{k-1}(H)$ for all $m, n \in \mathbb{Z}$. Then $g \colon \mathbb{Z} \longrightarrow N^k(H)/N^{k-1}(H)$ is a homomorphism, and so, if we put $T = g(1)$, then $g(n) = T^n h(n)$, where h is an $N^{k-1}(H)$-polynomial.

Let $u \in \mathcal{H}^{\mathrm{rat}}(H)$. If $k = 1$, that is, h is an H-polynomial, then by Proposition 1.21, $h(n)u$ periodically runs through a finite set of vectors in $\mathcal{H}^{\mathrm{rat}}(H)$. Let m be the period of the sequence $h(n)u$, $n \in \mathbb{Z}$. Since $T \in N(H) \setminus H$, T^m preserves $\mathcal{H}^{\mathrm{rat}}(H)$ and is ergodic on $\mathcal{H}^{\mathrm{rat}}(H)$, we have $\operatorname{C-lim}_n g(n)u = \operatorname{C-lim}_n T^n h(n)u = 0$. If $k > 1$, the vectors $T^n h(n)u$, $n \in \mathbb{Z}$, are pairwise orthogonal, and thus again $\operatorname{C-lim}_n g(n)u = 0$. ∎

2.10. We will now bring some facts pertaining to a "structure theory" of actions of a finitely generated nilpotent group of unitary operators with respect to "subspaces of rational spectrum" of its elements. (For more information about the structure theory see [L3].) Let $\mathcal{F}$ be the set of subgroups of G conjugate with H: $\mathcal{F} = \{THT^{-1},\ T \in G\}$. For $F \in \mathcal{F}$, $F = THT^{-1}$, we have $\mathcal{H}^{\mathrm{rat}}(F) = T\bigl(\mathcal{H}^{\mathrm{rat}}(H)\bigr)$, so by Proposition 2.8, the subspaces $\mathcal{H}^{\mathrm{rat}}(F)$, $F \in \mathcal{F}$, are pairwise orthogonal. Hence, G acts by permutations on the set $\bigl\{\mathcal{H}^{\mathrm{rat}}(F),\ F \in \mathcal{F}\bigr\}$ of pairwise orthogonal subspaces of $\mathcal{H}$. By Proposition 2.8, for every $F \in \mathcal{F}$ the stabilizer of $\mathcal{H}^{\mathrm{rat}}(F)$ is equal to $N(F)$, and the elements of $N(F) \setminus F$ are totally ergodic on $\mathcal{H}^{\mathrm{rat}}(F)$. Since H is a closed subgroup of G, $F \in \mathcal{F}$ is also closed, and by Proposition 1.18, $N(F)$ is closed. That is, for any $T \notin N(F)$, never $T^n \in N(F)$ for $n \neq 0$. Hence the orbit of $\mathcal{H}^{\mathrm{rat}}(F)$ under the action of $T \notin N(F)$ is infinite.

2.11. Definition. An action of a finitely generated nilpotent group of unitary operators on a Hilbert space $\mathcal{H}$ is *rationally-primitive* if there is a class $\mathcal{F}$ of conjugate subgroups of

G so that:

(i) the subspaces $\mathcal{H}^{\mathrm{rat}}(F)$, $F \in \mathcal{F}$, are pairwise orthogonal and span $\mathcal{H}$; elements of G permute the subspaces $\mathcal{H}^{\mathrm{rat}}(F)$, $F \in \mathcal{F}$, and the stabilizer of $\mathcal{H}^{\mathrm{rat}}(F)$ under this action is $N(F)$;

(ii) for every $F \in \mathcal{F}$, every $T \in N(F) \setminus F$ is totally ergodic on $\mathcal{H}^{\mathrm{rat}}(F)$;

(iii) for every $F \in \mathcal{F}$ and $T \notin N(F)$, $T^n \mathcal{H}^{\mathrm{rat}}(F) \neq \mathcal{H}^{\mathrm{rat}}(F)$ for all $n \in \mathbb{Z}$, $n \neq 0$.

2.12. The above considerations give us the following structure theorem:

Theorem. *For any unitary action of a finitely generated nilpotent group G on a Hilbert space $\mathcal{H}$, $\mathcal{H}$ is decomposable into the direct sum of pairwise orthogonal G-invariant subspaces with rationally-primitive action of G on each of them.*

Indeed, under the notation in 2.10 above, let $\mathcal{H}'$ be the closure in $\mathcal{H}$ of $\bigoplus_{F \in \mathcal{F}} \mathcal{H}^{\mathrm{rat}}(F)$. Then $\mathcal{H}'$ is a G-invariant subspace of $\mathcal{H}$, and the action of G on $\mathcal{H}'$ is rationally primitive. (Note that a totally ergodic action of G is also rationally-primitive: here $H = \{\mathbf{1}_G\}$ and $\mathcal{H}^{\mathrm{rat}}(H) = \mathcal{H}$.) Since the orthogonal complement of $\mathcal{H}'$ in $\mathcal{H}$ is G-invariant, an application of Zorn's lemma gives the result.

2.13. Example. Here is an example of a rationally-primitive action of a nilpotent group. Take for $\mathcal{H}$ the space $l_2(\mathbb{C})$. Let $\{u_j\}_{j \in \mathbb{Z}}$ be the standard basis in $\mathcal{H}$. Put $\mathcal{H}_j = \mathbb{C}u_j$, $j \in \mathbb{Z}$.

Let $\lambda = e^{2\pi i \alpha}$, where $\alpha \in \mathbb{R}$ is irrational. Define unitary operators P, T and S on $\mathcal{H}$ by $Pu_j = \lambda u_j$, $Tu_j = \lambda^j u_j$ and $Su_j = u_{j-1}$, $j \in \mathbb{Z}$. Then T and S commute with P and $[T, S] = P$, thus the group G generated by P, T and S is nilpotent of class 2.

Let $H = \langle T \rangle$ and $\mathcal{F}$ be the class of subgroups of G conjugate with H:

$$\mathcal{F} = \{H_k \mid H_k = S^k H S^{-k},\ k \in \mathbb{Z}\}.$$

For $k \in \mathbb{Z}$ put $T_k = S^k T S^{-k}$. Then $H_k = \langle T_k \rangle$, $T_k(\mathcal{H}_j) = \mathcal{H}_j$ for all $j \in \mathbb{Z}$, and $T_k\big|_{\mathcal{H}_j} = P^{j-k}\big|_{\mathcal{H}_j}$, $j \in \mathbb{Z}$. Thus $\mathcal{H}^{\mathrm{rat}}(H_k) = \mathcal{H}^{\mathrm{rat}}(T_k) = \mathcal{H}^{\mathrm{inv}}(T_k) = \mathcal{H}_k$. The normalizer $N(H_k)$ of H_k is generated by T_k and P, P preserves $\mathcal{H}_k$ and is totally ergodic on $\mathcal{H}_k$. Also, any $S^a T^b P^c \in G \setminus N(H_k)$, that is, with $a \neq 0$, maps $\mathcal{H}_k$ onto $\mathcal{H}_{k+a}$.

2.14. Proposition. *Let the action of G on $\mathcal{H}$ be rationally-primitive and let $\mathcal{F}$ be a class of conjugate subgroups of G described in 2.11. Then for any subgroup $E \subseteq G$, $\overline{\mathcal{H}^{\mathrm{rat}}(E)} = \overline{\bigoplus_{E \subseteq F \in \mathcal{F}} \mathcal{H}^{\mathrm{rat}}(F)}$.*

Proof. In light of Lemma 2.6, it suffices to consider the case $E = \langle T \rangle$, $T \in G$. Let $u \in \mathcal{H}$, $u = \sum_{F \in \mathcal{F}} u_F$ with $u_F \in \overline{\mathcal{H}^{\mathrm{rat}}(F)}$. Since T permutes the spaces $\mathcal{H}^{\mathrm{rat}}(F)$, $F \in \mathcal{F}$, u belongs to $\mathcal{H}^{\mathrm{rat}}(T)$ if and only if all u_F, $F \in \mathcal{F}$, are in $\mathcal{H}^{\mathrm{rat}}(T)$. By Proposition 2.8, $u_F \in \mathcal{H}^{\mathrm{rat}}(T)$ if and only if $T \in F$. So, $\overline{\mathcal{H}^{\mathrm{rat}}(T)} = \overline{\bigoplus_{T \in F \in \mathcal{F}} \mathcal{H}^{\mathrm{rat}}(F)}$. ∎

Remark. One can actually show that $\mathcal{H}^{\mathrm{rat}}(E)=\bigoplus_{E\subseteq F\in\mathcal{F}}\mathcal{H}^{\mathrm{rat}}(F)$ (see [L3]).

2.15. As another corollary, we obtain the following "nil"-polynomial generalization of the von Neumann theorem:

Theorem C. *For every G-polynomial g and every $u\in\mathcal{H}$, $\text{C-}\lim_n g(n)u$ exists.*

Proof. We may assume that the action of G on $\mathcal{H}$ is rationally-primitive and, moreover, that $u\in\mathcal{H}^{\mathrm{rat}}(F)$ for some $F\in\mathcal{F}$. After replacing g by $g(0)^{-1}g$, we may also assume that $g(0)=\mathbf{1}_G$. But then $\text{C-}\lim_n g(n)u$ exists by Proposition 2.9. ■

2.16. Let g be a G-polynomial with $g(0)=\mathbf{1}_G$ and let E be the group generated by elements of g: $E=\langle g(n),\ n\in\mathbb{Z}\rangle$. We define $\mathcal{H}^{\mathrm{rat}}(g)=\mathcal{H}^{\mathrm{rat}}(E)$; it is a (not necessarily closed) subspace of $\mathcal{H}$, invariant under the action of elements of g.

2.17. Theorem. *Let $u\in\mathcal{H}$ and let v be the orthogonal projection of u onto the closure of $\mathcal{H}^{\mathrm{rat}}(g)$. Then $\text{C-}\lim_n g(n)u=\text{C-}\lim_n g(n)v$. In particular, $\text{C-}\lim_n g(n)u\in\overline{\mathcal{H}^{\mathrm{rat}}(g)}$.*

Proof. We may assume that the action of G on $\mathcal{H}$ is rationally-primitive; let $\mathcal{F}$ be the corresponding class of conjugate subgroups of G. Let $E=\langle g(n),\ n\in\mathbb{Z}\rangle$. Then for $F\in\mathcal{F}$, $g\in\wp F$ if and only if $E\subseteq F$. It follows from Proposition 2.9 and Proposition 2.14 that $\text{C-}\lim_n g(n)u=0$ for any $u\perp\mathcal{H}^{\mathrm{rat}}(E)$. ■

2.18. Let $u\in\mathcal{H}$. We define *the period of G with respect to u*, $\mathrm{per}_u(G)$, as a minimal $d\in\mathbb{N}$ such that $T^d u=u$ for all $T\in G$ (if such d exists). Let G_u be the stabilizer of u under the action of G: $G_u=\{P\in G \mid Pu=u\}$. Then, clearly, $\mathrm{per}_u(G)=\mathrm{per}_{G_u}(G)$ (see 1.22). Similarly, for a G-polynomial g we define $\mathrm{per}_u(g)=\mathrm{per}_{G_u}(g)$ (see 1.26).

2.19. Proposition. *For every $r\in\mathbb{N}$ and $\varepsilon>0$ there is $M\in\mathbb{N}$ such that if g is a G-polynomial of degree $\le r$ with $\mathrm{per}_u(g)>M$ for all $u\in\mathcal{H}$, then for any $u\in\mathcal{H}$ one has $\|\text{C-}\lim_n g(n)u\|\le\varepsilon\|u\|$.*

We may assume that $g(0)=\mathbf{1}_G$ and that G is generated by the elements of g. Then by Proposition 1.31, $\mathrm{per}_u(G)$ is large if and only if $\mathrm{per}_u(g)$ is large. Thus, Proposition 2.19 can be reformulated in the following form:

2.20. Proposition. *For every $r\in\mathbb{N}$ and $\varepsilon>0$ there is $L\in\mathbb{N}$ such that if $\mathrm{per}_u(G)>L$ for all $u\in\mathcal{H}$ and g is a G-polynomial of degree $\le r$ with $g(0)=\mathbf{1}_G$ and such that the elements of g generate G, then for any $u\in\mathcal{H}$ one has $\|\text{C-}\lim_n g(n)u\|\le\varepsilon\|u\|$.*

Proof. Take any $u\in\mathcal{H}$; by Theorem 2.17 we may assume that $u\in\mathcal{H}^{\mathrm{rat}}(g)$. We may also assume that $\mathcal{H}=\mathrm{Span}(Gu)$ and so, is finite-dimensional. Furthermore, we may assume that $\mathcal{H}$ is a minimal G-invariant space. Then $\mathcal{H} = \mathrm{Span}(Gv)$ for any $v \in \mathcal{H}$.

For every $v \in \mathcal{H}$, let G_v be the stabilizer of v under the action of G. For any $S \in G$ we have $G_{Sv} = SG_vS^{-1}$. Since $\mathcal{H}$ is spanned by the elements Sv, $S \in G$, the group $\bigcap_{S\in G} SG_vS^{-1}$ acts trivially on $\mathcal{H}$. Let us factorize G by the subgroup which acts trivially on $\mathcal{H}$. Then $\bigcap_{S\in G} SG_vS^{-1} = \{\mathbf{1}_G\}$ for any $v \in \mathcal{H}$. Since $S\big(G_v/(G_2 \cap G_v)\big)S^{-1} = G_v/(G_2 \cap G_v)$, it follows that $G_v \subseteq G_2$ for all $v \in \mathcal{H}$.

Define $f(m,n) = g(m)^{-1}g(n)^{-1}g(n+m)$, $m,n \in \mathbb{Z}$. For every $m \in \mathbb{Z}$, let $g_m(n) = f(m,n)$; g_m is a G-polynomial of degree $\leq r-1$ with $g_m(0) = \mathbf{1}_G$. Let E be the subgroup of G generated by the values of f; by Lemma 1.34, $G_2 \subseteq E$.

We will prove the proposition by induction on r. If $r \leq 1$, g has form $g(n) = T^n$, $T \in G$, and if $\mathrm{per}_v(g) \geq 2$ for $v \in \mathcal{H}$, then $Tv \neq v$. Hence, if $\mathrm{per}_v(g) \geq 2$ for all $v \in \mathcal{H}$, then T is ergodic on $\mathcal{H}$, and $\operatorname*{C-lim}_n g(n)v = \operatorname*{C-lim}_n T^n v = 0$ for any $v \in \mathcal{H}$.

Now assume that Propositions 2.19 and 2.20 are true for $r-1$, and let M' be the constant corresponding to $r-1$ in Proposition 2.19. This means that if h is a G-polynomial of degree $\leq r-1$ and $\mathrm{per}_v(h) > M'$ for all $v \in \mathcal{H}$, then $\big\|\operatorname*{C-lim}_n h(n)v\big\| \leq \varepsilon\|v\|$ for any $v \in \mathcal{H}$. Let K be such that, by Proposition 1.33, applied to f and E defined above, if H is a subgroup of E with $\mathrm{per}_H(E) > K$, then the set $\big\{m \in \mathbb{Z} \bigm| \mathrm{per}_H(g_m) < M'\big\}$ has density $< \varepsilon$. Finally, by Proposition 1.36 we can choose N such that if $\mathrm{per}(E) > N$ then $\mathrm{per}_{G_2}(E) > K$.

Consider two cases. First, assume that $\mathrm{per}(E) > N$. Then by the choice of N, for all $v \in \mathcal{H}$ and for all $m \in \mathbb{Z}$ but a set of density $< \varepsilon$, we have $\mathrm{per}_{G_v}(g_m) \geq \mathrm{per}_{G_2}(g_m) > M'$ and so, $\big\|\operatorname*{C-lim}_n g_m(n)v\big\| \leq \varepsilon\|v\|$. Thus,

$$\begin{aligned}
&\operatorname*{C-lim\,sup}_m \big|\operatorname*{C-lim}_n \langle g(n)v, g(n+m)v\rangle\big| = \operatorname*{C-lim\,sup}_m \big|\operatorname*{C-lim}_n \langle v, g(n)^{-1}g(n+m)v\rangle\big| \\
&\quad \leq \|v\| \operatorname*{C-lim\,sup}_m \big\|\operatorname*{C-lim}_n g(n)^{-1}g(n+m)v\big\| = \|v\| \operatorname*{C-lim\,sup}_m \big\|\operatorname*{C-lim}_n g_m(n)v\big\| \leq 2\varepsilon\|v\|^2,
\end{aligned}$$

which implies $\big\|\operatorname*{C-lim}_n g(n)u\big\| \leq \sqrt{2\varepsilon}\|v\|$ by the van der Corput lemma.

Now consider the case $\mathrm{per}(E) = K \leq N$. Fix $b \in \mathbb{Z}$ and put $T_b = g(b)^{-1}g(b+K)$. Then by Lemma 1.35 we have $g(b+Kn) = g(b)T_b^n$ for all $n \in \mathbb{Z}$. If $T_bv = v$ for some v, then $g(b+Kn)v = g(b)v$ for all $n \in \mathbb{Z}$, which is impossible by Corollary 1.32 if $\mathrm{per}_v(g)$ is large enough. Thus, if $\mathrm{per}_g(v)$ is large enough for all $v \in \mathcal{H}$, then T_b is ergodic on $\mathcal{H}$ and so, $\operatorname*{C-lim}_n g(b+Kn)v = g(b)\operatorname*{C-lim}_n T^n v = 0$ for all $v \in \mathcal{H}$. Since it is true for all $b \in \mathbb{Z}$, we obtain

$$\operatorname*{C-lim}_n g(n)v = \frac{1}{K}\sum_{b=0}^{K-1} \operatorname*{C-lim}_n g(b+Kn)v = 0$$

for all $v \in \mathcal{H}$. ∎

2.21. Given a subgroup E of G and $l \in \mathbb{N}$, we define

$$\mathcal{H}^{(l)}(E) = \big\{u \in \mathcal{H} \bigm| P^l u = u \text{ for all } P \in E\big\} = \big\{u \in \mathcal{H} \bigm| \operatorname{per}_u(E) \text{ divides } l\big\}.$$

Lemma. *Let E be a subgroup of G and let $l \in \mathbb{N}$. Then $\mathcal{H}^{(l)}(E)$ is a closed E-invariant subspace of $\mathcal{H}$. If E is normal in G, then $\mathcal{H}^{(l)}(E)$ is also G-invariant.*

Proof. It is clear that the subspace $\mathcal{H}^{(l)}(E)$ is closed. Assume that T normalizes E: $T^{-1}ET = E$. Then for any $u \in \mathcal{H}^{(l)}(E)$ and $P \in E$ we have

$$P^l(Tu) = T(T^{-1}P^lT)u = T(T^{-1}PT)^l u = Tu.$$

■

2.22. Lemma. *For any subgroup E of G, $\mathcal{H}^{\mathrm{rat}}(E) = \bigcup_{l\in\mathbb{N}} \mathcal{H}^{(l)}(E)$.*

Proof. It is clear that $\mathcal{H}^{\mathrm{rat}}(E) \subseteq \bigcup_{l\in\mathbb{N}} \mathcal{H}^{(l)}(E)$. Let $u \in \mathcal{H}^{(l)}(E)$ and let G_u be the stabilizer of u. Then $P^l \in G_u$ for all $P \in E$. Since E, as a subgroup of G, is finitely generated and nilpotent, by Lemma 1.12 $G_u \cap E$ has finite index in E. It follows that Eu is finite and hence $u \in \mathcal{H}^{\mathrm{rat}}(E)$. ■

2.23. Let $g \in \wp G$ with $g(0) = \mathbf{1}_G$. For $l \in \mathbb{N}$, we define $\mathcal{H}^{(l)}(g) = \mathcal{H}^{(l)}(E)$, where E is the subgroup of G generated by the elements of g. We then have $\mathcal{H}^{\mathrm{rat}}(g) = \bigcup_{l\in\mathbb{N}} \mathcal{H}^{(l)}(g)$.

Theorem 2.17 tells us that for any vector $u \in \mathcal{H}$, the limit $\operatorname{C-lim}_n g(n)u$ lies in $\overline{\mathcal{H}^{\mathrm{rat}}(g)} = \overline{\bigcup_{l\in\mathbb{N}} \mathcal{H}^{(l)}(g)}$. It follows now from Proposition 2.19 that the main contribution to $\operatorname{C-lim}_n g(n)u$ is made by the components of u lying in $\mathcal{H}^{(l)}(g)$ with small l:

Theorem D. *For every $r \in \mathbb{N}$ and $\varepsilon > 0$ there is $L \in \mathbb{N}$ such that if g is a G-polynomial of degree $\leq r$ with $g(0) = \mathbf{1}_G$, and $u \in \mathcal{H}$ is such that $u \perp \mathcal{H}^{(l)}(g)$ for all $l \leq L$, then $\big\|\operatorname{C-lim}_n g(n)u\big\| \leq \varepsilon\|u\|$.*

Proof. We may assume that the elements of g generate G. Then the G-invariant space $\big(\operatorname{Span}\bigcup_{l=1}^{L} \mathcal{H}^{(l)}(g)\big)^{\perp}$ satisfies the assumptions of Proposition 2.20. ■

3. Existence of $\operatorname{C-lim}_n T^n u S^n v$ and joint ergodicity of two transformations

3.1. Throughout this section $\mathbf{X} = (X, \mathcal{B}, \mu)$ is a measure space with $\mu(X) = 1$. We will denote the Hilbert space $L^2(\mathbf{X})$ by $\mathcal{H}$ and identify the subspace of constants in $\mathcal{H}$ with $\mathbb{C}$. A measure preserving transformation of $\mathbf{X}$ induces a unitary operator on $\mathcal{H}$; by conventional abuse of notation, we will denote the transformation and the corresponding operator by the same symbol.

Let us now remark that without loss of generality we may assume that the measure space $\mathbf{X}$ is Lebesgue. This will allow us to freely use the measure theoretical apparatus developed in [F1], [Z1], [Z2] and [F3]. To see that we indeed can make such an assumption, we observe that, first, we clearly can assume that μ is non-atomic, and, second, given measure preserving transformations T and S of $\mathbf{X}$ and functions $u, v \in L^2(\mathbf{X})$, we can pass, if needed, to a T, S-invariant separable subalgebra of $\mathcal{B}$ with respect to which all the functions $T^n u, S^n v$, $n \in \mathbb{Z}$, are measurable. It remains to quote a well known theorem of Carathéodory (see, for example [Roy], Ch. 15, Theorem 4) which states that any separable atomless measure algebra $(X, \mathcal{B}, \mu)$ with $\mu(X) = 1$ is isomorphic to the measure algebra $\mathcal{L}$ induced by the Lebesgue measure on the unit interval. This isomorphism carries T and S into Lebesgue-measure preserving isomorphisms of $\mathcal{L}$, which by the classical von Neumann Theorem ([Roy], Ch. 15, Theorem 20) admit realization as point mappings.

3.2. A sub-σ-algebra $\mathcal{D}$ of $\mathcal{B}$ gives rise to *a factor* $\mathbf{Y} = (X, \mathcal{D}, \mu)$ *of* $\mathbf{X}$. The Hilbert space $L^2(\mathbf{Y}) = L^2(X, \mathcal{D}, \mu)$ is a closed subspace of $\mathcal{H}$. We denote by $E(\cdot|\mathbf{Y})$ the orthogonal projection from $L^2(\mathbf{X})$ onto $L^2(\mathbf{Y})$. $E(\cdot|\mathbf{Y})$ maps $L^\infty(\mathbf{X})$ onto $L^\infty(\mathbf{Y})$ and is extendible by continuity to a mapping $L^1(\mathbf{X}) \longrightarrow L^1(\mathbf{Y})$ ([F3], section 5.3). It is also clear that if u is a nonnegative function, then $E(u|\mathbf{Y})$ is nonnegative as well.

It is sometimes useful to interpret a factor $\mathbf{Y}$ as an underlying measure space $\mathbf{Y} = (Y, \mathcal{D}, \nu)$ provided with a mapping $\eta\colon X \longrightarrow Y$ satisfying $\mu\big(\eta^{-1}(B)\big) = \nu(B)$ for all $B \in \mathcal{D}$. Under the assumption that $(X, \mathcal{B}, \mu)$ is a *regular* space, there is *a decomposition of* μ *with respect to* Y, namely, a system of measures μ_y, $y \in Y$, with $\mu = \int \mu_y d\nu$ (see [F3], section 5.4). Then the projection $E(\cdot|\mathbf{Y})$ can be represented via the disintegration with respect to μ_y: for $u \in L^2(\mathbf{X})$, $E(u|\mathbf{Y})(y) = \int u d\mu_y$ for almost all $y \in Y$.

3.3. Let $\mathbf{Y} = (X, \mathcal{D}, \mu)$ be a factor of $\mathbf{X}$. The space $\mathcal{H} = L^2(\mathbf{X})$ has the structure of a module over the ring $L^\infty(\mathbf{Y})$. A closed subspace of $\mathcal{H}$ which is invariant under multiplication by functions from $L^\infty(\mathbf{Y})$ will be called *a submodule of* $\mathcal{H}$ *(over* Y*)*. A submodule $\mathcal{M}$ of $\mathcal{H}$ is said to be *finite-dimensional* if there are $u_1, \ldots, u_k \in \mathcal{M}$ such that the set $\big\{\varphi_1 u_1 + \ldots + \varphi_k u_k \bigm| \varphi_1, \ldots, \varphi_k \in L^\infty(\mathbf{Y})\big\}$ is dense in $\mathcal{M}$.

3.4. Let T be a measure preserving transformation of $\mathbf{X} = (X, \mathcal{B}, \mu)$. A factor $\mathbf{Y} = (X, \mathcal{D}, \mu)$ of $\mathbf{X}$ is said to be *T-invariant* if $\mathcal{D}$ is a T-invariant sub-σ-algebra of $\mathcal{B}$.

Let $\mathcal{M}$ be a T-invariant submodule of $\mathcal{H}$ over $\mathbf{Y}$. T is said to be *weakly mixing on* $\mathcal{M}$ *relative to* $\mathbf{Y}$ if for any $u, v \in \mathcal{M}$, the sequence of $L^1(\mathbf{Y})$-functions $E(uT^n v|\mathbf{Y})$, $n = 1, 2, \ldots$, converges to zero in density. (A sequence $\{u_n\}_{n\in\mathbb{N}}$ in a topological space converges to a point u *in density* if for any neighborhood U of u, the set $\#\big\{n \in \mathbb{N} \bigm| u_n \in U\big\}$ has density 1, that is, $\#\big\{n \in \{1, \ldots, N\} \bigm| u_n \in U\big\}/N \underset{N\to\infty}{\longrightarrow} 1$. If $\{u_n\}_{n\in\mathbb{N}}$ is a bounded sequence of vectors in a Hilbert space, then the convergence of u_n to zero in density is equivalent to $\operatorname{C-lim}_n \|u_n\| = 0$.) T is said to have on $\mathcal{M}$ *relatively discrete spectrum over* $\mathbf{Y}$

(or to be *compact on* $\mathcal{M}$ *relative to* $\mathbf{Y}$) if $\mathcal{M}$ is spanned by finite-dimensional T-invariant submodules.

Let $\mathbf{X}' = (X, \mathcal{B}', \mu)$ be an intermediate T-invariant factor of $\mathbf{X}$: $\mathcal{D} \subseteq \mathcal{B}' \subseteq \mathcal{B}$. Then T is said to have on $\mathbf{X}'$ *relatively discrete spectrum over* $\mathbf{Y}$ if T has relatively discrete spectrum over $\mathbf{Y}$ on the submodule $L^2(\mathbf{X}')$ of $\mathcal{H}$, and T is said to be *weakly mixing on* $\mathbf{X}'$ *relative to* $\mathbf{Y}$ if T is weakly mixing relative to $\mathbf{Y}$ on *the orthogonal complement of* $L^2(\mathbf{Y})$ *in* $L^2(\mathbf{X}')$.

3.5. We will need the following structure theorem:

Theorem. (Cf. [Z2], Theorem 7.3 and Corollary 7.10, and [F3], Lemma 7.3.) *Let T be a measure preserving transformation of $\mathbf{X}$ and let $\mathbf{Y}$ be a T-invariant factor of $\mathbf{X}$. Then $\mathcal{H}$ is the direct sum $\mathcal{H}^{\mathrm{wm}/Y}(T) \oplus \mathcal{H}^{\mathrm{ds}/Y}(T)$ of orthogonal T-invariant submodules such that T is weakly mixing on $\mathcal{H}^{\mathrm{wm}/Y}(T)$ relative to $\mathbf{Y}$ and T has on $\mathcal{H}^{\mathrm{ds}/Y}(T)$ relatively discrete spectrum over $\mathbf{Y}$. The space $\mathcal{H}^{\mathrm{ds}/Y}(T)$ contains $L^2(\mathbf{Y})$ and corresponds to a T-invariant factor $\mathbf{X}'$ of $\mathbf{X}$: $\mathcal{H}^{\mathrm{ds}/Y}(T) = L^2(\mathbf{X}')$.*

3.6. We will also use the following proposition:

Proposition. (Cf. [CL1], the proof of Proposition 5.) *Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be factors of $\mathbf{X}$ and let $\mathbf{Y}$ be a common factor of $\mathbf{X}_1$ and $\mathbf{X}_2$. Let T be a measure preserving transformation of $\mathbf{X}_1$ and S be a measure preserving transformation of $\mathbf{X}_2$ such that $\mathbf{Y}$ is T- and S-invariant and T and S coincide on $\mathbf{Y}$: $T\varphi = S\varphi$ for all $\varphi \in L^2(\mathbf{Y})$. Also assume that T and S have on $\mathbf{X}_1$ and, respectively, on $\mathbf{X}_2$ relatively discrete spectrum over $\mathbf{Y}$. Then for any $u \in L^\infty(\mathbf{X}_1)$ and $v \in L^\infty(\mathbf{X}_2)$, $\mathrm{C}\text{-}\lim_n T^n u S^n v$ exists in $L^2(\mathbf{X})$.*

Proof. Since the expression $T^n u S^n v$ is linear in u and v, we may assume that u lies in a finite-dimensional T-invariant submodule $\mathcal{M}$ of $\mathcal{H}$ over $\mathbf{Y}$, and v lies in a finite-dimensional S-invariant submodule $\mathcal{N}$ of $\mathcal{H}$ over $\mathbf{Y}$. We also may assume that the action of T $(= S)$ on $\mathbf{Y}$ is ergodic (otherwise, we can deal with the members of the ergodic decomposition of $\mathbf{Y}$ and induced decomposition of $\mathbf{X}$). Under this assumption, $\mathcal{M}$ possesses *an orthonormal basis over* $\mathbf{Y}$: a finite system of functions $u_1, \ldots, u_k \in L^\infty(\mathbf{X})$ spanning $\mathcal{M}$ as a module over $L^\infty(\mathbf{Y})$ and satisfying $E(u_i \overline{u_t} | \mathbf{Y}) = \delta_{i,t}$, $i, t = 1, \ldots, k$ (see [Z2] or [L2]). Let $v_1, \ldots, v_l$ be an orthonormal basis of $\mathcal{N}$ over $\mathbf{Y}$. Since for $\varphi \in L^2(\mathbf{Y})$, $S(\varphi v) = S\varphi S v = T\varphi S v$, it suffices to prove that $\mathrm{C}\text{-}\lim_n T^n \varphi T^n u_i S^n v_j$ exists for all i, j and all $\varphi \in L^2(\mathbf{Y})$.

Let $Tu_i(x) = \sum_{t=1}^k a_{i,t}(x) u_t(x)$, $i = 1, \ldots, k$, with $a_{i,t} \in L^\infty(\mathbf{Y})$, $i, t = 1, \ldots, k$. Let $A(x) = \big(a_{i,t}(x)\big)_{i,t=1}^k$. Since the operator induced by T on $L^2(\mathbf{X})$ is unitary, A is a unitary matrix: $AA^* = I$. Iterating the relation $T \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix} = A(x) \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix}$ one gets $T^n \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix} =$

$A(T^n x)\dots A(Tx)A(x)\begin{pmatrix}u_1\\ \vdots\\ u_k\end{pmatrix}$. Define $A^{(n)}(x)=A(T^n x)\dots A(Tx)A(x)$, $n\in\mathbb{N}$. Similarly, define a unitary matrix $B(x)=\big(b_{j,s}(x)\big)_{j,s=1}^{l}$ with entries from $L^\infty(\mathbf{Y})$ by $S\begin{pmatrix}v_1\\ \vdots\\ v_l\end{pmatrix}=B(x)\begin{pmatrix}v_1\\ \vdots\\ v_l\end{pmatrix}$, and put $B^{(n)}(x)=B(S^n x)\dots B(Sx)B(x)=B(T^n x)\dots B(Tx)B(x)$, $n\in\mathbb{N}$. Then

$$T^n u_i S^n v_j=\sum_{t=1}^{k}\sum_{s=1}^{l}a_{i,t}^{(n)}b_{j,s}^{(n)}u_t v_s,$$

where $a_{i,t}^{(n)},b_{j,s}^{(n)}\in L^\infty(\mathbf{Y})$ are the entries of the matrices $A^{(n)}$ and $B^{(n)}$ respectively. Thus we have only to prove that the sequence $(T^n\varphi)a_{i,t}^{(n)}b_{j,s}^{(n)}$ has C-limit in $L^2(\mathbf{Y})$ for all i,t,j,s and all $\varphi\in L^2(\mathbf{Y})$. We will do this by considering "an abstract model" of the action of $T\times S$ on $\mathcal{M}\otimes\mathcal{N}$. Let U be a k-dimensional vector space with orthonormal basis $e_1,\dots,e_k$, and let V be an l-dimensional vector space with orthonormal basis $e'_1,\dots,e'_l$. Consider the Hilbert space $W=L^2(\mathbf{Y})\otimes U\otimes V$ and define an operator R on W by

$$R(\varphi\otimes e_i\otimes e'_j)(x)=\sum_{t=1}^{k}\sum_{s=1}^{l}\varphi(Tx)a_{i,t}(x)b_{j,s}(x)\otimes e_t\otimes e'_s,$$
$$i=1,\dots,k,\ j=1,\dots,l,\ \varphi\in L^2(\mathbf{Y}).$$

Then R is a unitary operator on W, and $R^n(\varphi\otimes e_i\otimes e'_j)=\sum_{t=1}^{k}\sum_{s=1}^{l}(T^n\varphi)a_{i,t}^{(n)}b_{j,s}^{(n)}\otimes e_t\otimes e'_s$. Since by the ergodic theorem $\operatorname{C-lim}_n R^n(\varphi\otimes e_i\otimes e'_j)$ exists in W, the sequence $(T^n\varphi)a_{i,t}^{(n)}b_{j,s}^{(n)}$ has C-limit for all t,s, and we are done. ∎

3.7. We are now in position to prove our main result:

Theorem A. *Let T and S be measure preserving transformations of $\mathbf{X}$ generating a nilpotent group G. For any $u,v\in L^\infty(\mathbf{X})$, $\operatorname{C-lim}_n T^n uS^n v$ exists in $L^2(\mathbf{X})$.*

Proof. Define $g(n)=T^{-n}S^n$, $n\in\mathbb{Z}$. Then g is a G-polynomial with $\deg g\le c$ (to see this, note that the vector degree (see 1.4) of the G-polynomials T^n, S^n and so, of g does not exceed $(1,2,\dots,c)$) and $g(0)=\mathbf{1}_G$. Let H be the subgroup of G generated by g; one can check that

$$H=\big\{T^{a_1}S^{a_2}\dots T^{a_{t-1}}S^{a_t}\ \big|\ a_1+\dots+a_t=0\big\},$$

and that H is normal in G.

Let $u,v\in L^\infty(\mathbf{X})$ and let M be such that ess-sup$|u|$, ess-sup$|v|<M$. Fix $\varepsilon>0$. We will show that the sequence $\frac{1}{N}\sum_{n=1}^{N}T^n uS^n v$, $N\in\mathbb{N}$, is $4M^2\sqrt{\varepsilon}$-close to a sequence converging in $\mathcal{H}=L^2(\mathbf{X})$. Since ε is arbitrary, the theorem will follow.

Let L be the number whose existence, given ε and $r = c$, is guaranteed by Theorem 2.23. Define

$$\mathcal{H}' = \mathcal{H}^{(L!)}(H) = \{u \in \mathcal{H} \mid P^{L!}u = u \text{ for all } P \in H\}. \tag{3.1}$$

Since H is normal in G, by Lemma 2.21 $\mathcal{H}'$ is a G-invariant subspace of $\mathcal{H}$. Let $\mathcal{D}$ be the σ-algebra generated by functions from $\mathcal{H}'$ (that is, the minimal sub-σ-algebra of $\mathcal{B}$ with the property that all functions from $\mathcal{H}'$ are measurable with respect to $\mathcal{D}$). Let $\mathbf{Y} = (X, \mathcal{D}, \mu)$ be the corresponding G-invariant factor of $\mathbf{X}$. One can check that $\mathcal{H}' = L^2(\mathbf{Y})$.

Let $\mathcal{H}^{\mathrm{wm}/Y}(T)$ be the subspace of $\mathcal{H}$ on which T is weakly mixing relative to $\mathbf{Y}$, and let $\mathcal{H}^{\mathrm{ds}/Y}(T)$ be the subspace where T has relatively discrete spectrum over $\mathbf{Y}$. Then $\mathcal{H} = \mathcal{H}^{\mathrm{wm}/Y}(T) \oplus \mathcal{H}^{\mathrm{ds}/Y}(T)$, and $\mathcal{H}^{\mathrm{ds}/Y}(T) = L^2(\mathbf{X}_1)$ for a factor $\mathbf{X}_1 = (X, \mathcal{B}_T, \mu)$ of $\mathbf{X}$ (Theorem 3.5). Let $\mathcal{H} = \mathcal{H}^{\mathrm{wm}/Y}(S) \oplus \mathcal{H}^{\mathrm{ds}/Y}(S)$ be the analogous decomposition corresponding to S, and let $\mathbf{X}_2 = (X, \mathcal{B}_S, \mu)$ be such that $\mathcal{H}^{\mathrm{ds}/Y}(S) = L^2(\mathbf{X}_2)$. Decompose $u = u' + u''$ with $u' \in \mathcal{H}^{\mathrm{ds}/Y}(T)$ and $u'' \in \mathcal{H}^{\mathrm{wm}/Y}(T)$. Then $\text{ess-sup}|u'| < M$ and so, $\text{ess-sup}|u''| < 2M$. Similarly, let $v = v' + v''$ with $v' \in \mathcal{H}^{\mathrm{ds}/Y}(S)$, $v'' \in \mathcal{H}^{\mathrm{wm}/Y}(S)$, then $\text{ess-sup}|v'| < M$ and $\text{ess-sup}|v''| < 2M$.

Let us write $\frac{1}{N}\sum_{n=1}^{N} T^n u S^n v$ as

$$\frac{1}{N}\sum_{n=1}^{N} T^n u' S^n v' + \frac{1}{N}\sum_{n=1}^{N} T^n u' S^n v'' + \frac{1}{N}\sum_{n=1}^{N} T^n u'' S^n v.$$

We will show that the first average converges as N tends to infinity, and that the other two are asymptotically small: lim sup of the norm of each of them does not exceed $2M^2\sqrt{\varepsilon}$.

First, assume that at least one of u, v lies in the relatively-mixing component of the corresponding operator: say, $u = u''$ (the treatment of the case of $v = v''$ is completely analogous). Note that, under the made assumption, our $u = u''$ satisfies $\text{ess-sup}|u| < 2M$ (as for v, we have, as before, and $\text{ess-sup}|v| < M$).

We will use the van der Corput lemma (Lemma 2.1). For $n, m \in \mathbb{N}$, write

$$\int (T^n u S^n v)\overline{(T^{n+m} u S^{n+m} v)}\,d\mu = \int T^n(uT^m\bar{u}) S^n(vT^m\bar{v})\,d\mu = \int u_m g(n) v_m\,d\mu,$$

where $u_m = uT^m\bar{u}$ and $v_m = vS^m\bar{v}$.

Fix $m \in \mathbb{N}$. Since $\tilde{v}_m = v_m - E(v_m|\mathbf{Y})$ is orthogonal to $\mathcal{H}'$, $\tilde{v}_m$ is orthogonal to $\mathcal{H}^{(l)}(g)$ for all $l \le L$. By Theorem 2.23,

$$\big\|\text{C-}\lim_n g(n)\tilde{v}_m\big\| \le \varepsilon\|\tilde{v}_m\| \le \varepsilon\|v_m\|.$$

Hence,

$$\text{C-}\lim_n\Big|\int u_m g(n)\big(v_m - E(v_m|\mathbf{Y})\big)\,d\mu\Big| = \text{C-}\lim_n\Big|\int u_m g(n)\tilde{v}_m\,d\mu\Big| \le \varepsilon\|u_m\|\|v_m\| \le 4M^4\varepsilon.$$

On the other hand,

$$\Big|\int u_m g(n)\big(E(v_m|\mathbf{Y})\big)d\mu\Big| = \Big|\int E(u_m|\mathbf{Y})g(n)\big(E(v_m|\mathbf{Y})\big)d\mu\Big| \leq \big\|E(u_m|\mathbf{Y})\big\|\big\|E(v_m|\mathbf{Y})\big\|$$

for all $n \in \mathbb{Z}$. Since $\big\|E(u_m|\mathbf{Y})\big\| = \big\|E(uT^m\bar{u}|\mathbf{Y})\big\| \underset{m\to\infty}{\longrightarrow} 0$ in density, we have

$$\text{C-}\lim_m \sup\Big|\text{C-}\lim_n \int u_m g(n) v_m d\mu\Big| \leq 4M^4\varepsilon.$$

By the van der Corput lemma,

$$\limsup_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=1}^{N} T^n u S^n v\Big\| \leq 2M^2\sqrt{\varepsilon}.$$

Now assume that both $u \in \mathcal{H}^{\mathrm{ds}/Y}(T)$ and $v \in \mathcal{H}^{\mathrm{ds}/Y}(S)$. Since, by (3.1), the group H has finite period on the space $\mathcal{H}' = L^2(\mathbf{Y})$, the H-polynomial $g(n) = T^{-n}S^n$ has finite period on $\mathcal{H}'$ by Proposition 1.31(a). Let $K \in \mathbb{N}$ be such that $T^{-K}S^K$ is trivial on $\mathcal{H}'$, that is, T^K coincides with S^K (almost everywhere) on $\mathbf{Y}$.

The conditions of Proposition 3.6 are now satisfied: $\mathbf{Y}$ is a common factor of $\mathbf{X}_1$ and $\mathbf{X}_2$, T^K and S^K coincide on $\mathbf{Y}$, T^K has on $\mathbf{X}_T$ relatively discrete spectrum over $\mathbf{Y}$ and S^K has on $\mathbf{X}_S$ relatively discrete spectrum over $\mathbf{Y}$. Thus, the limit $\text{C-}\lim_n T^{Kn}\tilde{u}S^{Kn}\tilde{v}$ exists for any $\tilde{u} \in L^\infty(\mathbf{X}_T)$ and $\tilde{v} \in L^\infty(\mathbf{X}_S)$. In particular, $\text{C-}\lim_n T^{Kn}(T^m u)S^{Kn}(S^m v) = \text{C-}\lim_n T^{Kn+m}uS^{Kn+m}v$ exists for every $m = 0,\ldots,K-1$. So, $\text{C-}\lim_n T^n u S^n v$ exists in this case. ■

3.8. The proof of Theorem A supplies some information about the location of $\text{C-}\lim_n T^n u S^n v$. Let H be the normal subgroup of G introduced in the proof of Theorem A. For $l \in \mathbb{N}$, let $\mathcal{D}_l$ be the σ-algebra generated by functions from $\mathcal{H}^{(l)}(H)$, and let $\mathbf{Y}_l$ be the corresponding factor of $\mathbf{X}$: $\mathbf{Y}_l = (X,\mathcal{D}_l,\mu)$. Let $\mathcal{H}^{\mathrm{ds}/Y_l}(T)$ be the subspace of $\mathcal{H}$ on which T has relatively discrete spectrum over $\mathbf{Y}_l$, let $\mathcal{H}^{\mathrm{ds}/Y_l}(S)$ be the subspace of $\mathcal{H}$ on which S has relatively discrete spectrum over $\mathbf{Y}_l$, and let $\mathbf{X}_{T,l}$, $\mathbf{X}_{S,l}$ be the corresponding factors of $\mathbf{X}$. Let $\mathbf{Y}$ be the factor of $\mathbf{X}$ generated by all $\mathbf{Y}_l$, $l \in \mathbb{N}$, namely, $\mathbf{Y} = (X,\mathcal{D},\mu)$ where $\mathcal{D}$ is the σ-algebra generated by all $\mathcal{D}_l$, $l \in \mathbb{N}$. Finally, let $\mathbf{X}_T$ be the factor generated by all $\mathbf{X}_{T,l}$, $l \in \mathbb{N}$, let $\mathbf{X}_S$ be the factor generated by all $\mathbf{X}_{S,l}$, $l \in \mathbb{N}$, and let $\mathbf{X}_{(T,S)}$ be the factor generated by $\mathbf{X}_T$ and $\mathbf{X}_S$. Since H is normal in G, all the factors $\mathbf{Y}_l, \mathbf{X}_{T,l}, \mathbf{X}_{S,l}$, $l \in \mathbb{N}$, and $\mathbf{Y}, \mathbf{X}_T, \mathbf{X}_S, \mathbf{X}_{(T,S)}$ are G-invariant.

Corollary of the proof. *For $u, v \in L^\infty(\mathbf{X})$ let $u' = E(u|\mathbf{X}_T)$ and $v' = E(v|\mathbf{X}_S)$. Then $\text{C-}\lim_n T^n u S^n v = \text{C-}\lim_n T^n u' S^n v'$. In particular, $\text{C-}\lim_n T^n u S^n v \in L^2(\mathbf{X}_{(T,S)})$.*

3.9. It follows from a general nilpotent Szemerédi theorem proved in [L2] that if T and S generate a nilpotent group, then for any set $A \in \mathcal{B}$ with $\mu(A) > 0$ one has $\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu\big(A \cap T^{-n}A \cap S^{-n}A\big) > 0$. The following theorem gives a new, direct proof of this fact, and shows that one can actually replace lim inf by lim. (The same proof gives a little bit more, namely the existence and the positivity of the limit $\lim_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M+1}^{N} \mu(A \cap T^{-n}A \cap S^{-n}A)$; see Remark 2.2.)

Theorem E. *Let T and S be measure preserving transformations of $\mathbf{X}$ generating a nilpotent group. Then for any $A \in \mathcal{B}$ with $\mu(A) > 0$, $\operatorname{C-lim}_n \mu\big(A \cap T^{-n}A \cap S^{-n}A\big)$ exists and is positive.*

Proof. The limit

$$\operatorname{C-lim}_n \mu\big(A \cap T^{-n}A \cap S^{-n}A\big) = \operatorname{C-lim}_n \int \mathbf{1}_A T^n\mathbf{1}_A S^n\mathbf{1}_A d\mu = \int \mathbf{1}_A \operatorname{C-lim}_n (T^n\mathbf{1}_A S^n\mathbf{1}_A) d\mu$$

exists by Theorem A. We will use the notation introduced in 3.8. Let $u = E(\mathbf{1}_A|\mathbf{X}_T)$ and $v = E(\mathbf{1}_A|\mathbf{X}_S)$. Then $0 \le u, v \le 1$, and $u(x) > 0$, $v(x) > 0$ for almost all $x \in A$. By Corollary 3.8, $\operatorname{C-lim}_n \mu(A \cap T^{-n}A \cap S^{-n}A) = \operatorname{C-lim}_n \int \mathbf{1}_A T^n u S^n v d\mu$. Let $\varphi = E(\mathbf{1}_A uv|\mathbf{Y})$, then $\varphi(x) > 0$ for almost all $x \in A$. Let $a > 0$ be such that the set $B = \big\{x \in X \bigm| \varphi(x) > a\big\}$ has positive measure: $\nu(B) = b > 0$.

Choose $l \in \mathbb{N}$ so that, if we put $u' = E(u|\mathbf{X}_{T,l})$, $v' = E(v|\mathbf{X}_{S,l})$ and $\varphi' = E(\varphi|\mathbf{Y}_l)$, then $\|u'-u\|^2 < \frac{a^2b}{10^3}$, $\|v'-v\|^2 < \frac{a^2b}{10^3}$ and $\|\varphi'-\varphi\|^2 < \frac{a^2b}{10^3}$. To simplify notation, let us view $\mathbf{Y}_l$ as an underlying measure space $(Y_l, \mathcal{D}_l, \nu)$. Let $\mu = \int \mu_y d\nu$ be the decomposition of μ with respect to $\mathbf{Y}_l$. Then there is $B' \in \mathcal{D}_l$ with $\mu(B') = b' \ge \frac{7}{10}b$ such that $\varphi'(y) > \frac{9a}{10}$, $\|u'-u\|_y < \frac{a}{10}$ and $\|v'-v\|_y < \frac{a}{10}$ for all $y \in B'$. ($\|\cdot\|_y$ denotes the norm in $L^2(X, \mathcal{B}, \mu_y)$.)

The function u' lies in $L^2(\mathbf{X}_{T,l})$, the space on which T has relatively discrete spectrum over $\mathbf{Y}_l$. One can easily show that the orbit $\{T^n u'\}_{n\in\mathbb{Z}}$ is *relatively compact* with respect to $\mathbf{Y}_l$: for any $\varepsilon > 0$ there is a finite set $u_1, \ldots, u_k \in L^2(\mathbf{X}_{T,l})$ such that for almost every $y \in Y_l$ and every $n \in \mathbb{Z}$ there is $1 \le t(n,y) \le k$ for which $\big\|T^n u' - u_{t(n,y)}\big\|_y < \varepsilon$. Let $u_1, \ldots, u_{k_1}$ be such a set corresponding to $\varepsilon = \frac{a}{10}$. Analogously, let $v_1, \ldots, v_{k_2} \in L^2(\mathbf{X}_{S,l})$ be such that for almost all $y \in Y_l$, $\big\|S^n v' - v_{s(n,y)}\big\|_y < \frac{a}{10}$ for some $1 \le s(n,y) \le k_2$. Put $k = k_1k_2 + 1$.

Since, under the notation of 3.8, the group H has finite period on the space $\mathcal{H}^{(l)}(H) = L^2(\mathbf{Y}_l)$, the H-polynomial $g(n) = T^{-n}S^n$ has finite period on $\mathcal{H}'$. Let $K \in \mathbb{N}$ be such that $T^{-K}S^K$ is trivial on $\mathcal{H}'$; then T^K coincides with S^K (almost everywhere) on Y_l.

A set of integers is called *thick* if it contains arbitrarily long intervals in $\mathbb{Z}$. We need the following simple lemma, whose proof is analogous to that of Lemma 5.1 in [BMZ]:

Lemma. *Let T be a measure preserving transformation of a probability measure space $(Y, \mathcal{D}, \nu)$, let $D \in \mathcal{D}$, $\nu(D) = d > 0$. Then for any thick set $Q \subseteq \mathbb{Z}$ and any $k \in \mathbb{N}$ there* *are* $n_1, \dots, n_k \in Q$ *such that* $\nu\big(\bigcap_{i=1}^k T^{-n_i} D\big) > \frac{1}{2} d^k$ *and* $n_j - n_i \in Q$ *for all* $1 \le i < j \le k$.

Let Q be a thick set in $\mathbb{Z}$. Applying the above lemma to the set B' and (the thick) set $\{n \in \mathbb{Z} \mid Kn \in Q\}$, we can find $n_1, \dots, n_k \in K\mathbb{Z}$ such that $n_j - n_i \in Q$, $1 \le i < j \le k$, and for $C = \bigcap_{i=1}^k T^{-n_i} B'$ one has $\nu(C) = c > \frac{1}{2} b'^k$.

For any $y \in Y_k$, among k pairs of numbers $\big(t(n_i, y), s(n_i, y)\big)$, $1 \le i \le k$, there are at least two equal pairs. Let $i(y)$ and $j(y)$, $i(y) < j(y)$, be such that $t(n_{i(y)}, y) = t(n_{j(y)}, y)$ and $s(n_{i(y)}, y) = s(n_{j(y)}, y)$. Let $C' \subseteq C$ be such that $i = i(y)$ and $j = j(y)$ are constant on C' and $\nu(C') = c' \ge \frac{c}{k_1 k_2}$. Then

$$\big\|T^{n_j} u' - T^{n_i} u'\big\|_y < 2\frac{a}{10} = \frac{a}{5} \quad \text{and } \big\|S^{n_j} v' - S^{n_i} v'\big\|_y < \frac{a}{5}$$

for all $y \in C'$. It follows that

$$\big\|T^{n_j - n_i} u' - u'\big\|_{T^{n_i} y} < \frac{a}{5} \quad \text{and } \big\|S^{n_j - n_i} v' - v'\big\|_{S^{n_i} y} < \frac{a}{5}$$

for all $y \in C'$. Put $n = n_j - n_i$. Taking into account that $n, n_i, n_j \in K\mathbb{Z}$ and that the actions of T^K and S^K coincide on Y_l, we can rewrite the obtained inequalities as

$$\big\|T^n v' - v'\big\|_{T^{n_i} y} < \frac{a}{5} \quad \text{and } \big\|S^n v' - v'\big\|_{T^{n_i} y} < \frac{a}{5}.$$

Let $y \in C'$. Since $T^{n_i} y \in B'$, $\|u' - u\|_{T^{n_i} y} < \frac{a}{10}$. Since $T^{n_j} y \in B'$,

$$\big\|T^n u' - T^n u\big\|_{T^{n_i} y} = \big\|T^{n_j - n_i} u' - T^{n_j - n_i} u\big\|_{T^{n_i} y} = \big\|u' - u\big\|_{T^{n_j} y} < \frac{a}{10}.$$

Thus,

$$\big\|T^n u - u\big\|_{T^{n_i} y} < \frac{a}{5} + 2\frac{a}{10} = \frac{2a}{5}.$$

In the same way,

$$\big\|S^n u - u\big\|_{T^{n_i} y} < \frac{2a}{5}.$$

Since $T^{n_i} y \in B'$, $\int \mathbf{1}_A uv d\mu_{T^{n_i} y} = \varphi'(T^{n_i} y) > \frac{9a}{10}$. Taking into account that $0 \le u, v \le 1$, we obtain

$$\int \mathbf{1}_A T^n u S^n v d\mu_{T^{n_i} y} > \frac{9a}{10} - 2\frac{2a}{5} = \frac{a}{10}.$$

Since this is true for all y from the set C' of measure c', we have $\int \mathbf{1}_A T^n u S^n v d\mu > \frac{ac'}{10}$.

Now recall that $n = n_j - n_i$ is an element of Q, an arbitrarily chosen thick set. This shows that the set $P = \big\{n \in \mathbb{Z} \bigm| \int \mathbf{1}_A T^n u S^n v d\mu > \frac{cc'}{10}\big\}$ is *syndetic*, that is, has bounded gaps. Let $N \in \mathbb{N}$ be such that any interval of length N in $\mathbb{Z}$ contains an element of P. Then

$$\text{C-}\lim_n \mu(A \cap T^{-n} A \cap S^{-n} A) = \text{C-}\lim_n \int \mathbf{1}_A T^n u S^n v d\mu > \frac{ac'}{10N}.$$

■

3.10. Our next goal is to determine the conditions under which the limit in Theorem 3.7 is the "right" one. Two measure preserving transformations T and S of a measure space $\mathbf{X} = (X, \mathcal{B}, \mu)$, $\mu(X) = 1$, are said to be *jointly ergodic* if $\text{C-}\lim_n T^n u S^n v = \int u d\mu \int v d\mu$ in $L^2(\mathbf{X})$ for all $u, v \in L^\infty(\mathbf{X})$ (see [BB]).

Theorem B. *Let T and S be measure preserving transformations of $\mathbf{X}$ generating a nilpotent group. T and S are jointly ergodic if and only if the transformation $T \times S$ is ergodic on $\mathbf{X} \times \mathbf{X}$, and the group $H = \langle T^{-n}S^n,\ n \in \mathbb{Z}\rangle$ is ergodic on $\mathbf{X}$.*

Let us note that if G is nilpotent of class c, then the G-polynomials T^n, S^n and hence, $g(n) = T^{-n}S^n$ are of degree $\leq c$. It follows that H is generated by $T^{-1}S, \ldots, T^{-c}S^c$ (see 1.6). Thus H is ergodic if and only if the transformations $T^{-1}S, \ldots, T^{-c}S^c$ have no common invariant functions other than constants. In particular, for $c = 1$, that is, for commuting T and S, Theorem B is reduced to a special case of the joint ergodicity criterion in [BB] (see Theorem BB in the Introduction).

3.11. We start the proof of Theorem B with the following technical lemma.

Lemma. *Let G be a finitely generated nilpotent group of measure preserving transformations of $\mathbf{X} = (X, \mathcal{B}, \mu)$, and let H be a normal subgroup of G such that H is ergodic on $\mathbf{X}$ and G/H is abelian. If $\mathcal{H}^{\text{rat}}(H) \neq \mathbb{C}$, then $\mathcal{H}^{\text{rat}}(G) \neq \mathbb{C}$ as well. Moreover, there is $u \in \mathcal{H}$, $u \neq$ const, and a prime number r such that $T^r u = u$ for all $T \in G$.*

Proof. First, replace H by a maximal subgroup of G that contains H and has the property $\mathcal{H}^{\text{rat}}(H) \neq \mathbb{C}$. Let $w \in \mathcal{H}^{\text{rat}}(H)$, $w \neq$ const, let $l \in \mathbb{N}$ be such that $P^l w = w$ for all $P \in H$ and let $\mathcal{H}' = \mathcal{H}^{(l)}(H) = \big\{u \in \mathcal{H} \;\big|\; P^l u = u \text{ for all } P \in H\big\}$. Then $\mathcal{H}'$ corresponds to a nontrivial G-invariant factor $\mathbf{X}'$ of $\mathbf{X}$: $\mathcal{H}' = L^2(\mathbf{X}')$. After passing from $\mathbf{X}$ to $\mathbf{X}'$, we may assume that P^l is trivial on $\mathbf{X}$ for all $P \in H$.

Let $R \in H$, $R \neq \mathbf{1}_G$, be an element of the center of G: since G is nilpotent and $[G, G] \subseteq H$, such an element exists. If R is not ergodic on $\mathbf{X}$, we may replace $\mathbf{X}$ by its factor $\mathbf{X}''$ on which R is trivial, and replace G by $G/\langle R\rangle$: since H is ergodic on $\mathbf{X}$, H is ergodic on $\mathbf{X}''$ and so, the group $H/\langle R\rangle$ is nontrivial. By Lemma 1.9, this operation can be nontrivially performed only finitely many times. So, we may assume that R is ergodic on $\mathbf{X}$. Let $u \in \mathcal{H}$, $u \perp \mathbb{C}$, be an eigenvector of R: $Ru = \lambda u$. Then $|u| = \text{const} \neq 0$ and $\lambda^l = 1$.

Let $l' \in \mathbb{N}$ be the minimal integer for which $\lambda^{l'} = 1$, let r be any prime divisor of l' and $l' = rk$. Put $u' = uRu \ldots R^{k-1}u$, then $Ru' = \lambda^k u'$. Renaming u' by u and λ^k by λ, we have $Ru = \lambda u$ with $\lambda^r = 1$.

For $T \in G$, let $v = uTu \ldots T^{r-1}u$. Then $Rv = \lambda^r v = v$, so $v = \text{const}$. Hence $Tv = v$, that is, $TuT^2u \ldots T^r u = uTu \ldots T^{r-1}u$, and so, $T^r u = u$. ∎

3.12. Proof of Theorem B. *Necessity:* If $T \times S$ is not ergodic on $\mathbf{X} \times \mathbf{X}$, then $Tu = \lambda u$ and $Sv = \lambda v$ for some $\lambda \in \mathbb{C}$ and $u, v \in \mathcal{H}$ (see, for example, [F3], Lemma 4.18). If T is not ergodic, $\text{C-}\lim_n T^n w = w$ for a nonconstant $w \in \mathcal{H}$, then $\text{C-}\lim_n T^n w S^n 1 = w \neq \int w d\mu$. If both T and S are ergodic on $\mathbf{X}$, then we may assume that $|u| \equiv |v| \equiv 1$ and $\int u d\mu = \int v d\mu = 0$. Thus, $\text{C-}\lim_n T^n u S^n \bar{v} = u\bar{v} \neq \int u d\mu \int \bar{v} d\mu = 0$.

Now assume that H is not ergodic on $\mathbf{X}$, let $u \in \mathcal{H}$, $u \neq \text{const}$, be such that $T^{-n} S^n u = u$ for all $n \in \mathbb{Z}$. We may assume that $u \perp \mathbb{C}$, that is, $\int u d\mu = 0$. Then

$$\text{C-}\lim_n \int T^n \bar{u} S^n u d\mu = \text{C-}\lim_n \int \bar{u} T^{-n} S^n u d\mu = \int |u|^2 d\mu \neq \int \bar{u} d\mu \int u d\mu = 0.$$

Sufficiency: Let both $T \times S$ and H be ergodic. Then H satisfies the assumptions of Lemma 3.11, and, since T and S can not possess a common eigenvalue, $\mathcal{H}^{\text{rat}}(H) = \mathbb{C}$. Define $g(n) = T^{-n} S^n$. Then, by Theorem 2.17, $\text{C-}\lim_n g(n) u = \int u d\mu$ for all $u \in \mathcal{H}$.

Now, let $u, v \in \mathcal{H}$. Replacing u by $u - \int u d\mu$, we may assume that $\int u d\mu = 0$. Then

$$\begin{aligned}
&\text{C-}\lim_m \text{C-}\lim_n \int (T^n u S^n v) \overline{(T^{n+m} u S^{n+m} v)} d\mu \\
&\quad = \text{C-}\lim_m \text{C-}\lim_n \int T^n \big(u T^m \bar{u}\big) S^n \big(v S^m \bar{v}\big) d\mu \\
&\quad = \text{C-}\lim_m \text{C-}\lim_n \int \big(u T^m \bar{u}\big) g(n) \big(v S^m \bar{v}\big) d\mu = \text{C-}\lim_m \int u T^m \bar{u} d\mu \int v S^m \bar{v} d\mu \\
&\quad = \text{C-}\lim_m \int (u \otimes v)(T \times S)^m (\bar{u} \otimes \bar{v}) d\mu \times d\mu = \int u \otimes v d\mu \times d\mu \int \bar{u} \otimes \bar{v} d\mu \times d\mu \\
&\qquad\qquad\qquad\qquad\qquad = \int u d\mu \int v d\mu \int \bar{u} d\mu \int \bar{v} d\mu = 0.
\end{aligned}$$

By the van der Corput lemma, $\text{C-}\lim_n T^n u S^n v = 0$. ■

4. Counter-examples

In this section we will give three examples which demonstrate that neither the convergence of the expressions $\frac{1}{N} \sum_{n=1}^{N} u(T^n x) v(S^n x)$ nor the joint recurrence of T and S (i.e. the positivity of $\mu(T^{-n} A \cap S^{-n} A \cap A)$ for some $n > 0$) necessarily hold if the group $G = \langle T, S \rangle$ is a solvable non-nilpotent group. The three examples below correspond to "different types" of solvable groups, which hints that we may be dealing here with a general phenomenon. (See the discussion and conjectures in Section 5.)

Actually, the examples in this section show that hardly any of the convergence theorems proved in this paper holds true for solvable groups which do not contain a nilpotent subgroup of finite index. Indeed, the examples below show also that both the nil-polynomial ergodic theorem (Theorem C) and the joint ergodicity criterion (Theorem B) no longer hold. In all three examples we deal with ergodic transformations T, S such that for some $u, v \in L^2(X, \mathcal{B}, \mu)$ (these u and v are actually chosen to be characteristic functions of subsets of X) the weak limit $\text{C-}\lim_n \int (S^{-n}T^n u) v d\mu = \text{C-}\lim_n \int T^n u S^n v d\mu$ does not exist. It follows that $\text{C-}\lim_n S^{-n}T^n u$ does not exist in L^2-norm, which constitutes a counterexample to Theorem C. (Notice that when $\langle T, S\rangle$ is nilpotent, $T^{-n}S^n$ is a polynomial sequence.) The transformations T and S in example 4.2 have, in addition, the property that the action of the group generated by $T^{-n}S^n$, $n \in \mathbb{Z}$, is ergodic. This furnishes a counterexample to Theorem B.

Remark. A noncommutative counterexample to Theorem B was also given by Berend ([Be], Example 7.1). In Berend's example, the transformations T and S generate a nonsolvable group.

4.1. Our first construction is a modification of an example of Furstenberg ([F3], p. 40). Let $X = \{0,1\}^{\mathbb{Z}}$ equipped with the product measure determined by the weights $(\frac{1}{2}, \frac{1}{2})$ on $\{0,1\}$. For every $D \subseteq \mathbb{Z}$, let P_D be the transformation of X switching the coordinates corresponding to the elements of D: for $x = (x_j)_{j \in \mathbb{Z}}$ let $(xP_D)_j = \begin{cases} 1 - x_j, \ j \in D \\ x_j, \ j \notin D. \end{cases}$ (The transformations P_D act on X from the right while the corresponding induced action on functions on X will be from the left.) All such P_D, $D \subseteq \mathbb{Z}$, form an abelian group; we denote this group by H. Let S be the coordinate shift on X (also acting on X from the right): $(xS)_j = x_{j+1}$. Let G be the group of measure preserving transformations of X generated by H and S. It is easy to see that H is normal in G and so, G is a solvable group of class 2 (that is, *a metabelian group*).

Let $D = \{a_1, a_2, \ldots\} \subseteq \mathbb{Z}$ with $0 \le a_1 < a_2 < \ldots$. Put $T = SP_D$. Then for any $x \in X$ and $n \in \mathbb{Z}$,

$$(xS^n)_0 = x_n \quad \text{and} \quad (xT^n)_0 = \begin{cases} x_n, \ n \le a_1 \\ 1 - x_n, \ a_1 < n \le a_2 \\ x_n, \ a_2 < n \le a_3 \\ 1 - x_n, \ a_3 < n \le a_4 \\ \ldots \end{cases}$$

Put $A = \{x \in X \mid x_0 = 0\}$. Then $\mu(A) = \frac{1}{2}$, and we have $AT^{-n} = AS^{-n}$ for $n \le a_1$ and $a_{2k} < n \le a_{2k+1}$, $k \in \mathbb{N}$, and $AT^{-n} = (X \setminus A)S^{-n}$ for $a_{2k-1} < n \le a_{2k}$, $k \in \mathbb{N}$.

It follows that for any sequence $\{r_n\}_{n \in \mathbb{N}}$ made of 0's and $\frac{1}{2}$'s, one can find $D \subseteq \mathbb{Z}$ such that for $T = SP_D$ one has $r_n = \mu(AT^{-n} \cap AS^{-n})$, $n \in \mathbb{N}$. In Furstenberg's example $D = \{0\}$, so $\mu(AT^{-n} \cap AS^{-n}) = 0$ for all $n \in \mathbb{N}$. It gives a counterexample to the ergodic Szemerédi theorem (namely, to relation (0.3) in the Introduction) for solvable groups.

To get a counterexample to the ergodic Roth theorem (see (0.1)), take $D = \{a_1, a_1 + a_2, \ldots\}$ where $a_1, a_2, \ldots$ is a rapidly increasing sequence of integers, say $a_k = k!$.

Then $\mu(AT^{-n}\cap AS^{-n})$, $n\in\mathbb{N}$, is a sequence of the form

$$\big(\tfrac{1}{2},\underbrace{0,0}_{a_2},\underbrace{\tfrac{1}{2},\ldots,\tfrac{1}{2}}_{a_3},\underbrace{0,0,\ldots,0}_{a_4},\underbrace{\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\ldots,\tfrac{1}{2}}_{a_5},\underbrace{0,0,0,0,\ldots,0}_{a_6},\ldots\big),$$

and so, for $u=1_A$, the limit $\operatorname*{C-lim}_n\int T^nuS^nud\mu=\operatorname*{C-lim}_n\mu(AT^{-n}\cap AS^{-n})$ does not exist.

4.2. The group G in example 4.1 is solvable but is not polycyclic: its center H is the product of infinitely many copies of $\mathbb{Z}_2$ (a group is *polycyclic* if it has a finite subnormal series with cyclic factors). Our next example is called up to show that the ergodic Szemerédi and Roth theorems do not hold for polycyclic groups as well.

Let $X=\mathbb{T}^1\times\mathbb{T}^1$ be the torus equipped with Lebesgue measure, let for $x\in X$, $xP=x+\binom{\alpha}{0}$, $xQ=x+\begin{pmatrix}0\\ \alpha\end{pmatrix}$ and $xT=\Lambda x$, where $\alpha\in\mathbb{T}^1$ and $\Lambda=\begin{pmatrix}5&4\\6&5\end{pmatrix}$. Let G be the group generated by P, Q and T, acting on X on the right. We have $xT^{-1}PT=x+\Lambda\binom{\alpha}{0}=xP^5Q^6$ and $xT^{-1}QT=x+\Lambda\begin{pmatrix}0\\ \alpha\end{pmatrix}=xP^4Q^5$, so the subgroup $H=\langle P,Q\rangle$ is normal in G. Since $H\simeq\mathbb{Z}^2$ and $G/H\simeq\mathbb{Z}$, G is polycyclic.

Put $S=PTP^{-1}$, then $S^n=PT^nP^{-1}$. For $x\in X$ we have:

$$\begin{aligned}xT^{-n}S^n=xT^{-n}PT^nP^{-1}=\Lambda^n\big(\Lambda^{-n}x+\tbinom{\alpha}{0}\big)-\tbinom{\alpha}{0}&=x+\alpha(\Lambda^n\begin{pmatrix}1\\0\end{pmatrix}-\begin{pmatrix}1\\0\end{pmatrix})\\&=x+\alpha\begin{pmatrix}\lambda_n\\ \mu_n\end{pmatrix},\end{aligned}$$

where $\lambda_n=\frac{1}{2}(5+\sqrt{24})^n+\frac{1}{2}(5-\sqrt{24})^n-1$ and $\mu_n=\frac{\sqrt{6}}{4}(5+\sqrt{24})^n-\frac{\sqrt{6}}{4}(5-\sqrt{24})^n$. We have $\lambda_{n+1}/\lambda_n>5$ for all $n\in\mathbb{N}$.

Lemma. *Let $M\in\mathbb{N}$ and let $\lambda_1<\lambda_2<\ldots$ be a sequence of real numbers satisfying $\lambda_{n+1}/\lambda_n\geq M+1$ for all $n\in\mathbb{N}$. For any sequence of integers $c_1,c_2,\ldots$ with $0\leq c_i<M$ there is a real number α, $0\leq\alpha<1$, for which $\frac{c_n}{M}\leq\{\lambda_n\alpha\}<\frac{c_n+1}{M}$ for all $n\in\mathbb{N}$.*

Proof. For $\lambda\in\mathbb{R}$, let us call any interval $[\beta,\gamma]$ such that $\lambda[\beta,\gamma]\equiv[\frac{k}{M},\frac{k+1}{M}]\bmod 1$, $0\leq k\leq M-1$, a *(λ,k)-interval.* For every $n\in\mathbb{N}$, since $\lambda_{n+1}/\lambda_n\geq M+1$, any (λ_n,k)-interval contains a (λ_{n+1},l)-subinterval for each $l=0,\ldots,M-1$. So α can be taken as the common point of a nested sequence of (λ_n,c_n)-intervals, $n=1,2\ldots$. ■

We can now apply this lemma with $M=4$ to our sequence $\{\lambda_n\}$. To construct a counterexample to the ergodic Szemerédi theorem, we choose α corresponding to the sequence $2,2,\ldots$. Since the first coordinate of the point $0T^{-n}S^n$ is $\{\lambda_n\alpha\}$, we have $0T^{-n}S^n\in[\frac{1}{2},\frac{3}{4}]\times\mathbb{T}^1$ for all $n\in\mathbb{N}$. So for $A=[0,\frac{1}{4}]\times\mathbb{T}^1$ we have $AT^{-n}\cap AS^{-n}=\emptyset$ for all $n\in\mathbb{N}$.

For a counterexample to the ergodic Roth theorem, take α to be the number corresponding to the sequence $r = \big(2, \underbrace{0,0}_{a_2}, \underbrace{2,\dots,2}_{a_3}, \underbrace{0,0,\dots,0}_{a_4}, \underbrace{2,2,2,\dots,2}_{a_5}, \underbrace{0,0,0,0,\dots,0}_{a_6}, \dots\big)$, where, say, $a_k = k!$. Define $A = [0, \frac{1}{4}] \times \mathbb{T}^1$ and $B = [\frac{1}{2}, 1] \times \mathbb{T}^1$. Then either $AT^{-n}S^n \subset X \setminus B$ (if n is such that $r_n = 0$), or $AT^{-n}S^n \subset B$ (if $r_n = 2$). So, if we put $u = 1_A$ and $v = 1_B$, we will have $\int T^n u S^n v d\mu = \int (S^{-n}T^n u) v d\mu = \mu(AT^{-n}S^n \cap B) = \frac{1}{8} r_n$, $n \in \mathbb{N}$, a sequence that has no Cesàro limit.

Notice also that we simultaneously get a counterexample to Theorem B for polycyclic groups. Indeed, since both T and S are weakly (and even strongly) mixing, $T \times S$ is ergodic on $X \times X$. Transformations $T^{-n}S^n$ generate a dense subgroup of the group of all rotations of the torus X, which is, hence, ergodic on X. So, all assumptions of Theorem B are satisfied, but the limit $\text{C-}\lim_n T^n u S^n v$ does not exist.

4.3. Our last example is, in a sense, a hybrid of the preceding two. Let X be the infinite-dimensional torus $\mathbb{T}^{\mathbb{Z}}$, with the product measure. Let $\alpha \in \mathbb{T}^1$. For $x = (\dots, x_{-1}, x_0, x_1, \dots) \in X$, put $(xS)_j = x_{j+1}$, $(xP)_j \equiv x_j + 5^j \alpha \bmod 1$, and $T = P^{-1}SP$. Then

$$(xT^{-n}S^n)_0 = (xP^{-1}S^{-n}PS^n)_0 = x_0 + (5^n - 1)\alpha,$$

that is, $T^{-n}S^n$ acts on the zero-coordinate circle as a rotation by $(5^n - 1)\alpha$. Let $A = \big\{x \in X \mid x_0 \in [0, \frac{1}{4}]\big\}$. Then, similarly to example 4.2, one can choose α so that, for the corresponding T, $T^{-n}A \cap S^{-n}A = \emptyset$ for all $n \in \mathbb{N}$. And if we put $B = \big\{x \in X \mid x_0 \in [\frac{1}{2}, 1]\big\}$, then for any sequence $\{r_n\}_{n \in \mathbb{N}}$ made of 0's and $\frac{1}{4}$'s one can find α such that $r_n = \mu(AT^{-n} \cap BT^{-n})$ for all $n \in \mathbb{N}$.

5. Some conjectures

5.1. Theorems A and B on one hand and the counterexamples brought in Section 4 on the other, show that for solvable groups there is a sort of dichotomy related to the behavior of the expressions $\mu(T^{-n}A \cap S^{-n}A)$ and $\frac{1}{N}\sum_{n=1}^{N} f(T^n x) g(S^n x)$. Namely, the known convergence and recurrence results pertaining to commuting measure preserving transformations extend naturally to the case when T and S generate a nilpotent group. On the other hand, when T and S generate a solvable group of exponential growth, Theorems A and B are no longer valid. (We remark that, by Gromov's theorem, a finitely generated solvable group has exponential growth if and only if it is not virtually nilpotent.) Also, in the measure preserving system $(X, \mathcal{B}, \mu, \langle T, S \rangle)$ one can no longer expect to have the recurrence along the sequence $T^{-n}S^n$ (i.e. unlike the nilpotent case, it may occur that for a set $A \in \mathcal{B}$ with $\mu(A) > 0$ one has for all $n > 0$, $\mu(A \cap T^n S^{-n} A) = \mu(T^{-n}A \cap S^{-n}A) = 0$). The reason behind this dichotomy seems to be the fact that when two elements, T and S, generate a solvable group of exponential growth, the sequence $T^{-n}S^n$ is not necessarily a polynomial sequence (in the sense of definition in 1.1) whereas when $\langle T, S\rangle$ is nilpotent, this is always the case. The following example shows, however, that even if $\langle T, S\rangle$ is a solvable group of exponential growth, the sequence $T^{-n}S^n$ may be polynomial.

5.2. Let G be the group defined by generators T and S and the relations $S^2 = \mathbf{1}_G$ and $[S, T^{-n}ST^n] = \mathbf{1}_G$ for all $n \in \mathbb{Z}$. Then G is a solvable group: if we denote $S_n = T^{-n}ST^n$, then the subgroup H generated by $\{S_n,\ n \in \mathbb{Z}\}$ is abelian and normal in G, and G/H is generated by T. On the other hand, G is not nilpotent: the elements $[T, S] = S_1S$, $[T, [T, S]] = S_2S$, $\ldots$, $\big[T, [T, \ldots, [T, S]\ldots]\big] = S_nS$, $\ldots$ are all nontrivial. However, the sequence $g(n) = T^{-n}S^n$ is polynomial in G:

$$(Dg)(n) = \big(g(n)\big)^{-1} g(n+1) = S^{-n}T^nT^{-n-1}S^{n+1} = S^{-n}TS^{n+1} = TS_1^{-n}S^{n+1};$$
$$(D^2g)(n) \equiv S_1^{-1}S; \quad (D^3g)(n) \equiv \mathbf{1}_G.$$

Note that the limit $\text{C-}\lim_n T^nuS^nv = \frac{1}{2}\,\text{C-}\lim_n T^{2n}(u + TSu)$ clearly exists in $L^2(X)$ for any measure preserving action of G on a probability space $(X, \mathcal{B}, \mu)$ and any $u, v \in L^\infty(X)$.

5.3. Given a group G, a sequence $g\colon \mathbb{Z} \longrightarrow G$ and $m \in \mathbb{Z}$, let us define *the m-derivative of $g(n)$* by $D_mg(n) = g(n)^{-1}g(n + m)$. Notice that the 1-derivative D_1 coincides with derivative D as defined in 1.1. Let us say that a sequence $g(n)$ is *polynomial of degree $\le d$* if for any $m_1, \ldots, m_d \in \mathbb{Z}$, $D_{m_1} \ldots D_{m_d}g(n)$ is a constant sequence in G. One can show that for sequences in nilpotent groups this notion of polynomiality coincides with the one defined in 1.1. Note that the sequence $T^{-n}S^n$ featured in 5.2 satisfies this stronger definition of polynomiality as well. We have the following conjecture:

Conjecture. (i) *Let G be a group of unitary operators on a Hilbert space $\mathcal{H}$. If $g(n)$ is a polynomial sequence in G, then* $\text{C-}\lim_n g(n)u$ *exists for any $u \in \mathcal{H}$.*
(ii) *Let G be a group of measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$, and let $T, S \in G$ be such that the sequence $T^{-n}S^n$ is polynomial in G. Then for any $u, v \in L^\infty(X)$,* $\text{C-}\lim_n T^nuS^nv$ *exists in L^2-norm.*

5.4. On the other hand, counterexamples of Section 5 lead to the following conjecture:

Conjecture. *Let G be a finitely generated solvable group without a nilpotent subgroup of finite index. There is a measure preserving action of G on a probability measure space $(X, \mathcal{B}, \mu)$ such that*
(i) *there are $T, S \in G$ and $u, v \in L^\infty(X)$ such that* $\text{C-}\lim_n \int T^nuS^nv\,d\mu$ *does not exist.*
(ii) *there are $T, S \in G$ and $A \in \mathcal{B}$ with $\mu(A) > 0$ such that $\mu(T^{-n}A \cap S^{-n}A) = 0$ for all $n > 0$.*

5.5. A natural question is whether $\text{C-}\lim_n T_1^n u_1 \dots T_k^n u_k$ exists for all $k > 2$ and $T_1, \dots, T_k$ generating a nilpotent group. Though the problems seems to be very difficult, we believe that the following conjecture is true:

Conjecture. *Let G be a nilpotent group of measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$. Then for any G-polynomials $g_1, \dots, g_k$ and any $u_1, \dots, u_k \in L^\infty(X)$, $\text{C-}\lim_n g_1(n)u_1 \dots g_k(n)u_k$ exists in L^2-norm and almost everywhere.*

5.6. We pass now to a discussion of joint ergodicity. While in the case of commuting T, S Theorem B reduces to a special case of Theorem BB, we have, as yet, no nilpotent analog of Theorem BB for more than two transformations.

Conjecture. *Let $T_1, \dots, T_k$ be measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$, generating a nilpotent group. Then*

$$\text{C-}\lim_n T_1^n u_1 \dots T_k^n u_k = \int u_1 d\mu \dots \int u_k d\mu$$

in $L^2(X)$ for all $u_1, \dots, u_k \in L^\infty(X)$ if and only if $T_1 \times \dots \times T_k$ is ergodic on $X \times \dots \times X$ and for any $1 \le i < j \le k$, the group generated by $\{T_i^{-n} T_j^n,\ n \in \mathbb{Z}\}$ is ergodic on X.

5.7. It would be interesting to find a condition for joint ergodicity in more general groups. The counterexamples in Section 4 and the discussion in 5.3 lead us to the following conjecture (which for sake of conciseness we formulate for two transformations).

Conjecture. *Let T and S be measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$. Then $\text{C-}\lim_n T^n u S^n v = \int u d\mu \int v d\mu$ in $L^2(X)$ for all $u, v \in L^\infty(X)$ if and only if the following three conditions hold:*

(a) *$T \times S$ is ergodic on $X \times X$;*

(b) *the sequence $T^{-n} S^n$ is polynomial (in the sense of* 5.3*);*

(c) *the group generated by $\{T^{-n} S^n,\ n \in \mathbb{Z}\}$ is ergodic.*

5.8. We want to conclude by mentioning one more interesting problem. Namely, it would be nice to know to which extent the property of growth of a group alone is responsible for the validity of the results and counterexamples brought in the paper. It was Grigorchuk who constructed in [G] a large family of groups of *intermediate growth*, which occupy an intermediate place between the groups of polynomial and exponential growth.

Question. Which of the results obtained above extend to Grigorchuk's groups?

Bibliography

[Be] D. Berend, Joint ergodicity and mixing, *Journal d'Analyse Math.* **45** (1985), 255-284.

[BB] D. Berend and V. Bergelson, Jointly ergodic measure-preserving transformations, *Israel Journal of Math.* **49** No. 4 (1984), 307-314.

[B1] V. Bergelson, Weakly mixing PET, *Ergod. Th. and Dynam. Sys.* **7** (1987), 337-349.

[B2] V. Bergelson, Ergodic Ramsey Theory – an update, *Ergodic Theory of $\mathbb{Z}^d$-actions (edited by M. Pollicott and K. Schmidt), London Math. Soc. Lecture Note Series* **228** (1996), 1-61.

[BL1] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems, *Journal of AMS* **9** No. 3 (1996), 725-753.

[BL2] V. Bergelson and A. Leibman, Set-polynomials and a polynomial extension of the Hales-Jewett theorem, *Annals of Math.* **150** (1999), 33-75.

[BMZ] V. Bergelson, R. McCutcheon and Q. Zhang, Amenable Roth theorem, *Amer. Journal of Math.* **119** (1997), 1173-1211.

[Bo] J. Bourgain, Double recurrence and almost sure convergence, *J. Reine Angew. Math.* **404** (1990), 140-161.

[CL1] J.-P. Conze and E. Lesigne, Théorèmes ergodiques pour des mesures diagonales, *Bull. Soc. Math. France* **112** (1984), 143-175.

[CL2] J.-P. Conze and E. Lesigne, Sur un théorème ergodique pour des mesures diagonales, *Publ. Inst. Rech. Math. Rennes* 1987-1 (1988), 1-31.

[ET] P. Erdös and P. Turán, On some sequences of integers, *Journal of London Math. Soc.* **11** (1936), 261-264.

[F1] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *Journal d'Analyse Math.* **31** (1977), 204-256.

[F2] H. Furstenberg, Poincaré recurrence and number theory, *Bull. AMS (New Series)* **5** (1981), 211-234.

[F3] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, 1981.

[F4] H. Furstenberg, Nonconventional ergodic averages, *Proc. Sympos. Pure Math.* **50** (1990), 43-56.

[FK1] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, *Journal d'Analyse Math.* **34** (1978), 275-291.

[FK2] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for IP-systems and combinatorial theory, *Journal d'Analyse Math.* **45** (1985), 117-168.

[FK3] H. Furstenberg and Y. Katznelson, A density version of the Hales-Jewett theorem, *Journal d'Analyse Math.* **57** (1991), 64-119.

[FW] H. Furstenberg and B. Weiss, A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, *Con-*

vergence in Ergodic Theory and Probability, Walter de Gruyter, 1996, 193-227.

[G] R.I. Grigorchuk, Degrees of growth of finitely generated groups and the theory of invariant means, *Izv. Akad. Nauk SSSR Ser. Mat.* **48,** no. 5 (1984), 939-985.

[H] Ph. Hall, *Nilpotent Groups*, Queen Mary College Mathematical Notes, 1969.

[HK] B. Host and B. Kra, Convergence of Conze-Lesigne averages, *preprint.*

[Hua] L.K. Hua, *Additive Theory of Prime Numbers*, Amer. Math. Soc., 1965.

[L1] A. Leibman, Polynomial sequences in groups, *Journal of Algebra* **201** (1998), 189-206.

[L2] A. Leibman, Multiple recurrence theorem for measure preserving actions of a nilpotent group, *Geom. and Funct. Anal.* **8** (1998), 853-931.

[L3] A. Leibman, The structure of unitary action of a nilpotent group, *Ergodic Theory and Dynamical Systems* **20** (2000), 809-820.

[R] K. Roth, Sur quelques ensembles d'entiers, *C.R. Acad. Sci. Paris* **234** (1952), 388-390.

[Roy] H.L. Royden, *Real Analysis,* third edition, Prentice Hall, 1988.

[S] N. A. Shah, Limit distributions of polynomial trajectories on homogeneous spaces, *Duke Math. J.* **75** N3 (1994), 711-732.

[Sz] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199-245.

[Z1] R. Zimmer, Extensions of ergodic group actions, *Illinois J. Math.* **20** (1976), 373-409.

[Z2] R. Zimmer, Ergodic actions with generalized discrete spectrum, *Illinois J. Math.* **20** (1976), 555-588.

[Zh] Q. Zhang, On Convergence of the Averages $\frac{1}{N}\sum_{n=1}^{N} f_1(R^n x) f_2(S^n x) f_3(T^n x)$, *Monat. Math.* **122** (1996), 275-300.