POLYNOMIAL EXTENSIONS OF VAN DER WAERDEN'S AND SZEMERÉDI'S THEOREMS

V. Bergelson, A. Leibman

Abstract. An extension of the classical van der Waerden and Szemerédi theorems is proved for commuting operators whose exponents are polynomials. As a consequence, for example, one obtains the following result: Let $S \subseteq \mathbf{Z}^l$ be a set of positive upper Banach density, let $p_1(n), \dots, p_k(n)$ be polynomials with rational coefficients taking on integer values on the integers and satisfying $p_i(0) = 0$, $i = 1, \dots, k$; then for any $v_1, \dots, v_k \in \mathbf{Z}^l$ there exist an integer n and a vector $u \in \mathbf{Z}^l$ such that $u + p_i(n)v_i \in S$ for each $i \leq k$.

0. Introduction

0.1. In 1975 E. Szemerédi ([S]) confirmed a long standing conjecture of P. Erdős and P. Turán by showing that if a set $S \subseteq \mathbb{N}$ has positive upper density: $\overline{d}(S) = \limsup_{N\to\infty} \frac{|S \cap \{1,\dots,N\}|}{N} > 0$, then S contains arbitrarily long arithmetic progressions.

Szemerédi's proof was purely combinatorial and quite involved. In 1976 H. Furstenberg ([F1]) gave a completely different, ergodic theoretical proof of Szemerédi's theorem by proving a far reaching extension of the classical Poincare recurrence theorem and showing that Szemerédi's theorem is a consequence of it.

In 1978 H. Furstenberg and Y. Katznelson obtained a multidimensional extension of Szemerédi's theorem by deducing it from the following

Theorem.([FK1]A) *Let $(X, \mathcal{B}, \mu)$ be a measure space with $\mu(X) < \infty$, let T_1, $T_2, \dots, T_k$ be commuting measure preserving transformations of X and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\Big(T_1^{-n}A \cap T_2^{-n}A \cap \dots \cap T_k^{-n}A\Big) > 0.$$

A set $S \subseteq \mathbb{Z}^k$ is said to have *positive upper Banach density* if for a sequence of parallelepipeds $\Pi_n = [a_n^{(1)}, b_n^{(1)}] \times \dots \times [a_n^{(k)}, b_n^{(k)}] \subset \mathbb{Z}^k$, $n \in \mathbb{N}$, with $b_n^{(i)} - a_n^{(i)} \to \infty$, $i = 1, \dots, k$, one has:

$$\frac{|S \cap \Pi_n|}{|\Pi_n|} > \varepsilon \tag{0.1}$$

1991 *Mathematics Subject Classification.* Primary 11B83, 28D05, 54H20; Secondary 05A17, 05D10.

The first author gratefully acknowledges support received from the National Science Foundation (USA) via grants DMS-9103056 and DMS-9401093. The second author was supported by British Technion Society.

for some $\varepsilon > 0$.

Corollary.([FK1]B) *Let $S \subseteq \mathbb{Z}^k$ be a subset with positive upper Banach density and let $F \subset \mathbb{Z}^k$ be a finite configuration. Then there exist a positive integer n and a vector $u \in \mathbb{Z}^k$ such that $u + nF \subset S$.*

We remark that so far this corollary has no "conventional" combinatorial proof.

Theorem [FK1]A was extended further in [FK2] and recently Furstenberg and Katznelson ([FK3]) proved a density version of the Hales-Jewett theorem, which contains the results from [FK1] and [FK2] as quite special cases.

0.2. The purpose of this paper is to obtain an extension of Theorem [FK1]A in a different direction. What we are after is to give a joint extension of this theorem and of a theorem of Furstenberg-Sárközy which states that for any polynomial $p(n) \in \mathbb{Q}[n]$ taking on integer values on the integers and such that $p(0) = 0$, and for any $S \subseteq \mathbb{Z}$ with $\overline{d}(S) > 0$, there exist $n \in \mathbb{N}$, $x, y \in S$ such that $x - y = p(n)$. For example, one would like to know whether any set of positive upper density in $\mathbb{N}$ contains arbitrarily long arithmetic progressions whose difference is a perfect square. Such a theorem is indeed true and follows from a special case ($T_1 = T_2 = \ldots = T_k$, $p_j(n) = jn^2$, $j = 1, \ldots, k$) of the following

Theorem $\mathbf{A}_0$. *Let $(X, \mathcal{B}, \mu)$ be a probability space, $T_1, \ldots, T_k$ be commuting measure preserving invertible transformations of X, let $p_1(n), \ldots, p_k(n)$ be polynomials with rational coefficients taking on integer values on the integers and satisfying $p_i(0) = 0$, $i = 1, \ldots, k$, and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\Big(T_1^{-p_1(n)} A \cap T_2^{-p_2(n)} A \cap \ldots \cap T_k^{-p_k(n)} A\Big) > 0.$$

0.3. As a corollary of Theorem A_0 one gets the following

Theorem $\mathbf{B}_0$. *Let $S \subseteq \mathbb{Z}^l$, $l \in \mathbb{N}$, be a set of positive upper Banach density, let $p_1(n), \ldots, p_k(n)$ be polynomials with rational coefficients taking on integer values on the integers and satisfying $p_i(0) = 0$, $i = 1, \ldots, k$. Then for any $v_1, \ldots, v_k \in \mathbb{Z}^l$ there exist an integer n and a vector $u \in \mathbb{Z}^l$ such that $u + p_i(n)v_i \in S$ for each $i \leq k$.*

0.4. As a matter of fact, we prove an even more general result:

Theorem A. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $T_1, \ldots, T_t$ be commuting measure preserving invertible transformations of X, let $p_{1,1}(n), \ldots, p_{1,t}(n), p_{2,1}(n), \ldots, p_{2,t}(n), \ldots, p_{k,1}(n), \ldots, p_{k,t}(n)$ be polynomials with rational coefficients taking on integer values on the integers and satisfying $p_{i,j}(0) = 0$, $i = 1, \ldots, k$, $j = 1, \ldots, t$, and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\Big(\prod_{j=1}^{t} T_j^{-p_{1,j}(n)} A \cap \prod_{j=1}^{t} T_j^{-p_{2,j}(n)} A \cap \ldots \cap \prod_{j=1}^{t} T_j^{-p_{k,j}(n)} A\Big) > 0.$$

0.5. As a corollary, we get

Theorem B. *Let $S \subseteq \mathbb{Z}^l$, $l \in \mathbb{N}$, be a set of positive upper Banach density, let $p_{1,1}(n), \ldots, p_{1,t}(n), p_{2,1}(n), \ldots, p_{2,t}(n), \ldots, p_{k,1}(n), \ldots, p_{k,t}(n)$ be polynomials* *with rational coefficients taking on integer values on the integers and satisfying* $p_{i,j}(0) = 0$, $i = 1, \dots, k$, $j = 1, \dots, t$. *Then for any* $v_1, \dots, v_t \in \mathbb{Z}^l$ *there exist an integer* n *and a vector* $u \in \mathbb{Z}^l$ *such that* $u + \sum_{j=1}^t p_{i,j}(n) v_j \in S$ *for each* $i \le k$.

We can also express this theorem in an invariant form similar to [FK1]B:

Theorem B′. *Let* $P{:}\,\mathbb{Z}^r \longrightarrow \mathbb{Z}^l$, $r, l \in \mathbb{N}$, *be a polynomial mapping satisfying* $P(0) = 0$, *let* $F \subset \mathbb{Z}^r$ *be a finite set and let* $S \subseteq \mathbb{Z}^l$ *be a set of positive upper Banach density. Then for some* $n \in \mathbb{N}$ *and* $u \in \mathbb{Z}^l$ *one has* $u + P(nF) \subset S$.

Another corollary of Theorem A (which forms a polynomial generalization of Theorem 7.17, [F2]) is the following Theorem B″. The notion of positive upper Banach density in $\mathbb{R}^n$ (with respect to a sequence of blocks) is defined in complete analogy with formula (0.1). We remark that Theorems B′ and B″ are easily derivable one from another (cf. [F2], pp.152–153).

Theorem B″. *Let* $P{:}\,\mathbb{R}^r \longrightarrow \mathbb{R}^l$, $r, l \in \mathbb{N}$, *be a polynomial mapping satisfying* $P(0) = 0$, *let* $F \subset \mathbb{R}^r$ *be a finite set and let* $S \subseteq \mathbb{R}^l$ *be a set of positive upper Banach density. Then for some* $n \in \mathbb{N}$ *and* $u \in \mathbb{R}^l$ *one has* $u + P(nF) \subset S$.

0.6. The proof of Theorem A is similar in spirit to that of Theorem [FK1]A. Namely, given a dynamical system $\mathbf{X} = (X, \mathcal{B}, \mu, T_1, \dots, T_k)$ and a factor $\mathbf{Y}$ of $\mathbf{X}$ for which Theorem A holds true, one shows that Theorem A is valid for a non-trivial extension of $\mathbf{Y}$. One also shows that the set of factors of $\mathbf{X}$ for which Theorem A holds has a maximal element which therefore has to coincide with $\mathbf{X}$. As in [FK1], it is enough to deal with so called primitive extensions, in which relative compactness and relative weakly mixing properties are controllably combined.

0.7. Relative compactness is treated with the help of an appropriate coloring trick, which utilizes the following *polynomial van der Waerden theorem*, whose proof is given in Section 1:

Theorem C. *Let* (X, ρ) *be a compact metric space,* $T_1, \dots, T_t$ *commuting homeomorphisms of* X *and* $p_{1,1}(n), \dots, p_{1,t}(n), p_{2,1}(n), \dots, p_{2,t}(n), \dots, p_{k,1}(n), \dots, p_{k,t}(n)$ *be polynomials with rational coefficients taking on integer values on the integers and satisfying* $p_{i,j}(0) = 0$, $i = 1, \dots, k$, $j = 1, \dots, t$. *Then, for any positive* ε, *there exist* $x \in X$ *and* $n \in \mathbb{N}$ *such that* $\rho(T_1^{p_{i,1}(n)} T_2^{p_{i,2}(n)} \dots T_t^{p_{i,t}(n)} x, x) < \varepsilon$ *for all* $i = 1, \dots, k$ *simultaneously.*

The special case $k = t$, $p_{i,i}(n) = n$, $p_{i,j}(n) = 0$, $i \neq j$, corresponds to the "linear" topological van der Waerden theorem due to Furstenberg and Weiss [FW].

0.8. As for relative weak mixing, an appropriate generalization of the polynomial ergodic theorem in [B2] is needed (see Proposition 2.3 below). The flavor of it is conveyed by the following "absolute" case of it. (For the general case see Proposition 2.3 below.)

Theorem D. *Let* $(X, \mathcal{B}, \mu, \Gamma)$ *be a measure preserving system, where* Γ *is an abelian group, such that any* $T \in \Gamma$, $T \neq \mathbf{1}_\Gamma$, *is weakly mixing. Let* $T_1, \dots, T_k \in \Gamma$, *and* $p_{1,1}(n), \dots, p_{1,t}(n), p_{2,1}(n), \dots, p_{2,t}(n), \dots, p_{k,1}(n), \dots, p_{k,t}(n)$ *be polynomials with rational coefficients taking on integer values on the integers such that the expressions*

$$g_i(n) = T_1^{p_{i,1}(n)} \dots T_t^{p_{i,t}(n)}, \ i = 1, \dots, k,$$

and the expressions

$$g_i(n)g_l(n)^{-1}=T_1^{p_{i,1}(n)-p_{l,1}(n)}T_2^{p_{i,2}(n)-p_{l,2}(n)}\dots T_t^{p_{i,t}(n)-p_{l,t}(n)},\ i,l=1,\dots,k,\ i\neq l,$$

depend nontrivially on n (namely, all $g_i(n)$ and $g_i(n)g_l(n)^{-1}$ for $i\neq l$ are nonconstant mappings of $\mathbb{Z}$ into Γ). Then, for any $f_i\in L^\infty(X,\mu)$, $i=1,\dots,k$,

$$\lim_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1}\prod_{i=1}^{k}f_i(T_1^{p_{i,1}(n)}\dots T_t^{p_{i,t}(n)}x)-\prod_{i=1}^{k}\int f_i(x)\,d\mu\Big\|_{L^2(X,\mu)}=0.$$

0.9. Theorem C is proved in Section 1. Section 2 is devoted to the treatment of weakly mixing extensions. The proof of our main theorem, Theorem A, is given in Section 3. In Section 4 we treat its combinatorial corollaries, Theorem B and B$'$.

We shall freely use the apparatus of extensions developed in [F1] and [FK1]; see also [F2] and [FKO].

Acknowledgement. We thank N. Hindman, R. McCutcheon and H. Furstenberg for helpful comments regarding the earlier draft of the paper. We would also like to acknowledge a very helpful and constructive referee's report.

1. The polynomial van der Waerden theorem

Our first goal is to prove Theorem C, the "polynomial" version of the van der Waerden theorem. We follow the proof of the "linear" van der Waerden theorem due to Furstenberg and Weiss ([FW]), but instead of the ordinary induction process we shall use what we call *PET-induction* similar to that used in [B2].

1.1. To clarify some of the ideas of the proof of Theorem C we treat first two simple special cases. Recall first the "linear" van der Waerden theorem ([FW]).

Proposition. *Let (X,ρ) be a compact metric space and let T be a homeomorphism of X. Then for any $\varepsilon>0$, any $p\in\mathbb{N}$ and any $c_0,c_1,\dots,c_{p-1}\in\mathbb{Z}$ there exist $x\in X$ and $n\in\mathbb{N}$ such that $\rho(T^{c_in}x,x)<\varepsilon$, $i=0,\dots,p-1$.*

1.2. We shall need the following corollary of Proposition 1.1 (the routine proof of which is given for the convenience of the reader).

Corollary. *If (X,T) is minimal, then for each $\varepsilon>0$ the set of points satisfying the statement of Proposition 1.1 is dense in X.*

Proof. Take an arbitrary nonempty open $U\subseteq X$. Since (X,T) is assumed to be minimal, $X\setminus\bigcup_{m\in\mathbb{Z}}T^{-m}(U)$ is empty; so we can choose a finite covering $X=\bigcup_{j=1}^{k}T^{-m_j}(U)$. Let $\delta>0$ be such that the inequality $\rho(y_1,y_2)<\delta$, $y_1,y_2\in X$, implies $\rho(T^{m_j}y_1,T^{m_j}y_2)<\varepsilon$ for each $j=1,\dots,k$.

Let $y\in X$, $n\in\mathbb{N}$ satisfy $\rho(T^{c_in}y,y)<\delta$, $i=0,\dots,p-1$. Then, taking j for which $y\in T^{-m_j}U$ and $x=T^{m_j}y$, we have $x\in U$ and $\rho(T^{c_in}x,x)<\varepsilon$, $i=0,\dots,p-1$. ■

1.3. Let us consider first the simplest nonlinear case $k=t=1$, $p(n)=p_{1,1}(n)=n^2$. Let (X,ρ) be a compact metric space and let T be a homeomorphism of X. Without loss of generality we shall assume that the system (X,T) is minimal. Let $\varepsilon>0$; we have to find $x\in X$ and $n\in\mathbb{N}$ such that $\rho(T^{n^2}x,x)<\varepsilon$.

We shall find a sequence $x_0, x_1, x_2, \dots$ of points of X and a sequence $n_1, n_2, \dots$ of natural numbers such that

$$\rho(T^{(n_m+\dots+n_{l+1})^2} x_m, x_l) < \varepsilon/2 \text{ for every } l, m \in \mathbb{Z}_+,\ l < m \tag{1.1}$$

(where $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$). Since X is compact, for some $l < m$ one will have $\rho(x_m, x_l) < \varepsilon/2$; together with (1.1) this will give $\rho(T^{(n_m+\dots+n_{l+1})^2} x_m, x_m) < \varepsilon$.

Choose $x_0 \in X$ arbitrarily and put $n_1 = 1$, $x_1 = T^{-n_1^2} x_0$. Let $\varepsilon_1 < \varepsilon/2$ be such that $\rho(T^{n_1^2} y, x_0) < \varepsilon/2$ for every y for which $\rho(y, x_1) < \varepsilon_1$. Find, using Corollary 1.2 (with $\varepsilon = \varepsilon_1/2$, $p = 1$ and $c_0 = 2n_1$), $y_1 \in X$ and $n_2 \in \mathbb{N}$ such that $\rho(y_1, x_1) < \varepsilon_1/2$ and $\rho(T^{2n_1 n_2} y_1, y_1) < \varepsilon_1/2$. Put $x_2 = T^{-n_2^2} y_1$; then

$$\rho(T^{n_2^2} x_2, x_1) = \rho(y_1, x_1) < \varepsilon_1/2 < \varepsilon/2;$$

also,

$$\rho(T^{2n_1 n_2 + n_2^2} x_2, x_1) \le \rho(T^{2n_1 n_2} y_1, y_1) + \rho(y_1, x_1) < \varepsilon_1$$

and, hence, by the choice of ε_1,

$$\rho(T^{(n_1+n_2)^2} x_2, x_0) = \rho(T^{n_1^2} T^{2n_1 + n_2^2} x_2, x_0) < \varepsilon/2.$$

Suppose that x_m, n_m have been found; let us find x_{m+1}, n_{m+1}. Choose ε_m, $0 < \varepsilon_m < \varepsilon/2$, guaranteeing the implication

$$\rho(y, x_m) < \varepsilon_m \Longrightarrow \rho(T^{(n_m+\dots+n_{l+1})^2} y, x_l) < \varepsilon/2,\ l = 0, \dots, m-1,$$

and find (using Corollary 1.2 with $\varepsilon = \varepsilon_m/2$, $p = m$, $c_l = 2(n_m + \dots + n_{l+1})$, $l = 0, \dots, m-1$) y_m, n_{m+1} such that

$$\rho(y_m, x_m) < \varepsilon_m/2,\ \rho(T^{2(n_m+\dots+n_{l+1})n_{m+1}} y_m, y_m) < \varepsilon_m/2,\ l = 0, \dots, m-1.$$

Putting $x_{m+1} = T^{-n_{m+1}^2} y_m$, we obtain

$$\begin{aligned}\rho(T^{2(n_m+\dots+n_{l+1})n_{m+1}+n_{m+1}^2} x_{m+1}, x_m) &\le \rho(T^{2(n_m+\dots+n_{l+1})n_{m+1}} y_m, y_m) \\ &+ \rho(y_m, x_m) < \varepsilon_m, \quad l = 0, \dots, m-1\end{aligned}$$

and, hence, by the choice of ε_m,

$$\rho(T^{n_{m+1}^2} x_{m+1}, x_m) < \varepsilon/2$$

and

$$\rho(T^{(n_{m+1}+\dots+n_{l+1})^2} x_{m+1}, x_l) < \varepsilon/2 \text{ for } l = 0, \dots, m-1.$$

1.4. Our second example is $k = 2$, $t = 1$, $p_{1,1} = n^2$, $p_{1,2} = 2n^2$, that is, for any $\varepsilon > 0$, we want to find $x \in X$, $n \in \mathbb{N}$ for which $\rho(T^{n^2} x, x) < \varepsilon$, $\rho(T^{2n^2} x, x) < \varepsilon$. Consider the following statements (in all of them (X, T) is assumed to be a minimal system):

(i) (The linear case.) For any $\varepsilon > 0$, for any $q \in \mathbb{N}$ and any $c_0, \dots, c_{q-1} \in \mathbb{Z}$ there exist $x \in X$ and $n \in \mathbb{N}$ such that

$$\rho(T^{c_i n} x, x) < \varepsilon, \quad i = 0, \dots, q-1.$$

(ii)$_0$ For any $\varepsilon > 0$, for any $p \in \mathbb{N}$ and any $b_0, \dots, b_{p-1} \in \mathbb{Z}$ there exist $x \in X$ and $n \in \mathbb{N}$ such that

$$\rho(T^{n^2+b_i n}x, x) < \varepsilon, \quad i = 0, \dots, p-1.$$

(ii)$_q$, $q \in \mathbb{N}$. For any $\varepsilon > 0$, for any $p \in \mathbb{N}$ and any $b_0, \dots, b_{p-1} \in \mathbb{Z}$, $c_0, \dots, c_{q-1} \in \mathbb{Z}$ there exist $x \in X$ and $n \in \mathbb{N}$ such that

$$\rho(T^{n^2+b_i n}x, x) < \varepsilon, \quad i = 0, \dots, p-1,$$
$$\rho(T^{c_j n}x, x) < \varepsilon, \quad j = 0, \dots, q-1.$$

(iii) For any $\varepsilon > 0$ there exist $x \in X$ and $n \in \mathbb{N}$ such that

$$\rho(T^{n^2}x, x) < \varepsilon, \quad \rho(T^{2n^2}x, x) < \varepsilon.$$

We claim that the following implications hold:
(i)$\Longrightarrow$**(ii)**$_0$, **(ii)**$_{q-1}$$\Longrightarrow$**(ii)**$_q$ for any $q \in \mathbb{N}$, $\{$**(ii)**$_q$ for all $q \in \mathbb{Z}_+\}$ $\Longrightarrow$**(iii)**.

We remark that in each of the statements above the existence of one point satisfying it implies the existence of a dense set of such points for any $\varepsilon > 0$; see Corollary 1.2, or Corollary 1.8 below for a stronger statement.

(i)$\Longrightarrow$**(ii)**$_0$**.** We are going to find a sequence $x_0, x_1, x_2, \dots$ of points of X and a sequence $n_1, n_2, \dots$ of natural numbers such that

$$\rho(T^{(n_m+\dots+n_{l+1})^2+b_i(n_m+\dots+n_{l+1})}x_m, x_l) < \varepsilon/2, \quad i = 0, \dots, p-1, \tag{1.2}$$

for every $l, m \in \mathbb{Z}_+$, $l < m$. For some $l < m$ one will have $\rho(x_m, x_l) < \varepsilon/2$. Together with (1.2) this will ensure that

$$\rho(T^{(n_m+\dots+n_{l+1})^2+b_i(n_m+\dots+n_{l+1})}x_m, x_m) < \varepsilon, \quad i = 0, \dots, p-1.$$

Putting $n = n_m + \dots + n_{l+1}$ and $x = x_m$ we will be done.

Choose $x_0 \in X$ arbitrarily. Using statement **(i)** (with $q = p$), find $y_0 \in X$ and $n_1 \in \mathbb{N}$ such that $\rho(y_0, x_0) < \varepsilon/4$ and

$$\rho(T^{b_i n_1}y_0, y_0) < \varepsilon/4, \quad i = 0, \dots, p-1.$$

Put $x_1 = T^{-n_1^2}y_0$. Then for $i = 0, \dots, p-1$

$$\rho(T^{n_1^2+b_i n_1}x_1, x_0) = \rho(T^{b_i n_1}y_0, x_0) \le \rho(T^{b_i n_1}y_0, y_0) + \rho(y_0, x_0) < \varepsilon/2.$$

Suppose that x_m, n_m have been found; let us find x_{m+1}, n_{m+1}. Choose ε_m, $0 < \varepsilon_m < \varepsilon/2$, guaranteeing that

$$\rho(y, x_m) < \varepsilon_m \Longrightarrow \rho(T^{(n_m+\dots+n_{l+1})^2+b_i(n_m+\dots+n_{l+1})}y, x_l) < \varepsilon/2,$$
$$i = 0, \dots, p-1, \quad l = 0, \dots, m-1.$$

Now, using statement **(i)** (with $\varepsilon = \varepsilon_m/2$, $q = (m+1)p$, $c_{i,l} = 2(n_m+\dots+n_{l+1})+b_i$, $l = 0, \dots, m-1$, $c_{i,m} = b_i$, $i = 0, \dots, p-1$), find $y_m \in X$ and $n_{m+1} \in \mathbb{N}$ such that $\rho(y_m, x_m) < \varepsilon_m/2$ and

$$\rho(T^{(2(n_m+\dots+n_{l+1})+b_i)n_{m+1}}y_m, y_m) < \varepsilon_m/2, \ i = 0, \dots, p-1, \ l = 0, \dots, m-1,$$

$$\text{and } \rho(T^{b_i n_{m+1}}y_m, y_m) < \varepsilon_m/2 \ \ i = 0, \dots, p-1.$$

Putting $x_{m+1} = T^{-n_{m+1}^2} y_m$, we obtain

$$\rho(T^{(2(n_m+\ldots+n_{l+1})+b_i)n_{m+1}+n_{m+1}^2} x_{m+1}, x_m) \leq \rho(T^{(2(n_m+\ldots+n_{l+1})+b_i)n_{m+1}} y_m, y_m) \\ +\rho(y_m, x_m) < \varepsilon_m, \quad i = 0, \ldots, p-1, \quad l = 0, \ldots, m-1,$$

$$\rho(T^{b_i n_{m+1}+n_{m+1}^2} x_{m+1}, x_m) \leq \rho(T^{b_i n_{m+1}} y_m, y_m) + \rho(y_m, x_m) < \varepsilon_m, \\ i = 0, \ldots, p-1,$$

and, hence, by the choice of ε_m,

$$\rho(T^{(n_{m+1}+\ldots+n_{l+1})^2+b_i(n_{m+1}+\ldots+n_{l+1})} x_{m+1}, x_l) < \varepsilon/2, \\ i = 0, \ldots, p-1, \quad l = 0, \ldots, m.$$

(ii)$_{q-1}$$\Longrightarrow$**(ii)**$_q$**.** We are looking for a sequence $x_0, x_1, x_2, \ldots$ of points of X and a sequence $n_1, n_2, \ldots$ of natural numbers such that

$$\rho(T^{(n_m+\ldots+n_{l+1})^2+b_i(n_m+\ldots+n_{l+1})} x_m, x_l) < \varepsilon/2, \quad i = 0, \ldots, p-1, \\ \rho(T^{c_j(n_m+\ldots+n_{l+1})} x_m, x_l) < \varepsilon/2, \quad j = 0, \ldots, q-1, \tag{1.3}$$

for every $l, m \in \mathbb{Z}_+$, $l < m$. For some $l < m$ one will have $\rho(x_m, x_l) < \varepsilon/2$. Together with (1.3) this will ensure that

$$\rho(T^{n_m+\ldots+n_{l+1})^2+b_i(n_m+\ldots+n_{l+1})} x_m, x_m) < \varepsilon, \quad i = 0, \ldots, p-1, \\ \rho(T^{c_j(n_m+\ldots+n_{l+1})} x_m, x_m) < \varepsilon, \quad j = 0, \ldots, q-1.$$

Choose $x_0 \in X$ arbitrarily. By statement **(ii)**$_{q-1}$, there exist $y_0 \in X$ and $n_1 \in \mathbb{N}$ such that $\rho(y_0, x_0) < \varepsilon/4$ and

$$\rho(T^{n_1^2+(b_i-c_0)n_1} y_0, y_0) < \varepsilon/4, \quad i = 0, \ldots, p-1, \\ \rho(T^{(c_j-c_0)n_1} y_0, y_0) < \varepsilon/4, \quad j = 1, \ldots, q-1.$$

Put $x_1 = T^{-c_0 n_1} y_0$. Then

$$\rho(T^{n_1^2+b_i n_1} x_1, x_0) \leq \rho(T^{n_1^2+(b_i-c_0)n_1} y_0, y_0) + \rho(y_0, x_0) < \varepsilon/2, \quad i = 0, \ldots, p-1, \\ \rho(T^{c_j n_1} x_1, x_0) \leq \rho(T^{(c_j-c_0)n_1} y_0, y_0) + \rho(y_0, x_0) < \varepsilon/2, \quad j = 1, \ldots, q-1, \\ \rho(T^{c_0 n_1} x_1, x_0) = \rho(y_0, x_0) < \varepsilon/2.$$

Suppose that x_m, n_m have been found; let us find x_{m+1}, n_{m+1}. Choose ε_m, $0 < \varepsilon_m < \varepsilon/2$, guaranteeing the implication

$$\rho(y, x_m) < \varepsilon_m \Longrightarrow$$

$$\rho(T^{(n_m+\ldots+n_{l+1})^2+b_i(n_m+\ldots+n_{l+1})} y, x_l) < \varepsilon/2, \quad i = 0, \ldots, p-1, \quad \text{and} \\ \rho(T^{c_j(n_m+\ldots+n_{l+1})} y, x_l) < \varepsilon/2, \quad j = 0, \ldots, q-1, \quad \text{for } l = 0, \ldots, m-1.$$

Now, using statement **(ii)**$_{q-1}$ (with $\varepsilon = \varepsilon_m/2$), find $y_m \in X$ and $n_{m+1} \in \mathbb{N}$ such that $\rho(y_m, x_m) < \varepsilon_m/2$ and

$$\rho(T^{n_{m+1}^2+(2(n_m+\ldots+n_{l+1})+b_i-c_0)n_{m+1}} y_m, y_m) < \varepsilon_m/2, \\ i = 0, \ldots, p-1, \quad l = 0, \ldots, m-1,$$

$$\rho(T^{n_{m+1}^2+(b_i-c_0)n_{m+1}} y_m, y_m) < \varepsilon_m/2, \quad i = 0, \ldots, p-1,$$

$$\text{and} \quad \rho(T^{(c_j-c_0)n_{m+1}} y_m, y_m) < \varepsilon_m/2, \quad j = 1, \ldots, q-1.$$

Putting $x_{m+1}=T^{-c_0n_{m+1}}y_m$, we obtain

$$\rho(T^{n_{m+1}^2+2(n_m+\ldots+n_{l+1})n_{m+1}+b_in_{m+1}}x_{m+1},x_m)$$
$$\le \rho(T^{n_{m+1}^2+(2(n_m+\ldots+n_{l+1})+b_i-c_0)n_{m+1}}y_m,y_m)+\rho(y_m,x_m)<\varepsilon_m,$$
$$i=0,\ldots,p-1,\quad l=0,\ldots,m-1,$$

$$\rho(T^{n_{m+1}^2+b_in_{m+1}}x_{m+1},x_m)\le\rho(T^{n_{m+1}^2+(b_i-c_0)n_{m+1}}y_m,y_m)+\rho(y_m,x_m)<\varepsilon_m,$$
$$i=0,\ldots,p-1,$$

$$\rho(T^{c_jn_{m+1}}x_{m+1},x_m)\le\rho(T^{(c_j-c_0)n_{m+1}}y_m,y_m)+\rho(y_m,x_m)<\varepsilon_m,$$
$$j=1,\ldots,q-1,$$

$$\rho(T^{c_0n_{m+1}}x_{m+1},x_m)=\rho(y_m,x_m)<\varepsilon_m,$$

and, hence, by the choice of ε_m,

$$\rho(T^{(n_{m+1}+\ldots+n_{l+1})^2+b_i(n_{m+1}+\ldots+n_{l+1})}x_{m+1},x_l)<\varepsilon/2,$$
$$i=0,\ldots,p-1,\quad l=0,\ldots,m,$$

$$\rho(T^{c_j(n_{m+1}+\ldots+n_{l+1})}x_{m+1},x_l)<\varepsilon/2,\quad j=0,\ldots,q-1,\quad l=0,\ldots,m.$$

$\{$**(ii)**$_q$ for all $q\in\mathbb{Z}_+\}$ $\Longrightarrow$**(iii).** We are looking for a sequence $x_0,x_1,x_2,\ldots$ of points of X and a sequence $n_1,n_2,\ldots$ of natural numbers such that

$$\rho(T^{(n_m+\ldots+n_{l+1})^2}x_m,x_l)<\varepsilon/2 \text{ and}$$
(1.4)
$$\rho(T^{2(n_m+\ldots+n_{l+1})^2}x_m,x_l)<\varepsilon/2 \text{ for every } l,m\in\mathbb{Z}_+,\ l<m.$$

For some $l<m$ one will have $\rho(x_m,x_l)<\varepsilon/2$. Together with (1.4) this will ensure $\rho(T^{(n_m+\ldots+n_{l+1})^2}x_m,x_m)<\varepsilon$, $\rho(T^{2(n_m+\ldots+n_{l+1})^2}x_m,x_m)<\varepsilon$. Putting $n=n_m+\ldots+n_{l+1}$ and $x=x_m$ finishes the proof.

Choose $x_0\in X$ arbitrarily. By statement **(ii)**$_0$, there exist $y_0\in X$ and $n_1\in\mathbb{N}$ such that $\rho(y_0,x_0)<\varepsilon/4$ and $\rho(T^{n_1^2}y_0,y_0)<\varepsilon/4$. Put $x_1=T^{-n_1^2}y_0$. Then

$$\rho(T^{n_1^2}x_1,x_0)=\rho(y_0,x_0)<\varepsilon/2,$$

$$\rho(T^{2n_1^2}x_1,x_0)=\rho(T^{n_1^2}y_0,x_0)\le\rho(T^{n_1^2}y_0,y_0)+\rho(y_0,x_0)<\varepsilon/2.$$

Suppose that x_m, n_m have been found; let us find x_{m+1}, n_{m+1}. Choose ε_m, $0<\varepsilon_m<\varepsilon/2$, guaranteeing that

$$\rho(y,x_m)<\varepsilon_m\Longrightarrow\rho(T^{(n_m+\ldots+n_{l+1})^2}y,x_l)<\varepsilon/2\ \text{ and}$$

$$\rho(T^{2(n_m+\ldots+n_{l+1})^2}y,x_l)<\varepsilon/2\ \text{ for } l=0,\ldots,m-1.$$

Now, using statement **(ii)**$_m$ (with $\varepsilon=\varepsilon_m/2$, $p=m+1$, $c_l=2(n_m+\ldots+n_{l+1})$, $b_l=4(n_m+\ldots+n_{l+1})$, $l=0,\ldots,m-1$, $b_m=0$), find $y_m\in X$ and $n_{m+1}\in\mathbb{N}$ such that $\rho(y_m,x_m)<\varepsilon_m/2$ and

$$\rho(T^{2(n_m+\ldots+n_{l+1})n_{m+1}}y_m,y_m)<\varepsilon_m/2,\quad l=0,\ldots,m-1,$$

$$\rho(T^{n_{m+1}^2+4(n_m+\ldots+n_{l+1})n_{m+1}}y_m,y_m)<\varepsilon_m/2,\quad l=0,\ldots,m-1,$$

$$\rho(T^{n_{m+1}^2}y_m,y_m)<\varepsilon_m/2.$$

Putting $x_{m+1} = T^{-n_{m+1}^2} y_m$, we obtain

$$\rho(T^{2(n_m+\ldots+n_{l+1})n_{m+1}+n_{m+1}^2} x_{m+1}, x_m) \le \rho(T^{2(n_m+\ldots+n_{l+1})n_{m+1}} y_m, y_m)$$
$$+\rho(y_m, x_m) < \varepsilon_m, \quad l = 0, \ldots, m-1,$$

$$\rho(T^{4(n_m+\ldots+n_{l+1})n_{m+1}+2n_{m+1}^2} x_{m+1}, x_m) \le \rho(T^{4(n_m+\ldots+n_{l+1})n_{m+1}+n_{m+1}^2} y_m, y_m)$$
$$+\rho(y_m, x_m) < \varepsilon_m, \quad l = 0, \ldots, m-1,$$

$$\rho(T^{2n_{m+1}^2} x_{m+1}, x_m) \le \rho(T^{n_{m+1}^2} y_m, y_m) + \rho(y_m, x_m) < \varepsilon_m$$

and, hence, by the choice of ε_m,

$$\rho(T^{(n_{m+1}+\ldots+n_{l+1})^2} x_{m+1}, x_l) < \varepsilon/2, \quad \rho(T^{2(n_{m+1}+\ldots+n_{l+1})^2} x_{m+1}, x_l) < \varepsilon/2$$
$$\text{for } l = 0, \ldots, m.$$

1.5. Before embarking on the proof of Theorem C we shall introduce some technical definitions and notation. We shall call the polynomials we are working with, namely "the polynomials with rational coefficients taking on integer values on the integers and zero value at zero", *integral polynomials*. Throughout the following preliminary discussion and proof of Theorem C, the integer t, as well as D, the maximal degree of the polynomials $p_{i,j}(n)$, $i = 1, \ldots, k$, $j = 1, \ldots, t$, appearing in the formulation of Theorem C will be fixed. Expressions of the form $T_1^{p_1(n)} \ldots T_t^{p_t(n)}$, where $p_i(n)$ are integral polynomials with $\deg p_i(n) \le D$, $i = 1, \ldots, t$, will be called *polynomial expressions*. Products of polynomial expressions and their inverses are polynomial expressions as well:

$$g(n) = T_1^{p_1(n)} \ldots T_t^{p_t(n)}, \ h(n) = T_1^{q_1(n)} \ldots T_t^{q_t(n)} \implies$$
$$gh(n) = T_1^{p_1(n)+q_1(n)} \ldots T_t^{p_t(n)+q_t(n)}, \ g^{-1}(n) = T_1^{-p_1(n)} \ldots T_t^{-p_t(n)}.$$

The set of polynomial expressions is a group, denote this group by **PE**. Note also that polynomial expressions can be shifted along $\mathbb{Z}$:

$$g^{-1}(n_0) g(n + n_0) = T_1^{p_1(n+n_0)-p_1(n_0)} \ldots T_t^{p_t(n+n_0)-p_t(n_0)} \in \mathbf{PE} \text{ for any } n_0 \in \mathbb{Z}.$$

The degree, $\deg(g(n))$, of the polynomial expression $g(n) = T_1^{p_1(n)} \ldots T_t^{p_t(n)}$ is $\max_{i=1,\ldots,t}\{\deg(p_i(n))\}$; its *weight*, $w(g(n))$, is the pair of integers (r, d) defined by the condition $\deg p_{r+1}(n) = \ldots = \deg p_t(n) = 0$, $\deg p_r(n) = d \ge 1$. The weight (r, d) *is greater* than (s, e) if $r > s$ or if $r = s$ and $d > e$.

Examples. The polynomial expression $T_1^n T_2^0 \ldots T_t^0$ has degree 1 and weight $(1, 1)$; $T_1^{9n^2+4n} T_2^{3n^7+7n^4} T_3^{3n^2+19n} T_4^0 T_5^0$ has degree 7 and weight $(3, 2)$.

Two polynomial expressions, say $T_1^{p_1(n)} \ldots T_t^{p_t(n)}$ and $T_1^{q_1(n)} \ldots T_t^{q_t(n)}$, will be called *equivalent* if they have the same weight (r, d) and the leading coefficients of the polynomials $p_r(n)$, $q_r(n)$ coincide as well. If C is a set of equivalent polynomial expressions, its *weight*, $w(C)$ is by definition the weight of any of its members.

We shall call any finite subset of **PE** *a system*. *The degree* of a system is the maximal degree of its elements. For every system A form *the weight matrix* $\begin{pmatrix} N_{1,1} & \ldots & N_{1,D} \\ \vdots & \vdots & \vdots \\ N_{t,1} & \ldots & N_{t,D} \end{pmatrix}$, where $N_{s,d}$ is the number of equivalence classes formed by the elements of the system whose weights are (s, d).

Example. The system $\{T_1^{19n}T_2^0, T_1^{6n^2}T_2^0, T_1^{7n^2+19n}T_2^0, T_1^{7n^2}T_2^0, T_1^{4n^4}T_2^{n^2}, T_1^{n^2}T_2^{3n^3}, T_1^{n^2}T_2^{3n^3+2n}, T_1^nT_2^{2n^3+3n}, T_1^{10n^5}T_2^{n^3+4n^2+4n}, T_1^0T_2^{n^3+2n}, T_1^{n^5}T_2^{n^3+n^2}\}$ has weight matrix

$$\begin{pmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 & 0 \end{pmatrix}$$

(we assumed here $t = 2$ and $D = 5$).

1.6. We are going now to describe the PET-induction scheme (PET stands for Polynomial Ergodic Theorem.) Of course, this is just an induction over a particular well ordered set (of weight matrices).

Assume that a statement S is valid for the (trivial) system whose weight matrix is zero (this means that all of the polynomials in the exponents of the elements of the system are zeroes) and suppose that we were able to show that the truth of S for any system having a weight matrix of the form

$$M = \begin{pmatrix} 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & N_{r,d} & N_{r,d+1} & \dots & N_{r,D} \\ N_{r+1,1} & \dots & N_{r+1,d-1} & N_{r+1,d} & N_{r+1,d+1} & \dots & N_{r+1,D} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ N_{t,1} & \dots & N_{t,d-1} & N_{t,d} & N_{t,d+1} & \dots & N_{t,D} \end{pmatrix},$$

where $N_{r,d} \geq 1$, follows from its truth for all systems having a weight matrix of the form

$$M' = \begin{pmatrix} * & \dots & * & * & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & \dots & * & * & * & \dots & * \\ * & \dots & * & N_{r,d}-1 & N_{r,d+1} & \dots & N_{r,D} \\ N_{r+1,1} & \dots & N_{r+1,d-1} & N_{r+1,d} & N_{r+1,d+1} & \dots & N_{r+1,D} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ N_{t,1} & \dots & N_{t,d-1} & N_{t,d} & N_{t,d+1} & \dots & N_{t,D} \end{pmatrix},$$

where "$*$" means "any nonnegative integer". (We shall say that any weight matrix of the form M' *precedes* the weight matrix M). Then the statement is valid for all systems.

Indeed, starting with the trivial system and proceeding step by step, one checks in turn the validity of S for systems with weight matrices $\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\begin{pmatrix} 2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\dots$, $\begin{pmatrix} * & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\begin{pmatrix} 2 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\dots$, $\begin{pmatrix} * & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\dots$, $\begin{pmatrix} * & 2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\dots$, $\begin{pmatrix} * & * & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$, $\dots$, $\begin{pmatrix} * & * & \dots & * \\ * & * & \dots & * \\ \vdots & \vdots & \vdots & \vdots \\ * & * & \dots & * \end{pmatrix}$.

1.7. Proof of Theorem C. Denote by Γ the (commutative) group generated by $T_1, \dots, T_t$; passing if needed to a suitable subset of X, we may assume that the dynamical system (X, Γ) is minimal.

First of all notice that Theorem C holds trivially if all the polynomials in the exponents of T_j, $j = 1, \dots, t$, in the polynomial expressions $g_i(n) = T_1^{p_{i,1}(n)} \dots T_t^{p_{i,t}(n)}$, $i = 1, \dots, k$, are zeros. We shall prove (using PET-induction) that Theorem C is valid for any system by showing that its validity for arbitrary system $A = \{g_1(n), \dots, g_k(n)\}$ follows from its validity for all the systems whose weight matrices *precede* the weight matrix of A.

Choose $\varepsilon > 0$.

Let $g_1(n)$ be of the minimal weight in A; we may assume that A does not contain trivial polynomial expressions and, so, $w(g_1(n)) \geq (1, 1)$. Consider the system

$$A_0 = \{g_2(n)g_1^{-1}(n),\ g_3(n)g_1^{-1}(n),\ \dots,\ g_k(n)g_1^{-1}(n)\} \subset \mathbf{PE}$$

(if $k = 1$, A_0 is empty).

Notice that the elements of A nonequivalent to $g_1(n)$ do not change their weights and the equivalence of one to another after they have been multiplied by $g_1^{-1}(n)$; on the other hand, the weights of elements of A which are equivalent to $g_1(n)$ do decrease after these elements have been multiplied by $g_1^{-1}(n)$. Hence, the number of equivalence classes with the minimal weight in A decreases by 1 when we pass from A to A_0 (although some new equivalence classes with smaller weights can arise in A_0). This means that the weight matrix of A_0 precedes that of A, and by the PET-induction hypothesis, the statement of the theorem is valid for A_0.

Therefore, one can choose $y_0 \in X$, $n_1 \in \mathbb{N}$ such that

$$\rho(g_i(n_1)g_1^{-1}(n_1)y_0, y_0) < \varepsilon/2 \text{ for } i = 2, \dots, k$$

(if $k = 1$, let $y_0 \in X$ and $n_1 \in \mathbb{N}$ be arbitrary). Denote $x_0 = y_0$, $x_1 = g_1^{-1}(n_1)y_0$; then

$$g_1(n_1)x_1 = x_0,\ \rho(g_i(n_1)x_1, x_0) < \varepsilon/2 \text{ for } i = 2, \dots, k.$$

We will find a sequence of points $x_0, x_1, x_2, \dots \in X$ and a sequence of natural numbers $n_1, n_2, \dots$ such that for every l, m, $l < m$, one has

$$\rho(g_i(n_m + \dots + n_{l+1})x_m, x_l) < \varepsilon/2 \text{ for any } i = 1, \dots, k. \tag{1.5}$$

The points x_0, x_1 and the natural number n_1 have already been chosen; suppose that x_m, n_m have been chosen. The inequality (1.5) holds not only for x_m but also for all points of ε_m-neighborhood of x_m for some ε_m, $0 < \varepsilon_m < \varepsilon/2$. Since (X, Γ) is assumed to be minimal, there exists a finite set of elements of Γ, say $S_1, \dots, S_s \in \Gamma$, such that for every $y \in X$ there exists $t = t(y) \leq s$ such that $\rho(S_t y, x_m) < \varepsilon_m/2$. Choose δ_m such that, for every $y \in X$ there is some t, $1 \leq t \leq s$, so that the inequality $\rho(y, y') < \delta_m$ implies

$$\rho(S_t y', x_m) < \varepsilon_m. \tag{1.6}$$

Form the system

$$A_m = \left\{ \begin{array}{c} g_{1,m}(n) = g_1(n)g_1^{-1}(n), \\ g_{2,m}(n) = g_2(n)g_1^{-1}(n), \\ \vdots \\ g_{k,m}(n) = g_k(n)g_1^{-1}(n), \\ g_{1,m-1}(n) = g_1^{-1}(n_m)g_1(n+n_m)g_1^{-1}(n), \\ g_{2,m-1}(n) = g_2^{-1}(n_m)g_2(n+n_m)g_1^{-1}(n), \\ \vdots \\ g_{k,m-1}(n) = g_k^{-1}(n_m)g_k(n+n_m)g_1^{-1}(n), \\ \vdots \\ g_{1,0}(n) = g_1^{-1}(n_m+\ldots+n_1)g_1(n+n_m+\ldots+n_1)g_1^{-1}(n), \\ g_{2,0}(n) = g_2^{-1}(n_m+\ldots+n_1)g_2(n+n_m+\ldots+n_1)g_1^{-1}(n), \\ \vdots \\ g_{k,0}(n) = g_k^{-1}(n_m+\ldots+n_1)g_k(n+n_m+\ldots+n_1)g_1^{-1}(n) \end{array} \right\}$$

If $g_i(n)$ is not equivalent to $g_1(n)$, the polynomial expressions $g_{i,0}(n), \ldots, g_{i,m}(n) \in A_m$ have the same weights as $g_i(n)$ itself and their equivalence is preserved, that is, if $g_i(n)$ is equivalent to $g_l(n)$ then $g_{i,r}(n)$ is equivalent to $g_{l,s}(n)$ for every $r, s = 0, \ldots, m$. If $g_i(n)$ is equivalent to $g_1(n)$, the weights of these polynomial expressions decrease: $w(g_{i,r}(n)) < w(g_i(n)) = w(g_1(n))$. So, the number of equivalence classes having weights greater than $w(g_1(n))$ does not change whereas the number of equivalence classes of polynomial expressions having the minimal weight in A decreases by 1 when we pass from A to A_m. This means that the weight matrix of A_m precedes that of A and, by our PET-induction hypothesis, the conclusion of Theorem C holds for system A_m.

Hence, we can find $y_m \in X$, $n_{m+1} \in \mathbb{N}$ such that $\rho(h(n_{m+1})y_m, y_m) < \delta_m$ for every $h \in A_m$. Choose t, $1 \le t \le s$, such that (1.6) holds for all y' from the δ_m-neighborhood of y_m; denote $x_{m+1} = g_1^{-1}(n_{m+1})S_t y_m$. Then, since for every i, $1 \le i \le k$ and any l, $0 \le l \le m$,

$$\rho(g_{i,l}(n_{m+1})y_m, y_m) < \delta_m,$$

we have

$$\begin{gathered} \rho(S_t g_{i,l}(n_{m+1})y_m, x_m) \\ = \rho(g_i^{-1}(n_m+\ldots+n_{l+1})g_i(n_{m+1}+n_m+\ldots+n_{l+1})x_{m+1}, x_m) < \varepsilon_m, \\ l = 0, \ldots, m-1, \end{gathered}$$

and

$$\rho(g_i(n_{m+1})x_{m+1}, x_m) < \varepsilon_m.$$

Hence, by the choice of ε_m, one gets

$$\rho(g_i(n_{m+1}+n_m+\ldots+n_{l+1})x_{m+1}, x_l) < \varepsilon/2 \text{ for } l = 0, \ldots, m-1$$

and

$$\rho(g_i(n_{m+1})x_{m+1}, x_m) < \varepsilon/2.$$

Since X is compact, there exist l, m, $l < m$, such that $\rho(x_m, x_l) < \varepsilon/2$. Denote $n = n_m + \ldots + n_{l+1}$; then $\rho(g_i(n)x_m, x_l) < \varepsilon/2$ for every i, $1 \le i \le k$, and, hence, $\rho(g_i(n)x_m, x_m) < \varepsilon$. ■

1.8. Corollary. *If (X,Γ) is a minimal system, then for almost all (in sense of category) points $x \in X$ there exists a sequence $n_m \to \infty$ such that $g_i(n_m)x \to x$ simultaneously for all $i = 1,\dots,k$.*

Proof. Call a point $x \in X$ for which there exists $n \in \mathbb{N}$ such that $\rho(g_i(n)x, x) < \varepsilon$ for each $i = 1,\dots,k$, *ε-recurrent*. Call a point $x \in X$ *recurrent* if it is ε-recurrent for any $\varepsilon > 0$. We have to prove that the set of recurrent points is residual (that is, its complement in X is the union of countable number of closed nowhere dense sets).

Take an arbitrary nonempty open $U \subseteq X$. Since X is assumed to be minimal with respect to the action of Γ, $X \setminus \bigcup_{T\in\Gamma} T^{-1}(U)$ is empty; so we can choose a finite covering $X = \bigcup_{t=1}^{s} S_j^{-1}(U)$, $S_1,\dots,S_s \in \Gamma$. Let $\delta > 0$ be such that the inequality $\rho(y_1,y_2) < \delta$, $y_1,y_2 \in X$, implies $\rho(S_t y_1, S_t y_2) < \varepsilon$ for each $t = 1,\dots,s$. Theorem C says that there exist $y \in X$, $n \in \mathbb{N}$ satisfying $\rho(g_i(n)y, y) < \delta$, $i = 1,\dots,k$. Then, taking t for which $y \in S_t^{-1}U$ and $x = S_t y$, we have $x \in U$ and $\rho(g_i(n)x, x) < \varepsilon$, $i = 0,\dots,k$.

We have obtained that for any $\varepsilon > 0$ the set W_ε of ε-recurrent points is dense in X; it is clear also that W_ε is open. Therefore, the sets $Z_n = X \setminus W_{1/n}$, $n \in \mathbb{N}$, are closed and nowhere dense in X. Hence, the set $\bigcap_{n\in\mathbb{N}} W_{1/n} = X \setminus \bigcup_{n\in\mathbb{N}} Z_n$ of recurrent points is residual. ■

1.9. Before formulating the next corollary, recall that *the IP-set* generated by a sequence $\{s_m\}_{m\in\mathbb{N}}$ in $\mathbb{N}$ is defined by

$$FS(\{s_m\}) = \Big\{\sum_{m\in F} s_m :\ F \subset \mathbb{N},\ \#F < \infty\Big\}.$$

A set $P \subseteq \mathbb{N}$ is called *an IP*-set* if it has nontrivial intersection with any IP-subset of $\mathbb{N}$. It is not hard to see that any IP*-set has bounded gaps (since any set containing arbitrarily long intervals contains an IP-set).

Corollary.(of the proof) *For any IP-set I the integer n in Theorem C can be chosen from I. If (X,Γ) is a minimal system, the set*

$$P = \Big\{n :\ g_1^{-1}(n)U \cap \ldots \cap g_k^{-1}(n)U \cap U \neq \emptyset\Big\}$$

is an IP*-set for any nonempty open $U \subseteq X$.

Proof. Let $I = FS\Big(\{s_m\}_{m\in\mathbb{N}}\Big)$ be an IP-set. For $n \in I$, $n = s_{i_1} + \ldots + s_{i_m}$, define its *support*, $\sigma(n)$ by $\sigma(n) = \{i_1,\dots,i_m\}$ and let

$$I_n = FS\Big(\{s_m\}_{m\in\mathbb{N}\setminus\sigma(n)}\Big);$$

for $n_1,\dots,n_q \in I$ define

$$I_{n_1,\dots,n_q} = FS\Big(\{s_m\}_{m\in\mathbb{N}\setminus\bigcup_{j=1}^{q}\sigma(n_j)}\Big).$$

It is clear from the definition of IP-set that for any $n \in I$, $l \in I_n$ we have $n + l \in I$.

The statement of the corollary is trivial if all $g_i(n) \equiv \mathbf{1}_\Gamma$, that is, for trivial systems. We use PET-induction: assume that the statement is valid for the systems

$A_0, A_1, \dots$, defined in the proof of Theorem C. Then, in this proof, one can choose the numbers $n_1, n_2, \dots$ so that $n_1 \in I$, $n_{m+1} \in I_{n_1,\dots,n_m}$, $m \in \mathbb{N}$. It follows that for any integers $m > l \geq 0$, $n_m + \dots + n_{l+1} \in I$.

Hence, for any $\varepsilon > 0$, the set

$$\Bigl\{x \in X : \ \exists n \in I : \ \rho(g_i(n)x, x) < \varepsilon\Bigr\}$$

is nonempty. If (X, Γ) is minimal, the same arguments as in the proof of Corollary 1.8 show that it is residual. In this case, for any nonempty open $U \subseteq X$ and any IP-set I there exists $n \in I$ such that $g_i^{-1}(n)U \cap U \neq \emptyset$; this just means that P is an IP*-set. ∎

1.10. Theorem C admits an obvious formulation which is valid for non-metrizable compact spaces as well.

Proposition. *Let $\{U_1, \dots, U_r\}$ be an open covering of a compact topological space X, let $T_1, \dots, T_t$ be commuting homeomorphisms of X and let $p_{1,1}(n), \dots, p_{1,t}(n)$, $p_{2,1}(n), \dots, p_{2,t}(n), \dots, p_{k,1}(n), \dots, p_{k,t}(n)$ be polynomials with rational coefficients taking on integer values on the integers and satisfying $p_{i,j}(0) = 0$, $i = 1, \dots, k$, $j = 1, \dots, t$. Then there exist $1 \leq q \leq r$ and $n \in \mathbb{N}$ such that*

$$U_q \cap T_1^{p_{1,1}(n)} \dots T_t^{p_{1,t}(n)} U_q \cap \dots \cap T_1^{p_{k,1}(n)} \dots T_t^{p_{k,t}(n)} U_q \neq \emptyset.$$

We leave it to the reader to verify that the same proof as in 1.7, but written in language of neighborhoods, goes through (cf. [BPT] where this is done for the "linear" topological van der Waerden Theorem of Furstenberg and Weiss). We remark also that the non-metrizable fact follows from metrizable one by an application of Corollary 1.11 below.

1.11. Theorem C has a series of "chromatic" corollaries; the following two will be used in Section 3:

Corollary. *For any natural numbers K, k, t and l, for any integral polynomials $p_{1,1}(n), \dots, p_{1,t}(n), p_{2,1}(n), \dots, p_{2,t}(n), \dots, p_{k,1}(n), \dots, p_{k,t}(n)$ and for any vectors $v_1, \dots, v_t \in V = \mathbb{Z}^l$ there exist a constant M and a finite set $Q \subset V$ such that for every mapping $\chi\colon V \longrightarrow \{1, \dots, K\}$ there exist $u \in Q$, $m \leq M$ such that χ is constant on the set*

$$\Bigl\{u + \sum_{j=1}^{t} p_{i,j}(m) v_j, \ i = 1, \dots, k\Bigr\}.$$

Proof. Define on the set $\mathbf{K} = \{1, \dots, K\}^V$ of all mappings from V into $\{1, \dots, K\}$ a metric by

$$\rho(\chi_1, \chi_2) = \Bigl(\min\Bigl\{|u| : \ u \in V, \ \chi_1(u) \neq \chi_2(u)\Bigr\} + 1\Bigr)^{-1},$$

where $|u| = \sum_{i=1}^{l} |u_i|$. Clearly, $(\mathbf{K}, \rho)$ is compact. Define the homeomorphisms T_j, $j = 1, \dots, t$, of $\mathbf{K}$ by

$$T_j \chi(u) = \chi(u + v_j) \text{ for } \chi \in \mathbf{K}.$$

For each $1 \le i \le k$ let

$$g_i(n) = \prod_{j=1}^{t} T_j^{p_{i,j}(n)}.$$

Fix some $\chi \in \mathbf{K}$. Applying Theorem C to the closure X of the orbit of χ,

$$\{T_1^{a_1} \dots T_t^{a_t} \chi\}_{(a_1,\dots,a_t)\in\mathbb{Z}^t} \subseteq \mathbf{K},$$

the set of homeomorphisms $\{T_j|_X :\ j = 1,\dots,t\}$, the system $\{g_i(n),\ i = 1,\dots,k\}$ and $\varepsilon = 1$, we find $\chi' \in X$ such that, for some $m = m(\chi) \in \mathbb{N}$ and each $i = 1,\dots,k$,

$$\chi'(0) = g_i(m)\chi'(0) = \prod_{j=1}^{t} T_j^{p_{i,j}(m)} \chi'(0) = \chi'\Big(\sum_{j=1}^{t} p_{i,j}(m) v_j\Big).$$

Since $\chi' \in X$, for any $\varepsilon > 0$ there exist $n_1,\dots,n_t \in \mathbb{N}$ such that $\rho(\prod_{j=1}^{t} T_j^{n_j}\chi, \chi') < \varepsilon$. Taking

$$\varepsilon = \Big(\max_{1\le i\le k}\Big|\sum_{j=1}^{t} p_{i,j}(m) v_j\Big| + 1\Big)^{-1}$$

and putting $u = u(\chi) = \sum_{j=1}^{t} n_j v_j$ for corresponding $n_1,\dots,n_t$, we have

$$\chi\Big(u + \sum_{j=1}^{t} p_{i,j}(m) v_j\Big) = \chi'\Big(\sum_{j=1}^{t} p_{i,j}(m) v_j\Big) = \chi'(0) = \chi(u) \tag{1.7}$$

for any $i = 1,\dots,k$.

Let $\lambda\colon \mathbf{K} \longrightarrow \mathbb{N}$ be defined by

$$\lambda(\chi) = \min\Big\{|u| + m :\ u \in V,\ m \in \mathbb{N} \text{ satisfy (1.7)}\Big\}.$$

We claim that λ is continuous (and, moreover, locally constant). Indeed, let $\chi \in \mathbf{K}$ and let $u(\chi)$, $m(\chi)$ be u and m for which the minimum in the definition of $\lambda(\chi)$ is attained. Let $\chi_1 \in \mathbf{K}$ and $\rho(\chi_1, \chi)$ be so small that $\chi_1(v) = \chi(v)$ for all

$$v \in \Big\{u + \sum_{j=1}^{t} p_{i,j}(m) v_j,\quad i = 1,\dots,k,\ u \in V,\ m \in \mathbb{N} :\ |u| + m \le |u(\chi)| + m(\chi)\Big\}.$$

Then $\lambda(\chi_1) = \lambda(\chi)$, and the continuity of λ follows.

Since $\mathbf{K}$ is compact, λ is bounded. To finish the proof, put $M = \max_{\mathbf{K}} \lambda$, $Q = \Big\{u \in V :\ |u| \le \max_{\mathbf{K}} \lambda\Big\}$. ∎

1.12. Corollary. *For any natural numbers K, t, k and l, for any integral polynomials $p_{1,1}(n),\ \dots,\ p_{1,t}(n),\ p_{2,1}(n),\ \dots,\ p_{2,t}(n),\ \dots,\ p_{k,1}(n),\ \dots,\ p_{k,t}(n)$, for any vectors $v_1,\dots,v_t \in V = \mathbb{Z}^l$ and any mapping $\chi\colon \mathbb{Z}^l \longrightarrow \{1,\dots,K\}$ the set*

$$P = \Big\{m :\ \exists u \in V \text{ such that } \chi\Big(u + \sum_{j=1}^{t} p_{i,j}(m) v_j\Big) = \chi(u),\quad i = 1,\dots,k\Big\}$$

is an IP-set.*

Proof. Define $\mathbf{K}$, T_j, $j = 1, \dots, t$, in the same way as in the proof of Corollary 1.11. By Corollary 1.9, for any IP-set I the closure X of the orbit

$$\{T_1^{a_1} \dots T_t^{a_t} \chi\}_{(a_1, \dots, a_t) \in \mathbb{Z}^t} \subseteq \mathbf{K}$$

contains a mapping χ' such that for some $m \in I$

$$\chi'(0) = \prod_{j=1}^{t} T_j^{p_{i,j}(m)} \chi'(0) = \chi'\Big(\sum_{j=1}^{t} p_{i,j}(m)\Big) \quad \text{for every } i = 1, \dots, k.$$

Since $\chi' \in X$, the same holds also for an appropriate shift $\prod_{j=1}^{t} T_j^{n_j} \chi$ of χ; so, putting $u = \sum_{j=1}^{t} n_j v_j$, we obtain

$$\chi\Big(u + \sum_{j=1}^{t} p_{i,j}(m) v_j\Big) = \chi(u) \quad \text{for every } i = 1, \dots, k.$$

We have shown that P contains an element m of I; since I was an arbitrary IP-set, this proves the corollary. ∎

2. Weakly mixing extensions

As mentioned in the introduction, the method of proof of Theorem A which is analogous to the method of proof of Theorem [FK1]A, is that of exhausting the measure preserving system $(X, \mathcal{B}, \mu, T_1, \dots, T_t)$ by factors in which relative compactness and relative weakly mixing properties are combined. In this section we study polynomial recurrence of so called weakly mixing extensions. The information obtained in this section will be used in the proof of Theorem A given in Section 3. We start by recalling some relevant notions. For more information about the extensions see [F1], [F2], [FK1], [FKO].

2.1. Let Γ be an abelian group acting by measure preserving transformations on a probability measure space $(X, \mathcal{B}, \mu)$. The measure preserving system $(X, \mathcal{B}, \mu, \Gamma)$ is called weakly mixing relative to $T \in \Gamma$ if the diagonal action of Γ on the Cartesian square, $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu)$ is ergodic relative to T, that is the only measurable subsets of $X \times X$ which are invariant with respect to $T \times T$ are of measure 0 or 1. Weak mixing can be characterized in many equivalent ways. In particular, a measure preserving system $(X, \mathcal{B}, \mu, \Gamma)$ is weakly mixing relative to T if and only if $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, \Gamma)$ is weakly mixing relative to T, if and only if the action of T on $L^2(X, \mu)$ has no measurable eigenfunctions other than the constants, and also if and only if for any $f_0, f_1 \in L^\infty(X, \mu)$ one has

$$D\text{-}\lim_{n} \int f_0 \cdot T^n f_1 \, d\mu = \int f_0 \, d\mu \int f_1 \, d\mu.$$

$D\text{-}\lim_{n} a_n = a$ means that for any $\varepsilon > 0$ the set $\{n : \ |a_n - a| > \varepsilon\}$ has density zero.

2.2. $\mathbf{Y} = (Y, \mathcal{D}, \nu, \Gamma)$ is *a factor* of $\mathbf{X} = (X, \mathcal{B}, \mu, \Gamma)$ if we have a map $\alpha\colon X \longrightarrow Y$ preserving the measure:

$$A \in \mathcal{D} \Longrightarrow \alpha^{-1}(A) \in \mathcal{B} \text{ and } \mu(\alpha^{-1}(A)) = \nu(A)$$

and commuting with the action of Γ; when this is the case, $\mathbf{X}$ is called *an extension* of $\mathbf{Y}$. We denote by α^* the isometric embedding of $L^1(Y,\nu)$ into $L^1(X,\mu)$ determined by α; we shall identify $L^1(Y,\nu)$ with $\alpha^*(L^1(Y,\nu))$.

We assume that $(X,\mathcal{B},\mu)$ is a regular measure space. *The decomposition of the measure* $\mu = \int \mu_y \, d\nu$ corresponding to α is defined as a family of measures $\{\mu_y,\ y \in Y\}$ on $(X,\mathcal{B})$ measurably depending on y and satisfying

$$\int \Big(\int f \, d\mu_y\Big) \, d\nu = \int f \, d\mu \quad \text{for any } f \in L^1(X,\mu),$$
$$f(y) = \int \alpha^*(f) \, d\mu_y \quad \text{a. e. for any } f \in L^1(Y,\nu).$$

In addition, it commutes with the action of Γ: for any $T \in \Gamma$ and $f \in L^1(X,\mu)$ one has

$$T\int f \, d\mu_y = \int f \, d\mu_{Ty} = \int Tf \, d\mu_y \quad \text{for almost all } y \in Y.$$

The square of $\mathbf{X}$ *relative to* $\mathbf{Y}$, $\mathbf{X}\times_{\mathbf{Y}}\mathbf{X}$ is the dynamical system $(X\times_Y X, \mathcal{B}\times\mathcal{B}, \mu\times_Y\mu, \Gamma)$, where $X\times_Y X = \{(x_1,x_2) \in X\times X : \ \alpha(x_1) = \alpha(x_2)\}$ and $\mu\times_Y\mu$ is defined by

$$\int f\otimes g \, d\mu\times_Y\mu = \int \Big(\int f \, d\mu_y\Big)\Big(\int g \, d\mu_y\Big) \, d\nu$$

for any $f, g \in L^2(X,\mu)$ ($f\otimes g$ is defined by $(f\otimes g)(x_1,x_2) = f(x_1)g(x_2)$).

This gives, in particular, the decomposition $\mu\times_Y\mu = \int (\mu\times_Y\mu)_y \, d\nu$ corresponding to the extension $\alpha\times\alpha\colon \mathbf{X}\times_{\mathbf{Y}}\mathbf{X} \longrightarrow \mathbf{Y}$ where $(\mu\times_Y\mu)_y = \mu_y\times\mu_y$ for almost every $y \in Y$.

The extension $\mathbf{X} = (X,\mathcal{B},\mu,\Gamma) \longrightarrow \mathbf{Y} = (Y,\mathcal{D},\nu,\Gamma)$ is called *ergodic relative to* $T \in \Gamma$ if, modulo sets of zero measure, the only T-invariant sets in $\mathcal{B}$ are preimages of T-invariant sets in $\mathcal{D}$. The extension $\mathbf{X} \longrightarrow \mathbf{Y}$ is called *weakly mixing relative to* $T \in \Gamma$ if its relativized square $\mathbf{X}\times_{\mathbf{Y}}\mathbf{X} \longrightarrow \mathbf{Y}$ is ergodic relative to T.

We use the following properties of relatively weakly mixing extensions. If an extension is weakly mixing relative to T, its square is weakly mixing relative to T as well. If the extension is weakly mixing relative to T then every eigenvector of the action of T on $L^2(X,\mu)$ comes from $L^2(Y,\nu)$. Let $\mu = \int \mu_y \, d\nu$ be the decomposition of μ corresponding to an extension $(X,\mathcal{B},\mu,\Gamma) \longrightarrow (Y,\mathcal{D},\nu,\Gamma)$. The extension is weakly mixing relative to T if and only if for any $f_0, f_1 \in L^\infty(X,\mu)$

$$D\text{-}\lim_n \Big\| \int f_0 \cdot T^n f_1 \, d\mu_y - \int f_0 \, d\mu_y \cdot T^n\Big(\int f_1 \, d\mu_y\Big) \Big\|_{L^2(Y,\nu)} = 0$$

(see [F2], Proposition 6.2). In particular, when $\int f_1 \, d\mu_y = 0$ in $L^2(Y,\nu)$, one has

$$D\text{-}\lim_n \Big\| \int f_0 \cdot T^n f_1 \, d\mu_y \Big\|_{L^2(Y,\nu)} = 0.$$

A Γ-invariant extension is called *weakly mixing relative to a subgroup* $\Gamma' \subseteq \Gamma$ if it is weakly mixing relative to T for every $T \in \Gamma'$, $T \neq \mathbf{1}_{\Gamma'}$.

2.3. Theorem D admits a natural generalization to weakly mixing extensions. It is the following relativized version of Theorem D which we shall need in the proof of Theorem A in the next section. (Theorem D itself corresponds to the case of trivial $(Y,\mathcal{D},\nu,\Gamma)$.)

Proposition. *Let* $\alpha\colon(X,\mathcal{B},\mu,\Gamma)\longrightarrow(Y,\mathcal{D},\nu,\Gamma)$ *be a weakly mixing extension relative to* Γ *where* Γ *is an abelian group, let* $\mu=\int\mu_y\,d\nu(y)$*, let* $T_1,\dots,T_t\in\Gamma$*, and let* $g_i(n)=\prod_{j=1}^{t}T_j^{p_{i,j}(n)}$*,* $i=1,\dots,k$*, be such that* $g_i(n)$ *and* $g_i(n)g_l^{-1}(n)$*,* $i\neq l$*,* $i,l=1,\dots,k$*, depend nontrivially on* n*. Then for any* $f_1,\dots,f_k\in L^\infty(X,\mu)$

$$\lim_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1}\Big(\prod_{i=1}^{k}g_i(n)f_i-\prod_{i=1}^{k}g_i(n)\alpha^*\big(\int f_i\,d\mu_y\big)\Big)\Big\|_{L^2(X,\mu)}=0.$$

2.4. A convenient tool in the proof of Proposition 2.3 is the following "van der Corput" trick (see [B2], Theorem 1.5):

Lemma. *Let* $w_0,w_1,w_2,\dots$ *be a bounded sequence of elements of a Hilbert space (with the scalar product* $\langle,\rangle$ *and the norm* $\|\ \|$*). Assume that*

$$D\text{-}\lim_{h}\limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\langle w_n\,,\,w_{n+h}\rangle=0.$$

Then $\displaystyle\lim_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1}w_n\Big\|=0.$

2.5. We need also the following simple lemma:

Lemma. *Suppose* $\alpha\colon(X,\mathcal{B},\mu,\Gamma)\longrightarrow(Y,\mathcal{D},\nu,\Gamma)$ *be a nontrivial extension, and assume that* $T\in\Gamma$ *satisfies* $T^d=\mathbf{1}_\Gamma$ *for some* $d\in\mathbb{N}$*. Then* α *is not weakly mixing relative to* T*.*

Proof. Take an arbitrary $f\in L^2(X,\mu)\backslash L^2(Y,\nu)$ and consider the finite dimensional space $L=\mathrm{Span}\{T^if,\ i\in\mathbb{Z}\}$. L is invariant with respect to the action of T, and the orthogonal complement $M=(L^2(Y,\nu)\cap L)^\perp\subseteq L$ of $L^2(Y,\nu)$ in L is invariant as well and nonempty because of the choice of f; T has an eigenvector in M which is not contained in $L^2(Y,\nu)$. ∎

2.6. We want to start with some remarks. In Proposition 2.3 we deal with the expressions

$$g_i(n)f_i(x)=T_1^{p_{i,1}(n)}\dots T_t^{p_{i,t}(n)}f_i(x)=f_i\big(T_1^{p_{i,1}(n)}\dots T_t^{p_{i,t}(n)}x\big),\quad i=1,\dots,k.$$

Without loss of generality we may and will assume that $p_{i,j}(0)=0$, $j=1,\dots,t$, $i=1,\dots,k$ (since the functions f_i in the formulation of Proposition 2.3 are arbitrary, one can replace f_i by $T_1^{-p_{i,1}(0)}\dots T_t^{-p_{i,t}(0)}f_i$); this means that $g_i(n)$, $i=1,\dots,k$, are *polynomial expressions* and form *a system A* in the notation of Section 1.

Our next remark is that without loss of generality we may assume that $\int f_{i_0}\,d\mu_y=0$ in $L^2(Y,\nu)$ for some $1\leq i_0\leq k$. Indeed, the identity

$$\prod_{i=1}^{k}g_i(n)f_i=\sum_{E\subseteq\{1,\dots,k\}}\prod_{i\in E}g_i(n)\Big(f_i-\alpha^*\big(\int f_i\,d\mu_y\big)\Big)\prod_{i\notin E}g_i(n)\alpha^*\big(\int f_i\,d\mu_y\big)$$

shows that the treatment of the general case is reducible to dealing with finitely many expressions such that the functions h_i occurring in them either satisfy $\int h_i\,d\mu_y = 0$ a.e. or $h_i \in L^2(Y,\nu)$. We have to prove then that

$$\lim_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1}\prod_{i=1}^{k} g_i(n)f_i\Big\|_{L^2(X,\mu)} = 0. \tag{2.1}$$

Finally, we may assume that Γ is finitely generated (by $T_1,\ldots,T_t$); in light of Lemma 2.5, there is no loss of generality in assuming the group Γ to be free abelian. Choose a basis of Γ; every polynomial expression $g(n)$ can be expressed in terms of this basis. So, we may and shall assume that $T_1,\ldots,T_t$ are elements of this basis and are, consequently, linearly independent. Then the assumption $g(n) = \mathbf{1}_\Gamma$, where $g(n) = \prod_{j=1}^{t} T_j^{p_j(n)}$, implies $p_1(n) = \ldots = p_t(n) = 0$.

2.7. Proof of Proposition 2.3. Put $w_n = \prod_{i=1}^{k} g_i(n)f_i \in L^2(X,\mu)$; Lemma 2.4 says that (2.1) follows from

$$D\text{-}\lim_{h}\limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\Big\langle\prod_{i=1}^{k} g_i(n)f_i\,,\prod_{i=1}^{k} g_i(n+h)f_i\Big\rangle = 0. \tag{2.2}$$

Introduce the notation

$$L(n,h) = \Big\langle\prod_{i=1}^{k} g_i(n)f_i,\prod_{i=1}^{k} g_i(n+h)f_i\Big\rangle = \int\prod_{i=1}^{k} g_i(n)f_i\cdot g_i(n+h)f_i\,d\mu; \tag{2.3}$$

we have to prove that

$$D\text{-}\lim_{h}\limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} L(n,h) = 0.$$

Fix $h \in \mathbb{N}$ and consider the new system $\tilde{A}_h = \{g_i(n), g_i(n+h)g_i^{-1}(h),\ i = 1,\ldots,k\}$. Generally speaking, the elements of $\tilde{A}_h$ are not pairwise distinct: if $\deg g(n) = 1$, then $g(n) = g(n+h)g^{-1}(h)$. But if $\deg g(n) \geq 2$, then for any fixed polynomial expression $\tilde{g}(n)$ there exists at most one h such that $\tilde{g}(n) \equiv g(n+h)g^{-1}(h)$ (this follows from an analogous statement about polynomials of degree ≥ 2; recall that our T_j are assumed to be linearly independent and that if $\deg g(n) \geq 2$, where $g(n) = T_1^{p_1(n)}\ldots T_t^{p_t(n)}$, then for at least one of $p_j(n)$ one has $\deg p_j(n) \geq 2$).

Rearranging the polynomial expressions if needed, we can assume that $\deg g_1(n) = \ldots = \deg g_q(n) = 1$, $\deg g_i(n) \geq 2$, $q+1 \leq i \leq k$, for some $q \leq k$. Notice that for all but finitely many h

$$g_i(n) \not\equiv g_l(n+h) \quad \text{for } i = 1,\ldots,k,\quad l = q+1,\ldots,k. \tag{2.4}$$

The conditions $\deg g_i(n) = 1$, $1 \leq i \leq q$, imply that the polynomials $p_{i,j}(n)$ in the polynomial expressions $g_i(n) = T_1^{p_{i,1}(n)}\ldots T_t^{p_{i,t}(n)}$ are linear. So, for $i = 1,\ldots,q$, we have $g_i(n+h) = g_i(n)g_i(h)$. We can rewrite now (2.3) in the following way:

$$L(n,h) = \int \prod_{i=1}^{q} g_i(n)f_i \cdot g_i(n+h)f_i \prod_{i=q+1}^{k} g_i(n)f_i \prod_{i=q+1}^{k} g_i(n+h)f_i \, d\mu$$
$$= \int \prod_{i=1}^{q} g_i(n)(f_i \cdot g_i(h)f_i) \prod_{i=q+1}^{k} g_i(n)f_i \prod_{i=q+1}^{k} g_i(n+h)g_i^{-1}(h)(g_i(h)f_i)\, d\mu$$
$$= \int \prod_{i=1}^{k'} \tilde{g}_i(n)\tilde{f}_i \, d\mu$$

where $k' = 2k-q$, $\tilde{f}_i$ stands for either f_l, $g_l(h)f_l$, or $f_l \cdot g_l(h)f_l$ for some l, $1 \le l \le k$, and $\tilde{g}_i(n)$ stands either for $g_l(n)$ for some $1 \le l \le k$ or for $g_l(n+h)g_l^{-1}(h)$ for some $q+1 \le l \le k$. Notice that, generally speaking, $\tilde{f}_i$ and $\tilde{g}_i(n)$ depend on h.

Assume now that $\tilde{g}_1(n)$ has the minimal weight in $\tilde{A}_h$; since all $g_i(n) \not\equiv \mathbf{1}_\Gamma$ we have $w(\tilde{g}_1(n)) \ge (1,1)$. Since $\tilde{g}_1(n)$ is measure preserving, we may write

$$L(n,h) = \int \tilde{f}_1 \cdot \prod_{i=2}^{k'} \tilde{g}_i(n)\tilde{g}_1^{-1}(n)\tilde{f}_i \, d\mu.$$

Put $\hat{g}_i(n) = \tilde{g}_i(n)\tilde{g}_1^{-1}(n)$, $i = 1,\dots,k'$.

Recall that by the assumptions of the theorem

$$g_i(n) \not\equiv g_l(n),\ g_i(n+h) \not\equiv g_l(n+h) \text{ for } i,l = 1,\dots,k,\ i \ne l;$$

it follows from this and (2.4) that $\tilde{g}_i(n) \not\equiv \tilde{g}_l(n)$ for $i \ne l$ if h is big enough; so

$$\hat{g}_i(n) \not\equiv \mathbf{1}_\Gamma \text{ and } \hat{g}_i(n) \not\equiv \hat{g}_l(n) \text{ for } i,l = 2,\dots,k',\ i \ne l$$

for such h.

Let $A_h = \{\hat{g}_i(n),\ i = 2,\dots,k'\}$. Note that A_h has been obtained from A in the following way: we added to $A = \{g_i(n),\ i = 1,\dots,k\}$ polynomial expressions of the form $g_i(n+h)g_i^{-1}(h)$ where $g_i(n) \in A$, this did not change the family of the equivalence classes of A; then we multiplied all the elements of the new system $\tilde{A}_h$ by the inverse of an element of $\tilde{A}_h$ having the minimal weight. We have already dealt with such a situation in the proof of Theorem C: the polynomial expressions of $\tilde{A}_h$ nonequivalent to $\tilde{g}_1(n)$ do not change their weights and the equivalence of one to another after they have been multiplied by $\tilde{g}_1(n)$; the weights of elements of $\tilde{A}_h$ which are equivalent to $\tilde{g}_1(n)$ do decrease after these elements have been multiplied by $\tilde{g}_1(n)$. So, the number of the equivalence classes having any fixed weight greater than $w(\tilde{g}_1(n))$ does not change whereas the number of equivalence classes having the minimal weight in A decreases by 1 when we pass from $\tilde{A}_h$ to A_h. Hence, the weight matrix of A_h precedes that of A.

We shall now invoke PET-induction. Namely, assume that Proposition 2.3 has already been proved for all systems (and, in particular, for A_h) whose weight matrices precede that of A. So, we have for A_h,

$$\lim_{N\to\infty} \Big\| \frac{1}{N} \sum_{n=0}^{N-1} \Big(\hat{g}_2(n)\tilde{f}_2 \dots \hat{g}_{k'}(n)\tilde{f}_{k'}$$
$$- \hat{g}_2(n)\alpha^*(\int \tilde{f}_2 \, d\mu_y) \dots \hat{g}_{k'}(n)\alpha^*(\int \tilde{f}_{k'} \, d\mu_y) \Big) \Big\|_{L^2(X,\mu)} = 0$$

and, therefore, putting $L(h)=\limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}L(n,h)$, we get

$$
\begin{aligned}
L(h)&=\limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int\tilde{f}_1\cdot\prod_{i=2}^{k'}\hat{g}_i(n)\tilde{f}_i\,d\mu\\
&=\limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int\tilde{f}_1\cdot\prod_{i=2}^{k'}\hat{g}_i(n)\alpha^*\Big(\int\tilde{f}_i\,d\mu_y\Big)\,d\mu\\
&=\limsup_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int\Big(\prod_{i=1}^{k'}\hat{g}_i(n)\int\tilde{f}_i\,d\mu_y\Big)\,d\nu\le\prod_{i=1}^{k'}\Big\|\int\tilde{f}_i\,d\mu_y\Big\|_{L^2(Y,\nu)}
\end{aligned}
\tag{2.5}
$$

for h big enough.

If $\deg g_{i_0}(n)\ge 2$, we have $\tilde{f}_i=f_{i_0}$ for some i, and since $\int f_{i_0}\,d\mu_y=0$, the last product in (2.5) is equal to zero. Otherwise, if $g_{i_0}(n)=S^n$ for some $S\in\Gamma$, $S\ne\mathbf{1}_\Gamma$, we have $\tilde{f}_i=f_{i_0}\cdot g_{i_0}(h)f_{i_0}$ for some i, and for h big enough,

$$
L(h)\le C\Big\|\int f_{i_0}\cdot S^hf_{i_0}\,d\mu_y\Big\|_{L^2(Y,\nu)},
\tag{2.6}
$$

where $C=(1+\max_l\|f_l\|_{L^\infty(X)})^{2k-2}$. Since α is weakly mixing relative to S and $\int f_{i_0}\,d\mu_y=0$,

$$
D\text{-}\lim_h\Big\|\int f_{i_0}\cdot S^hf_{i_0}\,d\mu_y\Big\|_{L^2(Y,\nu)}=0.
$$

Hence, $D\text{-}\lim_h L(h)=0$. ■

2.8. Corollary. *Let $\alpha\colon(X,\mathcal{B},\mu,\Gamma)\longrightarrow(Y,\mathcal{D},\nu,\Gamma)$ be a weakly mixing extension where Γ is a free abelian group, let $g_i(n)=\prod_{j=1}^tT_j^{p_{i,j}(n)}$, $i=1,\dots,k$, be pairwise essentially distinct polynomial expressions (i.e. $(p_{i,1}(n),\dots,p_{i,t}(n))-(p_{l,1}(n),\dots,p_{l,t}(n))\ne const$ for $i\ne l$), where $T_1,\dots,T_t$ are linearly independent elements of Γ. Then for any $f_1,\dots,f_k\in L^\infty(X,\mu)$*

$$
\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int\Big|\int\prod_{i=1}^k g_i(n)f_i\,d\mu_y-\prod_{i=1}^k\int g_i(n)f_i\,d\mu_y\Big|\,d\nu=0.
$$

Proof. Again, we may reduce the proof to the proof of

$$
\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int\Big|\int\prod_{i=1}^k g_i(n)f_i\,d\mu_y\Big|\,d\nu=0
\tag{2.7}
$$

under the assumption that $\int f_l\,d\mu_y=0$ in $L^2(Y,\nu)$ for some $1\le i_0\le k$.

Since $T_1,\dots,T_t$ are linearly independent and $g_i(n)$, $i=1,\dots,k$, are pairwise essentially distinct as polynomial expressions, $g_i(n)$, $i=1,\dots,k$, are pairwise essentially distinct as mappings $\mathbb{Z}\longrightarrow\Gamma$, and at most one of them can be constant; assume without loss of generality that $g_1(n)=\mathbf{1}_\Gamma$. Proposition 2.3 gives then that

$$
\frac{1}{N}\sum_{n=0}^{N-1}\int\prod_{i=1}^k g_i(n)f_i\,d\mu=\frac{1}{N}\sum_{n=0}^{N-1}\int f_1\cdot\prod_{i=2}^k g_i(n)f_i\,d\mu\longrightarrow 0
$$

as $N \longrightarrow \infty$ and, so,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \int \prod_{i=1}^{k} g_i(n) f_i \, d\mu_y \, d\nu = 0. \tag{2.8}$$

If we apply (2.8) to the set of the functions $f_i \otimes \bar{f}_i \in L^2(X \times_Y X, \mu \times_Y \mu)$, $i = 1, \ldots, k$, (we may do this since $\int f_{i_0} \otimes \bar{f}_{i_0} \, d\mu_y \times d\mu_y = |\int f_{i_0} \, d\mu_y|^2 = 0$) we obtain

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \Big| \int \prod_{i=1}^{k} g_i(n) f_i \, d\mu_y \Big|^2 d\nu = 0.$$

This gives (2.7). ∎

2.9. In the next section we shall need the following special case of Corollary 2.8:

Corollary. *In the assumptions of Corollary 2.8, let $A \in \mathcal{B}$ and let $\varepsilon, \delta > 0$. Then the set of $n \in \mathbb{N}$ for which*

$$\nu\Big\{ y \in Y : \ \Big| \mu_y \big(\bigcap_{i=1}^{k} g_i^{-1}(n) A \big) - \prod_{i=1}^{k} \mu_y \big(g_i^{-1}(n) A \big) \Big| \geq \varepsilon \Big\} \geq \delta$$

has density 0.

Proof. Apply Corollary 2.8 to the set of functions $f_i = \mathbf{1}_A$, $i = 1, \ldots, k$; we obtain

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \Big| \mu_y \big(\bigcap_{i=1}^{k} g_i^{-1}(n) A \big) - \prod_{i=1}^{k} \mu_y \big(g_i^{-1}(n) A \big) \Big| \, d\nu = 0. \tag{2.9}$$

Denote $F_n(y) = \mu_y(\bigcap_{i=1}^{k} g_i^{-1}(n) A) - \prod_{i=1}^{k} \mu_y(g_i^{-1}(n) A)$. If the set

$$P_n = \Big\{ n : \ \nu\{ y \in Y : \ |F_n(y)| \geq \varepsilon \} \geq \delta \Big\}$$

were not of zero density, that is if there were c, $N_i \longrightarrow \infty$ such that

$$\frac{\#\{ n \in P_n, \ n < N_i \}}{N_i} \geq c, \quad i = 1, 2, \ldots,$$

we should have

$$\frac{1}{N_i} \sum_{n=0}^{N_i - 1} \int |F_n(y)| \, d\nu \geq c\varepsilon\delta, \quad i = 1, 2, \ldots$$

which would contradict (2.9). ∎

3. The polynomial Szemerédi Theorem

This section is devoted to the proof of Theorem A. Throughout this section Γ will stand for the measure preserving action of $\mathbb{Z}^t$ generated by $T_1, \ldots, T_t$ on $(X, \mathcal{B}, \mu)$.

3.1. We saw in the previous section that Theorem A holds in the special case when the system $(X, \mathcal{B}, \mu, \Gamma)$ is totally weakly mixing. It is not hard to see that Theorem A is true also when X is a compact abelian group and Γ acts by rotations on X. We leave to the reader the verification of the fact that in this case Theorem A follows from (appropriately applied) Weyl's theorem on uniform distribution of polynomials. Instead of this, we shall give now an alternative proof of this special case by using Corollary 1.12. A modification of this argument will be utilized in the proof of Theorem A.

A measure preserving system $(X, \mathcal{B}, \mu, \Gamma)$ is called *compact* if for any $f \in L^2(X, \mu)$ the orbit $\{Tf :\ T \in \Gamma\}$ is precompact in the strong topology of $L^2(X, \mu)$ (one can show that an ergodic measure preserving system is compact if and only if it is isomorphic to a system formed by rotations on a compact abelian group; we shall not use this fact).

Let $(X, \mathcal{B}, \mu, \Gamma)$ be compact, let $p_{1,1}(n), \dots, p_{1,t}(n), p_{2,1}(n), \dots, p_{2,t}(n), \dots, p_{k,1}(n), \dots, p_{k,t}(n)$ be integral polynomials, let $f = \mathbf{1}_A$ where $\mu(A) = a > 0$; put $\varepsilon = \sqrt{a/8k}$.

Since the set $\{Tf : T \in \Gamma\}$ is precompact, there exists a finite set $\{h_1, \dots, h_K\} \subset L^2(X, \mu)$ such that for any $T \in \Gamma$ there exists $\chi = \chi(T) \in \{1, \dots, K\}$ satisfying

$$\|Tf - h_\chi\|_{L^2(X,\mu)} < \varepsilon.$$

This gives the mapping $\chi\colon \Gamma \longrightarrow \{1, \dots, K\}$; by Corollary 1.12, applied to the set of vectors $v_j = T_j \in \Gamma$, the set

$$P = \Big\{ m \in \mathbb{N} :\ \exists T = T(m) \in \Gamma :\ \chi(T) = \chi(T \prod_{j=1}^{t} T_j^{p_{i,j}(m)}),\ i = 1, \dots, k \Big\}$$

is an IP*-set. Since any IP*-set is syndetic (i.e. has bounded gaps), its lower density is positive:

$$\underline{d}(P) = \liminf_{N \to \infty} \frac{\#(P \cap \{1, \dots, N\})}{N} > 0.$$

For any $m \in P$, we have

$$\|Tf - h_{\chi(T)}\|_{L^2(X,\mu)} < \varepsilon \quad \text{and} \quad \Big\|(T \prod_{j=1}^{t} T_j^{p_{i,j}(m)})f - h_{\chi(T)}\Big\|_{L^2(X,\mu)} < \varepsilon$$

for some $T = T(m) \in \Gamma$. Hence,

$$\Big\|Tf - (T \prod_{j=1}^{t} T_j^{p_{i,j}(m)})f\Big\|_{L^2(X,\mu)} = \Big\|f - (\prod_{j=1}^{t} T_j^{p_{i,j}(m)})f\Big\|_{L^2(X,\mu)} < 2\varepsilon.$$

Since $f = \mathbf{1}_A$, this means that

$$\mu\Big(A \triangle (\prod_{j=1}^{t} T_j^{-p_{i,j}(m)})A\Big) < 4\varepsilon^2, \quad i = 1, \dots, k,\ m \in P$$

and

$$\mu\Big(A\cap(\prod_{j=1}^{t}T_j^{-p_{1,j}(m)})A\cap\ldots\cap(\prod_{j=1}^{t}T_j^{-p_{k,j}(m)})A\Big)>a-4\varepsilon^2k=a/2.$$

Thus

$$\begin{aligned}\frac{1}{N}\sum_{n=0}^{N-1}\mu\Big(\bigcap_{i=1}^{k}(\prod_{j=1}^{t}T_j^{-p_{i,j}(n)})A\Big)&\geq\frac{1}{N}\sum_{\substack{m\in P\\ m\leq N}}\mu\Big(\bigcap_{i=1}^{k}(\prod_{j=1}^{t}T_j^{-p_{i,j}(m)})A\Big)\\&>\frac{1}{N}\cdot\#\Big(P\cap\{1,\ldots,N\}\Big)\cdot\frac{a}{2}\end{aligned}$$

and $\liminf_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mu\Big(\bigcap_{i=1}^{k}(\prod_{j=1}^{t}T_j^{-p_{i,j}(m)})A\Big)>\underline{d}(P)\cdot a/2.$ ■

3.2. Let $\mathbf{X}=(X,\mathcal{B},\mu,\Gamma)\longrightarrow\mathbf{Y}=(Y,\mathcal{D},\nu,\Gamma)$ be an extension, let $\mu=\int\mu_y\,d\nu$ be the corresponding decomposition of μ. Let $f\in L^2(X,\mu)$; then $\|f\|_y$ denotes the norm of f in $L^2(X,\mu_y)$.

3.3. An extension $\mathbf{X}=(X,\mathcal{B},\mu,\Gamma)\longrightarrow\mathbf{Y}=(Y,\mathcal{D},\nu,\Gamma)$ is called *compact relative to a subgroup* $\Gamma'\subseteq\Gamma$ if for every $f\in L^2(X,\mu)$ and any $\varepsilon,\delta>0$ there exist a set $b\in\mathcal{D}$ with $\nu(B)>1-\varepsilon$ and a finite set of functions $h_1,\ldots,h_K\in L^2(X,\mu)$ such that for each $R\in\Gamma'$ one has $\min_{1\leq l\leq K}\|R(f\cdot\mathbf{1}_{\alpha^{-1}(B)})-h_l\|_y<\delta$ for almost all $y\in Y$.

We shall use the following characterization of the relative compact extensions:

Lemma.(Lemma 7.10, [F2].) *Assume that the extension $\mathbf{X}\longrightarrow\mathbf{Y}$ is compact relative to $\Gamma'\subseteq\Gamma$ and let $A\in\mathcal{B}$, $\mu(A)>0$. One can find a subset $A'\subseteq A$ with $\mu(A')$ as close as one likes to $\mu(A)$ having the following property: For any $\varepsilon>0$ there exists a finite set of functions $h_1,\ldots,h_K\in L^2(X,\mu)$ such that for almost all $y\in Y$ and every $R\in\Gamma'$ there exists $1\leq l\leq K$ for which*

$$\|R\mathbf{1}_{A'}-h_l\|_y<\varepsilon.$$

3.4. An extension $\alpha:\mathbf{X}\longrightarrow\mathbf{Y}$ is called *primitive* if Γ is the direct product of two subgroups $\Gamma=\Gamma_c\times\Gamma_w$ such that α is compact relative to Γ_c and weakly mixing relative to Γ_w.

We will call the dynamical systems for which the conclusion of Theorem A is valid *SZP-systems*. The following two propositions show that, to prove Theorem A, it is enough to check that the property of being an SZP-system is preserved under passage to primitive extensions.

Proposition.(Theorem 6.16 in [F2].) *If $\gamma:\mathbf{X}\longrightarrow\mathbf{Z}$ is a nontrivial extension, one can find a system $\mathbf{Y}$ and homomorphisms $\alpha:\mathbf{X}\longrightarrow\mathbf{Y}$ and $\alpha':\mathbf{Y}\longrightarrow\mathbf{Z}$ with $\gamma=\alpha'\alpha$ and such that $\mathbf{Y}$ is a nontrivial primitive extension of $\mathbf{Z}$.*

Proposition. *The family of Γ-invariant factors which are SZP-systems has a maximal element (under inclusion).*

The proof of this proposition is completely analogous to that of Proposition 3.3 in [FK1] for SZ-systems.

3.5. Let $\alpha\colon (X,\mathcal{B},\mu,\Gamma) \longrightarrow (Y,\mathcal{D},\nu,\Gamma)$ be a primitive extension. We will assume that $\Gamma = \Gamma_c \times \Gamma_w$ is such that α is weakly mixing relative to Γ_w and compact relative to Γ_c. Now, Theorem A is the corollary of the following proposition:

Proposition. *If $(Y,\mathcal{D},\nu,\Gamma)$ is an SZP-system so is $(X,\mathcal{B},\mu,\Gamma)$.*

Proof. Let $A \in \mathcal{B}$ be of positive measure, let $T_1,\dots,T_t \in \Gamma$, let $p_{i,j}(n) = \sum_{d\ge 1} c_{i,j,d} n^d$, $1 \le i \le k$, $1 \le j \le t$, be integral polynomials and let $g_i(n) = T_1^{p_{i,1}(n)} \dots T_t^{p_{i,t}(n)}$, $1 \le i \le k$. Without loss of generality we may assume that $\{T_1,\dots,T_q\} \subset \Gamma_c$ and that $\{T_{q+1},\dots,T_t\}$ is a basis of Γ_w. Multiplying if needed the argument n of the polynomials $p_{i,j}(n)$ by a suitable natural number, we may assume that all $c_{i,b,d}$ are integers.

Let us write each $g_i(n)$ as the product of its compact and weakly mixing components:

$$g_i(n) = R_{(i)}(n) S_{(i)}(n), \quad i = 1,\dots,k,$$

$$R_{(i)}(n) = \prod_{j=1}^{q} T_j^{p_{i,j}(n)} \in \Gamma_c, \ \ S_{(i)}(n) = \prod_{j=q+1}^{t} T_j^{p_{i,j}(n)} \in \Gamma_w.$$

Let $\{R_1(n),\dots,R_r(n)\}$ be the set of all the compact components of $g_i(n)$, $i = 1,\dots,k$, including the identity and let $\{S_1(n),\dots,S_s(n)\}$ be the set of all **pairwise distinct** weakly mixing components of $g_i(n)$, $i = 1,\dots,k$. It is enough to find a set $P \subseteq \mathbb{N}$ of positive lower density (that is, $\liminf_{N\to\infty} \frac{\#(P\cap\{1,\dots,N\})}{N} > 0$) and $c > 0$ such that for $n \in P$

$$\mu\Big(\bigcap_{\substack{1\le i\le r \\ 1\le j\le s}} R_i(n)^{-1} S_j(n)^{-1} A \Big) > c.$$

Using Corollary 2.9 one can find a set $P' \in \mathbb{N}$ having positive lower density such that $\mu\left(\bigcap_{1\le j\le s} S_j(n)^{-1}A\right) > c'$ for some $c' > 0$ and every $n \in P'$; we shall use Corollary 1.11 to choose a subset A' of A of positive measure and a subset $P \subseteq P'$ of positive lower density consisting of $n \in \mathbb{N}$ for which the set $R_i(n)^{-1}S_j(n)^{-1}A'$ is "very close" to $S_j(n)^{-1}A'$ for each $1 \le i \le r$ and each $1 \le j \le s$.

3.6. Lemma. *Let $f, h_1,\dots,h_K \in L^2(X,\mu)$, $\varepsilon > 0$ be such that for almost all $y \in Y$ and every $R \in \Gamma_c$ there exists $1 \le l \le K$: $\|Rf - h_l\|_y < \varepsilon$; let $B \in \mathcal{D}$, $\nu(B) > 0$. Then there exist $P \subseteq \mathbb{N}$ with $\underline{d}(P) > 0$, a family of sets $\{B_n \in \mathcal{D},\ n \in P\}$ and a number $b > 0$ so that, for any $n \in P$, $1 \le j \le s$, $1 \le i \le r$ one has*
(i) $\nu(B_n) > b$,
(ii) $S_j(n)B_n \subseteq B$,
(iii) $\forall y \in B_n$, $\|R_i(n)S_j(n)f - S_j(n)f\|_y < 2\varepsilon$.

Proof. Let J be the set of all triples of integers (j,d,c) for which the term cn^d appears in one of the polynomials $p_{i,j}(n)$, $i = 1,\dots,k$, $j = 1,\dots,t$. Let $V = \mathbb{Z}^{\#J}$ be the lattice with the basis $\{v_{(j,d,c)},\ (j,d,c) \in J\}$. By Corollary 1.11, there exist $M \in \mathbb{N}$, $Q \subset V$, $\#Q < \infty$, such that, for any $\chi\colon V \longrightarrow \{1,\dots,K\}$, there exist $m \le M$ and $u \in Q$ such that $\chi(u + \sum_{(j,d,c)\in E} cm^d v_{(j,d,c)}) = \chi(u)$ for any $E \subseteq J$. (We apply the corollary to the polynomials $p_{E,(j,d,c)}(n) = \begin{cases} cn^d, \ (j,d,c) \in E \\ \ \ 0 \text{ otherwise.} \end{cases}$)

Denote

$$\mathbf{R}(u,p)=\prod_{\substack{(j,d,c)\in J\\ 1\le j\le q}} T_j^{p^d\cdot u_{(j,d,c)}}\in\Gamma_c,\ \ \mathbf{S}(u,p)=\prod_{\substack{(j,d,c)\in J\\ q+1\le j\le t}} T_j^{p^d\cdot u_{(j,d,c)}}\in\Gamma_w,$$
$$\mathbf{T}(u,p)=\mathbf{R}(u,p)\mathbf{S}(u,p)\in\Gamma \text{ for } u=(u_{(j,d,c)},\ (j,d,c)\in J)\in V,\ p\in\mathbb{N}.$$

Fix $p\in\mathbb{N}$, $y\in Y$ and define $\chi\colon V\longrightarrow\{1,\ \dots,\ K\}$ by the rule

$$\chi(u)=l\implies \|\mathbf{R}(u,p)f-h_l\|_{\mathbf{S}(u,p)y}<\varepsilon.$$

Applying Corollary 1.11, find $h=h(p,y)\in L^2(X,\mu)$, $m=m(p,y)\le M$ and $u=u(p,y)\in Q$ such that

$$\Big\|\mathbf{R}(u+\sum_{(j,d,c)\in E} cm^d\cdot v_{(j,d,c)},p)f-h\Big\|_{\mathbf{S}(u+\sum_{(j,d,c)\in E} cm^d\cdot v_{(j,d,c)},p)y}<\varepsilon \tag{3.1}$$

for any $E\subseteq J$.

Let

$$R(n)=\prod_{j=1}^{q} T_j^{\sum_d c_{j,d}n^d},\quad S(n)=\prod_{j=q+1}^{t} T_j^{\sum_d c_{j,d}n^d}. \tag{3.2}$$

Taking $E=\{(j,d,c_{j,d})\}$ to be the set of those triples $(j,d,c_{j,d})$ which appear in (3.2), we have

$$\mathbf{R}\Big(u+\sum_{(j,d,c)\in E} cm^d\cdot v_{(j,d,c)},p\Big)=\mathbf{R}(u,p)R(pm),$$
$$\mathbf{S}\Big(u+\sum_{(j,d,c)\in E} cm^d\cdot v_{(j,d,c)},p\Big)=\mathbf{S}(u,p)S(pm),$$

and, when $E\subseteq J$, (3.1) gives

$$\begin{aligned}&\|\mathbf{R}(u,p)R(pm)f-h\|_{\mathbf{S}(u,p)S(pm)y}\\ &=\|R(pm)S(pm)f-S(pm)(\mathbf{R}(u,p)^{-1}h)\|_{\mathbf{T}(u,p)y}<\varepsilon.\end{aligned} \tag{3.3}$$

In particular, (3.3) is valid for $R=R_i$, $S=S_j$ for every $i=1,\dots,r$, $j=1,\dots,s$; since one of R_i was supposed to be identity, this implies

$$\|R_i(pm)S_j(pm)f-S_j(pm)f\|_{\mathbf{T}(u,p)y}<2\varepsilon,\quad i=1,\dots,r,\ j=1,\dots,s. \tag{3.4}$$

Put

$$C_p=\bigcap_{\substack{m\le M\\ u\in Q\\ 1\le j\le s}}\mathbf{T}(u,p)^{-1}S_j(pm)^{-1}B. \tag{3.5}$$

Since $(Y, \mathcal{D}, \nu, \Gamma)$ is an SZP-system, there exist $b' > 0$, $P' \subseteq \mathbb{N}$, $\underline{d}(P') > 0$, such that $\nu(C_p) > b'$ for $p \in P'$.

For $m \leq M$, $u \in Q$ define

$$C_p(m, u) = \Big\{ y \in C_p : \ m(p, y) = m, \ u(p, y) = u \Big\}$$

in the notation of (3.1). Then for every $p \in P'$ there exist $m_p \leq M$, $u_p \in Q$ for which $\nu(C_p(m_p, u_p)) > b$, where we have denoted $b = b'/(M \cdot \#Q)$. Put $P = \{pm_p, \ p \in P'\}$; then $\underline{d}(P) \geq \underline{d}(P')/M^2 > 0$.

For $n = pm_p \in P$, $p \in P'$, define

$$B_n = \mathbf{T}(u_p, p) C_p(m_p, u_p);$$

then $\nu(B_n) > b$ and $S_j(n) B_n \subseteq B$ by (3.5), that is we have (i) and (ii). Furthermore, from the definition of $C_p(m, u)$, (3.4) is valid for $y \in C_p(m_p, u_p)$, $m = m_p$, $u = u_p$; this gives (iii). ■

3.7. The end of the proof of Proposition 3.5. Let $0 < a < \mu(A)$. Passing if needed to a smaller subset, we shall assume without loss of generality that there exist $h_1, \ \ldots, \ h_K \in L^2(X, \mu)$ such that for almost all $y \in Y$ and every $R \in \Gamma_c$ one has $\|R\mathbf{1}_A - h_l\|_y < \varepsilon$ for some $1 \leq l \leq K$, where we have put $\varepsilon = \sqrt{a^s/16rs}$.

Put $B = \{y \in Y : \ \mu_y(A) > a\}$; then $\nu(B) > 0$. By Lemma 3.6, applied to $f = \mathbf{1}_A$, there exist $P \in \mathbb{N}$ of positive lower density, a number $b > 0$ and a set $\{B_n \in \mathcal{D}, \ n \in P\}$ with $\nu(B_n) > b$ such that $S_j(n) B_n \subseteq B$ and

$$\|R_i(n) S_j(n) \mathbf{1}_A - S_j(n) \mathbf{1}_A\|_y < 2\varepsilon, \quad 1 \leq i \leq r, \ 1 \leq j \leq s, \ n \in P, \ y \in B_n.$$

This gives $\mu_y(S_j(n)^{-1} A) > a$ and

$$\mu_y\Big(R_i(n)^{-1} S_j^{-1}(n) A \,\triangle\, S_j(n)^{-1} A\Big) < 4\varepsilon^2, \quad 1 \leq i \leq r, \ 1 \leq j \leq s, \quad n \in P, \ y \in B_n. \tag{3.6}$$

By Corollary 2.9,

$$\mu_y\Big(\bigcap_{j=1}^{s} S_j(n)^{-1} A\Big) > \frac{1}{2} \prod_{j=1}^{s} \mu_y\Big(S_j(n)^{-1} A\Big) > a^s/2 \tag{3.7}$$

for all $y \in B_n$ except a subset of the measure $< b/2$ and all $n \in \mathbb{N}$ except a subset of $\mathbb{N}$ of density zero; passing if needed to appropriate subsets of B_n and P we shall assume that (3.7) holds for all $y \in B_n$ and any $n \in P$.

Then, for $y \in B_n$ and $n \in P$, (3.6) and (3.7) give:

$$\mu_y\Big(\bigcap_{\substack{1 \leq i \leq r \\ 1 \leq j \leq s}} R_i(n)^{-1} S_j(n)^{-1} A\Big) > a^s/2 - rs \cdot 4\varepsilon^2 = a^s/4.$$

Since $\nu(B_n) > b/2$, we have for $n \in P$

$$\mu\Big(\bigcap_{\substack{1 \leq i \leq r \\ 1 \leq j \leq s}} R_i(n)^{-1} S_j(n)^{-1} A\Big) > a^s b/8. \ ■$$

4. Combinatorial corollaries

4.1. Since the derivation of Theorem B from Theorem A is completely analogous to the derivation of the by now classical Furstenberg-Katznelson's multidimensional Szemerédi theorem (Theorem [FK1]A of the introduction), we shall confine ourselves to few explanatory remarks.

Given a set $S \subseteq \mathbb{Z}^l$ and a sequence of parallelepipeds $\Pi_n = [a_n^{(1)}, b_n^{(1)}] \times \ldots \times [a_n^{(l)}, b_n^{(l)}] \subset \mathbb{Z}^l$, let

$$\bar{d}_{\{\Pi_n\}}(S) = \limsup_{n\to\infty} \frac{|S \cap \Pi_n|}{|\Pi_n|}.$$

The upper Banach density of S is defined by

$$d^*(S) = \sup \bar{d}_{\{\Pi_n\}}(S),$$

where the supremum is taken over all sequences $\{\Pi_n\}$ satisfying

$$|b_n^{(j)} - a_n^{(j)}| \to \infty,\ j = 1, \ldots, l,\ \text{as } n \to \infty.$$

According to Furstenberg's correspondence principle, given a set $S \subseteq \mathbb{Z}^l$ with $d^*(S) > 0$ there exist a probability space $(X, \mathcal{B}, \mu)$, commuting measure preserving transformations $T_j \colon X \longrightarrow X$, $j = 1, \ldots, l$, and a set $A \in \mathcal{B}$ satisfying $\mu(A) = d^*(S)$ such that for any $k \in \mathbb{N}$ and any $u_1, \ldots, u_k \in \mathbb{Z}^l$

$$d^*\Big(\bigcap_{i=1}^{k}(S - u_i)\Big) \geq \mu\Big(\bigcap_{i=1}^{k} T_1^{u_i^{(1)}} \ldots T_l^{u_i^{(l)}} A\Big)$$

(cf. [F1], p.152; see also [B1] where Furstenberg's correspondence principle is discussed in detail for $l = 1$).

It should be clear now why Theorem B follows from Theorem A. We remark in passing that one can also show that Theorem A follows from Theorem B.

4.2. We shall show now the equivalence of Theorems B and B$'$. Let $S \subseteq \mathbb{Z}^l$ be a set of positive upper Banach density.

To see that Theorem B implies Theorem B$'$, let $P \colon \mathbb{Z}^r \longrightarrow \mathbb{Z}^l$ be a polynomial mapping satisfying $P(0) = 0$ and let $F = \{w_1, \ldots, w_k\} \subset \mathbb{Z}^r$ be a finite set. Taking in Theorem B $t = l$ and $v_1, \ldots, v_t$ the basis vectors of $\mathbb{Z}^l$ and applying it to the polynomials defined by

$$p_{i,j}(n) = P(w_i n)_j, \quad n \in \mathbb{N}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, l,$$

one gets for some $u \in \mathbb{Z}^l$

$$u + P(w_i n) = u + \sum_{j=1}^{t} P(w_i n)_j v_j \in S, \quad i = 1, \ldots, k,$$

that is $u + P(nF) \subset S$.

To see that Theorem B follows from Theorem B$'$ one takes $k = r$ and applies Theorem B$'$ to the polynomial mapping $P \colon \mathbb{Z}^r \longrightarrow \mathbb{Z}^l$ defined by

$$P(n_1, \ldots, n_r) = \sum_{j=1}^{t} \sum_{i=1}^{r} p_{i,j}(n_i) v_j$$

and the finite configuration

$$F = \Big\{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)\Big\} \subset \mathbb{Z}^r.$$

References

[**B1**] Bergelson, V., *Ergodic Ramsey theory*, Cont. Math. **65** (1987), 63–87.

[**B2**] Bergelson, V., *Weakly mixing PET*, Ergod. Th. and Dynam. Sys. **7** (1987), 337–349.

[**BPT**] Blaszczyk, A., Plewik, S., Turek, S., *Topological multidimensional van der Waerden theorem*, Comment. Math. Univ. Carolinae **30, N4** (1989), 783–787.

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

[**F2**] Furstenberg, H., *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, 1981.

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

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

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

[**FKO**] Furstenberg, H., Katznelson, Y., Ornstein, D., *The ergodic theoretical proof of Szemerédi's theorem*, Bull. of the Amer. Math. Soc. **7, N3** (1982), 527–552.

[**FW**] Furstenberg, H., Weiss, B., *Topological dynamics and combinatorial number theory*, J. d'Analyse Math. **34** (1978), 61–85.

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

V. Bergelson. Department of Mathematics, Ohio State University, Columbus, OHIO 43210, USA.
E-mail address: vitaly@math.ohio-state.edu

A. Leibman. Department of Mathematics, Technion, Haifa 23000, Israel.
E-mail address: sashal@techunix.technion.ac.il