Combinatorial and Diophantine Applications of Ergodic Theory

VITALY BERGELSON*
The Ohio State University
Columbus, OH 43210, USA
E-mail: *vitaly@math.ohio-state.edu*

CONTENTS

*The author acknowledges support received from the National Science Foundation (USA) via grant DMS-0345350

1 Introduction

The main focus of this survey is the mutually perpetuating interplay between ergodic theory, combinatorics and Diophantine analysis.

Ergodic theory has its roots in statistical and celestial mechanics. In studying the long time behavior of dynamical systems, ergodic theory deals first of all with such phenomena as recurrence and uniform distribution of orbits.

Ramsey theory, a branch of combinatorics, is concerned with the phenomenon of preservation of highly organized structures under finite partitions.

Diophantine analysis concerns itself with integer and rational solutions of systems of polynomial equations.

To get a feeling about possible connections between these three quite distinct areas of mathematics, let us consider some examples.

Our first example is related to Fermat's last theorem. Given $n \in \mathbb{N}$, where $\mathbb{N}$, here and throughout this survey, represents the set of positive integers, and a prime p, consider the equation $x^n + y^n \equiv z^n (\mod p)$. This equation (as well as its more general version $ax^n + by^n + cz^n \equiv 0(\mod p)$) was extensively studied in the 19th and early 20th centuries. (See [D2], vol. 2, Ch. 26 for information on the early work and [Ri], Ch. XII for more recent developments and extensions.) We are going to prove, with the help of ergodic and combinatorial considerations, the following theorem.

Theorem 1.1. *For fixed $n \in \mathbb{N}$ and a large enough prime p, the polynomial $f(z, y) = z^n - y^n$ represents the finite field $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$. In other words, for any $c \in \mathbb{Z}_p$ there exist $z, y \in \mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$, such that $c = z^n - y^n$.*

Putting $c = x^n$ immediately gives the following result, which was proved by Schur in 1916. (See also [D1].)

Corollary 1.2. ([Schur]) *For fixed $n \in \mathbb{N}$ and large enough prime p, the equation $x^n + y^n \equiv z^n (\mod p)$ has nontrivial solutions.*

In the course of the proof of Theorem 1.1 we shall utilize the following classical fact due to F. Ramsey ([Ram]). For a nice discussion which puts Ramsey's theorem into the perspective of *Ramsey theory*, see [GraRS]. In what follows, $|A|$ denotes the cardinality of a set A.

Theorem 1.3. *For any $n, r \in \mathbb{N}$ there exists a constant $c = c(n, r)$ such that if a set A satisfies $|A| \geq c$ and the set $[A]^2$ of two-element subsets of A is partitioned into r cells (or, as we will often say, is r-colored): $[A]^2 = \bigcup_{i=1}^r C_i$, then there exists a subset $B \subset A$ satisfying $|B| > n$ and such that for some i, $1 \leq i \leq r$, $[B]^2 \subset C_i$. (In this case we say that $[B]^2$ is* monochromatic.*)*

We shall also be using the following abstract version of the Poincaré recurrence theorem (cf. [Po], pp. 69–72).

Theorem 1.4. *Assume that μ is a finitely-additive probability measure on a measurable space $(X, \mathcal{B})$, and let a group G (which is not necessarily infinite or commutative) act on $(X, \mathcal{B}, \mu)$ by measure preserving transformations T_g, $g \in G$. Let $A \in \mathcal{B}$ with $\mu(A) = a > 0$ and let an integer k satisfy $k > \lfloor \frac{1}{a} \rfloor$. If $|G| \geq k$, then for any k distinct elements $g_1, g_2, \ldots, g_k \in G$ there exist $1 \leq i < j \leq k$ such that $\mu(A \cap T_{g_i g_j^{-1}} A) > 0$.*

Proof. If the statement does not hold, then for any $i \neq j$, $\mu(T_{g_i} A \cap T_{g_j} A) = 0$. But then $\mu\big(\bigcup_{i=1}^k T_{g_i} A\big) = \sum_{i=1}^k \mu(T_{g_i} A) = ka > 1$, in contradiction with $\mu(X) = 1$. □

Remark 1.5. If one measures the triviality of a mathematical statement by the triviality of its proof, one can only wonder how and why a statement as trivial as Theorem 1.4 can lead to interesting applications. Yet it does! In particular, we shall utilize it in the proof of Theorem 1.1 and, at least implicitly, on few more occasions. (See Theorems 1.11 and 1.12 below. See also [Be4] for additional examples and more discussion.)

Proof of Theorem 1.1. Let ν be the normalized counting measure on $\mathbb{Z}_p$. Noting that the index r of the multiplicative subgroup $\Gamma = \{x^n : x \in \mathbb{Z}_p^*\}$ in $\mathbb{Z}_p^*$ is at most n, we get $\nu(\Gamma) \geq \dfrac{1}{n + \frac{1}{n}}$. (The quantity $\frac{1}{n}$ in the denominator accounts for the neutral element $0 \in \mathbb{Z}_p$.) Let $a_1, \ldots, a_r \in \mathbb{Z}_p^*$, where $r \leq n$, be such that $\mathbb{Z}_p^* = \bigcup_{i=1}^r \Gamma a_i$ is the partition of $\mathbb{Z}_p^*$ into disjoint cosets of Γ. Let $A = \{2^j,\ 1 \leq j < \log_2 p\}$. Interpreting A as a subset of $\mathbb{Z}_p^*$, we note that since all the differences $2^j - 2^i$, $1 \leq i < j < \log_2 p$ are distinct, there is a natural bijection between the set $[A]^2$ of two-element subsets of A and the set $\triangle(A) = \{2^j - 2^i,\ 1 \leq i < j < \log_2 p\} \subseteq \mathbb{Z}_p$. The partition $\mathbb{Z}_p^* = \bigcup_{i=1}^r \Gamma a_i$ naturally induces a partition (coloring) of $\triangle(A)$. Assuming that p is large enough, we get by Theorem 1.3 a subset $B \subset A$ with the property that

$|B| > n$ and such that the set of differences of distinct elements from B, $\triangle(B)$, is monochromatic, i.e. for some $i_0 \in \{1, 2, \ldots, r\}$, $\triangle(B) \subset \Gamma a_{i_0}$. But then Γ itself also contains a set of differences, namely $\triangle(Ba_{i_0}^{-1})$.

Let us apply now Theorem 1.4 to the action of $\mathbb{Z}_p$ on itself by translations: $x \to x + g$, $g \in \mathbb{Z}_p$. Let $c \in \mathbb{Z}_p^*$ be arbitrary. Consider the set $Ba_{i_0}^{-1}c \subset \Gamma c$. Since $|Ba_{i_0}^{-1}c| = |B| > n$, we have by Theorem 1.4 that there is an element x in the set of differences $\triangle(Ba_{i_0}^{-1}c)$, such that $\nu(\Gamma \cap \Gamma - x) > 0$. Noting that x is of the form gc, where $g \in \triangle(Ba_{i_0}^{-1}c) \subset \Gamma$, we have $(\Gamma \cap \Gamma - gc) \neq 0$, which implies $gc \in \Gamma - \Gamma = \{z^n - y^n : z, y \in \mathbb{Z}_p^*\}$. Utilizing the fact that $g \in \Gamma$, we get $c \in \Gamma - \Gamma$. Since $c \in \mathbb{Z}_p^*$ was arbitrary (and since, trivially, $0 \in \Gamma - \Gamma$) we finally get $\mathbb{Z}_p = \Gamma - \Gamma$.

□

We leave it to the reader to check that routine adaptation of the proof above allows one to show that for fixed n the polynomial $f(z, y) = z^n - y^n$ represents any large enough finite field. While this result has also a more traditional number-theoretical proof (see [Schm]), the "soft" method utilized in the proof of Theorem 1.1, gives, after appropriate modifications, the following more general result, which so far has no conventional proof. We shall provide the proof at the end of Section 5.

Theorem 1.6. ([BeS]) *Let F be an infinite field and let Γ be a multiplicative subgroup of finite index in $F^* = F \setminus \{0\}$. Then*

$$\Gamma - \Gamma = \{x - y : x, y \in \Gamma\} = F.$$

While Theorem 1.1 is stronger than Schur's result (Corollary 1.2), the following key lemma from [Schur] is of independent interest as one of the earliest results of Ramsey theory.

Theorem 1.7. *For any $r \in \mathbb{N}$, there exists a positive constant $c = c(r)$ such that for any integer $N \geq c$, any r-coloring $\{1, 2, \ldots, N\} = \bigcup_{i=1}^r C_i$ yields a monochromatic solution of the equation $x + y = z$.*

Proof. The result almost immediately follows from Ramsey's theorem (Theorem 1.3 above) via an argument similar to the one utilized in the proof of Theorem 1.1. (Schur's original proof was somewhat longer, but completely elementary.) Observe that if r is fixed and N is sufficiently large, then one of the C_i contains the set of differences of a 3-element set $A = \{a_1, a_2, a_3\}$. The desired result then follows by setting $x = a_3 - a_2$, $y = a_2 - a_1$, $z = a_3 - a_1$. □

To derive Corollary 1.2 from Theorem 1.7, one considers the partition of $\{1, 2, \ldots, p-1\}$ induced by the partition of $\mathbb{Z}_p^*$ into disjoint cosets of the multiplicative group $\Gamma = \{x^n : x \in \mathbb{Z}_p^*\}$. It then follows from Theorem 1.7 that there exists a coset Γc and $x, y, z \in \Gamma c$ such that (both in $\mathbb{N}$ and in $\mathbb{Z}_p$) $x + y = z$. Writing, for some $x_1, y_1, z_1 \in \Gamma$, $x = x_1 c$, $y = y_1 c$, $z = z_1 c$, we get, after the cancellation, $x_1^n + y_1^n = z_1^n (\mod p)$.

Arguably, the earliest nontrivial result of Ramsey theory is the following theorem which D. Hilbert utilized in [Hil] in order to show that if the polynomial $p(x, y) \in \mathbb{Z}[x, y]$ is irreducible, then there exists $n \in \mathbb{N}$ such that $p(x, n) \in \mathbb{Z}[x]$ is also irreducible. Given d distinct integers $x_1, \ldots, x_d$, define the *d-cube generated by* $x_1, \ldots, x_d$ by $Q(x_1, \ldots, x_d) = \{\sum_{i=1}^d \varepsilon_i x_i, \varepsilon_i \in \{0, 1\}\}$.

Theorem 1.8. ([Hil]) *For any $d, r \in \mathbb{N}$ and any partition $\mathbb{N} = \bigcup_{i=1}^r C_i$, one of the C_i contains infinitely many translates of a d-cube.*

We shall see below that Hilbert's theorem admits a very simple proof based on a version of Poincaré recurrence theorem. But first we are going to formulate and discuss Hindman's classical Finite Sums Theorem, proved in [Hin1], which contains both Schur's and Hilbert's theorems as very special cases.

Definition 1.9. *Let $(x_i)_{i=1}^\infty \subset \mathbb{N}$. The IP set generated by the sequence $(x_i)_{i=1}^\infty$ is the set $FS(x_i)_{i=1}^\infty$ of finite sums of elements of $(x_i)_{i=1}^\infty$ with distinct indices:*

$$FS(x_i)_{i=1}^\infty = \Big\{x_\alpha = \sum_{i \in \alpha} x_i, \alpha \subset \mathbb{N}, 1 \leq |\alpha| < \infty\Big\}.$$

IP sets can be viewed as a natural generalization of the notion of a d-cube (if one disregards the following subtle distinction: while the vertices x_i of the d-cube are supposed to be distinct, no such assumption is made in Definition 1.9.) This explains the term IP (coined by H. Furstenberg and B. Weiss in [FuW1]): Infinite-dimensional Parallelepiped.

Theorem 1.10. ([Hin1]) *For any finite partition of $\mathbb{N}$, one of the cells of the partition contains an IP set.*

The original proof of Theorem 1.10 in [Hin1] was, in Hindman's own words, "horrendously complicated." It therefore comes as a pleasant surprise that Hindman's theorem admits a short and easy proof. The following simple proposition is the key to proofs of Hindman's and many other results of a similar nature.

Theorem 1.11. *Let $\mathcal{S}$ be a family of nonempty sets in $\mathbb{N}$. If $\mathcal{S}$ has the following property:*

(i) *for any $A \in \mathcal{S}$ there exist arbitrarily large $t \in \mathbb{N}$ such that*
$$A \cap (A - t) \in \mathcal{S},$$

then for any $A \in \mathcal{S}$ and any $d \in \mathbb{N}$, there exist $t_1 < t_2 < \ldots < t_d$ such that A contains infinitely many translates of the d-cube $Q(t_1, t_2, \ldots, t_d)$. If the following stronger property holds:

(ii) *for any $A \in \mathcal{S}$ there exist arbitrarily large $t \in A$ such that*
$$A \cap (A - t) \in \mathcal{S},$$

then each $A \in \mathcal{S}$ contains an IP set.

Proof. Let $A \in \mathcal{S}$ and let t_1 be such that $A_1 = A \cap (A - t_1) \in \mathcal{S}$. By assumption, there exists $t_2 > t_1$ such that $A_2 = A_1 \cap (A_1 - t_2) \in \mathcal{S}$. But $A_2 = A \cap (A - t_1) \cap (A - t_2) \cap (A - (t_1 + t_2))$ and so it is clear that, for any $a \in A_2$, one has $a + Q(t_1, t_2) \subset A$. Continuing in this fashion one gets, after d steps, $t_1 < t_2 < \ldots < t_d$ such that $A_d = \bigcap_{\alpha \in \mathcal{F}_d} (A - t_\alpha) \in \mathcal{S}$, where $\mathcal{F}_d$ is the set of all subsets of $\{1, 2, \ldots, d\}$ and $t_\alpha = \sum_{\alpha \in \mathcal{F}_d} t_i$. Then any $a \in A_d$ has the property that $a + Q(t_1, t_2, \ldots, t_d) \subset A_d \subset A$, which proves the first assertion of the theorem. Now, let us assume that property (ii) holds. It is easy to see that by choosing at each step $t_i \in A_{i-1}$, where $A_0 = A$, one gets, for any $d \in \mathbb{N}$, $Q(t_1, t_2, \ldots, t_d) \subset A$. This clearly implies that $FS(t_i)_{i=1}^{\infty} \subset A$ and we are done. $\square$

Recall that, for a set $A \subset \mathbb{N}$, the upper density $\overline{d}(A)$ is defined by $\overline{d}(A) = \limsup_{N \to \infty} \frac{|A \cap \{1,2,\ldots,N\}|}{N}$. It is easy to see, by trivial adaptation of the proof of Theorem 1.4 above, that if $\overline{d}(A) > 0$ then there exist arbitrarily large $t \in \mathbb{N}$ such that $\overline{d}(A \cap (A - t)) > 0$. Applying Theorem 1.11, we have now the following result which, in view of the fact that for any finite partition $\mathbb{N} = \bigcup_{i=1}^{r} C_i$ at least one of the C_i has positive upper density, may be considered as a strengthening of Hilbert's Theorem 1.8.

Theorem 1.12. *Let $A \subset \mathbb{N}$ have positive upper density. Then for any $d \in \mathbb{N}$, there exist $t_1 < t_2 < \ldots < t_d$ such that the set*

$$\{a \in A : a + Q(t_1, t_2, \ldots, t_d) \subset A\}$$

has positive upper density. In particular, A contains infinitely many translates $a + Q(t_1, t_2, \ldots, t_d)$ with $a \in A$.

Remark 1.13. One says that Theorem 1.12 is a *density* version of Theorem 1.8, which is a result about *partitions*. While we were lucky to produce a rather trivial proof of this density result, usually this is not the case. As we shall see in detail in Section 4, the density versions of partition results are much deeper and have rather involved and sophisticated proofs.

As we shall momentarily see, Hindman's theorem also follows from Theorem 1.11. To make the derivation possible, one needs only to find a family $\mathcal{S}$ of subsets of $\mathbb{N}$ which satisfies condition (ii) and has the property that for any finite partition $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, one of the C_i belongs to $\mathcal{S}$. This is best achieved by utilizing $\beta\mathbb{N}$, the Stone-Čech compactification of $\mathbb{N}$ interpreted as the space of ultrafilters on $\mathbb{N}$. To be more precise, one utilizes the fact that, with respect to a naturally inherited operation extending the addition in $\mathbb{N}$, $\beta\mathbb{N}$ is a compact semitopological semigroup and, as such, has an idempotent. Any such idempotent allows one to introduce a certain $\{0,1\}$-valued measure μ on the power set $\mathcal{P}(\mathbb{N})$ which, in turn, provides the sought after family $\mathcal{S}$ by the rule $A \in \mathcal{S} \Leftrightarrow \mu(A) = 1$.

The properties of such measures are described in the following proposition, the proof of which will be given in Section 3. (See Theorem 3.3 below.)

Proposition 1.14. *There exists a finitely additive* $\{0,1\}$*-valued probability measure* μ *on the space* $\mathcal{P}(\mathbb{N})$ *of all subsets of* $\mathbb{N}$ *which is "almost shift-invariant" in the following sense. For any* $C \subset \mathbb{N}$ *with* $\mu(C) = 1$*, the set*

$$T_C = \big\{n \in \mathbb{N} : \mu(C - n) = 1\big\} \tag{1.1}$$

satisfies $\mu(T_C) = 1$*.*

We are now in a position to give a proof of Hindman's theorem.

Proof of Theorem 1.10. Let μ be an almost shift-invariant measure as described in Proposition 1.14, and let $\mathbb{N} = \bigcup_{i=1}^{r} C_i$ be a finite partition. Since μ is a probability measure, $\mu\big(\bigcup_{i=1}^{r} C_i\big) = 1$, which, by finite additivity and $\{0,1\}$-valuedness, implies that one of the C_i, call it C, satisfies $\mu(C) = 1$. By (1) we have $\mu(T_C) = 1$, which implies that $\mu(C \cap T_C) = 1$. It follows that the set $\big\{n \in C : \mu(C - n) = 1\big\}$ is of full measure and, in particular, that property (ii) in Theorem 1.11 is satisfied. Hence C contains an IP set and we are done.

□

Remark 1.15. Hindman's theorem finds numerous applications in ergodic theory, topological dynamics, and Diophantine analysis. Some of these will be discussed in this survey. Before moving on with our discussion, we want to record here the following equivalent version of Hindman's theorem, which can be interpreted as "indestructibility" of IP sets under finite partitions.

Theorem 1.16. *For any finite partition of an IP set, one of the cells of the partition contains an IP set.*

We leave the elementary derivation of Theorem 1.16 from Hindman's theorem to the reader. (The other direction is trivial due to the fact that $\mathbb{N} = FS(2^i)_{i=1}^{\infty}$.) On a more sophisticated level, offered by the familiarity with $\beta\mathbb{N}$, Theorem 1.16 becomes an immediate consequence of the proof of Hindman's theorem given above. Indeed, one can show that any IP set in $\mathbb{N}$ is the support of an almost shift-invariant measure. (See Theorem 3.4 below.)

Our next example is the celebrated van der Waerden theorem.

Theorem 1.17. ([vdW1],[vdW2]) *For any $r \in \mathbb{N}$ and any finite partition $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, one of the C_i contains arbitrarily long arithmetic progressions.*

We remark that one cannot, in general, expect to get in Theorem 1.17 an infinite arithmetic progression in one of the C_i. Indeed, let us represent $\mathbb{N}$ as the union of disjoint intervals of increasing length and alternately color them red and blue. This obviously gives a two-coloring $\mathbb{N} = R \cup B$ without an infinite monochromatic progression.

The idea behind this remark can also be utilized to show that Theorem 1.17 implies the following, ostensibly stronger, finitistic version.

Theorem 1.18. *For any $r, l \in \mathbb{N}$ there exists $c = c(r, l)$ such that if $N \geq c$, then for any partition $\{1, 2, \ldots, N\} = \bigcup_{i=1}^{r} C_i$, one of the C_i contains an arithmetic progression of length l.*

Proof of Theorem 1.18 via Theorem 1.17. Assume by way of contradiction that Theorem 1.18 fails. Then there exist natural numbers r, l and, for any $N \in \mathbb{N}$, an interval I with $|I| \geq N$ and an r-coloring of I, which we will find convenient to view as a mapping $f : I \to \{1, 2, \ldots, r\}$, such that I contains no monochromatic progression of length l. Calling such r-colorings (and the corresponding intervals) AP_l-free, we may assume without loss of generality that AP_l-free intervals I_n, $n \in \mathbb{N}$ tile $\mathbb{N}$ and satisfy $|I_{n+1}| \geq 2|I_n|$.

Given an r-coloring $f : I \to \{1, 2, \ldots, r\}$ of an interval I, let us call the r-coloring defined by $\tilde{f} : I \to \{r+1, r+2, \ldots, 2r\}$ a *disjoint copy of* f if for all $k \in I$, $f(k) = \tilde{f}(k) - r$. To finish the argument, let us replace, for every $n \in \mathbb{N}$, the AP_l-free colorings $f_{2n} : I_{2n} \to \{1, 2, \ldots, r\}$ by their disjoint copies $\tilde{f_{2n}} : I_{2n} \to \{r+1, r+2, \ldots, 2r\}$. This results in a $2r$-coloring of $\mathbb{N}$ which has no monochromatic arithmetic progressions of length l, which contradicts Theorem 1.17.

□

While in Khintchine's book ([Kh]) van der Waerden's theorem is called a "pearl of number theory", it should, perhaps, be more properly called a pearl of geometry. Indeed, it is not hard to see that van der Waerden's theorem is equivalent to the following result, which not only has an apparent geometric flavor, but also is suggestive of natural multidimensional extensions.

Theorem 1.19. *For any finite partition $\mathbb{Z} = \bigcup_{i=1}^{r} C_i$, one of the C_i has the property that for any finite set $F \subset \mathbb{Z}$, there exist $a \in \mathbb{Z}$, and $b \in \mathbb{N}$ such that $aF + b = \{ax + b : x \in F\} \subset C_i$. In other words, one of the C_i contains a homothetic copy of any finite set.*

Here is the formulation of the multidimensional analogue of Theorem 1.19. It was first proved by Grünwald (Gallai), who apparently never published his proof. (Grünwald's authorship is acknowledged in [Rado], p. 123.)

Theorem 1.20. *For any $d \in \mathbb{N}$ and any finite partition $\mathbb{Z}^d = \bigcup_{i=1}^{r} C_i$, one of the C_i has the property that for any finite set $F \subset \mathbb{Z}^d$, there exist $n \in \mathbb{N}$, and $v \in \mathbb{Z}^d$ such that $nF + v = \{nx + v : x \in F\} \subset C_i$.*

We shall now formulate yet another, dynamical, version of the (multidimensional) van der Waerden theorem. The idea to apply the methods of topological dynamics to partition results is due to H. Furstenberg and B. Weiss. (See [FuW1].)

Theorem 1.21. (Cf. [FuW1], Theorem 1.4.) *Let $d \in \mathbb{N}$, $\varepsilon > 0$, and let X be a compact metric space. For any finite set of commuting homeomorphisms $T_i : X \to X$, $i = 1, 2, \ldots, k$, there exist $x \in X$ and $n \in \mathbb{N}$ such that*

$$\operatorname{diam}\{x, T_1^n x, T_2^n x, \ldots, T_k^n x\} < \varepsilon.$$

The reader will find various proofs of Theorem 1.21 in Sections 2 and 3. For now, we shall confine ourselves to the proof of the equivalence of Theorems 1.20 and 1.21.

Theorem 1.20 ⇒ Theorem 1.21. Let $y \in X$ be arbitrary. For a vector $m = (m_1, m_2, \ldots, m_k) \in \mathbb{Z}^k$, write $T^m y = T_1^{m_1} T_2^{m_2} \ldots T_k^{m_k} y$. Since X is compact, there exists a finite family of open balls of radius $\frac{\varepsilon}{2}$, call it $\{B_i\}_{i=1}^r$, which covers X. Assign to each $m \in \mathbb{Z}^k$ the minimal i for which $T^m y \in B_i$. This produces a finite coloring $\mathbb{Z}^k = \bigcup_{i=1}^{r'} C_i$ (where $r' \leq r$.) Let $S = \{0, e_1, \ldots, e_k\}$, where e_i are the standard unit vectors. By Theorem 1.20, there exist $n \in \mathbb{N}$ and $v \in \mathbb{Z}^k$ such that $nS + v$ is monochromatic. But this means that $T^v y, T^{v+ne_1} y, \ldots, T^{v+ne_k} y$ all belong to the same $\frac{\varepsilon}{2}$-ball. Writing $x = T_y^v$ and noting that $T^{ne_i} = T_i^n$, $i = 1, 2, \ldots, k$, we get $\operatorname{diam}\{x, T_1^n x, T_2^n x, \ldots, T_k^n x\} < \varepsilon$.

Theorem 1.21 ⇒ Theorem 1.20. The r-colorings of $\mathbb{Z}^d$ (viewed as mappings from $\mathbb{Z}^d$ to $\{1, 2, \ldots, r\}$) are naturally identified with the points of the compact product space $\Omega = \{1, 2, \ldots, r\}^{\mathbb{Z}^d}$. For $m = (m_1, m_2, \ldots, m_d) \in \mathbb{Z}^d$, let $|m| = \max_{1 \leq i \leq d} |m_i|$. Introduce a metric on Ω by defining, for any pair $x, y \in \Omega$, $\rho(x, y) = \inf\left\{\frac{1}{n} : x(m) = y(m) \text{ for } m \text{ with } |m| < n\right\}$. It is easy to see that the metric ρ is compatible with the product topology and that $\rho(x, y) < 1 \Leftrightarrow x(0) = y(0)$. Let $F = \left\{a_1, a_2, \ldots, a_k\right\} \subset \mathbb{Z}^d$. Define the homeomorphisms $T_i : \Omega \to \Omega, i = 1, 2, \ldots, k$, by $(T_i x)(m) = x(m + a_i)$, and set, for $n = (n_1, n_2, \ldots, n_k) \in \mathbb{Z}^k$, $T^n = T_1^{n_1} T_2^{n_2} \ldots T_k^{n_k}$. Let now $x(m)$ be the element of Ω corresponding to the coloring $\mathbb{Z}^d = \bigcup_{i=1}^r C_i$ (in other words, for any $m \in \mathbb{Z}^d$, $x(m) = i$ iff $m \in C_i$.) Let $X = \overline{\{T^n x\}_{n \in \mathbb{Z}^k}}$ be the orbital closure of x in Ω. Note that for any $\delta > 0$ and any $y \in X$, there exists $m \in \mathbb{Z}^k$ with $\rho(T^m x, y) < \delta$. Setting $\varepsilon = 1$ in Theorem 1.21, we can find $y \in X$ and $n \in \mathbb{N}$ such that $\operatorname{diam}\{y, T_1^n y, \ldots, T_k^n y\} < 1$. Choosing $u = (u_1, \ldots, u_k) \in \mathbb{Z}^k$ so that the element of the orbit $T^u x$ is close enough to y, and also such that $T_i^n(T^m x)$ are close enough to $T_i^n y$ for $i = 1, 2, \ldots, k$, we shall still have $\operatorname{diam}\{T^u x, T^u T_1^n x, \ldots, T^u T_k^n x\} < 1$. This implies that, for $v = u_1 a_1 + u_2 a_2 + \ldots + u_k a_k$, $x(v) = x(v + na_1) = \ldots = x(v + na_k)$, which means that the set $v + nF$ is monochromatic.

□

In accordance with the general philosophy of Ramsey Theory (see [Be3] for more discussion), one should expect the density version of Theorem 1.20 to hold true as well. While the proof of this density version is far from being trivial, its formulation is easily guessable (see Theorem 1.23 below.) It is also natural to expect that the dynamical form of the multidimensional van der Waerden theorem, our Theorem 1.21, can be "upgraded" in such a way that it gives a dynamical equivalent to the density version of Theorem 1.20. Theorem 1.24 below, proved in [FuK1], confirms these expectations. To present the historical development in its natural order, we should mention that already the density version of the one-dimensional van der Waerden theorem, conjectured by Erdős and Turán in the mid-thirties in [ET], proved quite recalcitrant and was settled only in 1975 by Szemerédi ([Sz]). A few years later, Furstenberg ([Fu2]) gave a completely different, ergodic-theoretical proof of Szemerédi's theorem, thereby starting a new area of dynamics which is today called Ergodic Ramsey Theory. The multidimensional Szemerédi theorem proved in [FuK1] was the first result in the long and impressive line of dynamical proofs of various combinatorial and number-theoretical results, most of which still do not have a conventional proof. Many of these results will be discussed in the subsequent sections. (See [FuW1], [FuK1], [FuK3], [FuK3], [FuK4], [Le1], [Le2], [BeL1], [BeL2], [BeL3], [BeM1], [BeM3], [BeMZ].)

Definition 1.22. *Let $d \in \mathbb{N}$ and $E \subset \mathbb{Z}^d$.*
(i) *The upper density of E, $\overline{d}(E)$, is defined by*

$$\overline{d}(E) = \limsup_{N\to\infty} \frac{|E \cap [-N, N]^d|}{(2N+1)^d}.$$

(ii) *The upper Banach density of E, $d^*(E)$, is defined by*

$$d^*(E) = \limsup_{N_i - M_i \to \infty, 1\le i\le d} \frac{|E \cap \prod_{i=1}^d [M_i, N_i - 1]|}{\prod_{i=1}^d (N_i - M_i)}$$

Here are now combinatorial and dynamical formulations of the density version of the multidimensional van der Waerden theorem. (Cf. [FuK1].)

Theorem 1.23. *Let $d \in \mathbb{N}$, and let $E \subset \mathbb{Z}^d$ have positive upper Banach density. For any finite set $F \subset \mathbb{Z}^d$, there exist $n \in \mathbb{N}$ and $v \in \mathbb{Z}^d$ such that $nF + v \subset E$.*

Theorem 1.24. *Let $(X, \mathcal{B}, \mu)$ be a probability measure space. For any finite set $\{T_1, \dots, T_k\}$ of commuting measure preserving transformations of X and for any $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists $n \in \mathbb{N}$ such that*

$$\mu\big(A \cap T_1^{-n}A \cap T_2^{-n}A \cap \ldots \cap T_k^{-n}A\big) > 0.$$

To see that Theorem 1.23 follows from Theorem 1.24, one can use a correspondence principle, introduced by Furstenberg in [Fu2] in order to derive

Szemerédi's theorem from an ergodic multiple recurrence result which he established in [Fu2] and which corresponds to taking $T_i = T^i, i = 1, 2, \ldots, k$ in Theorem 1.24.

For the proof of the following version of Furstenberg's correspondence principle, see [BeM3], Prop. 7.2. See also Theorem 5.8 in Section 5 for a general form of Furstenberg's correspondence principle for amenable (semi)groups.

Theorem 1.25. *Let $d \in \mathbb{N}$. For any set $E \subset \mathbb{Z}^d$ with $d^*(E) > 0$, there exists a probability measure preserving system $(X, \mathcal{B}, \mu, \{T^{\boldsymbol{n}}\}_{\boldsymbol{n}\in\mathbb{Z}^d})$ and a set $A \in \mathcal{B}$ with $\mu(A) = d^*(E)$ such that for all $k \in \mathbb{N}$ and $\boldsymbol{n}_1, \boldsymbol{n}_2, \ldots, \boldsymbol{n}_k \in \mathbb{Z}^d$ one has*

$$d^*\big(E \cap (E - \boldsymbol{n}_1) \cap \ldots \cap (E - \boldsymbol{n}_k)\big) \geq \mu\big(A \cap T^{\boldsymbol{n}_1} A \cap \ldots \cap T^{\boldsymbol{n}_k} A\big).$$

We leave it to the reader to verify (with the help of Theorem 1.25) that Theorem 1.24 implies Theorem 1.23. Here is a simple proof of the other implication (the idea behind this proof will also be used in the proof of Lemma 5.10 in Section 5.)

Theorem 1.23 $\Rightarrow$ Theorem 1.24. Assume, by way of contradiction, that there exist a probability measure space $(X, \mathcal{B}, \mu)$, commuting measure preserving transformations $T_1, T_2, \ldots, T_k$ of X, and a set $A \in \mathcal{B}$ with $\mu(A) > 0$ such that for all $n \in \mathbb{N}$, $\mu(A \cap T_1^{-n} A \cap \ldots \cap T_k^{-n} A) = 0$. Deleting, if needed, a set of measure zero from A, we may and will assume that one actually has $A \cap T_1^{-n} A \cap \ldots \cap T_k^{-n} A = \emptyset$ for all $n \in \mathbb{N}$. For $\boldsymbol{n} = (n_1, n_2, \ldots, n_k)$, write $T^{\boldsymbol{n}} = T_1^{n_1} T_2^{n_2} \ldots T_k^{n_k}$ and let

$$f_N(x) = \frac{1}{(2N+1)^k} \sum_{\boldsymbol{n}\in[-N,N]^k} 1_A(T^{\boldsymbol{n}} x),\ \ N = 1, 2, \ldots$$

Note that $0 \leq f_N(x) \leq 1$ for all $x \in X$ and $N \in \mathbb{N}$ and that $\int f_N d\mu = \mu(A)$. Let $f(x) = \limsup_{N\to\infty} f_N(x)$. By Fatou's lemma, we have

$$\int f d\mu = \int \limsup_{N\to\infty} f_N d\mu \geq \limsup_{N\to\infty} \int f_N d\mu = \mu(A).$$

It follows that there exists $x_0 \in X$ such that $\limsup f_N(x_0) = f(x_0) \geq \mu(A)$. Hence for some increasing sequence $N_i \to \infty$, one has

$$\lim_{i\to\infty} f_{N_i}(x_0) = \lim_{i\to\infty} \frac{1}{(2N_i+1)^k} \sum_{\boldsymbol{n}\in[-N_i,N_i]^k} 1_A(T^{\boldsymbol{n}} x_0) = f(x_0) \geq \mu(A).$$

This implies that the set $E = \{\boldsymbol{n} \in \mathbb{Z}^k : T^{\boldsymbol{n}} x_0 \in A\}$ has positive upper density. (We actually showed that $\overline{d}(E) \geq \mu(A)$.) By Theorem 1.23, the set E contains a configuration of the form $nF + v$, where $F = \{0, e_1, e_2, \ldots, e_k\}$, $n \in \mathbb{N}$, and $v \in \mathbb{Z}^k$ (where e_i are the standard unit vectors.) This implies that $A \cap T_1^{-n} A \cap \ldots \cap T_k^{-n} A \neq \emptyset$, which contradicts the assumption made above.

$\square$

It is perhaps of interest to observe that while the combinatorial version of the multidimensional van der Waerden theorem follows immediately from Theorem 1.23 by the observation that for any finite partition $\mathbb{Z}^d = \bigcup_{i=1}^r C_i$, at least one of the C_i satisfies $\overline{d}(C_i) \geq \frac{1}{r}$, the derivation of Theorem 1.21 from Theorem 1.24 is less trivial, and depends on the fact that for any $\mathbb{Z}^d$-action by homeomorphisms of a compact space, there exists an invariant measure.

We would like to formulate still another important extension of van der Waerden's theorem, the powerful Hales-Jewett theorem.

Consider the following generalization of tic-tac-toe: there are r players which are taking turns in placing the symbols $s_1, \ldots, s_r$ in the $k \times k \times \ldots \times k$ (n times) array, which one views as the n-th cartesian power A^n of a k-element set $A = \{a_1, a_2, \ldots, a_k\}$. (In the classical tic-tac-toe, we have $r = 2$, $k = 3$, $n = 2$.) It is convenient to think of the symbols $s_1, \ldots, s_r$ as colors, and to identify the elements of the array A^n as the set $W_n(A)$ of words of length n over the alphabet A. We are going to define now the notion of a *combinatorial line* in A^n. Let $\tilde{A} = A \cup \{t\}$ be an extension of the alphabet A, obtained by adding a new symbol t. Let $W_n(t)$ be the set of words of length n over $\tilde{A}$ in which the symbol t occurs. Given a word $w(t) \in W_n(t)$, let us define a combinatorial line as a set $\{w(a_1), w(a_2), \ldots, w(a_k)\}$ obtained by substituting for t the elements of A. For example, the word $13t241t2$ over the alphabet $\{1, 2, 3, 4, 5\} \cup \{t\}$ gives rise to the combinatorial line $\{13124112, 13224122, 13324132, 13424142, 13524152\}$. The goal of the players is to obtain a monochromatic combinatorial line. The following celebrated theorem of Hales and Jewett ([HaJ]) implies that for fixed r, k and large enough n, the first player can always win.

Theorem 1.26. *Let $r, k \in \mathbb{N}$. There exists $c = c(k, r)$ such that if $n \geq c$, then, for any r-coloring of the set $W_n(A)$ of words of length n over the k-letter alphabet $A = \{a_1, a_2, \ldots, a_k\}$, there is a monochromatic combinatorial line.*

Taking $A = \{0, 1, \ldots, l - 1\}$ and interpreting $W_n(A)$ as integers in base l

having at most n digits in their base l expansion, we see that in this situation, the elements of a combinatorial line form an arithmetic progression of length l (with difference of the form $d = \sum_{i=0}^{k-1} \varepsilon_i l^i$, where $\varepsilon_i = 0$ or 1.) Thus van der Waerden's theorem is a corollary of Theorem 1.26.

Take now A to be a finite field F. Then $W_n(F) = F^n$ has the natural structure of an n-dimensional vector space over F. It is easy to see that, in this case, a combinatorial line is an affine linear one-dimensional subspace of F^n. We have therefore the following corollary of the Hales-Jewett theorem.

Theorem 1.27. *Let F be a finite field. For any $r \in \mathbb{N}$ there exists $c = c(r)$ such that if V is a vector space over F having dimension at least c, then for any r-coloring $V = \bigcup_{i=1}^{r} C_i$, one of the C_i contains an affine line.*

One of the signs of the fundamental nature of the Hales-Jewett theorem is that one can easily derive from it its multidimensional version. (This fact will be especially appreciated by anyone who tried to derive from van der Waerden's theorem its multidimensional version.) Let $t_1, t_2, \ldots, t_m$ be m variables and let $w(t_1, t_2, \ldots, t_m)$ be a word of length n over the alphabet $A \cup \{t_1, \ldots, t_m\}$. (We assume, of course, that the letters t_i do not belong to A.) If, for some n, $w(t_1, \ldots, t_m)$ is a word of length n in which all of the variables $t_1, t_2, \ldots, t_m$ occur, the result of the substitution $\big\{w(t_1, t_2, \ldots, t_m)\big\}_{(t_1,t_2,\ldots,t_m)\in A^m} = \big\{w(a_{i_1}, a_{i_2}, \ldots, a_{i_m}) : a_{i_j} \in A, j = 1, 2, \ldots, m\big\}$ is called a combinatorial m-space.

Observe now that if we replace the original alphabet A by A^m, then a combinatorial line in $W_n(A^m)$ can be interpreted as an m-space in $W_{nm}(A)$. Thus, we have the following ostensibly stronger theorem as a corollary of Theorem 1.26.

Theorem 1.28. *Let $r, k, m \in \mathbb{N}$. There exists $c = c(r, k, m)$ such that if $n \geq c$, then for any r-coloring of the set $W_n(A)$ of words of length n over the k-letter alphabet A, there exists a monochromatic m-space.*

Theorem 1.28 obviously implies the following multidimensional extension of Theorem 1.27.

Theorem 1.29. *Let F be a finite field. For any $r, m \in \mathbb{N}$, there exists $c = c(r, m)$ such that if V is a vector space over F having dimension at least c, then for any r-coloring $V = \bigcup_{i=1}^{r} C_i$, one of the C_i contains an m-dimensional affine space.*

We leave it to the reader to derive from Theorem 1.28 the multidimensional van der Waerden theorem and an extension of Theorem 1.27 pertaining to m-dimensional affine subspaces of F^n. See Section 2 for a proof of the Hales-Jewett theorem and for more discussion and applications.

Our next example is the following surprising theorem, which was proved independently by Sárközy and Furstenberg, and which has interesting links with spectral theory, Diophantine approximations, combinatorics, and dynamical systems. (See [Sa], [Fu2], [Fu3], [Fu4], [KM].)

Theorem 1.30. *Let $E \subset \mathbb{N}$ be a set of positive upper density, and let $p(n) \in \mathbb{Z}[n]$ be a polynomial with $p(0) = 0$. Then there exist $x, y \in E$ and $n \in \mathbb{N}$ such that $x - y = p(n)$.*

This result is perhaps more surprising than any of the theorems formulated above. One can surely expect the set of differences of a large set to be even larger. For example, if, for $E \subset \mathbb{N}$, $d^*(E) > 0$, then it is not hard to show that the set of differences $E - E = \{x - y : x, y \in E\}$ is syndetic, i.e. has bounded gaps. (See, for example, [Fu3], Prop. 3.19, or [Be3], pp. 8–9.) But there is, a priori, no obvious reason for the set $E - E$ to be so "well spread" as to nontrivially intersect the set of values taken by any integer polynomial vanishing at zero. The following dynamical counterpart of Theorem 1.30, due to Furstenberg, is just as striking. (See [Fu2], Prop. 1.3 and [Fu3], Thm. 3.16.)

Theorem 1.31. *For any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ with $\mu(A) > 0$, and any polynomial $p(n) \in \mathbb{Z}[n]$ with $p(0) = 0$, there exists $n \in \mathbb{N}$ such that $\mu(A \cap T^{p(n)}A) > 0$.*

Remarks.

1. One can derive Theorem 1.30 from Theorem 1.31 by utilizing Furstenberg's correspondence principle. In the other direction, one can, for example, mimic the argument that was used above to derive Theorem 1.24 from Theorem 1.23.

2. One should, of course, view Theorem 1.31 as a refinement of the Poincaré recurrence theorem. While the classical Poincaré recurrence theorem only tells us that a typical point returns, under the evolution laws of the dynamical system, to a set of positive volume in the phase space, Theorem 1.31 tells us that this will happen along any prescribed in advance sequence of "polynomial" times. However, when compared with the Poincaré recurrence theorem, Theorem 1.31 is a rather deep result. This is, in particular, manifested by the fact that all the known proofs ot the Theorem 1.31 prove actually more than stated.

Furstenberg's proof of Theorem 1.31 utilizes the spectral theorem for unitary operators. The proof that we have chosen to present here is "softer" in the sense that it avoids the usage of the spectral theorem and thereby is susceptible to further generalizations. (See Theorems 4.2.10 and 4.2.13 below.)

We shall need the following useful result, which can be viewed as a Hilbert space version of the classical van der Corput difference theorem in the theory of uniform distribution.

Theorem 1.32. (van der Corput trick) *Let $(u_n)_{n\in\mathbb{N}}$ be a bounded sequence in a Hilbert space $\mathcal{H}$. If for every $h \in \mathbb{N}$ it is the case that*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h}, u_n\rangle = 0,$$

then $\lim_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=1}^{N} u_n\Big\| = 0.$

Proof. Observe that for any $\varepsilon > 0$ and any $H \in \mathbb{N}$, if N is large enough then

$$\Big\|\frac{1}{N}\sum_{n=1}^{N} u_n - \frac{1}{N}\frac{1}{H}\sum_{n=1}^{N}\sum_{h=0}^{H-1} u_{n+h}\Big\| < \varepsilon.$$

But,

$$\begin{aligned}\limsup_{N\to\infty}\Big\|\frac{1}{N}\frac{1}{H}\sum_{n=1}^{N}\sum_{h=0}^{H-1} u_{n+h}\Big\|^2 &\le \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\Big\|\frac{1}{H}\sum_{h=0}^{H-1} u_{n+h}\Big\|^2\\ &= \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\frac{1}{H^2}\sum_{h_1,h_2=0}^{H-1}\langle u_{n+h_1}, u_{n+h_2}\rangle \le \frac{B}{H},\end{aligned}$$

where $B = \sup_{n\in\mathbb{N}}\|u_n\|^2$. Since H was arbitrary, we are done. □

Proof of Theorem 1.31. Let $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$ and let $U : \mathcal{H} \to \mathcal{H}$ be the unitary operator induced by T:

$$(Uf)(x) = f(Tx), f \in L^2(X, \mathcal{B}, \mu).$$

Let, for each $a \in \mathbb{N}$,

$$\mathcal{H}_a = \{f : U^a f = f\}$$

$$\mathcal{H}_{erg}^{(a)} = \{f : \Big|\Big|\frac{1}{N}\sum_{n=0}^{N-1} U^{an} f\Big|\Big| \to 0\}$$

The classical ergodic splitting (with respect to U^a), $\mathcal{H} = \mathcal{H}_a \oplus \mathcal{H}_{erg}^{(a)}$, leads to the following, more suitable for our goals, splitting of $\mathcal{H}$ into "rational spectrum" and "totally ergodic" parts. Let

$$\mathcal{H}_{rat} = \overline{\{f : \exists a \in \mathbb{N} : U^a f = f\}} = \overline{\bigcup_{a=1}^{\infty} \mathcal{H}_a}$$

$$\mathcal{H}_{tot.erg.} = \{f : \forall a \in \mathbb{N}, \Big|\Big|\frac{1}{N}\sum_{n=0}^{N-1} U^{an} f\Big|\Big| \to 0\} = \bigcap_{a=1}^{\infty} \mathcal{H}_{erg}^{(a)}$$

It is easy to check now that $\mathcal{H}_{rat}^{\perp} = \mathcal{H}_{tot.erg.}$ and that $\mathcal{H} = \mathcal{H}_{rat} \oplus \mathcal{H}_{tot.erg.}$ Let $1_A = f + g$, where $f \in \mathcal{H}_{rat}$, $g \in \mathcal{H}_{tot.erg.}$ We remark that since $1_A \geq 0$ and $\int 1_A d\mu = \mu(A) > 0$, one has $f \geq 0$, $f \neq 0$. Indeed, f minimizes the distance from $\mathcal{H}_{rat}$ to 1_A, and the function $\max\{f, 0\}$ (which, as is not too hard to check, also belongs to $\mathcal{H}_{rat}$) would do at least as well in minimizing this distance. This remark equally applies, for any $a \in \mathbb{N}$, to the orthogonal projection f_a of 1_A onto $\mathcal{H}_a$. Note also that $\int f_a d\mu = \mu(A)$.

We are going to show that $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \mu(A \cap T^{p(n)}A)$ exists and is positive. Note that, in view of the orthogonal decomposition $1_A = f + g$, we have:

$$\mu(A \cap T^{p(n)}A) = \int (f+g)U^{p(n)}(f+g)d\mu = \int fU^{p(n)}fd\mu + \int gU^{p(n)}gd\mu.$$

We shall show first that $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \int gU^{p(n)}gd\mu = 0$ (and hence it will remain to show that

$$\lim_{N\to\infty} \tfrac{1}{N}\textstyle\sum_{n=0}^{N-1} \mu(A \cap T^{p(n)}A) = \lim_{N\to\infty} \tfrac{1}{N}\sum_{n=0}^{N-1} \int fU^{p(n)}fd\mu > 0.)$$

Note that since $g \in \mathcal{H}_{tot.erg.}$, one has, for any linear polynomial $p(n)$ with integer coefficients, $\lim_{N\to\infty} \| \frac{1}{N} \sum_{n=0}^{N-1} U^{p(n)} g \| = 0$.

We shall use the van der Corput trick to inductively reduce the situation to this linear case. Let $u_n = U^{p(n)} g, n \in \mathbb{N}$. We have:

$$\langle u_{n+h}, u_n \rangle = \langle U^{p(n+h)} g, U^{p(n)} g \rangle = \langle U^{p(n+h)-p(n)} g, g \rangle.$$

Notice that, for any fixed $h \in \mathbb{N}$, the degree of the polynomial $p(n+h) - p(n)$ equals $\deg p(n) - 1$. Using the fact that strong convergence implies weak convergence, we have by the induction hypothesis:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle u_{n+h}, u_n \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle U^{p(n+h)-p(n)} g, g \rangle = 0.$$

It follows from Theorem 1.31 that $\lim_{N\to\infty} \| \frac{1}{N} \sum_{n=0}^{N-1} U^{p(n)} g \| = 0$ and hence $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int g U^{p(n)} g d\mu = 0$. It remains now to prove that $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f U^{p(n)} f d\mu > 0$. Note first that the existence of this limit is almost obvious. Indeed, since $f \in \mathcal{H}_{rat}$, it is enough to check only the case when f belongs to one of the $\mathcal{H}_a$, in which case there is practically nothing to check since, for such f, the sequence $U^{p(n)} f$, $n \in \mathbb{N}$, is periodic.

To see that the limit in question is strictly positive, choose $a \in \mathbb{N}$ so that $\|f - f_a\|$ is close to zero (where f_a is the orthogonal projection on $\mathcal{H}_a$). Note now that if $n \in a\mathbb{N}$ then $p(n)$ is divisible by a, and hence $\int f_a U^{p(n)} f_a d\mu = \int f_a^2 d\mu \geq (\mu(A))^2$. This implies that for $n \in a\mathbb{N}$, $\int f U^{p(n)} f d\mu$ is close to $\mu^2(A)$, which clearly implies the positivity of the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f U^{p(n)} f d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T^{p(n)} A).$$

□

We will conclude the introductory section here. Each of the examples above is a small fragment of a much bigger picture. In the subsequent sections, we shall try to supply more facts and details so that the reader will be able to see better both the multiple interconnections between various theorems of Ergodic Ramsey Theory and the general direction of the flow of current developments.

Acknowledgement. The author would like to thank Ronnie Pavlov and Sasha Leibman for their assistance in preparing this survey, and Neil Hindman, Emmanuel Lesigne, and Randall McCutcheon for helpful comments on

the preliminary version of this survey. Special thanks go also to the editors of this volume, Anatole Katok and Boris Hasselblatt for their infinite patience.

2 Topological dynamics and partition Ramsey theory

In various applications, one would like not only to be able to find certain types of monochromatic configurations for any finite coloring of a highly organized structure, such as $\mathbb{N}$ or an infinite vector space over a finite field, but also to know that these configurations are plentiful. There are many notions of largeness which one can use to measure the abundance of sought-after configurations. One of these notions is that of *syndeticity.* A subset S in $\mathbb{N}$ is syndetic if finitely many translates of S cover $\mathbb{N}$, i.e. for some k and $a_1, a_2, \ldots, a_k \in \mathbb{N}$, one has $\bigcup_{i=1}^{k}(S - a_i) = \mathbb{N}$. (This definition can be easily adapted to make sense in any semigroup.)

A stronger notion of largeness which we will presently introduce with the help of IP sets (see Definition 1.9) not only implies syndeticity, but has also the finite intersection property.

Definition 2.1. *A set $E \subseteq \mathbb{N}$ is said to be IP* if it has nontrivial intersection with any IP set.*

It is not too hard to see that any IP* set is syndetic. Indeed, if an IP* set S were not syndetic, then its complement would contain an infinite union of intervals $[a_n, b_n]$ with $b_n - a_n \to \infty$, and it is not hard to show that any such union of intervals contains an IP set, which leads to a contradiction with the assumption that S is an IP* set.

Let us show now that the family of IP* sets has the finite intersection property. It is enough to prove that if S_1, S_2 are IP* sets, then $S_1 \cap S_2$ is as well. Let E be an arbitrary IP set and consider the partition $E = (E \cap S_1) \cup (E \cap S_1^c)$. By Hindman's theorem (see Theorem 1.10), either S_1 or S_1^c has to contain an IP set E_1. But, it is clear that it has to be S_1, since S_1 is IP*, and hence $S_1 \cap E_1 \neq \emptyset$, which implies that $E_1 \subset E \cap S_1$. Now, since S_2 is also an IP* set, we have $E_1 \cap S_2 \neq \emptyset$, which implies that $(E \cap S_1) \cap S_2 = E \cap (S_1 \cap S_2) \neq \emptyset$.

We are going to formulate and prove now the so-called IP van der Waerden theorem (proved first in [FuW1]) which, in particular, will tell us that the set of differences of monochromatic arithmetic progressions always to be found in any finite coloring of $\mathbb{Z}$ is an IP* set. At the same time, this IP van der

Waerden theorem is powerful enough to imply not only Theorem 1.20, but also Theorem 1.29. The proof presented below is taken from [Be5] and is based on the proof of the multidimensional van der Waerden's theorem in [BPT]. We need first to introduce a few more definitions, and some notation.

An *$\mathcal{F}$-sequence* in an arbitrary space Y is a sequence $\{y_\alpha\}_{\alpha\in\mathcal{F}}$ indexed by the set $\mathcal{F}$ of the finite nonempty subsets of $\mathbb{N}$. If Y is a (multiplicative) semigroup, one says that an $\mathcal{F}$-sequence defines an *IP-system* if for any $\alpha = \{i_1, i_2, ..., i_k\} \in \mathcal{F}$, one has $y_\alpha = y_{i_1}y_{i_2}\dots y_{i_k}$. IP-systems should be viewed as generalized semigroups. Indeed, if $\alpha\cap\beta=\emptyset$, then $y_{\alpha\cup\beta} = y_\alpha y_\beta$. We shall often use this formula for sets α, β satisfying $\alpha < \beta$.

We will be working with IP-systems generated by homeomorphisms belonging to a commutative group G acting minimally on a compact space X. (Recall that (X, G) is a *minimal* dynamical system if for each nonempty open set $V \subset X$ there exist $S_1, ..., S_r \in G$ so that $\bigcup\limits_{i=1}^{r} S_iV = X$.)

Theorem 2.2. *Let X be a compact topological space and G a commutative group of its homeomorphisms such that the dynamical system (X, G) is minimal. For any nonempty open set $V \subseteq X$, any $k \in \mathbb{N}$, any IP-systems $\{T^{(1)}_\alpha\}_{\alpha\in\mathcal{F}}, ..., \{T^{(k)}_\alpha\}_{\alpha\in\mathcal{F}}$ in G and any $\alpha_0 \in \mathcal{F}$, there exists $\alpha \in \mathcal{F}, \alpha > \alpha_0$, such that $V \cap T^{(1)}_\alpha V \cap ... \cap T^{(k)}_\alpha V \neq \emptyset$.*

Proof. We fix a nonempty open $V \subseteq X$ and $S_1, ..., S_r \in G$ with the property that $S_1V \cup S_2V \cup \ldots \cup S_rV = X$. (The existence of $S_1, ..., S_r$ is guaranteed by the minimality of (X, G).) The proof proceeds by induction on k. The case $k = 1$ is almost trivial, but we shall do it in detail to set up the notation in a way that indicates the general idea.

So, let $\{T_i\}_{i=1}^{\infty}$ be a fixed sequence of elements in G and $\{T_\alpha\}_{\alpha\in\mathcal{F}}$ the IP-system generated by $\{T_i\}_{i=1}^{\infty}$. (This means of course that for any finite nonempty set $\alpha = \{i_1, i_2, ..., i_m\} \subset \mathbb{N}$, one has $T_\alpha = T_{i_1}T_{i_2}\cdots T_{i_m}$.)

Now we construct a sequence $W_0, W_1, ...$ of nonempty open sets in X so that:

(i) $W_0 = V$;

(ii) $T_n^{-1}W_n \subseteq W_{n-1}, \forall n \geq 1$;

(iii) each $W_n, n \geq 1$, is contained in one of the sets $S_1V, S_2V, ..., S_rV$. (We recall that $S_1V \cup S_2V \cup ... \cup S_rV = X$.)

To define W_1, let t_1, $1 \le t_1 \le r$, be such that $T_1 V \cap S_{t_1} V = T_1 W_0 \cap S_{t_1} V \ne \emptyset$; let $W_1 = T_1 W_0 \cap S_{t_1} V$. If W_n was already defined, then let t_{n+1} be such that $1 \le t_{n+1} \le r$ and $T_{n+1} W_n \cap S_{t_{n+1}} V \ne \emptyset$, and let $W_{n+1} = T_{n+1} W_n \cap S_{t_{n+1}} V$. By the construction, each W_n is contained in one of the $S_1 V, ..., S_r V$, so there will necessarily be two natural numbers $i < j$ and $1 \le t \le r$ such that $W_i \cup W_j \subseteq S_t V$ (pigeonhole principle!). Let $U = S_t^{-1} W_j$ and $\alpha = \{i+1, i+2, ..., j\}$. We have

$$\begin{aligned} T_\alpha^{-1} U &= T_{i+1}^{-1} T_{i+2}^{-1} \dots T_j^{-1} S_t^{-1} W_j = S_t^{-1} T_{i+1}^{-1} T_{i+2}^{-1} \dots T_j^{-1} W_j \\ &\subseteq S_t^{-1} T_{i+1}^{-1} T_{i+2}^{-1} \dots T_{j-1}^{-1} W_{j-1} \subseteq \dots \subseteq S_t^{-1} T_{i+1}^{-1} W_{i+1} \subseteq S_t^{-1} W_i \subseteq V. \end{aligned}$$

So, $U \subseteq T_\alpha V$ and $U \subseteq V$ which implies $V \cap T_\alpha V \ne \emptyset$.

Notice that since the pair $i < j$ for which there exists t with the property $W_i \cup W_j \subseteq S_t V$ could be chosen with arbitrarily large i, it follows that the set $\alpha = \{i+1, ..., j\}$ for which $V \cap T_\alpha V \ne \emptyset$ could be chosen so that $\alpha > \alpha_0$.

Assume now that the theorem holds for any k IP-systems in G. Fix a nonempty set V and $k+1$ IP-systems $\{T_\alpha^{(1)}\}_{\alpha \in \mathcal{F}}, ..., \{T_\alpha^{(k+1)}\}_{\alpha \in \mathcal{F}}$. We shall also fix the homeomorphisms $S_1, ..., S_r \in G$ (whose existence is guaranteed by minimality) satisfying $S_1 V \cup ... \cup S_r V = G$. We shall inductively construct a sequence $W_0, W_1, ...$ of nonempty open sets in X and an increasing sequence $\alpha_1 < \alpha_2 < ...$ in $\mathcal{F}$ so that

(a) $W_0 = V$,

(b) $(T_{\alpha_n}^{(1)})^{-1} W_n \cup (T_{\alpha_n}^{(2)})^{-1} W_n \cup \dots \cup (T_{\alpha_n}^{(k+1)})^{-1} W_n \cup \subseteq W_{n-1}$ for all $n \ge 1$, and

(c) each $W_n, n \ge 1$ is contained in one of the sets $S_1 V, ..., S_r V$.

To define W_1, apply the induction assumption to the nonempty open set $W_0 = V$ and IP-systems

$$\{(T_\alpha^{(k+1)})^{-1} T_\alpha^{(1)}\}_{\alpha \in \mathcal{F}}, ..., \{(T_\alpha^{(k+1)})^{-1} T_\alpha^{(k)}\}_{\alpha \in \mathcal{F}}.$$

There exists $\alpha_1 \in \mathcal{F}$ such that

$$\begin{aligned} &V \cap (T_{\alpha_1}^{(k+1)})^{-1} T_{\alpha_1}^{(1)} V \cap \dots \cap (T_{\alpha_1}^{(k+1)})^{-1} T_{\alpha_1}^{(k)} V \\ &\quad = W_0 \cap (T_{\alpha_1}^{(k+1)})^{-1} T_{\alpha_1}^{(1)} W_0 \cap \dots \cap (T_{\alpha_1}^{(k+1)})^{-1} T_{\alpha_1}^{(k)} W_0 \ne \emptyset. \end{aligned}$$

Applying $T_{\alpha_1}^{(k+1)}$, we get

$$T_{\alpha_1}^{(k+1)} W_0 \cap T_{\alpha_1}^{(1)} W_0 \cap \dots \cap T_{\alpha_1}^{(k)} W_0 \ne \emptyset.$$

It follows that for some $1 \le t_1 \le r$

$$W_1 := T^{(1)}_{\alpha_1} W_0 \cap T^{(2)}_{\alpha_1} W_0 \cap \ldots \cap T^{(k+1)}_{\alpha_1} W_0 \cap S_{t_1} V \ne \emptyset.$$

Clearly, W_0 and W_1 satisfy (b) and (c) above for $n = 1$

If W_{n-1} and $\alpha_{n-1} \in \mathcal{F}$ have already been defined, apply the induction assumption to the nonempty open set W_{n-1} (and the IP-systems $\{(T^{(k+1)}_\alpha)^{-1} T^{(1)}_\alpha\}_{\alpha\in\mathcal{F}}$,..., $\{(T^{(k+1)}_\alpha)^{-1} T^{(k)}_\alpha\}_{\alpha\in\mathcal{F}}$) to get $\alpha_n > \alpha_{n-1}$ such that

$$W_{n-1} \cap (T^{(k+1)}_{\alpha_n})^{-1} T^{(1)}_{\alpha_n} W_{n-1} \cap \ldots \cap (T^{(k+1)}_{\alpha_n})^{-1} T^{(1)}_{\alpha_n} W_{n-1} \ne \emptyset,$$

and hence, for some $1 \le t_n \le r$,

$$W_n := T^{(1)}_{\alpha_n} W_{n-1} \cap ... \cap T^{(k+1)}_{\alpha_n} W_{n-1} \cap S_{t_n} V \ne \emptyset.$$

Again, this W_n clearly satisfies the conditions (b) and (c).

Since, by the construction, each W_n is contained in one of the sets $S_1 V, ..., S_r V$, there is $1 \le t \le r$ such that infinitely many of the W_n are contained in $S_t V$. In particular, there exists i as large as we please and $j > i$ so that $W_i \cup W_j \subseteq S_t V$. Let $U = S_t^{-1} W_j$ and $\alpha = \alpha_{i+1} \cup ... \cup \alpha_j$.

Notice that $U \subseteq V$, and for any $1 \le m \le k+1$, $(T^{(m)}_\alpha)^{-1} U \subseteq V$. Indeed,

$$\begin{aligned}(T^{(m)}_\alpha)^{-1} U = (T^{(m)}_{\alpha_{i+1}\cup\ldots\cup\alpha_j})^{-1} S_t^{-1} W_j = S_t^{-1} (T^{(m)}_{\alpha_{i+1}})^{-1} \ldots (T^{(m)}_{\alpha_j})^{-1} W_j \\ \subseteq S_t^{-1} (T^{(m)}_{\alpha_{i+1}})^{-1} \ldots (T^{(m)}_{\alpha_{j-1}})^{-1} W_{j-1} \subseteq \ldots \subseteq S_t^{-1} (T^{(m)}_{\alpha_{i+1}})^{-1} W_{i+1} \subseteq S_t^{-1} W_i \subseteq V\end{aligned}$$

It follows that $U \cup (T^{(1)}_\alpha)^{-1} U \cup \ldots \cup (T^{(n+1)}_\alpha)^{-1} U \subseteq V$, and this, in turn, implies $V \cap T^{(1)}_\alpha V \cap \ldots \cap T^{(k+1)}_\alpha V \ne \emptyset$. □

Corollary 2.3. *If X is a compact metric space and G a commutative group of its homeomorphisms, then for any k IP-systems $\{T^{(1)}_\alpha\}_{\alpha\in\mathcal{F}}, ..., \{T^{(k)}_\alpha\}_{\alpha\in\mathcal{F}}$ in G, any $\alpha_0 \in \mathcal{F}$, and any $\varepsilon > 0$ there exist $\alpha > \alpha_0$ and $x \in X$ such that the diameter of the set $\{x, T^{(1)}_\alpha x, ..., T^{(k)}_\alpha x\}$ is smaller than ε.*

Proof. If (X, G) is minimal, then the claim follows immediately from Theorem 2.2. If not, then pass to a minimal, nonempty, closed G-invariant subset of X. (Such a subset always exists by Zorn's lemma.) □

Corollary 2.4. *Under the conditions of Corollary 2.3, one can find, for any $m \in \mathbb{N}$, finite sets $\alpha_1 < \alpha_2 < \ldots < \alpha_m$ and $x \in X$ such that x and all the points $T^{(i_1)}_{\alpha_1} x, T^{(i_2)}_{\alpha_2} x, \ldots, T^{(i_m)}_{\alpha_m} x$, $i_1, \ldots, i_m \in \{1, 2, \ldots, k\}$ belong to the same open ball of radius ε.*

Proof. The result follows by simple iteration. Assume that the group generated by $T_\alpha^{(i)}, i = 1, 2, \ldots, k$ acts on X in a minimal fashion. Let V be an open ball of radius ε. By Theorem 2.2, for any $\alpha_0 \in \mathcal{F}$, there exists $\alpha_1 > \alpha_0$ such that

$$V_1 = V \cap \bigcap_{i=1}^{k} \big(T_{\alpha_1}^{(i)}\big)^{-1} V \neq \emptyset.$$

Applying Theorem 2.2 again, one gets $\alpha_2 > \alpha_1$ such that

$$V_2 = V_1 \cap \bigcap_{i=1}^{k} \big(T_{\alpha_1}^{(i)}\big)^{-1} V_1 = V \cap \bigcap_{i_1, i_2 = 1}^{k} \big(T_{\alpha_1}^{i_1} T_{\alpha_2}^{i_2}\big)^{-1} V \neq \emptyset.$$

Let V_k be the nonempty set obtained as the result of k iterations of this procedure. It is easy to see that any $x \in V_k$ satisfies the claim of the Corollary. $\square$

The following corollary of Theorem 2.2 is a refinement of the multidimensional van der Waerden theorem.

Corollary 2.5. *For any $r, d, k \in \mathbb{N}$, any IP sets $\big(n_\alpha^{(1)}\big)_{\alpha \in \mathcal{F}}$, $\big(n_\alpha^{(2)}\big)_{\alpha \in \mathcal{F}}$, $\ldots$, $\big(n_\alpha^{(k)}\big)_{\alpha \in \mathcal{F}}$ in $\mathbb{N}$, any finite set $F = \{u_1, u_2, \ldots, u_k\} \subset \mathbb{Z}^d$ and any partition $\mathbb{Z}^d = \bigcup_{i=1}^r C_i$, there exist $i \in \{1, 2, \ldots, r\}$, $\alpha \in \mathcal{F}$ and $v \in \mathbb{Z}^d$ such that*

$$v + \{n_\alpha^{(1)} u_1, n_\alpha^{(2)} u_2, \ldots, n_\alpha^{(k)} u_k\} \subset C_i.$$

Remark. Taking $d = 1$, all $\big(n_\alpha^{(i)}\big)_{\alpha \in \mathcal{F}}$ identical and $F = \{0, 1, \ldots, k\}$, one obtains the fact that for any finite coloring $\mathbb{N} = \bigcup_{i=1}^r C_i$ and any $k \in \mathbb{N}$, the set $\big\{n \in \mathbb{N} :$ for some $a \in \mathbb{Z}, \{a, a+n, \ldots, a+(k-1)n\}$ is monochromatic$\big\}$ is IP*.

Proof of Corollary 2.5. Notice that Corollary 2.3 implies that for any commuting homeomorphisms $T_1, T_2, \ldots, T_k$ of a compact space X, any $\varepsilon > 0$ and any IP sets $\big(n_\alpha^{(1)}\big)_{\alpha \in \mathcal{F}}$, $\big(n_\alpha^{(2)}\big)_{\alpha \in \mathcal{F}}$, $\ldots$, $\big(n_\alpha^{(k)}\big)_{\alpha \in \mathcal{F}}$ in $\mathbb{N}$, there exists $x \in X$ and $\alpha \in \mathcal{F}$ such that $\operatorname{diam}\big\{x, T_1^{n_\alpha^{(1)}} x, \ldots, T_k^{n_\alpha^{(k)}} x\big\} < \varepsilon$. The desired combinatorial result follows now by the argument which is practically identical to one used in the Introduction in the derivation of Theorem 1.20 from Theorem 1.21. $\square$

Let us show that Theorem 1.29 is also derivable from Theorem 2.2. We shall find it more convenient to deal with the following equivalent form of Theorem 1.29.

Theorem 2.6. *Let F be a finite field and V_F an infinite vector space over F. For any finite coloring $V_F = \bigcup_{i=1}^r C_i$ and any $m \in \mathbb{N}$, there exists a monochromatic affine m-space, that is, an m-dimensional affine subspace.*

Before embarking on the proof of Theorem 2.6, let us briefly explain why Theorems 1.29 and 2.6 are equivalent. Clearly one has only to show that Theorem 2.6 implies Theorem 1.29. This follows from the compactness of the space of r-colorings of V_F. Assuming without loss of generality that V_F is countably infinite, observe that, as an abelian groups, V_F is isomorphic to the direct sum F_∞ of countably many copies of F:

$$F_\infty = \big\{g = (a_1, a_2, \ldots) : a_i \in F \text{ and all but finitely many } a_i = 0\big\} = \bigcup_{n=1}^{\infty} F_n,$$

where $F_n = \big\{g = (a_1, a_2, \ldots) : a_i = 0 \text{ for } i > n\big\} \cong F \oplus \ldots \oplus F$ (n times.)

For $g = (a_1, a_2, \ldots) \in F_\infty$, let $|g|$ be the minimal natural number such that $a_i = 0$ for all $i \geq |g|$. Note that $|g| = 0$ if and only if $g = \mathbf{0} = (0, 0, \ldots)$. We will identify the space of r-colorings of V_F with $\Omega = \{1, 2, \ldots, r\}^{F_\infty}$. For any pair $x = x(g)$, $y = y(g)$, $g \in F_\infty$, of elements of Ω, let

$$\rho(x, y) = \inf_{n \in \mathbb{N}} \big\{\tfrac{1}{n} : x(g) = y(g) \text{ for } g \text{ with } |g| < n\big\}.$$

One readily checks that ρ is a metric on Ω with the property $\rho(x, y) = 1 \Leftrightarrow x(\mathbf{0}) \neq y(\mathbf{0})$. Moreover, (Ω, ρ) is a compact space, and it is the compactness of (Ω, ρ) which, as we shall now see, is behind the fact that Theorem 2.6 implies Theorem 1.29.

Assume that Theorem 2.6 holds true but Theorem 1.29 does not. Then, there exist $r, m \in \mathbb{N}$ such that for any $n \in \mathbb{N}$, there exists an r-coloring $F_n = \bigcup_{i=1}^r C_i$ with no monochromatic affine m-subspace. Viewing each such coloring as a map $f_n : F_n \to \{1, 2, \ldots, r\}$ and extending f_n, for each $n \in \mathbb{N}$, arbitrarily to a map $g_n : F_\infty \to \{1, 2, \ldots, r\}$, we obtain the sequence $(g_n)_{n\in\mathbb{N}}$ of elements of the compact space $\{1, 2, \ldots, r\}^{F_\infty}$, which, by compactness, has a convergent subsequence $(g_{n_i})_{i\in\mathbb{N}}$. The limiting coloring $g = \lim_{i\to\infty} g_{n_i}$ will also not have monochromatic affine m-subspaces, which contradicts Theorem 2.6.

Proof of Theorem 2.6. Fix m IP-systems $\big\{g_\alpha^{(i)}\big\}_{\alpha\in\mathcal{F}}$, $i = 1, 2, \ldots, m$ such that, for each i, $\text{Span}\big\{g_\alpha^{(i)}, \alpha \in \mathcal{F}\big\}$ is an infinite subset in V_F. We will show a stronger fact that, for any partition $V_F = \bigcup_{i=1}^r C_i$, one of the C_i contains an affine m-space of the form $h + \mathrm{Span}\{g_1, \ldots, g_m\}$ with $g_j \in \{g_\alpha^{(j)}\}_{\alpha\in\mathcal{F}}$. In other words, we will show that the set of ordered m-tuples $(g_1, \ldots, g_m)$ such that, for some h, $h + \mathrm{Span}\{g_1, \ldots, g_m\} \subset C_i$ is an IP* set in the m-fold direct sum $F_\infty \oplus \ldots \oplus F_\infty$ (where the notion of IP* is defined in the obvious sense.)

We start with showing that one can always find $\alpha_1 \in \mathcal{F}$ and $i \in \{1, 2, \ldots, r\}$ so that the one-dimensional affine subspace $h + \{cg_{\alpha_1}^{(i)}, c \in F\}$ is contained in C_i. For $h \in V_F$, let $T_h : \Omega \to \Omega$ be defined by $(T_h x)(g) = x(g + h)$. Clearly T_h is a homeomorphism of Ω for every $h \in V_F$. Let $\xi \in \Omega$ be the element in Ω corresponding to the partition $V_F = \bigcup_{i=1}^r C_i$, i.e. $\xi(g) = i \Leftrightarrow g \in C_i$. Finally, let $X \subseteq \Omega$ be the orbital closure of $\xi(g)$: $X = \overline{\{T_h\xi, h \in V_f\}}$.

Use now the IP system $\{g_\alpha^{(1)}\}_{\alpha\in\mathcal{F}}$ to define, for every $c \in F$, $c \neq 0$, an IP system of homeomorphisms $T_\alpha^{(c)} := T_{cg_\alpha^{(1)}}, \alpha \in \mathcal{F}$. In this way, we get $|F| - 1$ IP systems of commuting homeomorphisms of X. Applying Corollary 2.3 to the space X and the IP systems $T_\alpha^{(c)}$ and taking $\varepsilon < 1$, we get a point $x_1 \in X$ and $\alpha_1 \in \mathcal{F}$ such that the diameter of $\{T_{cg_{\alpha_1}^{(1)}}, c \in F\}$ is less than 1. This implies $x_1(0) = x_1(cg_{\alpha_1}^{(1)})$ for every $c \in F$. Since the orbit $\{T_h\xi, h \in V_F\}$ is dense in X, there exists $h_1 \in V_F$ such that $(T_{h_1}\xi)(g)$ and $x_1(g)$ agree on all g satisfying $|g| \leq |g_{\alpha_1}^{(1)}|$. If $\xi(h_1) = i$, then C_i contains the affine line $h_1 + \{cg_{\alpha_1}^{(1)}, c \in F\}$. (We, of course, took care in choosing α_1 so that $g_{\alpha_1}^{(1)} \neq \mathbf{0}$, which is possible in view of our assumptions on $\{g_\alpha^{(i)}\}_{\alpha\in\mathcal{F}}$.) Introducing now the IP systems $T_{cg_\alpha^{(2)}}, c \in F, c \neq 0$, and applying Corollary 2.4, we will find $x_2 \in X$ and $\alpha_2 > \alpha_1$ such that

$$\mathrm{diam}\{T_{c_1 g_{\alpha_1}^{(1)} + c_2 g_{\alpha_2}^{(2)}} x_2 : c_1, c_2 \in F\} < 1$$

(again, our assumption allows us to choose α_2 so that $g_{\alpha_1}^{(1)}$ and $g_{\alpha_2}^{(2)}$ are linearly independent in V_F.) Similarly to the argument above, it follows now that for some $h_2 \in V_F$ the affine 2-space $h_2 + \mathrm{Span}\{g_{\alpha_1}^{(1)}, g_{\alpha_2}^{(2)}\}$ is monochromatic. After repeating this procedure $m - 2$ more times, we will get the desired monochromatic affine m-space.

□

We are going now to give still another proof of van der Waerden's theorem. This proof has the advantage that, when properly interpreted, it gives also a proof of the Hales-Jewett theorem. To stress the affinity between the van der Waerden theorem and that of Hales-Jewett, this "double" proof is given in two parallel columns having many identical portions. To ease the presentation and to emphasize the correspondence between the number-theoretical and set-theoretical notions, we will abide by the following notational agreement: "$+$" will be used both for addition in $\mathbb{N}$ and for operation of taking (disjoint) unions of sets, "$-$" will be used not only for subtraction in $\mathbb{N}$ (when the minuend is not smaller than the subtrahend) but also instead of the set-theoretical difference "$\setminus$" in expressions of the form $A \setminus B$ where $B \subseteq A$. The sign "$\cdot$" will be used for both the multiplication in $\mathbb{N}$ and for the operation of taking the Cartesian products (and will often be omitted). The sign "$\preceq$" will mean either the usual inequality "$\leq$" or the set-theoretical containment "$\subseteq$". "0" will mean either zero or the empty set $\emptyset$. For any set E, $\mathcal{F}(E)$ will mean the set of finite subsets (including the empty set) of E.

Let (X, ρ) be a compact metric space. Let $q \in \mathbb{N}$.

Denote the set of nonnegative integers by $\boldsymbol{\mathcal{F}}$. Let T be a continuous self-mapping of X. Let A be a set consisting of q pairwise distinct natural numbers

$$A = \{p_i \in \mathbb{N} : i = 1, \ldots, q\};$$

assume without loss of generality that $p_1 < p_2 < \ldots < p_q$.

Let S be an infinite set, denote $\mathcal{F}(S)$ by $\boldsymbol{\mathcal{F}}$. Let $V = \{1, \ldots, q\} \times S$ and let $(T^a)_{a \in \mathcal{F}(V)}$ be an action of $\mathcal{F}(V)$ on X. (That is, T is a mapping from $\mathcal{F}(V)$ into the set of continuous self-mappings of X satisfying the following condition: if $a \cap b = \emptyset$, then $T^{a \cup b} = T^a T^b$. Put $p_i = \{1, \ldots, i\}$, $i = 1, \ldots, q$, and $a = \{p_1, \ldots, p_q\}$.

We are going to prove the following (two) proposition(s):

Proposition 2.7. ([BeL2], Prop. L) *For any $\varepsilon > 0$ there exists $N \in \boldsymbol{\mathcal{F}}$, such that for any $x \in X$ there exist $n \preceq N$, $n \neq 0$, and $a \preceq p_q(N - n)$ such that for any $p \in A$,*

$$\rho(T^{a+pn}x, T^a x) < \varepsilon.$$

Remark 2.8. Let us show how Proposition 2.7 implies the "classical" Hales-Jewett theorem. (See Theorem 1.26 above.) First, we pass to the combinatorial version of Proposition 2.7: let $r, q \in \mathbb{N}$; there exists $M \in \mathbb{N}$ such that for $N = \{1, \ldots, M\}$ and $V = \{1, \ldots, q\} \times N$, given an r-coloring of $\mathcal{F}(V)$ one can find a nonempty $n \preceq N$ and $a \preceq \{1, \ldots, q\} \times (N - n)$ such that the set $L = \Big\{a \cup (\{1\} \times n), a \cup (\{1, 2\} \times n), \ldots, a \cup (\{1, \ldots, q\} \times n)\Big\}$ is monochromatic. Second, we identify $\mathcal{F}(V)$ with $\mathcal{F}(\{1, \ldots, q\})^M$, $D \leftrightarrow (D_1, \ldots, D_M)$ where $D_i = D \cap (\{1, \ldots, q\} \times \{j\})$, $j = 1, \ldots, M$, and define a mapping φ from

$\mathcal{F}(V)$ to the "M-dimensional cube" $Q = \{0, \ldots, q\}^M$ by $\varphi(D_1, \ldots, D_M) = (|D_1|, \ldots, |D_M|)$. Now, any r-coloring of Q induces an r-coloring of $\mathcal{F}(V)$, and the φ-image of a monochromatic set L as in the assertion above is just a monochromatic line in Q.

Proof. We will prove this proposition by induction on q. Define B by

$$B = \{p_i - p_1,\ i = 2, \ldots, q\}.$$

Since B contains $q-1$ elements, we may assume that the statement to prove is valid for B, that is, for any $\varepsilon > 0$, there exists $N \in \mathscr{F}$ such that for any $x \in X$ there exist $n \preceq N$, $a \preceq (p_q - p_1)(N - n)$, such that $n \neq 0$ and for every $r \in B$ one has $\rho(T^{a+rn}x, T^a x) < \varepsilon$.

Let $\varepsilon > 0$. Let $k \in \mathbb{N}$ be such that among any $k+1$ points of X there are two points at a distance less than $\varepsilon/2$.

Put $\varepsilon_0 = \varepsilon/2k$. By the induction hypothesis, there exists $N_0 \in \mathscr{F}$ such that for any $x \in X$ there exist $n \preceq N_0$ and $a \preceq (p_q - p_1) \cdot (N_0 - n)$ such that $n \neq 0$ and for every $r \in B$ one has $\rho(T^{a+rn}x, T^a x) < \varepsilon_0$.

Let $\varepsilon_1 > 0$ be such that the inequality $\rho(y_1, y_2) < \varepsilon_1$ implies the inequality

$$\rho(T^b y_1, T^b y_2) < \varepsilon/2k$$

for any $b \preceq p_q N_0$. Let $N_1 \in \mathscr{F}$ be such that $N_1 \cap N_0 = 0$ (this disjointness condition concerns the part of Proposition 2.7 dealing with the Hales-Jewett theorem only) and for any $x \in X$ there exist $n \preceq N_1$ and $a \preceq (p_q - p_1)(N_1 - n)$ such that $n \neq 0$ and for every $r \in B$ one has $\rho(T^{a+rn}x, T^a x) < \varepsilon_1$.

Continue this process: assume that $\varepsilon_0, \ldots, \varepsilon_i$ and $N_0, \ldots, N_i \in \mathscr{F}$ have been already chosen. Let $\varepsilon_{i+1} > 0$ be such that the inequality $\rho(y_1, y_2) < \varepsilon_{i+1}$ implies the inequality

$$\rho(T^b y_1, T^b y_2) < \varepsilon/2k$$

for any $b \preceq p_q(N_0 + \ldots + N_i)$. Let $N_{i+1} \in \mathscr{F}$ be such that $N_{i+1} \cap (N_0 \cup \ldots \cup N_i) = 0$ (again, this disjointness condition is relevant for the Hales-Jewett part of Proposition 2.7 only) and for any $x \in X$ there exist $n \preceq N_{i+1}$ and $a \preceq (p_q - p_1)(N_{i+1} - n)$, such that $n \neq 0$ and for every $r \in B$ one has $\rho(T^{a+rn}x, T^a x) < \varepsilon_{i+1}$.

Continue the process of choosing ε_i, N_i up to $i = k$, and put $N = N_0 + \ldots + N_k$.

Now fix an arbitrary point $x \in X$.

Applying the definition of N_k to the point

$$y_k = T^{p_1 N_k} x,$$

find $n_k \preceq N_k$, $n_k \neq 0$, and $a_k \preceq (p_q - p_1)(N_k - n_k)$ such that for every $r \in B$ we have

$$\rho(T^{rn_k + a_k} y_k, T^{a_k} y_k) < \varepsilon_k.$$

Then, applying the definition of N_{k-1} to the point

$$y_{k-1} = T^{p_1(N_k + N_{k-1}) + a_k} x,$$

find $n_{k-1} \preceq N_{k-1}$, $n_{k-1} \neq 0$, and $a_{k-1} \preceq (p_q - p_1)(N_{k-1} - n_{k-1})$ such that for every $r \in B$ we have

$$\rho(T^{rn_{k-1} + a_{k-1}} y_{k-1}, T^{a_{k-1}} y_{k-1}) < \varepsilon_{k-1}.$$

Continue this process: suppose that we have already found $n_k, \ldots, n_i$, $a_k, \ldots, a_i$. Applying the definition of N_{i-1} to the point

$$y_{i-1} = T^{p_1(N_k + \ldots + N_{i-1}) + a_k + \ldots + a_i} x,$$

find $n_{i-1} \preceq N_{i-1}$, $n_{i-1} \neq 0$, and $a_{i-1} \preceq (p_q - p_1)(N_{i-1} - n_{i-1})$ such that for every $r \in B$ we have

$$\rho(T^{rn_{i-1} + a_{i-1}} y_{i-1}, T^{a_{i-1}} y_{i-1}) < \varepsilon_{i-1}.$$

Continue the process of choosing n_i, a_i up to $i = 0$.

For every $0 \leq i \leq k$ we have $0 \neq n_i \preceq N_i$ and $a_i \preceq (p_q - p_1)(N_i - n_i)$; therefore, for any $0 \leq i \leq j \leq k$ we have

$$\begin{gathered} p(n_j + \ldots + n_i) + a_j + \ldots + a_0 \\ +p_1(N_j + \ldots + N_0 - n_j - \ldots - n_0) \\ \preceq p_q(n_j + \ldots + n_{i+1}) \\ +(p_q - p_1)(N_j - n_j + \ldots + N_0 - n_0) \\ +p_1(N_j + \ldots + N_0 - n_j - \ldots - n_0) \\ = p_q(N_0 + \ldots + N_j). \end{gathered} \tag{2.1}$$

besides, $N_i \cap N_l = 0$ for $i \neq l$. Therefore, for any $0 \leq i \leq j \leq k$ we have

$$\begin{gathered} \big(p(n_j + \ldots + n_i) + a_j + \ldots + a_0 \\ +p_1(N_j + \ldots + N_0 - n_j - \ldots - n_0)\big) \\ \cap\big(a_{j+1} + (p - p_1)n_{j+1}\big) = 0. \end{gathered} \tag{2.2}$$

And, for any $0 \le j \le k$,

$$\begin{aligned}
a_k + \ldots + a_0 + p_1(N_k + \ldots + N_0 - n_j - \ldots - n_0) \\
\preceq (p_q - p_1)(N_k - n_k + \ldots + N_0 - n_0) + p_1(N_k + \ldots + N_0 - n_j - \ldots - n_0) \\
\preceq p_q(N_k + \ldots + N_0 - n_j - \ldots - n_0).
\end{aligned} \tag{2.3}$$

Define points x_i, $i = 0, \ldots, k$, by

$$x_i = T^{a_k+\ldots+a_0+p_1(N_k+\ldots+N_0-n_i-\ldots-n_0)}x.$$

We are going to show that for any $0 \le i \le j \le k$ and any $p \in A$,

$$\rho(T^{p(n_j+\ldots+n_{i+1})}x_j, x_i) \le \frac{\varepsilon}{2k}(j-i). \tag{2.4}$$

We will prove this by induction on $j - i$; when $j = i$ the statement is trivial. We will derive the validity of (2.4) for i, j, where $i < j$, from its validity for $i, j-1$.

By the definition of n_j,

$$\rho(T^{a_j+(p-p_1)n_j}y_j, T^{a_j}y_j) < \varepsilon_j,$$

where

$$y_j = T^{p_1(N_k+\ldots+N_j)+a_k+\ldots+a_{j+1}}x.$$

So, by the choice of ε_j (and (2.1)),

$$\begin{aligned}
&\rho(T^{p(n_{j-1}+\ldots+n_{i+1})+a_{j-1}+\ldots+a_0+p_1(N_{j-1}+\ldots+N_0-n_{j-1}-\ldots-n_0)}T^{a_j+(p-p_1)n_j}y_j, \\
&\quad T^{p(n_{j-1}+\ldots+n_{i+1})+a_{j-1}+\ldots+a_0+p_1(N_{j-1}+\ldots+N_0-n_{j-1}-\ldots-n_0)}T^{a_j}y_j) < \varepsilon/2k.
\end{aligned}$$

Using the definition of y_j, x_j (and (2.2)), we see

$$\begin{aligned}
&T^{p(n_{j-1}+\ldots+n_{i+1})+a_{j-1}+\ldots+a_0+p_1(N_{j-1}+\ldots+N_0-n_{j-1}-\ldots-n_0)}T^{a_j+(p-p_1)n_j}y_j \\
&= T^{p(n_j+\ldots+n_{i+1})+a_k+\ldots+a_0+p_1(N_k+\ldots+N_0-n_j-\ldots-n_0)}x = T^{p(n_j+\ldots+n_{i+1})}x_j
\end{aligned}$$

and

$$\begin{aligned}
&T^{p(n_{j-1}+\ldots+n_{i+1})+a_{j-1}+\ldots+a_0+p_1(N_{j-1}+\ldots+N_0-n_{j-1}-\ldots-n_0)}T^{a_j}y_j \\
&= T^{p(n_{j-1}+\ldots+n_{i+1})+a_k+\ldots+a_0+p_1(N_k+\ldots+N_0-n_{j-1}-\ldots-n_0)}x = T^{p(n_{j-1}+\ldots+n_{i+1})}x_{j-1}.
\end{aligned}$$

Since, by the induction hypothesis,

$$\rho(T^{p(n_{j-1}+\ldots+n_{i+1})}x_{j-1}, x_i) \le \frac{\varepsilon}{2k}(j-i-1),$$

we obtain (2.4).

By the choice of k, among the $k+1$ points $x_0, \ldots, x_k$ there are two, say x_i, x_j, $0 \le i < j \le k$, for which $\rho(x_i, x_j) < \varepsilon/2$. Put

$$\begin{gathered} n = n_j + \ldots + n_{i+1}, \\ a = a_k + \ldots + a_0 + p_1(N_k + \ldots + N_0 - n_j - \ldots - n_0). \end{gathered}$$

Then $x_j = T^a x$ and

$$\begin{gathered} \rho(T^{a+pn}x, T^a x) = \rho(T^{pn}x_j, x_j) \\ \le \rho(T^{pn}x_j, x_i) + \rho(x_j, x_i) < \varepsilon(j-i)/2k + \varepsilon/2 \le \varepsilon. \end{gathered}$$

Furthermore, $n \preceq N$, $n \neq 0$ and $a \preceq p_q(N-n)$ by (2.3). This proves Proposition 2.7. $\square$

We shall formulate now a polynomial extension of the multidimensional van der Waerden's theorem which was obtained in [BeL1]. We leave it to the reader to formulate the combinatorial equivalent of this result. (See also Theorems 2.12 and 2.14 below.)

Theorem 2.9. *Let (X, ρ) be a compact metric space, let $T_1, \ldots, T_t$ be commuting homeomorphisms of X and let $p_{i,j}$, $i = 1, \ldots, k$, $j = 1, \ldots, t$, be polynomials taking on integer values on the integers and vanishing at zero. Then, for any positive ε, there exist $x \in X$ and $n \in \mathbb{N}$ such that*

$$\rho\big(T_1^{p_{i,1}(n)} T_2^{p_{i,2}(n)} \ldots T_t^{p_{i,t}(n)} x, x\big) < \varepsilon \tag{2.5}$$

for all $i = 1, \ldots, k$ simultaneously. Moreover, the set $\{n \in \mathbb{Z} : \forall \epsilon > 0, \exists x \in X$ such that $\forall i \in \{1, 2, \ldots, k\}, (2.5)$ is satisfied$\}$ is an IP set.*

We provide now a proof of a special case. Let (X, ρ) be a compact metric space and let T be a homeomorphism of X. Let $\varepsilon > 0$; we will find $x \in X$ and $n \in \mathbb{N}$ such that $\rho(T^{n^2}x, x) < \varepsilon$.

Without loss of generality we will assume that the system (X, T) is minimal. We shall find 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

$$\rho(T^{(n_m+\ldots+n_{l+1})^2}x_m, x_l) < \varepsilon/2 \text{ for every } l, m \in \mathbb{Z}_+,\ l < m \tag{2.6}$$

(where $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$). Since X is compact, for some $l < m$ one will have $\rho(x_m, x_l) < \varepsilon/2$; together with (2.6) this will give $\rho(T^{(n_m+\ldots+n_{l+1})^2}x_m, x_m) < \varepsilon$.

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 $\rho(y, x_1) < \varepsilon_1$. Using the "linear" van der Waerden theorem, find $y_1 \in X$ 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

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

and (using the linear van der Waerden theorem) find y_m, n_{m+1} such that

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

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

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

and, hence, by the choice of ε_m,

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

and

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

Remark. We leave it to the reader to check that the proof above shows actually that a number n with the property that, for some x, $\rho(T^{n^2} x, x) < \epsilon$ can be chosen from any IP set.

We are going to formulate now the polynomial extension of the Hales-Jewett theorem which was obtained in [BeL2]. Like its "linear" special case, the polynomial Hales-Jewett theorem has many equivalent formulations. The one we have chosen to present here is a natural extension of Proposition 2.7.

Theorem 2.10. *Let (X, ρ) be a compact metric space. For fixed $d, q \in \mathbb{N}$, let $\mathcal{P}_N$ be the set of subsets of $\{1, 2, \dots, N\}^d \times \{1, 2, \dots, q\}$. Let $T(c)$, $c \in \mathcal{P}_N$ be a family of self-mappings of X such that $a \cap b = \emptyset$ implies $T(a \cup b) = T(a)T(b)$. Then for any $x \in X$ and any $\epsilon > 0$ there exist $N \in \mathbb{N}$, $a \in \mathcal{P}_N$ and a nonempty set $\gamma \subseteq \{1, 2, \dots, N\}$ such that $a \cap \big(\gamma^d \times \{1, 2, \dots q\}\big) \neq \emptyset$ and $\rho\Big(T\big(a \cup \big(\gamma^d \times \{i\}\big)x, T(a)x\Big) < \epsilon$ for every $i = 1, 2, \dots, q$.*

Here is the combinatorial version of Theorem 2.10. We leave it to the reader to verify that it is indeed equivalent to Theorem 2.10. (In one direction, the argument is similar to that in Remark 2.8 above. See also [Be3], pp.45-47 and [BeL2], Prop. 3.3.)

Theorem 2.11. ([BeL2], Thm. PHJ) *For any $r, d, q \in \mathbb{N}$ there exists $N = N(r, d, q)$ such that for any r-coloring of the set $\mathcal{P}_N$ of subsets of $\{1, 2, \dots, N\}^d \times \{1, 2, \dots, q\}$ there exist $a \in \mathcal{P}_N$ and a nonempty set $\gamma \subseteq \{1, 2, \dots, N\}$ with $a \cap \big(\gamma^d \times \{1, 2, \dots, q\}\big) = \emptyset$ and such that the sets*

$$a, a \cup (\gamma^d \times \{1\}), a \cup (\gamma^d \times \{2\}), \dots, a \cup (\gamma^d \times \{q\})$$

are all of the same color.

We will formulate now some corollaries of the polynomial Hales-Jewett theorem. The density versions of these results (or, rather, the ergodic counterparts of these density versions) will be discussed in Section 4. The following result extends and refines the polynomial van der Waerden theorem.

Theorem 2.12. ([BeL2], Thm. 0.14) *For any $t, m \in \mathbb{N}$, any polynomial mapping $P : \mathbb{Z}^t \to \mathbb{Z}^m$ satisfying $P(0) = 0$, any finite set $F \subset \mathbb{Z}^t$ and any finite coloring $\chi : \mathbb{Z}^m \to \{1, 2, \dots, r\}$, there is $l \in \{1, 2, \dots, r\}$ such that the set*

$$\Big\{(n_1, \dots, n_t) : \text{ there is } a \in \mathbb{Z}^m \text{ such that } \chi\big(a + P(n_1 v_1, \dots n_t v_t)\big) = l \\ \text{for all } v = (v_1, \dots, v_t) \in F\Big\}$$

is an IP set in $\mathbb{Z}^t$.*

The following proposition corresponds to the special case $t = 2q$, $m = 1$, $P(n_1, k_1, n_2, k_2, \dots, n_q, k_q) = \sum_{i=1}^{q} n_i k_i$ and $F = \big\{(1,1,0,\dots,0), (0,0,1,1,0,\dots,0), \dots, (0,0,\dots,0,0,1,1)\big\}$.

Proposition 2.13. *Let $\big(n_i^{(1)}\big)_{i\in\mathbb{N}}, \big(k_i^{(1)}\big)_{i\in\mathbb{N}}, \dots, \big(n_i^{(q)}\big)_{i\in\mathbb{N}}, \big(k_i^{(q)}\big)_{i\in\mathbb{N}}$ be sequences in $\mathbb{Z}$ and let $\big(n_\alpha^{(1)}\big)_{\alpha\in\mathcal{F}}, \big(k_\alpha^{(1)}\big)_{\alpha\in\mathcal{F}}, \dots, \big(n_\alpha^{(q)}\big)_{\alpha\in\mathcal{F}}, \big(k_\alpha^{(q)}\big)_{\alpha\in\mathcal{F}}$ be (additive) IP sets generated by these sequences. Then for any finite coloring of $\mathbb{Z}$ there exists a monochromatic set of the form*

$$\big\{a, a + n_\gamma^{(1)} k_\gamma^{(1)}, a + n_\gamma^{(2)} k_\gamma^{(2)}, \dots, a + n_\gamma^{(q)} k_\gamma^{(q)}\big\}$$

for some $a \in \mathbb{N}$ and a finite nonempty $\gamma \subset \mathbb{N}$.

One can also derive from the polynomial Hales-Jewett theorem an analogue of the polynomial van der Waerden theorem which is valid in any commutative ring. The following corollary of Theorem 2.11 contains the combinatorial counterpart of Theorem 2.9 as a special case.

Theorem 2.14. ([BeL2], Thm. 0.17) *Let W and V be vector spaces over an infinite field K, let $P : W \to V$ be a polynomial mapping with $P(0) = 0$, and let $F \subset W$ be a finite set. If $V = \bigcup_{i=1}^{r} C_i$ is a finite coloring of V, then:*

(i) *There exist $a \in V$ and $n \in K$, $n \neq 0$ such that the set*

$$a + P(nF) = \big\{a + P(nv) : v \in F\big\}$$

is monochromatic.

(ii) *For some $l \in \{1, 2, \dots, r\}$ the set*

$$\big\{n \in K : \text{ there exists } a \in V \text{ such that } a + P(nv) \in C_i \text{ for all } v \in F\big\}$$

is an IP set.*

Observing that the various versions of van der Waerden's theorem are linked with recurrence theorems for commuting homeomorphisms of a compact metric space, one is naturally inclined to inquire whether these recurrence results can be generalized to a non-commutative situation. The answer, in general, is NO. (See [Fu3], p. 40 and [BeH2].) The following theorem, due to A. Leibman ([Le1]), shows that, when the homeomorphisms generate a nilpotent group, the answer is YES. Note that Leibman's theorem is, at the same time, a generalization of the polynomial van der Waerden theorem, Theorem 2.9 above.

Theorem 2.15. *Let (X, ρ) be a compact metric space, let homeomorphisms $T_1, \dots, T_t$ of X generate a nilpotent group and let $p_{i,j}$, $i = 1, \dots, k$, $j = 1, \dots, t$, be polynomials taking on integer values on the integers and vanishing at zero. Then, for any positive ε, there exist $x \in X$ and $n \in \mathbb{N}$ such that $\rho\big(T_1^{p_{i,1}(n)} T_2^{p_{i,2}(n)} \dots T_t^{p_{i,t}(n)} x, x\big) < \varepsilon$ for all $i = 1, \dots, k$ simultaneously.*

To get the feeling of some of the ideas behind the proof of Theorem 2.15 let us consider the simplest noncommutative situation. Let (X, ρ) be a compact metric space and let homeomorphisms T and S of X do not commute but be such that $R = [T, S]$ commute with both T and S; the group G generated by T and S is then two-step nilpotent. Let $\varepsilon > 0$; our goal is to find $x \in X$ and $n \in \mathbb{N}$ such that both $\rho(T^n x, x) < \varepsilon$ and $\rho(S^n x, x) < \varepsilon$.

Without loss of generality we will assume that X is minimal with respect to the action of the group G. The sequence $S^n T^{-n}$, $n \in \mathbb{N}$, in G can be written as a "polynomial sequence" $(ST^{-1})^n R^{n(n-1)/2}$. Since the homeomorphisms ST^{-1} and R commute, we have, by the polynomial van der Waerden theorem that, for any $\delta > 0$ there exists $y \in X$ such that, for some $n \in \mathbb{N}$, one has $\rho(S^n T^{-n} y, y) < \delta$. We will first show that the set of points with this property is dense in X: for any open set $V \subseteq X$ we will find $y \in V$ such that, for some $n \in \mathbb{N}$, both $T^n y, S^n y \in V$. Since X is minimal with respect to the action of G and compact, there exist $P_1, \dots, P_k \in G$ such that $\bigcup_{i=1}^k P_i^{-1} V = X$. Let δ be a Lebesgue number for this cover. For any $P = T^a S^b R^c \in G$ we have

$$\begin{aligned} P^{-1} S^n T^{-n} P = S^n T^{-n} [S^n T^{-n}, P] &= (ST^{-1})^n R^{n(n-1)/2} [S, T]^{an} [T, S]^{-bn} \\ &= (ST^{-1})^n R^{n(n-1)/2 - (a+b)n}, \ n \in \mathbb{N}. \end{aligned}$$

All these "polynomial sequences" lie in the commutative group generated by ST^{-1} and R. Thus, by the polynomial van der Waerden theorem, there exist $z \in X$ and $n \in \mathbb{N}$ such that $\rho(P_i^{-1} S^n T^{-n} P_i z, z) < \delta$ for all $i = 1, \dots, k$. Hence, there exists $i \in \{1, \dots, k\}$ such that $z \in P_i^{-1} V$ and $P_i^{-1} S^n T^{-n} P_i z \in P_i^{-1} V$, $i = 1, \dots, k$. It remains to put $y = P_i z$.

We will now construct a sequence $x_0, x_1, x_2, \dots$ of points of X and a sequence $n_1, n_2, \dots$ of positive integers such that

$$\begin{aligned} \rho(T^{n_m + \dots + n_{l+1}} x_m, x_l) < \varepsilon/2 \text{ and } \rho(S^{n_m + \dots + n_{l+1}} x_m, x_l) < \varepsilon/2 \\ \text{for every } l, m \in \mathbb{Z}_+, \ l < m. \end{aligned} \tag{2.7}$$

Since X is compact, for some $l < m$ we will have $\rho(x_m, x_l) < \varepsilon/2$; together with (2.7) this implies $\rho(T^{n_m + \dots + n_{l+1}} x_m, x_m) < \varepsilon$ and $\rho(S^{n_m + \dots + n_{l+1}} x_m, x_m) < \varepsilon$.

Using the polynomial van der Waerden theorem find a point $x_0 \in X$ and an integer $n_1 \in \mathbb{N}$ such that $\rho(S^{n_1}T^{-n_1}x_0, x_0) < \varepsilon/2$. Put $x_1 = T^{-n_1}x_0$, then $\rho(T^{n_1}x_1, x_0) = 0 < \varepsilon/2$ and $\rho(S^{n_1}x_1, x_0) < \varepsilon/2$.

Suppose now that x_m, n_m satisfying (2.7) have been found for all $m \le k$; we will find x_{k+1}, n_{k+1}. Choose δ, $0 < \delta < \varepsilon/2$, such that $\rho(y, x_k) < \delta$ implies $\rho(T^{n_k+\dots+n_{l+1}}y, x_l) < \varepsilon/2$ and $\rho(S^{n_k+\dots+n_{l+1}}y, x_l) < \varepsilon/2$ for all $l = 0, \dots, k-1$. Find y_k in the $\delta/2$-neighborhood of x_k and $n_{k+1} \in \mathbb{N}$ such that $\rho(S^{n_{k+1}}T^{-n_{k+1}}y_k, y_k) < \delta/2$. Put $x_{k+1} = T^{-n_{k+1}}y_k$, then $\rho(T^{n_{k+1}}x_{k+1}, x_k) < \delta/2 < \varepsilon/2$ and $\rho(S^{n_{k+1}}x_{k+1}, x_k) < \delta < \varepsilon/2$. By the choice of δ this implies $\rho(T^{n_{k+1}+\dots+n_{l+1}}x_{k+1}, x_l) < \varepsilon/2$ and $\rho(S^{n_{k+1}+\dots+n_{l+1}}x_{k+1}, x_l) < \varepsilon/2$ for all $l = 0, \dots, k-1$.

We want to conclude this section by discussing a nilpotent version of the polynomial Hales-Jewett theorem, which was obtained in [BeL4]. But first we want to give a formulation of a corollary of the polynomial Hales-Jewett theorem, which will be suggestive of the further, nilpotent generalization.

Write $\mathcal{F}' = \mathcal{F} \cup \emptyset$ (where $\mathcal{F}$, as before, denotes the set of finite nonempty subsets of $\mathbb{N}$). Let G be a commutative (semi)group. A mapping $P : \mathcal{F}' \to G$ is an *IP polynomial of degree 0* if P is constant, and, inductively, is an *IP polynomial of degree* $\le d$ if for any $\beta \in \mathcal{F}'$ there exists an IP polynomial $D_\beta P : \mathcal{F}'(\mathbb{N} \setminus \beta) \to G$ of degree $\le d-1$ (where $\mathcal{F}'(\mathbb{N} \setminus \beta)$ is the set of finite subsets of $\mathbb{N} \setminus \beta$), such that $P(\alpha \cup \beta) = P(\alpha) \cup (D_\beta P)(\alpha)$ for every $\alpha \in \mathcal{F}'$ with $\alpha \cap \beta = \emptyset$. We have the following theorem.

Theorem 2.16. ([BeL2]) *Let G be an abelian group of self-homeomorphisms of a compact metric space (X, ρ) and let $P_1, P_2, \dots, P_k$ be IP polynomials mapping $\mathcal{F}'$ into G and satisfying $P_i(\emptyset) = 1_G$ for all $i \in \{1, \dots, k\}$. Then for any $\epsilon > 0$ there exist $x \in X$ and a nonempty $\alpha \in \mathcal{F}'$ such that $\rho\big(P_i(\alpha)x, x\big) < \epsilon$ for $i = 1, \dots, k$.*

It is proved in [BeL2], Thm. 8.3, that if G is an abelian group then a mapping $P : \mathcal{F}' \to G$ is an IP polynomial of degree $\le d$ if and only if there exists a family $\big\{g_{(j_1,\dots,j_d)}\big\}_{(j_1,j_2,\dots,j_d)\in\mathbb{N}^d}$ of elements of G such that for any $\alpha \in \mathcal{F}'$ one has $P(\alpha) = \prod_{(j_1,\dots,j_d)\in\alpha^d} g_{(j_1,\dots,j_d)}$. This characterization of IP polynomials makes sense in the nilpotent setup as well. Given a nilpotent group G, let us call a mapping $P : \mathcal{F}' \to G$ an IP polynomial if for some $d \in \mathbb{N}$ there exists a family $\big\{g_{(j_1,\dots,j_d)}\big\}_{(j_1,\dots,j_d)\in\mathbb{N}^d}$ of elements of G and a linear order $<$ on $\mathbb{N}^d$ such that for any $\alpha \in \mathcal{F}'$ one has $P(\alpha) = \prod^{<}_{(j_1,\dots,j_d)\in\alpha^d} g_{(j_1,\dots,j_d)}$. (The entries in the product $\prod^{<}$ are multiplied in accordance with the order

$<$.) We can formulate now the nilpotent version of the polynomial Hales-Jewett theorem, which contains many results formulated above, in particular Theorems 1.29, 2.9 and 2.15, as special cases.

Theorem 2.17. ([BeL4], Thm. 0.24) *Let G be a nilpotent group of self-homeomorphisms of a compact metric space (X, ρ) and let $P_1, \ldots, P_k : \mathcal{F}' \to G$ be polynomial mappings satisfying $P_1(\emptyset) = \ldots = P_k(\emptyset) = 1_G$. Then, for any $\epsilon > 0$, there exist $x \in X$ and a nonempty $\alpha \in \mathcal{F}'$ such that $\rho(P_i(\alpha)x, x) < \epsilon$ for all $i = 1, 2, \ldots, k$.*

The following corollary of Theorem 2.17 can be viewed as the nilpotent generalization of Hilbert's theorem (See Theorems 1.8 and 1.12). It is worth noting that, unlike Hilbert's theorem which had an easy proof, the nilpotent version of it is far from being trivial.

Theorem 2.18. ([BeL4], Thm. 5.5) *Let G be an infinite nilpotent group. For any $k, r \in \mathbb{N}$ there exist $N \in \mathbb{N}$ such that for any $g_j^{(i)} \in G$, $1 \leq i \leq k$, $1 \leq j \leq N$, and any r-coloring of G there exist a nonempty set $\alpha \subseteq \{1, 2, \ldots, N\}$ and infinitely many $h \in G$ such that for $h_i = \prod_{j \in \alpha} g_j^{(i)}$, $i = 1, \ldots, k$, (where the entries are multiplied in the natural order of $j \in \mathbb{N}$), all the products $hh_{i_1}h_{i_2} \ldots h_{i_l}$ with $0 \leq l \leq k$ and distinct $i_1, i_2, \ldots, i_l$ are of the same color.*

Finally, we formulate a corollary of the nilpotent Hales-Jewett theorem which may be viewed as an extension of Theorem 1.29. See [BeL4], Thm. 5.9 for yet another nilpotent extension of Theorem 1.29.

Theorem 2.19. ([BeL4], Thm. 5.8) *For any $r, q, c \in \mathbb{N}$ and prime integer p there exists $k \in \mathbb{N}$ such that if F is a field of characteristic p and of cardinality at least k, then for any r-coloring of the group G of $q \times q$ upper triangular matrices over F with unit diagonal, there exist a subgroup H of G with $|H_q| \geq c$ and $h \in G$ such that the coset hH is monochromatic.*

3 Dynamical, combinatorial, and Diophantine applications of $\beta\mathbb{N}$

In this section, we shall discuss briefly the Stone-Čech compactification of the natural numbers, $\beta\mathbb{N}$, and indicate some of its connections with and applications to topological dynamics, combinatorics, and the theory of Diophantine approximations.

We start with some general definitions and facts. The reader will find the missing details in [Be3], Section 3, and [Be6]. (See also [HinS] for a comprehensive treatment of topological algebra in Stone-Čech compactifications and applications thereof.)

An ultrafilter p on $\mathbb{N}$ is a *maximal filter*, namely a family of subsets of $\mathbb{N}$ satisfying the following conditions: (the first three of which constitute, for a nonempty family of sets, the definition of a *filter*.)

(i) $\emptyset \notin p$

(ii) $A \in p$ and $A \subset B$ imply $B \in p$

(iii) $A \in p$ and $B \in p$ imply $A \cap B \in p$

(iv) (maximality) if $r \in \mathbb{N}$ and $\mathbb{N} = \bigcup_{i=1}^r A_i$ then, for some $i \in \{1, 2, \ldots, r\}$, $A_i \in p$.

The space of ultrafilters on $\mathbb{N}$ has a natural topology which turns it into a universal compactification of $\mathbb{N}$, the so-called Stone-Čech compactification. (See more on that in [HinS].)

A convenient way of looking at ultrafilters is to identify each ultrafilter p with a finitely additive $\{0, 1\}$-valued probability measure μ_p on the power set $\mathcal{P}(\mathbb{N})$. This measure μ_p is naturally defined by the requirement $\mu_p(A) = 1$ iff $A \in p$ and it follows immediately from the conditions (i) through (iv) above that $\mu_p(\emptyset) = 0$, $\mu_p(\mathbb{N}) = 1$ and that for any finite disjoint collection $A_1, \ldots, A_r$ of subsets of $\mathbb{N}$, one has $\mu_p\Big(\bigcup_{i=1}^r A_i\Big) = \sum_{i=1}^r \mu_p(A_i)$. Without saying so explicitly, we will always think of ultrafilters as such measures, but we will prefer to write $A \in p$ rather that $\mu_p(A) = 1$.

Any $n \in \mathbb{N}$ naturally defines an ultrafilter $\{A \subset \mathbb{N} : n \in A\}$. Such ultrafilters, which can be viewed as "delta measures" concentrated at points of $\mathbb{N}$, are called *principal* and, alas, are the only ones which can be constructed without the use of Zorn's lemma (see [CN], pp. 161-162.) Since many of the constructions in topological dynamics and ergodic theory use this or that equivalent of Zorn's lemma, we will not be bothered by this, notwithstanding the fact that there is certainly some importance in knowing which mathematical results are Zorn lemma free.

Suppose that $\mathcal{C}$ is a family of subsets of $\mathbb{N}$ which has the finite intersection property. Then there is some $p \in \beta\mathbb{N}$ such that $C \in p$ for each $C \in \mathcal{C}$. Indeed, let

$$\tilde{\mathcal{C}} = \{\mathcal{B} \subset \mathcal{P}(\mathbb{N}) : \mathcal{B} \text{ has the finite intersection property and } \mathcal{C} \subset \mathcal{B}\}.$$

Clearly, $\tilde{\mathcal{C}} \neq \emptyset$ (since $\mathcal{C} \in \tilde{\mathcal{C}}$). Also, the union of any chain in $\tilde{\mathcal{C}}$ is a member of $\tilde{\mathcal{C}}$. By Zorn's Lemma there is a maximal member p of $\tilde{\mathcal{C}}$, which is actually maximal with respect to the finite intersection property and hence a member of $\beta\mathbb{N}$.

To see that non-principal ultrafilters exist, take for example

$$\mathcal{C} = \{A \subset \mathbb{N} : A^c = \mathbb{N} \setminus A \text{ is finite}\}.$$

Clearly $\mathcal{C}$ has the finite intersection property, so there is an ultrafilter $p \in \beta\mathbb{N}$ such that $C \in p$ for all $C \in \mathcal{C}$. It is easy to see that such p cannot be principal.

For another example, take

$$\mathcal{D} = \{A \subset \mathbb{N} : d(A) = \lim_{n\to\infty} \frac{|A \cap \{1, 2, \dots, n\}|}{n} = 1\}.$$

Again, $\mathcal{D}$ clearly satisfies the finite intersection property. If p is any ultrafilter for which $\mathcal{D} \subset p$, then any member of p has positive upper density. (If $d(A) = 0$, then $A^c = (\mathbb{N} \setminus A) \in \mathcal{D}$.)

These examples hint that the space $\beta\mathbb{N}$ is quite large. It is indeed: the cardinality of $\beta\mathbb{N}$ equals that of $\mathcal{P}\big(\mathcal{P}(\mathbb{N})\big)$ ([GiJ], 6.10(a)).

Let us say now a few words about topology in $\beta\mathbb{N}$. Given $A \subset \mathbb{N}$, let $\overline{A} = \{p \in \beta\mathbb{N} : A \in p\}$. The set $\mathcal{G} = \{\overline{A} : A \subset \mathbb{N}\}$ forms a basis for the open sets (and a basis for the closed sets). To see that $\mathcal{G}$ is indeed a basis for a topology on $\beta\mathbb{N}$ observe that if $A, B \subset \mathbb{N}$, then $\overline{A} \cap \overline{B} = \overline{A \cap B}$. Also, $\overline{\mathbb{N}} = \beta\mathbb{N}$ and hence $\bigcup_{\overline{A}\in\mathcal{G}} \overline{A} = \beta\mathbb{N}$. (Notice also that $\overline{A} \cup \overline{B} = \overline{A \cup B}$.) With this topology, $\beta\mathbb{N}$ satisfies the following.

Theorem 3.1. *$\beta\mathbb{N}$ is a compact Hausdorff space.*

Proof. Let $\mathcal{K}$ be a cover of $\beta\mathbb{N}$ by sets belonging to the base $\mathcal{G} = \{\overline{A} : A \subset \mathbb{N}\}$. Let $\mathcal{C} \subset \mathcal{P}(\mathbb{N})$ be such that $\mathcal{K} = \{\overline{A} : A \in \mathcal{C}\}$. Assume that $\mathcal{K}$ has no finite subcover. Consider the family $\mathcal{D} = \{A^c : A \in \mathcal{C}\}$. There are two possibilites (each leading to a contradiction):

(i) $\mathcal{D}$ has the finite intersection property. Then, as shown above, there exists an ultrafilter p such that $A^c \in p$ for each $A^c \in \mathcal{D}$. Since p is an ultrafilter, $A^c \in p$ if and only if $A \notin p$. On the other hand, since $\mathcal{K}$ covers $\beta\mathbb{N}$, for some element $\overline{A}$ of the cover $p \in \overline{A}$, or equivalently $A \in p$, a contradiction.

(ii) $\mathcal{D}$ does not have the finite intersection property. Then for some $A_1, \cdots, A_r \in \mathcal{C}$ one has $\bigcap_{i=1}^r A_i^c = \emptyset$, or $\bigcup_{i=1}^r A_i = \mathbb{N}$, which implies that $\bigcup_{i=1}^r \overline{A_i} = \beta\mathbb{N}$. Again, this is a contradiction, as we assumed that $\mathcal{K}$ has no finite subcover.

As for the Hausdorff property, notice that if $p, q \in \beta\mathbb{N}$ are distinct ultrafilters then since each of them is maximal with respect to the finite intersection property, neither of them is contained in the other. If $A \in p \backslash q$, then $A^c \in q \backslash p$, which means that $\overline{A}$ and $\overline{A^c}$ are disjoint neighborhoods of p and q. $\square$

Remark. Being a nice compact Hausdorff space, $\beta\mathbb{N}$ is in many respects quite a strange object. We mentioned already that its cardinality is that of $\mathcal{P}\big(\mathcal{P}(\mathbb{N})\big)$. It follows that $\beta\mathbb{N}$ is not metrizable, as otherwise, being a compact and hence separable metric space, it would have cardinality not exceeding that of $\mathcal{P}(\mathbb{N})$. Another curious feature of $\beta\mathbb{N}$ is that any infinite closed subset of $\beta\mathbb{N}$ contains a copy of all of $\beta\mathbb{N}$.

Since $\overline{\mathbb{N}} = \beta\mathbb{N}$, it is natural to attempt to extend the operation of addition from (the densely embedded) $\mathbb{N}$ to $\beta\mathbb{N}$. Since ultrafilters are measures (principal ultrafilters being just the *point measures* corresponding to the elements of $\mathbb{N}$), it comes as no surprise that the extension we look for takes the form of a *convolution.* What is surprising, however, is that the algebraic structure of $\beta\mathbb{N}$ was explicitly introduced only relatively recently (in [CY]). In the following definition, $A - n$ (where $A \subset \mathbb{N}$, $n \in \mathbb{N}$) is the set of all m for which $m + n \in A$. For $p, q \in \mathbb{N}$, define

$$p + q = \big\{A \subset \mathbb{N} : \{n \in \mathbb{N} : (A - n) \in p\} \in q\big\}.$$

Remarks.
1. Note that in much of the literature, including [HinS], what we have written as $p + q$ is denoted as $q + p$.
2. It is not hard to check that for principal ultrafilters the operation $+$ corresponds to addition in $\mathbb{N}$.
3. Despite the somewhat forbiding phrasing of the operation just introduced in set-theoretical terms, the perspicacious reader will notice the direct analogy between this definition and the usual formulas for convolution of measures μ, ν on a locally compact group G (cf. [HeR], 19.11):

$$\mu * \nu(A) = \int_G \nu(x^{-1}A)\, d\mu(x) = \int_G \mu(Ay^{-1})\, d\nu(y).$$

4. Before checking the correctness of the definition, a word of warning: the introduced operation $+$ (which will turn out to be well defined and associative) is badly noncommutative. This seems to contradict our intuition since $(\mathbb{N}, +)$ is commutative and in the case of σ-additive measures on abelian semi-groups convolution *is* commutative. The explanation: our ultrafilters, being only *finitely* additive measures, do not obey the Fubini theorem, which is behind the commutativity of the usual convolution.

Let us show that $p+q$ is an ultrafilter. Clearly $\emptyset \notin p+q$. Let $A, B \in p+q$. This means that $\{n \in \mathbb{N} : (A-n) \in p\} \in q$ and $\{n \in \mathbb{N} : (B-n) \in p\} \in q$. Since p and q are ultrafilters, we have:

$$\begin{aligned}\{n \in \mathbb{N} \ &: (A \cap B) - n \in p\} \\ &= \{n \in \mathbb{N} : (A-n) \in p\} \cap \{n \in \mathbb{N} : (B-n) \in p\} \in q.\end{aligned}$$

Assume now that $A \subset \mathbb{N}$, $A \notin p+q$. We want to show that $A^c \in p+q$. Since $A \notin p+q$, we know that $\{n \in \mathbb{N} : (A-n) \in p\} \notin q$, or, equivalently, $\{n \in \mathbb{N} : (A-n) \in p\}^c \in q$. But this is true precisely when $\{n \in \mathbb{N} : (A^c - n) \in p\} \in q$, which is the same as $A^c \in p+q$. It follows that $p+q \in \beta\mathbb{N}$.

Let us now check the associativity of the operation $+$. Let $A \subset \mathbb{N}$ and $p, q, r \in \beta\mathbb{N}$. One has:

$$\begin{aligned}& A \in p+(q+r) \Leftrightarrow \{n \in \mathbb{N} : (A-n) \in p\} \in q+r \\ \Leftrightarrow \quad & \{m \in \mathbb{N} : \big(\{n \in \mathbb{N} : (A-n) \in p\} - m\big) \in q\} \in r \\ \Leftrightarrow \quad & \big\{m \in \mathbb{N} : \{n \in \mathbb{N} : (A-m-n) \in p\} \in q\big\} \in r \\ \Leftrightarrow \quad & \{m \in \mathbb{N} : (A-m) \in p+q\} \in r \Leftrightarrow A \in (p+q)+r.\end{aligned}$$

Theorem 3.2. *For any fixed $p \in \beta\mathbb{N}$ the function $\lambda_p(q) = p+q$ is a continuous self map of $\beta\mathbb{N}$.*

Proof. Let $q \in \beta\mathbb{N}$ and let $\mathcal{U}$ be a neighborhood of $\lambda_p(q)$. We will show that there exists a neighborhood $\overline{B}$ of q such that for any $r \in \overline{B}$, $\lambda_p(r) \in \mathcal{U}$. Let $A \subset \mathbb{N}$ be such that $\lambda_p(q) = p+q \in \overline{A} \subset \mathcal{U}$. Then $A \in p+q$. Let us show that the set

$$B = \{n \in \mathbb{N} : (A-n) \in p\}$$

will do for our purposes. Indeed, by the definition of $p+q$, $B \in q$, or, in other words, $q \in \overline{B}$. If $r \in \overline{B}$ then $B = \{n \in \mathbb{N} : (A-n) \in p\} \in r$. This means that $A \in p+r = \lambda_p(r)$, or $\lambda_p(r) \in \overline{A} \in \mathcal{U}$. □

With the operation $+$, $\beta\mathbb{N}$ becomes, in view of Theorem 3.2, a compact *left topological* semigroup.

Theorem 3.3. *If $(G, *)$ is a compact left topological semigroup (i.e. for any $x \in G$ the function $\lambda_x(y) = x * y$ is continuous) then G has an idempotent.*

Remark. For compact *topological semigroups* (i.e. with an operation which is continuous in both variables), this result is due to Numakura, [Nu]; for left topological semigroups the result is due to Ellis, [E1].

Proof. Let

$$\mathcal{G} = \big\{A \subset G : A \neq \emptyset, A \text{ is compact}, A * A = \{x * y : x, y \in A\} \subset A\big\}.$$

Since $G \in \mathcal{G}$, $\mathcal{G} \neq \emptyset$. By Zorn's Lemma, there exists a minimal element $A \in \mathcal{G}$. If $x \in A$, then $x * A$ is compact and satisfies

$$(x * A) * (x * A) \subset (x * A) * (A * A) \subset (x * A) * A \subset x * (A * A) \subset x * A.$$

Hence $x * A \in \mathcal{G}$. But $x * A \subset A * A \subset A$, which implies that $x * A = A$. Thus $x \in x * A$, which implies that $x = x * y$ for some $y \in A$. Now consider $B = \{z \in A : x * z = x\}$. The set B is closed (since $B = \lambda_x^{-1}(\{x\})$), and we have just shown that B is nonempty. If $z_1, z_2 \in B$ then $z_1 * z_2 \in A * A \subset A$ and $x * (z_1 * z_2) = (x * z_1) * z_2 = x * z_2 = x$. So $B \in \mathcal{G}$. But $B \subset A$ and hence $B = A$. So $x \in B$ which gives $x * x = x$. $\square$

For a fixed $p \in \beta\mathbb{N}$ we shall call a set $C \subset \mathbb{N}$ *p-big* if $C \in p$. The notion of largeness induced by idempotent ultrafilters is special (and promising) in that it inherently has a shift-invariance property. Indeed, if $p \in \beta\mathbb{N}$ with $p + p = p$ then

$$A \in p \Leftrightarrow A \in p + p \Leftrightarrow \{n \in \mathbb{N} : (A - n) \in p\} \in p.$$

A way of interpreting this is that if p is an idempotent ultrafilter, then A is p-big if and only if for p-many $n \in \mathbb{N}$ the shifted set $(A - n)$ is p-big. Or, still somewhat differently: $A \subset \mathbb{N}$ is p-big if for p-almost all $n \in \mathbb{N}$ the set $(A - n)$ is p-big. This is the reason why specialists in ultrafilters called such idempotent ultrafilters "almost shift invariant" in the early seventies (even before the existence of such ultrafilters was established).

Remark. The reader is invited to check that if p is an idempotent ultrafilter, then for any $a \in \mathbb{N}$, $a\mathbb{N} \in p$. This, in particular, implies that such p cannot be a principal ultrafilter. (This can also be deduced from the fact that $(\mathbb{N}, +)$ has no idempotents.)

Each idempotent ultrafilter $p \in \beta\mathbb{N}$ induces a "measure preserving dynamical system" with the phase space $\mathbb{N}$, σ-algebra $\mathcal{P}(\mathbb{N})$, measure p, and "time" being the "p-preserving" $\mathbb{N}$-action induced by the shift. The two peculiarities about such a measure preserving system are that the phase space is countable and that the "invariant measure" is only finitely additive and is preserved by our action not for all, but for almost all instances of "time." Notice that the "Poincaré recurrence theorem" trivially holds: If $A \in p$ then, since there are p-many n for which $(A - n) \in p$, one has, for any such n, $A \cap (A - n) \in p$.

As we saw in the introduction, it is this defining property of idempotent ultrafilters (arranged there as Proposition 1.14) which is all that one needs for the proof of Hindman's theorem.

The following result gives the "ultrafilter explanation" of Theorem 1.16 in the Introduction. We shall also need it in the proof of Theorem 3.5 below.

Theorem 3.4. *For any sequence $(x_i)_{i\in\mathbb{N}}$ in $\mathbb{N}$ there is an idempotent $p \in \beta\mathbb{N}$ such that $FS((x_i)_{i\in\mathbb{N}}) \in p$.*

Sketch of the proof. Let $\Gamma = \bigcap_{n=1}^{\infty} \overline{FS((x_i)_{i=n}^{\infty})}$. (The closures are taken in the natural topology of $\beta\mathbb{N}$.) Clearly, Γ is compact and nonempty. It is not hard to show that Γ is a subsemigroup of $(\beta\mathbb{N}, +)$. Being a compact left-topological semigroup, Γ has an idempotent. If $p \in \Gamma$ is an idempotent, then $\overline{\Gamma} = \Gamma \ni p$ which, in particular, implies $FS((x_i)_{i=1}^{\infty}) \in p$.

□

The space $\beta\mathbb{N}$ has also another natural semigroup structure, namely, the one inherited from the multiplicative semigroup $(\mathbb{N}, \cdot)$, and is a left topological compact semigroup with respect to this structure too. In particular, there are (many) multiplicative idempotents, namely ultrafilters q with the property

$$A \in q \Leftrightarrow \{n \in \mathbb{N} : A/n \in q\} \in q$$

(where $A/n := \{m \in \mathbb{N} : mn \in A\}$.) By complete analogy with the proof of (the additive version of) Hindman's theorem, one can show that any member of a multiplicative idempotent contains a multiplicative IP set, namely a set of finite products of the form

$$FP(y_n)_{n=1}^{\infty} = \Big\{\prod_{i\in\alpha} y_i : \alpha \subset \mathbb{N},\ 1 \leq |\alpha| < \infty\Big\}.$$

It follows that for any finite partition $\mathbb{N} = \bigcup_{i=1}^{r} C_i$ there are $i, j \in \{1, 2, ..., r\}$ such that C_i contains an additive IP set and C_j contains a multiplicative IP set. The following theorem due to Hindman shows that one can always have $i = j$.

Theorem 3.5. [Hin2] *For any finite partition* $\mathbb{N} = \bigcup_{i=1}^{r} C_i$*, there exists* $i \in \{1, 2, ..., r\}$ *and sequences* $(x_n)_{n=1}^{\infty}$ *and* $(y_n)_{n=1}^{\infty}$ *in* $\mathbb{N}$ *such that*

$$FS\big((x_n)_{n=1}^{\infty}) \cup FP((y_n)_{n=1}^{\infty}\big) \subseteq C_i.$$

Proof. Let Γ be the closure in $\beta\mathbb{N}$ of the set of additive idempotents. We claim that $p \in \Gamma$ if and only if every p-large set A contains an additive IP set. Indeed, if $A \in p \in \Gamma$, then $\overline{A}$ is a (clopen) neighborhood of p. It follows that there exists $q \in \overline{A}$ with $q + q = q$. Then $A \in q$ and by Hindman's theorem, A contains an IP set. Conversely, if $\overline{A}$ is a basic neighborhood of p and for some $(x_n)_{n=1}^{\infty}$, $FS((x_n)_{n=1}^{\infty}) \subseteq A$, then by Theorem 3.4, there exists an idempotent q with $FS((x_n)_{n=1}^{\infty}) \in q$, which implies $A \in q$, and hence $p \in \Gamma$.

We will show now that Γ is a right ideal in $(\beta\mathbb{N}, \cdot)$. Let $p \in \Gamma,\ q \in \beta\mathbb{N}$, and let $A \in p \cdot q$. Then $\{x \in \mathbb{N} :\ Ax^{-1} \in p\} \in q$ and, in particular, $\{x \in \mathbb{N} :\ Ax^{-1} \in p\}$ is nonempty. Let x be such that $Ax^{-1} \in p$. Since $p \in \Gamma$, there exists a sequence $(y_n)_{n=1}^{\infty}$ with $FS((y_n)_{n=1}^{\infty}) \subseteq Ax^{-1}$, which implies $FS((xy_n)_{n=1}^{\infty}) \subseteq A$ and so $p \cdot q \in \Gamma$. We see that Γ is a compact subsemigroup in $(\beta\mathbb{N}, \cdot)$ and hence contains a multiplicative idempotent. To finish the proof, let $\cup_{i=1}^{r} C_i = \mathbb{N}$ and let $p \in \Gamma$ satisfy $p \cdot p = p$. Let $i \in \{1, 2, ..., r\}$ be such that $C_i \in p$. Then, since $p \in \Gamma,\ C_i$ contains an additive IP set. Also, since p is a multiplicative idempotent, C_i contains a multiplicative IP set. We are done. □

Remarks.
1. For an elementary proof of Theorem 3.5, see [BeH4].
2. Theorem 3.13 below shows that for any finite partition $\bigcup_{i=1}^{r} C_i = \mathbb{N}$ one of C_i has interesting additional properties. In particular, one of C_i can be shown to contain in addition to an additive and a multiplicative IP sets, also arbitrarily long arithmetic and arbitrarily long geometric progressions.

Seeing how much mileage one can get by sheer analogy between idempotent ultrafilters and measure preserving systems, it would be natural to inquire (in a hope that this can lead to interesting new results) whether there is a class of idempotents which could be likened to a minimal topological system (with an invariant measure.)

To answer this question, let us extend the shift operation $\sigma : n \to n+1, n \in \mathbb{N}$, from $\mathbb{N}$ to $\beta\mathbb{N}$, by the rule $q \to q+1$ (where 1 denotes the principal ultrafilter of sets containing the integer 1), and consider the topological dynamical system $(\beta\mathbb{N}, \sigma)$.

The following theorem establishes the connection between minimal subsystems of $(\beta\mathbb{N}, \sigma)$ and minimal right ideals in $(\beta\mathbb{N}, +)$.

Theorem 3.6. *The minimal closed invariant subsets of the dynamical system* $(\beta\mathbb{N}, \sigma)$ *are precisely the minimal right ideals of* $(\beta\mathbb{N}, +)$.

Proof. We first observe that closed σ-invariant sets in $\beta\mathbb{N}$ coincide with right ideals. Indeed if I is a right ideal, i.e. satisfies $I + \beta\mathbb{N} \subseteq I$, then for any $p \in I$ one has $p+1 \in I + \beta\mathbb{N} \subseteq I$, so that I is σ-invariant. On the other hand, if S is a closed σ-invariant set in $\beta\mathbb{N}$ and $p \in S$, then $p + \beta\mathbb{N} = p + \overline{\mathbb{N}} = \overline{p + \mathbb{N}} \subseteq \overline{S} = S$, which implies $S + \beta\mathbb{N} \subseteq S$.

Now the theorem follows from a simple general fact that any minimal right ideal in a compact left-topological semigroup $(G, \cdot)$ is closed. Indeed, if R is a right ideal in $(G, \cdot)$ and $x \in R$, then xG is compact as the continuous image of G and is an ideal. Hence the minimal ideal containing x is compact as well. (The fact that R contains a minimal ideal follows by an application of Zorn's lemma to the nonempty family $\{I : \ I \text{ is a closed right ideal of } G \text{ and } I \subseteq R\}$). $\square$

Our next step is to observe that any minimal right ideal in $(\beta\mathbb{N}, +)$, being a compact left-topological semigroup, contains, by Theorem 3.3, an idempotent.

Definition 3.7. *An idempotent p in* $(\beta\mathbb{N}, +)$ *is called* minimal *if p belongs to a minimal right ideal.*

Theorem 3.8. *Any minimal subsystem of* $(\beta\mathbb{N}, \sigma)$ *is of the form* $(p + \beta\mathbb{N}, \sigma)$ *where p is a minimal idempotent in* $(\beta\mathbb{N}, +)$.

Proof. It is obvious that, for any $p \in (\beta\mathbb{N}, +)$, $p + \beta\mathbb{N}$ is a right ideal. To see that any minimal right ideal is of this form, take any $q \in R$ and observe that $q + \beta\mathbb{N} \subseteq R + \beta\mathbb{N} \subseteq R$. Since R is minimal, we get $q + \beta\mathbb{N} = R$. In particular, one can take q to be an idempotent. $\square$

We shall need the following definition in order to formulate some immediate corollaries of Theorem 3.8.

Definition 3.9. *A set $A \subseteq \mathbb{N}$ is piecewise syndetic if it can be represented as an intersection of a syndetic set with an infinite union of intervals $[a_n, b_n]$, where $b_n - a_n \to \infty$.*

Remark. It is not hard to see that $A \subseteq \mathbb{N}$ is piecewise syndetic if and only if there exists a finite set $F \subset \mathbb{N}$ such that the family

$$\left\{\bigcup_{t\in F}(A-t) - n : n \in \mathbb{N}\right\}$$

has the finite intersection property. While this description of piecewise syndeticity looks somewhat forbidding, it has the advantage of making sense in any semigroup. As we shall see in the proof of Corollary 3.10 below, it is this form of the definition of piecewise syndeticity which is much easier to check when dealing with minimal idempotents.

Corollary 3.10. *Let p be a minimal idempotent in $(\beta\mathbb{N}, +)$.*
(i) *For any $A \in p$ the set $B = \{n : (A-n) \in p\}$ is syndetic.*
(ii) *Any $A \in p$ is piecewise syndetic.*

Proof. Statement (i) follows immediately from the fact that $(p + \beta\mathbb{N}, \sigma)$ is a minimal system. Indeed, note that the assumption $A \in p$ just means that $p \in \overline{A}$, i.e. $\overline{A}$ is a (clopen) neighborhood of p. Now, in a minimal dynamical system every point x is *uniformly recurrent*, i.e. visits any of its neighborhoods V along a syndetic set. This implies that the set $\{n : p+n \in \overline{A}\} = \{n : A \in p+n\} = \{n : A-n \in p\}$ is syndetic.

(ii) Since the set $B = \{n : A-n \in p\}$ is syndetic, the union of finitely many shifts of B covers $\mathbb{N}$, i.e. for some finite set $F \subset \mathbb{N}$ one has $\bigcup_{t\in F}(B-t) = \mathbb{N}$. So, for any $n \in \mathbb{N}$ there exists $t \in F$ such that $n \in B - t$, or $n + t \in B$. By the definition of B this implies $(A - (n+t)) \in p$. It follows that for any n the set $\bigcup_{t\in F}(A-t) - n$ belongs to p, and consequently, the family $\{\bigcup_{t\in F}(A-t) - n : n \in \mathbb{N}\}$ has the finite intersection property. By the remark above, this is equivalent to piecewise syndeticity of A. $\square$

Remark. It follows from part (ii) of Corollary 3.10 that for any finite partition $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, one of the C_i is piecewise syndetic, and moreover for any finite partition of a piecewise syndetic set, one of the cells of the partition is again piecewise syndetic. One can show that (with the appropriately arranged definition of piecewise syndeticity) this result holds for any infinite semigroup. (In the case of the semigroup $(\mathbb{N}, +)$, this fact can be proved in an elementary fashion, and is apparently originally due to T. Brown, [Br].)

Note that it follows from the definition above that if A is a syndetic set in $\mathbb{N}$, then, for some finite set $F \subset \mathbb{N}$, the set $\bigcup_{t\in F}(A-t)$ contains arbitrarily long intervals. It follows now from Theorem 1.18 that any piecewise syndetic set contains arbitrarily long arithmetic progressions. (Since any piecewise syndetic set has positive upper Banach density, this fact also follows from Szemerédi's theorem, but this would be an overkill.)

On the other hand, it is clear that since for any minimal idempotent $p \in \beta\mathbb{N}$ and any finite partition $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, one of the C_i belongs to p, van der Waerden's theorem follows from the following result. The proof below is a slight modification of the proof in [BeFHK]. (Cf. also [FuK3].)

Theorem 3.11. *Let $p \in (\beta\mathbb{N}, +)$ be a minimal idempotent and let $A \in p$. Then A contains arbitrarily long arithmetic progressions.*

Proof. Fix $k \in \mathbb{N}$ and let $G = (\beta\mathbb{N})^k$. Clearly, G is a compact left topological semigroup with respect to the product topology and coordinatewise addition. Let

$$E_0 = \{(a, a+d, ..., a+(k-1)d) : \ a \in \mathbb{N},\ d \in \mathbb{N} \cup \{0\}\},$$
$$I_0 = \{(a, a+d, ..., a+(k-1)d) : \ a, d \in \mathbb{N}\}.$$

Clearly, E_0 is a semigroup in $\mathbb{N}^k$ and I_0 is an ideal of E_0. Let $E = cl_G E_0$ and $I = cl_G I_0$ be, respectively, the closures of E_0 and I_0 in G. It follows by an easy argument, which we leave to the reader, that E is a compact subsemigroup of G and I is a two-sided ideal of E. Let now $p \in (\beta\mathbb{N}, +)$ be a minimal idempotent and let $\tilde{p} = (p, p, ..., p) \in G$. We claim that $\tilde{p} \in I$ and that this implies that each member of p contains a length k arithmetic progression. Indeed, assume that $\tilde{p} \in I$ and let $A \in p$. Then $\overline{A} \times ... \times \overline{A} = (\overline{A})^k$ is a neighborhood of $\tilde{p}$. Hence $\tilde{p} \in (\overline{A})^k \cap cl_G I_0 = cl_G(A^k \cap I_0)$, which implies $A^k \cap I_0 \neq \emptyset$. It follows that for some $a, d \in \mathbb{N}$ $(a, a+d, ..., a+(k-1)d) \in A^k$ which finally implies $\{a, a+d, ..., a+(k-1)d\} \subset A$.

So it remains to show that $\tilde{p} \in I$. We check first that $\tilde{p} \in E$. Let $A_1, A_2, ..., A_k \in p$. Then $\overline{A}_1 \times \overline{A}_2 \times ... \times \overline{A}_k \ni \tilde{p}$. If $a \in \bigcap_{i=1}^{k} A_i$ then $(a, a, ..., a) \in (\overline{A}_1 \times \overline{A}_2 \times ... \times \overline{A}_k) \cap E_0$ which implies $\tilde{p} \in E$.

Now, since p is a minimal idempotent, there is a minimal right ideal R of $(\beta\mathbb{N}, +)$ such that $p \in R$. Since $\tilde{p} \in E$, $\tilde{p}+E$ is a right ideal of E and there is a minimal right ideal $\tilde{R}$ of E such that $\tilde{R} \subseteq \tilde{p}+E$. Let $\tilde{q} = (q_1, q_2, ..., q_k)$ be an idempotent in $\tilde{R}$. Then $\tilde{q} \in \tilde{p}+E$ and for some $\tilde{s} = (s_1, s_2, ..., s_k)$ in E we get $\tilde{q} = \tilde{p} + \tilde{s}$. We shall show now that $\tilde{p} = \tilde{q} + \tilde{p}$. Indeed, from $\tilde{q} = \tilde{p} + \tilde{s}$ we get, for each $i = 1, 2, ..., k$, $q_i = p+s_i$. This implies $q_i \in R$ and since R is minimal,

$q_i + \beta\mathbb{N} = R$. Hence $p \in q_i + \beta\mathbb{N}$. Let , for each $i = 1, 2, ..., k,\ t_i \in \beta\mathbb{N}$ be such that $p = q_i + t_i$. Then $q_i + p = q_i + q_i + t_i = q_i + t_i = p$ and so we obtained $\tilde{p} = \tilde{q} + \tilde{p}$.

To finish the proof, we observe that $\tilde{p} = \tilde{q} + \tilde{p}$ implies $\tilde{p} \in \tilde{q} + E = \tilde{R}$ which, in its turn, implies $\tilde{p} \in I$ (since, as it is not hard to see, any minimal right ideal is contained in a two-sided ideal). We are done. □

Definition 3.12. *A set $A \subseteq \mathbb{N}$ is called additively (respectively, multiplicatively) central if there is a minimal idempotent $p \in (\beta\mathbb{N}, +)$ (respectively, $p \in (\beta\mathbb{N}, \cdot)$), such that $A \in p$.*

As theorems above indicate, central sets are an ideal object for Ramsey-theoretical applications. For example, central sets in $(\mathbb{N}, +)$ not only are large (i.e. piecewise syndetic) but also are combinatorially rich and, in particular, contain IP sets and arbitrarily long arithmetic progressions. Similarly, the multiplicative central sets in $(\mathbb{N}, \cdot)$ (namely, the members of minimal idempotents in $(\beta\mathbb{N}, \cdot)$) are multiplicatively piecewise syndetic, contain finite products sets (i.e. the multiplicative IP sets), arbitrarily long geometric progressions etc.

The following theorem obtained in collaboration with N. Hindman may be viewed as an enhancement of Theorem 3.5 above.

Theorem 3.13. ([BeH1], p. 312) *For any finite partition $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, one of C_i is both additively and multiplicatively central.*

Sketch of the proof. Let $M = cl\{p :\ p$ is a minimal idempotent in $(\beta\mathbb{N}, +)\}$. Then one can show that M is a right ideal in $(\beta\mathbb{N}, \cdot)$ (see [BH1], Theorem 5.4, p.311). Let $R \subseteq M$ be a minimal right ideal and pick an idempotent $q = q \cdot q$ in R. Let $i \in \{1, 2, ..., r\}$ be such that $C_i \in q$. Since q is a minimal idempotent in $(\beta\mathbb{N}, \cdot)$, C_i is central in $(\mathbb{N}, \cdot)$. Since $C_i \in q$ and $q \in M$, there is some minimal idempotent p in $(\beta\mathbb{N}, +)$ with $C_i \in p$. Hence C_i is also central in $(\mathbb{N}, +)$.

□

The following theorem supplies a useful family of examples of additively and multiplicatively central sets in $\mathbb{N}$.

Theorem 3.14. ([BeH4], Lemma 3.3) *For any sequence $(a_n)_{n=1}^{\infty}$ and an increasing sequence $(b_n)_{n=1}^{\infty}$ in $\mathbb{N}$, $\bigcup_{n=1}^{\infty}\{a_n, a_n+1, a_n+2, ..., a_n+b_n\}$ is additively central and $\bigcup_{n=1}^{\infty}\{a_n \cdot 1, a_n \cdot 2, ..., a_n \cdot b_n\}$ is multiplicatively central.*

The original definition of central sets in $(\mathbb{N},+)$, due to H. Furstenberg, was made in the language of topological dynamics. Before introducing Furstenberg's definition of centrality, we want first to recall some relevant dynamical notions.

Given a compact metric space (X,d), a continuous map $T: X \to X$ and not necessarily distinct points $x_1, x_2 \in X$, one says that x_1, x_2 are *proximal*, if for some sequence $n_k \to \infty$ one has $d(T^{n_k}x_1, T^{n_k}x_2) \to 0$.

A point which is proximal only to itself is called *distal*. In case all the points of X are distal T is called a distal transformation and (X,T) is called a distal system.

Recall that a point x in a dynamical system (X,T) is called *uniformly recurrent* if for any neighborhood V of x the set $\{n : T^n x \in V\}$ is syndetic. Since in a minimal system any point is uniformly recurrent and since any compact topological system has a minimal subsystem, any topological system has a uniformly recurrent point. A stronger statement, due to J. Auslander ([A]) and R. Ellis ([E2]) says that in a dynamical system on a compact metric space, any point is proximal to a uniformly recurrent point. (Note that this, in particular, implies that any distal point is uniformly recurrent.)

We are now ready to formulate Furstenberg's original definition of central sets in $(\mathbb{N},+)$. For the proof of the equivalence of this definition to Definition 3.12 above, see Theorem 3.22.

Definition 3.15. (see [Fu3], p.161.) *A subset $S \subseteq \mathbb{N}$ is a* central *set if there exists a system (X,T), a point $x \in X$, a uniformly recurrent point y proximal to x, and a neighborhood U_y of y such that $S = \{n : \; T^n x \in U_y\}$.*

In order to prove the equivalence of the two definitions of centrality, we need to introduce first the notion of convergence along ultrafilters. As we shall see, this notion allows one to better understand distality, proximality, and recurrence in topological dynamical systems. We would like to point out that some proofs involving ultrafilters are similar to known proofs involving the so-called Ellis enveloping semigroup. This is not surprising in view of the fact that the Ellis semigroup is a particular type of compactification and, as such, is in many respects similar to the universal object, the Stone-Čech compactification. In particular, it allows one to much more easily deal with combinatorial applications of topological dynamics.

Given an ultrafilter $p \in \beta\mathbb{N}$ and a sequence $(x_n)_{n\in\mathbb{N}}$ in a topological space X, one writes $p\text{-}\lim_{n\in\mathbb{N}} x_n = y$ if, for every neighborhood U of y, one has

$\{n : x_n \in U\} \in p$. It is easy to see that if X is a compact Hausdorff space, then $p\text{-}\lim_{n\in\mathbb{N}} x_n$ exists and is unique for any sequence $(x_n)_{n\in\mathbb{N}}$ in X.

Theorem 3.16. *Let X be a compact Hausdorff space and let $p, q \in \beta\mathbb{N}$. Then for any sequence $(x_n)_{n\in\mathbb{N}}$ in X one has*

$$(q+p)\text{-}\lim_{r\in\mathbb{N}} x_r = p\text{-}\lim_{t\in\mathbb{N}} q\text{-}\lim_{s\in\mathbb{N}} x_{s+t}. \tag{3.1}$$

In particular, if p is an idempotent, and $q = p$, one has

$$p\text{-}\lim_{r\in\mathbb{N}} x_r = p\text{-}\lim_{t\in\mathbb{N}} p\text{-}\lim_{s\in\mathbb{N}} x_{s+t}.$$

Proof. Let $x = (q+p)\text{-}\lim_{r\in\mathbb{N}} x_r$. Given a neighborhood U of x we have $\{r : x_r \in U\} \in q+p$. Recalling that a set $A \subseteq \mathbb{N}$ is a member of ultrafilter $q+p$ if and only if $\{n \in \mathbb{N} : (A-n) \in q\} \in p$, we get

$$\{t : (\{s : x_s \in U\} - t) \in q\} = \{t : \{s : x_{s+t} \in U\} \in q\} \in p$$

This means that, for p-many t, $q\text{-}\lim_{s\in\mathbb{N}} x_{s+t} \in U$ and we are done. $\square$

Proposition 3.17. *Let (X,T) be a topological system and let $x \in X$ be an arbitrary point. Given an idempotent ultrafilter $p \in \beta\mathbb{N}$, let $p\text{-}\lim_{n\in\mathbb{N}} T^n x = y$. Then $p\text{-}\lim_{n\in\mathbb{N}} T^n y = y$. If x is a distal point (i.e. x is proximal only to itself) then $p\text{-}\lim_{n\in\mathbb{N}} T^n x = x$.*

Proof. Applying Theorem 3.16 (and the fact that $p+p=p$), we have

$$p\text{-}\lim_{n\in\mathbb{N}} T^n y = p\text{-}\lim_{n\in\mathbb{N}} T^n \; p\text{-}\lim_{m\in\mathbb{N}} T^m x = p\text{-}\lim_{n\in\mathbb{N}} \; p\text{-}\lim_{m\in\mathbb{N}} T^{m+n} x = p\text{-}\lim_{n\in\mathbb{N}} T^n x = y.$$

If x is a distal point, then the relations $p\text{-}\lim_{n\in\mathbb{N}} T^n x = y = p\text{-}\lim_{n\in\mathbb{N}} T^n y$ clearly imply $x = y$ and we are done. $\square$

Remark. Note that Proposition 3.17 implies that a continuous distal self-map T of a compact metric space is onto. It follows that T is invertible and T^{-1} is also distal.

Let R be a minimal right ideal in $\beta\mathbb{N}$. By Theorem 3.8 above, (R,σ), where $\sigma : p \to p+1$, is a minimal (nonmetrizable) system. Given a topological system (X,T) and a point $x \in X$, let $\varphi : R \to X$ be defined by $\varphi(p) = p\text{-}\lim_{n\in\mathbb{N}} T^n x$. Observe that if the set $Y \subseteq X$ is defined by $Y = \left\{p\text{-}\lim_{n\in\mathbb{N}} T^n x : p \in R\right\}$, then the following diagram is commutative:

$$\begin{array}{ccc} R & \xrightarrow{\sigma} & R \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\varphi} \\ Y & \xrightarrow{T} & Y \end{array}$$

It follows that (Y,T) is a minimal system. We will use this observation in the proof of the following result.

Proposition 3.18. *If (X,T) is a minimal system then for any $x \in X$ and any minimal right ideal R in $\beta\mathbb{N}$ there exists a minimal idempotent $p \in R$ such that $p\text{-}\lim T^n x = x$.*

Proof. By the observation above, $X = \{p\text{-}\lim_{n\in\mathbb{N}} T^n x,\ p \in R\}$. It follows that the set $\Gamma = \{p \in R : p\text{-}\lim_{n\in R} T^n x = x\}$ is nonempty and closed. We claim that Γ is a semigroup. Indeed, if $p, q \in \Gamma$, one has :

$$(p+q)\text{-}\lim_{n\in\mathbb{N}} T^n x = q\text{-}\lim_{n\in\mathbb{N}} T^n\, p\text{-}\lim_{m\in\mathbb{N}} T^m x = x.$$

By Theorem 3.3, Γ contains an idempotent which has to be minimal since it belongs to R. We are done. $\square$

We shall need the following simple fact in the proofs below. The proof is immediate and is left as an exercise for the reader.

Theorem 3.19. *Let (X,T) be a topological system, R a minimal right ideal in $\beta\mathbb{N}$, and let $x \in X$ be a point in X. The following are equivalent:*
(i) *x is uniformly reccurent;*
(ii) *there exists a minimal idempotent $p \in R$ such that $p\text{-}\lim_{n\in\mathbb{N}} T^n x = x$.*

It follows from Proposition 3.17 that for any topological system (X,T), any $x \in X$, and any idempotent ultrafilter p, the points x and $y = p\text{-}\lim_{n\in\mathbb{N}} T^n x$ are proximal. (If (X,T) is a distal system then $y = x$.) The following theorem gives a partial converse of Proposition 3.17.

Theorem 3.20. *If (X,T) is a topological system and x_1, x_2 are proximal, not necessarily distinct points and if x_2 is uniformly reccurent, then there exists a minimal idempotent $p \in \beta\mathbb{N}$ such that $p\text{-}\lim_{n\in\mathbb{N}} T^n x_1 = x_2$.*

Proof. Let $I = \{p \in \beta\mathbb{N} : p\text{-}\lim_{n\in\mathbb{N}} T^n x_1 = p\text{-}\lim_{n\in\mathbb{N}} T^n x_2\}$. It is not hard to see that I is a nonempty closed subset of $\beta\mathbb{N}$. One immediately checks that I is a right ideal. Let R be a minimal right ideal in I. Since x_2 is uniformly recurrent, its orbital closure is a minimal system. By Proposition 3.18, there exists a minimal idempotent $p \in R$ such that $p\text{-}\lim T^n x_2 = x_2$. Then $p\text{-}\lim_{n\in\mathbb{N}} T^n x_1 = p\text{-}\lim_{n\in\mathbb{N}} T^n x_2 = x_2$ and we are done. $\square$

One can give a similar proof to the following classical result due to J. Auslander [A] and R. Ellis [E2].

Theorem 3.21. *Let (X,T) be a topological system. For any $x \in X$ there exists a uniformly recurrent point y in the orbital closure $\overline{\{T^n x\}}_{n\in N}$, such that x is proximal to y. Moreover, for any minimal right ideal $R \subset \beta\mathbb{N}$ there exists a minimal idempotent $p \in R$ such that $p\text{-}\lim_{n\in\mathbb{N}} T^n x = y$.*

Proof. Let R be a minimal ideal in $\beta\mathbb{N}$ and let p be a (minimal) idempotent in R. Let $y = p\text{-}\lim_{n\in\mathbb{N}} T^n x$. Clearly, y belongs to the orbital closure of x. By Proposition 3.17, the points x and y are proximal. By Theorem 3.19, y is uniformly recurrent. We are done. $\square$

We are in position now to establish the equivalence of two notions of central that were discussed above.

Theorem 3.22. *The following properties of a set $A \subseteq \mathbb{N}$ are equivalent:*

(i) (cf. [Fu3], Definition 8.3) *There exists a topological system (X,T), and a pair of (not necessarily distinct) points $x, y \in X$ where y is uniformly recurrent and proximal to x, such that for some neighborhood U of y one has:*

$$A = \{n \in \mathbb{N} : T^n x \in U\}$$

(ii) (See Definition 3.12 above, see also [BeH1], Definition 3.1) *There exists a minimal idempotent $p \in (\beta\mathbb{N}, +)$ such that $A \in p$.*

Proof. (i) $\Rightarrow$ (ii) By Theorem 3.20, there exists a minimal idempotent p, such that $p\text{-}\lim_{n\in\mathbb{N}} T^n x = y$. This implies that for any neighborhood U of y the set $\{n \in \mathbb{N} : T^n x \in U\}$ belongs to p.

(ii) $\Rightarrow$(i) The idea of the following proof is due to B. Weiss. Let A be a member of a minimal idempotent $p \in \beta\mathbb{N}$. Let $X = \{0,1\}^{\mathbb{Z}}$, the space of bilateral 0-1 sequences. Endow X with the standard metric, which turns it into a compact space:

$$d(\omega_1, \omega_2) = \inf\left\{\tfrac{1}{n+1} : \omega_1(i) = \omega_2(i) \text{ for } |i| < n\right\}$$

Let $T : X \to X$ be the shift operator: $T(\omega)(n) = \omega(n+1)$. Then T is a homeomorphism of X and (X,T) is a topological dynamical system. Viewing A as a subset of $\mathbb{Z}$, let $x = 1_A \in X$. Finally, let $y = p\text{-}\lim_{n\in\mathbb{N}} T^n x$. By Proposition 3.17, x and y are proximal. Also, since p is minimal, y is, by Theorem 3.19, a uniformly recurrent point. We claim that $y(\mathbf{0}) = 1$. Indeed, define $U = \{z \in X : z(\mathbf{0}) = y(\mathbf{0})\}$, and note that, since $y = p\text{-}\lim_{n\in\mathbb{N}} T^n x$ and $A \in p$, one can find $n \in A$ such that $T^n x \in U$. But since $x = 1_A$, $(T^n x)(\mathbf{0}) = 1$. But then, given $n \in \mathbb{Z}$, we have: $T^n x \in U \Leftrightarrow (T^n x)(\mathbf{0}) = 1 \Leftrightarrow x(n) = 1 \Leftrightarrow n \in A$. It follows that $A = \{n \in \mathbb{Z} : T^n x \in U\}$ and we are done. $\square$

Let (X,T) be a topological system. In [Fu3], a point $x \in X$ is called IP* recurrent if for any neighborhood U of x, the set $\{n \in \mathbb{N} : T^n x \in U\}$ is an IP* set. It is easy to see that a point x is IP* recurrent if and only if for any idempotent $p \in \beta\mathbb{N}$, one has $p\text{-}\lim_{n\in\mathbb{N}} T^n x = x$. Note that the property of a point x being IP* recurrent is much stronger than that of uniform recurrence (which, by Theorem 3.19, is equivalent to the fact that for *some* minimal idempotent p, one has $p\text{-}\lim_{n\in\mathbb{N}} T^n x = x$.) While, in a minimal system, every point is uniformly recurrent, there are minimal systems having no IP* recurrent points. For example, any minimal topologically weakly mixing system has this property. (See [Fu3], Theorem 9.12.) The following theorem shows that distal points (and no others) are IP* recurrent.

Theorem 3.23. *Let (X,T) be a dynamical system and $x \in X$. The following are equivalent:*
(i) *x is a distal point;*
(ii) *x is IP* recurrent.*

Proof. (i) $\Rightarrow$ (ii). By Proposition 3.17, for any idempotent p, the points x and $p\text{-}\lim_{n\in\mathbb{N}} T^n x$ are proximal. Since x is distal, this may happen only if $x = p\text{-}\lim T^n x$. But this means that x is an IP* recurrent point.

(ii) $\Rightarrow$ (i). If x is not distal, then there exists $y \neq x$, such that x and y are proximal. But then, by Theorem 3.20, there exists an idemponent p such that $p\text{-}\lim T^n x = y$. Since $y \neq x$, this contradicts (ii). $\square$

We shall conclude this section with some Diophantine applications of distal minimal systems. The results which we are going to describe can be viewed as enhancements of classical theorems due to Kronecker, Hardy–Littlewood, and Weyl, and will be based on the following characterization of distal systems. A set $E \subset \mathbb{N}$ is called IP^*_+ if it is a translation of an IP^* set.

Theorem 3.24. *Assume that (X,T) is a minimal system. Then it is distal if and only if for any $x \in X$ and any open set $U \subseteq X$ the set $\{n : T^n x \in U\}$ is IP^*_+.*

Proof. Assume that (X,T) is distal. By minimality, there exists $n_0 \in \mathbb{N}$ such that $T^{n_0}x \in U$. By Theorem 3.23, the set $\{n : T^n(T^{n_0}x) \in U\}$ is IP^* which, of course, implies that the set $\{n : T^n x \in U\}$ is IP^*_+.

Assume now that for any x_1, x_2 and a neighborhood U of x_2 the set $\{n : T^n x_1 \in U\}$ is IP^*_+. We will find it convenient to call an IP^*_+ set $A \subseteq \mathbb{N}$ *proper* if A is not IP^* (i.e. A is a nontrivial shift of an IP^* set and, moreover, this shifted IP^* set is not IP^*). If T were not distal, then for some distinct points x_1, x_2 and idempotents p, q one would have: $p\text{-}\lim_{n\in\mathbb{N}} T^n x_1 = x_2$, $q\text{-}\lim_{n\in\mathbb{N}} T^n x_2 = x_1$ and also $p\text{-}\lim_{n\in\mathbb{N}} T^n x_2 = x_2$, $q\text{-}\lim_{n\in\mathbb{N}} T^n x_1 = x_1$ (see Theorem 3.20 and Proposition 3.17). Let U be a small enough neighborhood of x_2. Then, since $p\text{-}\lim_{n\in\mathbb{N}} T^n x_1 = x_2$, the set $S = \{n : T^n x_1 \in U\}$ is a member of p, and hence cannot be a proper IP^*_+ set. But, since $q\text{-}\lim_{n\in\mathbb{N}} T^n x_1 = x_1$, the set S cannot be an improper IP^*_+ set (that is, an IP^* set) either: if U is small enough, $S \notin q$. So T has to be distal. We are done. □

The following theorem was obtained by Hardy and Littlewood in [HaL] and may be viewed as a polynomial extension of a similar "linear" theorem due to Kronecker ([Kro].)

Theorem 3.25. *If the numbers $1, \alpha_1, ..., \alpha_k$ are linearly independent over $\mathbb{Q}$, then for any $d \in \mathbb{N}$ and any kd intervals $I_{lj} \subset [0,1]$, $l = 1, ..., d; j = 1, ..., k$ the set*

$$\Gamma_{dk} = \{n \in \mathbb{N} : n^l \alpha_j \mod 1 \in I_{lj},\ l = 1, ..., d;\ j = 1, ..., k\}$$

is infinite.

In 1916, H. Weyl ([Weyl]) introduced the notion of uniform distribution and obtained many strong results extending and enhancing the earlier work of Kronecker, Hardy–Littlewood and others on Diophantine approximations.

Perhaps the most famous result obtained in [Weyl] was the theorem on uniform distribution of the sequence $p(n) \pmod 1$, $n = 1, 2, \ldots$, where $p(n)$ is a real polynomial having at least one irrational coefficient other than the constant term. This theorem also admits a nice ergodic proof, via the study of a class of affine transformations of the torus, due to Furstenberg ([Fu1]).

In connection to the Hardy–Littlewood theorem, Weyl was able to show in [Weyl] that the set Γ_{dk} has positive density equal to the product of the lengths of I_{ij}. This also can be shown by using the dynamical approach of Furstenberg. In the following theorem, we show that the affine transformations of the kind treated by Furstenberg in [Fu1] can also be utilized to prove the following strengthening of the Hardy–Littlewood theorem.

Theorem 3.26. *Under the assumptions and notation of Theorem 3.25, the set Γ_{dk} is IP^*_+.*

Proof. To make the formulas more transparent we shall put $d = 3$. It will be clear that the same proof gives the general case.

We start with the easily checkable claim that if $T_\alpha : \mathbb{T}^3 \to \mathbb{T}^3$ is defined by $T_\alpha(x, y, z) = (x + \alpha, y + 2x + \alpha, z + 3x + 3y + \alpha)$ then $T^n_\alpha(0, 0, 0) = (n\alpha, n^2\alpha, n^3\alpha)$. This transformation T is distal (easy) and minimal. The last assertion can actually be derived from the case $k = 1$ of Hardy–Littlewood theorem above, but also can be proved directly. (For example, this fact is a special case of Lemma 1.25, p.36 in [Fu3]). Our next claim is that if the numbers $1, \alpha_1, \alpha_2, ..., \alpha_k$ are linearly independent over Q, then the product map $T = T_{\alpha_1} \times \cdots \times T_{\alpha_k}$ (acting on $\mathbb{T}^{3k}$) is distal and minimal as well. (The distality is obvious, and the minimality follows, again, from an appropriately modified Lemma 1.25 in [Fu3]). By minimality of T, the orbit of zero in $\mathbb{T}^{3k}$ is dense, and this, together with Theorem 3.24, gives the desired result. □

We conclude this section by formulating a general result which may be proved by refining the techniques used above.

Theorem 3.27. *If real polynomials $p_1(t), p_2(t), ..., p_k(t)$ have the property that for any non-zero vector $(h_1, h_2, ..., h_k) \in \mathbb{Z}^k$ the linear combination $\sum_{i=1}^k h_i p_i(t)$ is a polynomial with at least one irrational coefficient other than the constant term then for any k subintervals $I_j \subset [0, 1]$, $j = 1, ..., k$, the set*

$$\{n \in \mathbb{N} : p_j(n) \mod 1 \in I_j,\ j = 1, ..., k\}$$

is IP^*_+.

4 Multiple recurrence

One of the common features of the topological multiple recurrence results which were discussed in the previous sections is that they have streamlined, and often relatively short, proofs. In particular, in proving these theorems, one does not have to analyze and distinguish between various types of dynamical behavior which the topological system may possess. In other words, the proofs evolve without taking into account the possibly intricate structure of the system. The situation with measure-theoretical multiple recurrence is, at least at present, quite different. All of the known proofs of dynamical theorems such as the ergodic Szemerédi theorem and other more recent and stronger multiple recurrence results, which will be discussed in this section, are complicated by the fact that systems with different types of dynamical behavior require different types of arguments. Yet, these proofs have a certain (and, in the opinion of the author, quite beautiful) structure which, in some "big" sense, is the same in different proofs.

Our plan for this section is as follows. In subsection 4.1, we shall analyze the proof of Furstenberg's ergodic Szemerédi theorem, and, in particular, provide complete proofs of some important special cases.

In subsection 4.2, we will give an overview of (the proofs of) the major multiple recurrence results (as well as their density counterparts) which have appeared since the publication of Furstenberg's groundbreaking paper [Fu2]. An attempt will be made to emphasize the common features of these proofs and to amplify the subtle points. The flow of the discussion in subsection 4.2 will eventually lead us to some quite recent results and natural open problems.

4.1 Furstenberg's ergodic Szemerédi theorem

This subsection is devoted to the thorough discussion of the proof of Furstenberg's ergodic Szemerédi theorem, which corresponds to the case $T_i = T^i$ in Theorem 1.24 formulated in the introduction.

Theorem 4.1.1. ([Fu2]) *For any probability measure preserving system* $(X, \mathcal{B}, \mu, T)$*, any* $A \in \mathcal{B}$ *with* $\mu(A) > 0$ *and any* $k \in \mathbb{N}$*, there exists* $n \in \mathbb{N}$ *such that* $\mu(A \cap T^{-n}A \cap \ldots \cap T^{-kn}A) > 0$.

Actually, we do not know how to prove Theorem 4.1.1 without proving, at least superficially, a little bit more. (The situation here is analogous to what we encountered when discussing and proving the Sárközy-Furstenberg theorem – see Theorem 1.31 and the subsequent remarks.)

Here is the version of Theorem 4.1.1 which we will find convenient to work with.

Theorem 4.1.2. *For any probability measure preserving system* $(X, \mathcal{B}, \mu, T)$*, any* $A \in \mathcal{B}$ *with* $\mu(A) > 0$*, and any* $k \in \mathbb{N}$*, one has:*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T^{-n}A \cap \ldots \cap T^{-kn}A) > 0$$

Remark 4.1.3. As a matter of fact, the result proved by Furstenberg in [Fu2] establishes that

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

This implies (via the Furstenberg correspondence principle) not only that any set of positive upper density in $\mathbb{N}$ contains arithmetic progressions but that the differences of these progressions form a syndetic set. This fact, in turn, follows from a much stronger IP Szemerédi theorem proved by Furstenberg and Katznelson in [FuK3]. (See Theorems 4.2.14 and 4.2.15 below.) One of the reasons we have chosen to deal with the formulation as in Theorem 4.1.2 is that it has a simpler proof which nevertheless will allow us to stress the main ideas and will naturally serve as the basis for a discussion of possible extensions.

There are a few assumptions that we may make without loss of generality.

First, we can assume that the measure μ is non-atomic. (This follows from the fact that the atoms of μ generate an invariant sub-σ-algebra, and Theorems 4.1.1 and 4.1.2 are trivially satisfied in the case of atomic measure spaces.)

Second, we can assume that the space $(X, \mathcal{B}, \mu)$ is Lebesgue, i.e. is isomorphic to the unit interval with Lebesgue measure. Indeed, given the set $A \in \mathcal{B}$, we can pass, if needed, to a T-invariant separable sub-σ-algebra of $\mathcal{B}$ with respect to which all of the functions $f_n = 1_A(T^n x)$ and their finite products are measurable. By Carathéodory's theorem, (see [Roy], Chap. 15, Theorem 4) 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 into a Lebesgue-measure preserving isomorphism of $\mathcal{L}$, which by the classical theorem due to von Neumann (see [Roy], Ch. 15, Th. 20) admits realization as a point mapping.

Finally, we can assume that the measure preserving systems that we are dealing with are invertible. Indeed, assuming the invertibility of the measure preserving transformations occurring in the formulations of multiple recurrence theorems such as Theorem 1.24 or Theorem 4.1.2, not only makes the proofs more convenient, but also is sufficient for combinatorial applications. On the other hand, it is not hard to show that in the case of measure preserving actions of commutative semigroups with cancellation, the general case follows from the invertible one. See, for example, [Fu3], Ch. 7, Section 4.

These remarks apply also to the other multiple recurrence results formulated below, and we will tacitly keep the above assumptions throughout the rest of this survey.

We could also assume that the measure preserving transformation T in the formulation of Theorems 4.1.1 and 4.1.2 is ergodic. Despite the fact that this would make some of the arguments somewhat simpler, we have chosen not to do so since, in more general situations such as, say, the multidimensional ergodic Szemerédi theorem (Theorem 1.24 above), one can assume only that one of the transformations involved is ergodic, which does not help things too much. We will however allow ourselves to assume ergodicity of T in dealing with a particular case of Theorem 4.1.2, namely the case $k = 2$, where, as we shall see below, one can thereby get a short proof via a special argument.

To get a better insight we begin by discussing some pertinent special cases.

Theorem 4.1.2 is clearly trivial if T is periodic, i.e. if for some m, $T^m = Id$. The next case, in order of complexity, is that of T being almost perioidic, say a translation by an irrational α on the unit circle. Let $\|x\|$ denote the distance from a real number x to the nearest integer. If α is irrational, then, as it is easy to see, for any $\varepsilon > 0$, the set $\{n \in \mathbb{Z} : \|n\alpha\| < \varepsilon\}$ is syndetic. (It is actually IP*.) Hence for a syndetic set of n, the operator T^n is ε-close to the identity operator (in the strong topology on the space of operators.) It follows that, in this case, for any $\varepsilon > 0$ the set $\big\{n : \big|\mu\big(A \cap T^{-n}A \cap \ldots \cap T^{-kn}A\big) - \mu(A)\big| < \varepsilon\big\}$ is syndetic, which is clearly more than enough for our purposes.

A slightly more general class of measure preserving systems, for which a similar argument works, is the class of so-called *compact* systems (another term: systems with discrete spectrum.) These are defined by the requirement that any $f \in L^2(X, \mathcal{B}, \mu)$ is *compact*, i.e. the closure of the orbit $\{T^n f\}_{n\in\mathbb{Z}}$ in L^2 is compact. To see that this is indeed only a slightly more general situation, note that one can show (see [HalN], Theorem 4) that if $(X, \mathcal{B}, \mu, T)$ is a compact ergodic system, then it is conjugate to a translation on a compact abelian group. Now, if $(X, \mathcal{B}, \mu, T)$ is a compact system and $A \in \mathcal{B}$ with $\mu(A) > 0$, then, as before, the set $\big\{n : \|T^n f - f\| < \varepsilon\big\}$ is syndetic and we see that in this case Theorem 4.1.2 holds for the same reason as in the case of the irrational translation.

Let us assume now that the system $(X, \mathcal{B}, \mu, T)$ is such that no nonconstant function $f \in L^2$ is compact. In particular, this means that the unitary operator induced on L^2 by T (and which, by the customary abuse of notation, we will often be denoting also by T) has no nontrivial eigenfunctions. Measure preserving systems with this property were introduced, under the name *dynamical systems with continuous spectra*, in [KoN] and form one of the most important classes of measure preserving systems. Today, such systems are called *weakly mixing systems*. As we shall see below, Theorem 4.1.2 can be verified for weakly mixing transformations with relative ease. Before showing this, we want to summarize various equivalent forms of weak mixing in the following theorem. For the proofs see [KoN] (where the stress is placed on measure preserving $\mathbb{R}$-actions), [Hopf], or any of the more modern texts such as [Hal], [Wal], or [Pe]. Note that in most books, either (i) or (ii) below is taken as the "official" definition of weak mixing, whereas the original definition in [KoN] corresponds to condition (vi).

Theorem 4.1.4. *Let T be an invertible measure preserving transformation of a probability measure space $(X, \mathcal{B}, \mu)$. Let U_T denote the operator defined on measurable functions by $(U_T f)(x) = f(Tx)$. The following conditions are equivalent:*

(i) *For any $A, B \in \mathcal{B}$*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \big|\mu(A \cap T^{-n}B) - \mu(A)\mu(B)\big| = 0;$$

(ii) *For any $A, B \in \mathcal{B}$ there is a set $P \subset \mathbb{N}$ of density zero such that*

$$\lim_{n\to\infty,\, n\notin P} \mu(A \cap T^{-n}B) = \mu(A)\mu(B);$$

(iii) *$T\times T$ is ergodic on the Cartesian square of $(X,\mathcal{B},\mu)$;*

(iv) *For any ergodic probability measure preserving system $(Y,\mathcal{D},\nu,S)$ the transformation $T\times S$ is ergodic on $X\times Y$;*

(v) *If f is a measurable function such that for some $\lambda\in\mathbb{C}$, $U_Tf=\lambda f$ a.e., then $f=const$ a.e.;*

(vi) *For $f\in L^2(X,\mathcal{B},\mu)$ with $\int f=0$ consider the representation of the positive definite sequence $\langle U_T^n f,f\rangle, n\in\mathbb{Z}$, as a Fourier transform of a measure ν on $\mathbb{T}$:*

$$\langle U_T^n f,f\rangle=\int_{\mathbb{T}}e^{2\pi inx}d\nu,\quad n\in\mathbb{Z}$$

(this representation is guaranteed by Herglotz theorem, see [He]). Then ν has no atoms.

As we shall see below, it is the relativized version of weak mixing, that is, the notion of *weak mixing relative to a factor*, that plays an important role in the analysis of the structure of an arbitrary dynamical system and which is behind the proof of Theorem 4.1.2. First, let us verify the validity of Theorem 4.1.2 for weakly mixing systems.

Theorem 4.1.5. *If $(X,\mathcal{B},\mu,T)$ is a weakly mixing system, then for any $k\in\mathbb{N}$ and any $f_i\in L^\infty(X,\mathcal{B},\mu)$, $i=1,2,\dots,k$, one has:*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}T^nf_1T^{2n}f_2\dots T^{kn}f_k=\int f_1d\mu\int f_2d\mu\dots\int f_kd\mu$$

in the L^2-norm.

Theorem 4.1.5 implies that for any $f_i\in L^\infty$, $i=0,1,\dots,k$, one has

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int f_0T^nf_1\dots T^{kn}f_k=\int f_0d\mu\int f_1d\mu\dots\int f_kd\mu.$$

Putting $f_i=1_A$, $i=0,1,\dots,k$, gives us

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mu\big(A\cap T^{-n}A\cap\dots\cap T^{-kn}A\big)=\big(\mu(A)\big)^{k+1}.$$

As a matter of fact, Theorem 4.1.5 implies that for some set $E\subset\mathbb{N}$ having zero density, one has

$$\lim_{\substack{n\to\infty\\ n\notin E}} \mu\big(A\cap T^{-n}A\cap\ldots\cap T^{-kn}A\big)=\mu(A)^{k+1} \tag{4.1}$$

To see this, note first that, since T is weakly mixing, it follows from Theorem 4.1.4, (iii) and (iv), that not only $T\times T$ is ergodic, but also $(T\times T)\times T=T\times T\times T$ and $(T\times T\times T)\times T=(T\times T)\times(T\times T)$ are ergodic (on X^3 and X^4 respectively.) But then $T\times T$ is weakly mixing. Applying Theorem 4.1.5 to $T\times T$ and performing routine manipulations one gets, for any $f_i\in L^\infty$, $i=0,1,\ldots,k$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\Big(\int f_0T^nf_1\ldots T^{kn}f_kd\mu-\int f_0d\mu\int f_1d\mu\ldots\int f_kd\mu\Big)^2=0,$$

which implies (4.1).

In the proof of Theorem 4.1.5, we shall utilize the following version of the van der Corput trick (cf. Theorem 1.32.) For the proof see, for example, [BeL3], p. 445.

Theorem 4.1.6. *Let $(u_n)_{n\in\mathbb{N}}$ be a bounded sequence in a Hilbert space $\mathcal{H}$. If for every $h\in\mathbb{N}$ it is the case that $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h},u_n\rangle$ exists and if $\lim_{H\to\infty}\frac{1}{H}\sum_{h=1}^{H}\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h},u_n\rangle=0$, then $\lim_{N\to\infty}\big\|\frac{1}{N}\sum_{n=1}^{N}u_n\big\|=0$.*

Proof of Theorem 4.1.5. Since any weakly mixing system is ergodic, the claim of the theorem trivially holds for $k=1$. To see how the induction works, consider the case $k=2$. Since the case when one of f_1,f_2 is constant brings us back to $k=1$, we can assume, in view of the identity $f=(f-\int f)+\int f$, that $\int f_1d\mu=0$. Let now $u_n=T^nf_1T^{2n}f_2$. By the ergodicity of T, we have:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h},u_n\rangle=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int T^{n+h}f_1T^{2n+2h}f_2T^nf_1T^{2n}f_2d\mu$$
$$=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int T^hf_1T^{n+2h}f_2f_1T^nf_2d\mu$$
$$=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int\big(f_1T^hf_1\big)T^n\big(f_2T^{2h}f_2\big)d\mu=\int f_1T^hf_1d\mu\int f_2T^{2h}f_2d\mu.$$

We remark now that if T is weakly mixing, then T^2 is also weakly mixing and hence $T \times T^2$ is ergodic (on the product space $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu)$). Writing $f_1 \otimes f_2$ for $f_1(x)f_2(y)$ and using the ergodicity of $T \times T^2$ we have:

$$\begin{aligned}
&\lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \int f_1 T^h f_1 \, d\mu \int f_2 T^{2h} f_2 \, d\mu \\
&= \lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \int (f_1 \otimes f_2)(T \times T^2)^h (f_1 \otimes f_2) d(\mu \times \mu) \\
&= \Big(\int f_1 \otimes f_2 \, d(\mu \times \mu)\Big)^2 = \Big(\int f_1 d\mu\Big)^2 \Big(\int f_2 d\mu\Big)^2 = 0.
\end{aligned}$$

The result now follows from Theorem 4.1.6. Note now that the same argument (in which one uses the ergodicity of $T \times T^2 \times \ldots \times T^k$) works for general k. We are done.

□

Remark 4.1.7.
1. It is not hard to show that if the system $(X, \mathcal{B}, \mu, T)$ is such that for any $f_1, f_2 \in L^\infty(X, \mathcal{B}, \mu)$, one has $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f_1 T^{2n} f_2 = \int f_1 d\mu \int f_2 d\mu$ in the L^2-norm, then T is weakly mixing.
2. By using a modification of Theorem 4.1.6, which pertains to the uniform Cesàro averages $\frac{1}{N-M} \sum_{n=M}^{N-1} x_n$, (see, for example, Remark 2.2 in [BeL3]) one can show that in Theorem 4.1.5 one actually has

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} T^n f_1 T^{2n} f_2 \ldots T^{kn} f_k = \int f_1 \, d\mu \int f_2 \, d\mu \ldots \int f_k \, d\mu,$$

which implies

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu\big(A \cap T^{-n} A \cap \ldots \cap T^{-kn} A\big) = \mu(A)^{k+1}.$$

3. The reader is invited to check that the proof above actually gives the following more general result, first proved in [BB].

Theorem 4.1.8. *Assume that, for $k \geq 2$, $T_1, T_2, \ldots, T_k$ are commuting measure preserving transformations on a probability space $(X, \mathcal{B}, \mu)$. Then the following are equivalent:*

(i) *For any* $f_1, f_2, \ldots, f_k \in L^\infty(X, \mathcal{B}, \mu)$ *one has*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} T_1^n f_1 T_2^n f_2 \ldots T_k^n f_k = \int f_1 d\mu \int f_2 d\mu \ldots \int f_k d\mu \text{ in } L^2.$$

(ii) *For any* $i \neq j$, $T_i T_j^{-1}$ *is ergodic on* X *and* $T_1 \times T_2 \times \ldots \times T_k$ *is ergodic on* X^k.

The two special cases of Theorem 4.1.2, which we verified above, correspond on a spectral level to two complementary classes of unitary operators, namely those having discrete spectrum and continuous spectrum. While these two cases are much too special to allow us to conclude the proof of Theorem 4.1.2 for general k, they are sufficient for $k = 2$ (which constitutes the first non-trivial case of Theorem 4.1.2.) These two special cases are also important in that they indicate a possible line of attack which we will discuss after first completing the proof for $k = 2$.

Proof of Theorem 4.1.2 for k=2. Assume first that T is ergodic, and consider the following splitting of $L^2(X, \mathcal{B}, \mu) = \mathcal{H}$. $\mathcal{H} = \mathcal{H}_c \oplus \mathcal{H}_{wm}$, where the T-invariant subspaces $\mathcal{H}_c$ and $\mathcal{H}_{wm}$ are defined as follows:

$$\begin{aligned}
\mathcal{H}_c &= \overline{\mathrm{Span}\{f \in \mathcal{H} : \text{ there exists } \lambda \in \mathbb{C} \text{ with } Tf = \lambda f\}} \\
&= \{f \in \mathcal{H} : \text{ the orbit } (T^n f)_{n\in\mathbb{Z}} \text{ is precompact in norm topology}\}, \\
\mathcal{H}_{wm} &= \mathcal{H}_c^\perp = \{f \in \mathcal{H} : \forall g \in \mathcal{H}, \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} |\langle Tf, g\rangle| = 0\}.
\end{aligned}$$

(We remark in passing that this splitting is valid for any unitary operator, and moreover, can be defined in such a way that it makes sense for any group of unitary operators.)

Writing f for 1_A let $f = f_c + f_{wm}$ where $f_c \in \mathcal{H}_c, f_{wm} \in \mathcal{H}_{wm}$. Note that $f_c \geq 0$ and $\int f_c d\mu = \mu(A)$, while $\int f_{wm} d\mu = 0$. The non-negativity of f_c follows from an argument similar to the one used in the proof of Theorem 1.31 to establish the non-negativity of the projection of 1_A on the space $\mathcal{H}_{rat}$. Note also that applying this argument again to $1 - f_c$ gives us that $0 \leq f_c \leq 1$. Since $f_{wm} = 1_A - f_c$, it follows also that f_{wm} satisfies $|f_{wm}| \leq 1$.

Using the decomposition $f = f_c + f_{wm}$, we have

$$\frac{1}{N}\sum_{n=0}^{N-1} T^n f T^{2n} f = \frac{1}{N}\sum_{n=0}^{N-1} T^n f_c T^{2n} f_c + \sum_{n=0}^{N-1} T^n f_c T^{2n} f_{wm}$$
$$+\frac{1}{N}\sum_{n=0}^{N-1} T^n f_{wm} T^{2n} f_c + \frac{1}{N}\sum_{n=0}^{N-1} T^n f_{wm} T^{2n} f_{wm}.$$

We claim that the last three expressions (in which f_{wm} occurs) have zero limit in L^2 as $N \to \infty$. To see this the reader is invited to reexamine the proof of Theorem 4.1.5 and to observe that it was actually shown there that if $\varphi \in \mathcal{H}_{wm}$, then for any $\psi \in \mathcal{H} = L^2(X, \mathcal{B}, \mu)$, one has (assuming that at least one of φ, ψ is bounded):

$$\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=0}^{N-1} T^n \varphi T^{2n} \psi \right\| = \lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=0}^{N-1} T^n \psi T^{2n} \varphi \right\| = 0.$$

So, we see that in L^2,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T^n f T^{2n} f = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T^n f_c T^{2n} f_c,$$

if the latter limit exists. But the existence of this limit clearly follows from the fact that this is certainly the case when one substitutes for f_c a finite linear combination of eigenfunctions of T and that eigenfunctions span the space $\mathcal{H}_c$. So we have that $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T^n f T^{2n} f$ also exists in L^2, and hence

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \int f T^n f T^{2n} f d\mu = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \int 1_A T^n 1_A T^{2n} 1_A d\mu$$
$$= \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \mu(A \cap T^{-n} A \cap T^{-2n} A)$$

exists as well.

It remains to establish the positivity of the limit in question. Note that, for bounded $g_1, g_2 \in \mathcal{H}_c$, one has $g_1 \cdot g_2 \in \mathcal{H}_c$, and hence f_{wm} is orthogonal to $T^n f_c T^{2n} f_c$. We have:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mu(A\cap T^{-n}A\cap T^{-2n}A)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int fT^n f_c T^{2n} f_c d\mu$$
$$=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int (f_c+f_{wm})T^n f_c T^{2n} f_c d\mu=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int f_c T^n f_c T^{2n} f_c d\mu.$$

Note now that since f_c is a compact function, the set

$$\{n\in\mathbb{Z}:\|T^n f_c-f_c\|<\varepsilon\}$$

is syndetic for every $\varepsilon>0$, and hence the set

$$S_\varepsilon=\{n\in\mathbb{Z}:\Big|\int f_c T^n f_c T^{2n} f_c d\mu-\int f_c^3 d\mu\Big|<\varepsilon\}$$

is also syndetic. Note also that $\int f_c^3 d\mu\ge\big(\int f_c\big)^3=\big(\mu(A)\big)^3$.

Therefore, if ε is small enough, we shall have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mu(A\cap T^{-n}A\cap T^{-2n}A)$$
$$\ge\overline{\lim_{N\to\infty}}\frac{1}{N}\sum_{n\in S_\varepsilon\cap[0,N-1]}\mu(A\cap T^{-n}A\cap T^{-2n}A)>0.$$

To finish the proof of this special case one uses the ergodic decomposition. It is not hard to see that in the non-ergodic case, both the convergence and the positivity of the limit hold as well. We omit the details.

□

As a bonus, we have obtained the fact that for any measure preserving system $(X,\mathcal{B},\mu,T)$ and any $f,g\in L^\infty$, $\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}T^n fT^{2n}g$ exists in L^2 and equals $\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}T^n f_c T^{2n} g_c$, where f_c,g_c denote the orthogonal projections of f,g on $\mathcal{H}_c$. One, naturally, would like to know whether, in general, one has the convergence of the expressions of the form $\frac{1}{N}\sum_{n=0}^{N-1}T^n f_1 T^{2n} f_2\dots T^{kn} f_k$, where $f_i\in L^\infty(X,\mathcal{B},\mu)$, $i=1,2,\dots,k$. For $k=3$ the positive answer to this question was provided in [CL2] for totally ergodic T and in full generality in [FuW2], [Zh], and [HoK1]. The recalcitrant problem of establishing the convergence for general k was solved only recently, in the remarkable work of B. Host and B. Kra [HoK2] and T.

Ziegler, [Zie]. See Section 5 below for a discussion of various convergence results which are suggested by combinatorial applications of ergodic theory. See also appendices A and B written by A. Leibman and A. Quas and M. Wierdl which deal with convergence issues. Note, however, that while the study of convergence is more fundamental from the point of view of ergodic theory, it is (multiple) recurrence, i.e. the positivity of the expressions like $\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \ldots \cap T^{-kn}A)$, which is needed for combinatorial and number-theoretic applications. (The situation is, of course, more complex. To be somewhat imprecise, convergence results may, in some cases, provide the shortest path to establishing recurrence. This point is certainly supported by the proof of Theorem 1.31 and the above discussion of the $k = 2$ case of Theorem 4.1.2. See also Theorem 5.20(i) below.)

We return now to our discussion of Theorem 4.1.5. It turns out that for $k > 2$, the Hilbertian splitting utilized above for $k = 2$ is no longer sufficient, and in order to establish multiple recurrence one has to undertake a deeper study of the structure of general measure preserving systems. In order to describe the main points of Furstenberg's approach, we will review first some general facts. For more information and missing details, the reader is encouraged to consult [Fu2], [Fu3], and [FuKO]. Our presentation below follows mainly [FuKO] where a simplified proof of Theorem 4.1.2 is presented. The only significant point of departure from [FuKO] is in the treatment of compact extensions (see Definition 4.1.15 below), where we will use a "soft" argument based on van der Waerden's theorem. One of the reasons for this choice is that coloring theorems seem to be indispensable in proving more sophisticated multiple recurrence results and we want to use this opportunity to acquaint the reader with this technique.

Given two probability measure spaces $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{D}, \nu)$ and a map $\pi : X \to Y$ such that $\pi^{-1}(\mathcal{D}) \subset \mathcal{B}$ and $\pi\mu = \nu$, we say that $(X, \mathcal{B}, \mu)$ is an *extension* of $(Y, \mathcal{D}, \nu)$, and that $(Y, \mathcal{D}, \nu)$ is a *factor* of $(X, \mathcal{B}, \mu)$.

Under mild conditions on the regularity of the space $(X, \mathcal{B}, \mu)$ (which are usually satisfied in the case of Lebesgue spaces – our standing assumption), one can associate with the factor $(Y, \mathcal{D}, \nu)$ a family of measures $\{\mu_y\}_{y \in Y}$ on $(X, \mathcal{B})$ with the following properties:

(i) For each $f \in L^1(X, \mathcal{B}, \mu)$ one has $f \in L^1(X, \mathcal{B}, \mu_y)$ for a.e. $y \in Y$.

(ii) The function $g(y) = \int f d\mu_y$ belongs to $L^1(Y, \mathcal{D}, \nu)$ and

$$\int \Big(\int f(x) d\mu_y(x) \Big) d\nu = \int f(x) d\mu(x).$$

(iii) If f is measurable with respect to $\pi^{-1}(\mathcal{D})$, then

$$\int f d\mu_{\pi(x)} = f(x) \text{ a.e.}$$

Using the family $\{\mu_y\}_{y\in Y}$, we will write $\mu = \int \mu_y d\nu(y)$ (which means that, for any $A \in \mathcal{B}$, $\mu(A) = \int \mu_y(A)d\nu(y)$) and refer to this decomposition as the *disintegration* of μ with respect to the factor $(Y, \mathcal{D}, \nu)$.

For any $1 \leq p \leq \infty$ one can define the *conditional expectation* operator $E(\cdot|Y)$ from $L^p(X, \mathcal{B}, \mu)$ to $L^p(Y, \mathcal{D}, \nu)$ by the formula

$$E(f|Y)(y) = \int f d\mu_y, f \in L^2(X, \mathcal{B}, \mu).$$

Clearly, for $f \geq 0$, one has $E(f|Y) \geq 0$ and $E(1|Y) = 1$. Also, by property (ii) above, one has $\int f d\mu = \int E(f|Y)d\nu$.

Note that, given the measure space $(X, \mathcal{B}, \mu)$, there is a natural 1-1 correspondence between its factors and sub-σ-algebras of $\mathcal{B}$. This correspondence allows one to identify the space $L^2(Y, \mathcal{D}, \nu)$ with a closed subspace of $L^2(X, \mathcal{B}, \mu)$ which is of the form $L^2(X, \mathcal{B}_1, \mu)$, where $\mathcal{B}_1 = \pi^{-1}(\mathcal{D})$. This, in turn, leads to a convenient interpretation of conditional expectation operator as the orthogonal projection $L^2(X, \mathcal{B}, \mu) \to L^2(X, \mathcal{B}_1, \mu) \cong L^2(Y, \mathcal{D}, \nu)$.

For any $f \in L^\infty(Y, \mathcal{D}, \nu)$ (viewed as a bounded function in $L^2(X, \mathcal{B}, \mu)$ which is measurable with respect to $\mathcal{B}_1$), one has $E(gf|Y) = fE(g|Y)$. For more details on conditional expectation operators see [Fu3], Ch. 5, Section 3 or [Bi], Section 34.

Suppose now that $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ are extensions of $(Y, \mathcal{D}, \nu)$ and $\pi_1 : X_1 \to Y$, $\pi_2 : X_2 \to Y$ are the corresponding measure preserving mappings. One can form the *fibre product* space $(X, \mathcal{B}, \mu)$, where

$$X = X_1 \times_Y X_2 = \big\{(x_1, x_2) \in X_1 \times X_2 : \pi_1(x_1) = \pi_2(x_2)\big\},$$

$\mathcal{B}$ is the restriction of $\mathcal{B}_1 \times \mathcal{B}_2$ to X, and μ is defined via the disintegrations $\{\mu_y^{(1)}\}_{y\in Y}$, $\{\mu_y^{(2)}\}_{y\in Y}$ by the formula

$$\mu(A) = (\mu_1 \times_Y \mu_2)(A) = \int (\mu_y^{(1)} \times \mu_y^{(2)})(A)d\nu(y).$$

The notions of extension, factor, and fibre product are naturally extended to measure preserving systems. Given two probability measure preserving systems $\boldsymbol{X} = (X, \mathcal{B}, \mu, T)$ and $\boldsymbol{Y} = (Y, \mathcal{D}, \nu, S)$, one says that $\boldsymbol{X}$ is an extension of $\boldsymbol{Y}$, and $\boldsymbol{Y}$ a factor of $\boldsymbol{X}$, if the corresponding map $\pi : X \to Y$ is not only measure preserving but also satisfies $S\pi(x) = \pi T(x)$ for a.e. $x \in X$. We have now the following formulas:

(iii) for almost every $y \in Y$, $T\mu_y = \mu_{Sy}$, meaning $\mu_y(T^{-1}A) = \mu_{Sy}(A)$ for any $A \in \mathcal{B}$.

(iv) For any $f \in L^2(X, \mathcal{B}, \mu), SE(f|Y) = E(Tf|Y)$.

When the system $\boldsymbol{X} = (X, \mathcal{B}, \mu, T)$ is not ergodic, it has a natural nontrivial factor $\boldsymbol{X}_{inv} = (X, \mathcal{B}_{inv}, \mu, T)$, where $\mathcal{B}_{inv}$ is the σ-algebra of T-invariant sets in $\mathcal{B}$. It is not hard to see that the disintegration of μ corresponding to this factor is nothing but the classical ergodic decomposition of μ (which was treated first in [Ne2].) Another natural example of a factor, which we have, implicitly, encountered already, is associated with the space $\mathcal{H}_c$ of compact functions. Indeed, one has the following theorem.

Theorem 4.1.9. (cf. [Kre1], Theorem 2.2.) *Let $(X, \mathcal{B}, \mu, T)$ be a probability measure preserving system and let $\mathcal{B}_c$ be the smallest σ-algebra in $\mathcal{B}$ with respect to which the elements of $\mathcal{H}_c$ are measurable. Then $\mathcal{B}_c$ is T-invariant and $\mathcal{H}_c \cong L^2(X, \mathcal{B}_c, \mu)$.*

Clearly, $(X, \mathcal{B}_c, \mu, T)$ is a maximal compact factor. As we have already mentioned above, if T is ergodic, the system $(X, \mathcal{B}_c, \mu, T)$ is conjugate to a translation on a compact abelian group (see [HalN], Theorem 4.) In this case, $(X, \mathcal{B}_c, \mu, T)$ is often called the (maximal) Kronecker factor.

We can formulate now a criterion in terms of factors for a system to be weakly mixing. (Note that this is just a new way of expressing a familiar concept.)

Theorem 4.1.10. *A probability measure preserving system is weakly mixing if and only if it has no nontrivial compact factors.*

In order to prove Theorem 4.1.2, one has to study the relativized notions of weak mixing and compactness with respect to a factor.

To define a weak mixing extension, one needs the notion of a relative product of measure preserving systems. Let $\boldsymbol{X}_1 = (X_1, \mathcal{B}_1, \mu_1, T_1)$ and $\boldsymbol{X}_2 = (X_2, \mathcal{B}_2, \mu_2, T_2)$ be extensions of $\boldsymbol{Y} = (Y, \mathcal{D}, \nu, S)$. We claim that the measure $\mu_1 \times_Y \mu_2$ is $(T_1 \times T_2)$-invariant, that is, for any measurable $A \subseteq X_1 \times_Y X_2$

one has $\mu_1 \times_Y \mu_2\big((T_1 \times T_2)^{-1}A\big) = \mu_1 \times_Y \mu_2(A)$. One needs only to verify this for the sets of the form $A = A_1 \times A_2$ where $A_i \in \mathcal{B}_i$. By definition of $\mu_1 \times_Y \mu_2$ we have

$$\mu_1 \times_Y \mu_2\big((T_1 \times T_2)^{-1}(A_1 \times A_2)\big) = \int \mu_y^{(1)} \times \mu_y^{(2)}(T_1^{-1}A_1 \times T_2^{-1}A_2)d\nu(y)$$
$$= \int T_1\mu_y^{(1)} \times T_2\mu_y^{(2)}(A_1 \times A_2)d\nu(y) = \int \mu_{Sy}^{(1)} \times \mu_{Sy}^{(2)}(A_1 \times A_2)d\nu(y)$$
$$= \int \mu_y^{(1)} \times \mu_y^{(2)}(A_1 \times A_2)dS\nu(y) = \int \mu_y^{(1)} \times \mu_y^{(2)}(A_1 \times A_2)d\nu(y)$$
$$= \mu_1 \times_Y \mu_2(A_1 \times A_2),$$

and so $\boldsymbol{X}_1 \times_{\boldsymbol{Y}} \boldsymbol{X}_2 = (X_1 \times X_2, \mathcal{B}_1 \times \mathcal{B}_2, \mu_1 \times_Y \mu_2, T_1 \times T_2)$ is a measure preserving system, which is called the relative product of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ (with respect to $\boldsymbol{Y}$.)

Definition 4.1.11. *The system $\boldsymbol{X} = (X, \mathcal{B}, \mu, T)$ is an ergodic extension of $\boldsymbol{Y} = (Y, \mathcal{D}, \nu, S)$ if the only T-invariant sets in $\mathcal{B}$ are preimages of the invariant sets in $\mathcal{D}$. The system $\boldsymbol{X}$ is a weakly mixing extension of $\boldsymbol{Y}$ if $\boldsymbol{X} \times_{\boldsymbol{Y}} \boldsymbol{X}$ is an ergodic extension of $\boldsymbol{Y}$.*

One can show that most properties of the "absolute" weak mixing (and in particular, items (i) through (iv) in Theorem 4.1.4) extend, with obvious modifications, to statements about relative weak mixing. For example, one has the following fact.

Theorem 4.1.12. (Cf. [Fu3], Proposition 6.2.) *A measure preserving system $(X, \mathcal{B}, \mu, T)$ is a weak mixing extension of $(Y, \mathcal{D}, \nu, S)$ if and only if for any $A_1, A_2 \in \mathcal{B}$ one has*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \big(\mu_y(A_1 \cap T^{-n}A_2) - \mu_y(A_1)\mu_y(T^{-n}A_2)\big)^2 d\nu(y) = 0.$$

Moreover, by using Theorem 4.1.6, one can obtain (by an argument analogous to the one used in the proof of Theorem 4.1.5) the following result.

Theorem 4.1.13. *If $(X, \mathcal{B}, \mu, T)$ is a weak mixing extension of $(Y, \mathcal{D}, \nu, S)$, then for any $A_0, A_1, \ldots, A_k \in \mathcal{B}$ one has*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \big(\mu_y(A_0 \cap T^{-n}A_1 \cap T^{-2n}A_2 \ldots \cap T^{-kn}A_k) -$$
$$\mu_y(A_0)\mu_y(T^{-n}A_1) \ldots \mu_y(T^{-kn}A_k)\big)^2 d\nu(y) = 0.$$

Let us say, following [FuKO], that a system $\boldsymbol{X} = (X, \mathcal{B}, \mu, T)$ has the SZ property, or that T is SZ, if Theorem 4.1.2 holds for $\boldsymbol{X}$. (For example, as we have already seen above, compact and weakly mixing systems do have the SZ property)

We have now the following corollary of Theorem 4.1.13:

Theorem 4.1.14. ([FuKO], Theorem 8.4) *If $(X, \mathcal{B}, \mu, T)$ is weakly mixing extension of $(Y, \mathcal{D}, \nu, S)$ and the transformation S is SZ, then $(X, \mathcal{B}, \mu, T)$ has the SZ property.*

Proof. Let $A \in \mathcal{B}$ with $\mu(A) > 0$ and denote $f = 1_A$. Note that if $a > 0$ is small enough then the set $A_1 = \{y : E(1_A|Y)(y) \geq a\}$ satisfies $\nu(A_1) > 0$. It follows now from Theorem 4.1.13 (and the formula $E(f|Y)(y) = \int f d\mu_y$) that, since $E(1_A|Y) \geq a \cdot 1_{A_1}$,

$$\frac{1}{N}\sum_{n=0}^{N-1} \mu(A \cap T^{-n}A \cap \ldots \cap T^{-kn}A) > \frac{1}{2}a^{k+1} \cdot \frac{1}{N}\sum_{n=0}^{N-1} \nu(A_1 \cap S^{-n}A_1 \cap \ldots \cap S^{-kn}A_1)$$

for all large enough N. The result now follows from the assumption that $(Y, \mathcal{D}, \nu, S)$ has the SZ property. $\square$

We will define now relatively compact extensions and show that an analogue of Theorem 4.1.14 holds. Unlike Theorem 4.1.14, which is a more or less straightforward extension of Theorem 4.1.5, the argument needed for the treatment of compact extensions (the mere definition of which is, in our opinion, much less trivial than that of weak mixing extensions) is perhaps the most subtle part of the proof of Theorem 4.1.2.

Definition 4.1.15. *Let $\boldsymbol{X} = (X, \mathcal{B}, \mu, T)$ be an extension of $\boldsymbol{Y} = (Y, \mathcal{D}, \nu, S)$. Call a function $f \in L^2(X, \mathcal{B}, \mu)$ almost periodic, or an AP-function, relative to $\boldsymbol{Y}$ if for any $\varepsilon > 0$ there exist $r \in \mathbb{N}$ and functions $g_1, g_2, \ldots, g_r \in L^2(X, \mathcal{B}, \mu)$ such that, for every $n \in \mathbb{Z}$, $\min_{1 \leq s \leq r} \|T^n f - g_s\|_{L^2(\mu_y)} < \varepsilon$ for almost every $y \in Y$. We say that $\boldsymbol{X}$ is a compact extension of $\boldsymbol{Y}$ if AP-functions are dense in $L^2(X, \mathcal{B}, \mu)$.*

Clearly any compact system (i.e. a system for which the subspace $\mathcal{H}_c$ coincides with $L^2(X, \mathcal{B}, \mu)$) is a compact extension of the trivial one-point system. Note that in this case every element of $L^2(X, \mathcal{B}, \mu)$ is an AP-function.

A less trivial example and, in a sense, a typical one is given by so-called isometric extensions. Let $\boldsymbol{Y} = (Y, \mathcal{D}, \nu, S)$ be an arbitrary system and let Z

be a compact metric space equipped with a probability measure η on $\mathcal{B}_Z$, the (completion of the) σ-algebra of Borel sets in Z. Suppose that G is a compact group of isometries of Z and define for some measurable family $\sigma(y)$ of elements of G a transformation T on $X = Y \times Z$ by

$$T(y, z) = (Sy, \sigma(y)z).$$

One can verify that the system $\boldsymbol{X} = (Y \times Z, \mathcal{D} \times \mathcal{B}_Z, \nu \times \eta, T)$ is a compact extension of $\boldsymbol{Y}$. Perhaps the shortest parth to this verification is to consider (as a special case) the familiar skew product on the 2-torus given by $R : (y, z) \to (y + \alpha, z + y)$ and to observe that a similar proof works for general isometric extensions. Note that for non-trivial isometric extensions it is no longer true that every function is AP. For example, it is not hard to see that in the case of the transformation R above, if α is irrational then for any $f \in L^2(Y) \setminus L^\infty(Y)$, functions of the form $f(y)e^{2\pi i z}$ cannot be AP relative to the factor corresponding to the translation of the first coordinate by α.

The importance of weakly mixing and compact extensions lies with the fundamental fact, established by Furstenberg in the course of his proof of Szemerédi's theorem, that any system $\boldsymbol{X} = (X, \mathcal{B}, \mu, T)$ appears in a chain (possibly transfinite), $\boldsymbol{X} \to \ldots \to \boldsymbol{X}_{\alpha+1} \to \boldsymbol{X}_\alpha \to \ldots \to \boldsymbol{X}_1 \to \boldsymbol{X}_0$, in which the individual links $\boldsymbol{X}_{\alpha+1} \to \boldsymbol{X}_\alpha$ are either compact or weakly mixing extensions. (As a matter of fact, one can take all of the extensions, with the possible exception of the last link $\boldsymbol{X} = \boldsymbol{X}_{\eta+1} \to \boldsymbol{X}_\eta$, to be compact.)

The topological predecessor of this ergodic-theoretical structure theorem is a similar structure theorem, also due to Furstenberg, for distal systems, which states that any distal system can be seen as a tower of *isometric* extensions. See [Fu1] for details. The structure theory of distal systems works for general locally compact group actions, which hints that measure theoretical structure theory can also be established in this generality. This was done in independent and deep work of Zimmer. (See [Zim1], [Zim2].)

Returning to the discussion of the proof of Theorem 4.1.2, we are in position now to describe the general scheme of the proof. As we shall show in detail below, if $\boldsymbol{X}$ is a compact extension of $\boldsymbol{Y}$ and $\boldsymbol{Y}$ has the SZ property, then $\boldsymbol{X}$ also does. Now, one can show by a routine argument that any totally ordered by inclusion family of factors of a system $(X, \mathcal{B}, \mu, T)$ which have the SZ property has a maximal element. (See [FuKO], Proposition 7.1.) But then it follows from the structure theorem cited above that this maximal factor has to be $(X, \mathcal{B}, \mu, T)$ itself, which gives Theorem 4.1.2. A somewhat shorter (or, rather, less involved) path, which avoids the full strength of the structure theorem, is via the following proposition.

Theorem 4.1.16. ([FuKO], Theorem 5.10) *If $\boldsymbol{X} = (X, \mathcal{B}, \mu, T)$ is an extension of $\boldsymbol{Y} = (Y, \mathcal{D}, \nu, S)$ which is not relatively weak mixing, then there exists a strictly intermediate factor $\boldsymbol{X}^*$ between $\boldsymbol{Y}$ and $\boldsymbol{X}$ such that $\boldsymbol{X}^*$ is a compact extension of $\boldsymbol{Y}$.*

Either way, all that is needed now to bring the proof of Theorem 4.1.2 to conclusion is the following result.

Theorem 4.1.17. *If $\boldsymbol{X} = (X, \mathcal{B}, \mu, T)$ is a compact extension of $\boldsymbol{Y} = (Y, \mathcal{D}, \nu, S)$ and $\boldsymbol{Y}$ has the SZ property, then $\boldsymbol{X}$ also does.*

Proof. We shall utilize van der Waerden's theorem on arithmetic progressions. To make the ideas clear (and to stress the relevance of van der Waerden's theorem), let us first go back to the "absolute" case and show how the proof works when $(X, \mathcal{B}, \mu, T)$ is a compact system. Let $A \in \mathcal{B}$ with $\mu(A) > 0$, let $f = 1_A$ and let, for a given $\varepsilon > 0$, $g_1, g_2, \ldots, g_r$ be elements of the compact set $K = \overline{\{T^n f\}_{n\in\mathbb{Z}}}$ such that for any $n \in \mathbb{Z}$ there is $j = j(n)$ in $\{1, 2, \ldots, r\}$ satisfying $\|T^n f - g_{j(n)}\| < \varepsilon$. This naturally defines an r-coloring $\mathbb{Z} = \bigcup_{i=1}^r C_i$, and by van der Waerden's theorem, there exists $j \in \{1, 2, \ldots, r\}$ such that for some $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ one has $\|T^{m+in} f - g_j\| < \varepsilon, i = 0, 1, \ldots, k$, which implies $\operatorname{diam}\big\{T^m f, T^{m+n} f, \ldots, T^{m+kn} f\big\} < 2\varepsilon$ (in $L^2(X, \mathcal{B}, \mu)$.) Note that the set of possible n with this property has positive lower density (and, in fact, is syndetic and even IP* – see the remark following Corollary 2.5.) Since T is an isometry, we have $\operatorname{diam}\big\{f, T^n f, \ldots, T^{kn} f\big\} < 2\varepsilon$, and hence by choosing ε small enough, we see that for a "large" set of n, $\int f T^n f \ldots T^{kn} f d\mu = \mu(A \cap T^{-n} A \cap \ldots \cap T^{-kn} A)$ is arbitrarily close to $\mu(A)$. This certainly implies that

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

The scheme of usage of van der Waerden's theorem in the case of relatively compact extensions is similar but a little bit more sophisticated. Before embarking on the proof, let us make some convenient reductions. First, note that deleting from a given set $A \in \mathcal{B}$ with $\mu(A) > 0$ portions for which $\mu_y(A) \leq \frac{1}{2}\mu(A)$ removes less than half of the measure from A, and hence we can assume without loss of generality that there exists a set $A_1 \in \mathcal{D}$ with $\nu(A_1) \geq \frac{1}{2}\mu(A)$ and such that, for $y \in A_1$, $\mu_y(A) \geq \frac{1}{2}\mu(A)$ and for $y \notin A_1$, $\mu_y(A) = 0$. Second, one can show that, by removing additional arbitrarily small portions from A, one can assume that $f = 1_A$ is compact relative to $\boldsymbol{Y}$. (See for the details [FuKO], p. 548 or [Fu3], Theorem 6.13.)

Fix a small enough $\varepsilon > 0$ and functions $g_1, g_2, \ldots, g_r$ (one of which is assumed to be 0) such that for any $n \in \mathbb{Z}$,

$$\min_{1 \leq s \leq r} \|T^n f - g_s\|_y < \varepsilon \text{ for a.e. } y \in Y.$$

Let N be such that for any r-coloring of $\{1, 2, \ldots, N\}$ one has a monochromatic progression of length $k+1$, and assume that for the set $A_1 \in \mathcal{D}$ described above and some $c_1 > 0$, the set $R_N = \big\{n \in \mathbb{N} : \nu(A_1 \cap S^{-n}A_1 \cap \ldots \cap S^{-nN}A_1) > c_1\big\}$ is of positive lower density. We shall show that there exist constants $c_2 > 0$ and $M \in \mathbb{N}$ such that for any $n \in R_N$ there exists $d \in \{1, 2, \ldots, M\}$ with $\mu(A \cap T^{-dn}A \cap \ldots \cap T^{-k(dn)}A) > c_2$. This, clearly, will imply that $\boldsymbol{X}$ has the SZ property. Note that for every $y \in A_1$ and $n \in R_N$ one has $S^{in}y \in A_1, i = 0, 1, \ldots, N$. Now, for each $y \in A_1$ and each $n \in R_N$, the inequalities $\min_{1 \leq s \leq r} \|T^{in}f - g_s\|_y < \varepsilon, i = 1, 2, \ldots, N$, define an r-coloring of $\{1, 2, \ldots, N\}$. By van der Waerden's theorem, there exists a monochromatic arithmetic progression $\{i, i+d, \ldots, i+kd\} \subset \{1, 2, \ldots, N\}$ which implies that, for some $g_{s(y)} = g$, $\|T^{(i+jd)n}f - g\|_y < \varepsilon$ for $j = 0, 1, \ldots, k$. This, in turn, implies that $\|T^{jdn}f - g\|_{S^{in}y} < \varepsilon$. It remains now to choose a progression $\{i, i+d, \ldots, i+kd\}$ which occurs for a set A_2 of y of measure at least $\frac{\nu(A_1)}{P}$, where P is the total number of possibilities for the choice of a $(k+1)$-element progression from $\{1, 2, \ldots, n\}$. Note that, since $A_2 \subset A_1$, for each $y \in A_2$, one has $\mu_{S^{in}y}(A) \geq \frac{1}{2}\mu(A)$. This implies that (if ε is small enough)

$$\mu_{S^{in}y}\big(A \cap T^{-dn}A \cap \ldots \cap T^{-k(dn)}A\big) > \mu_{S^{in}y}(A) - (k+1)\varepsilon > \frac{1}{3}\mu(A).$$

Integrating over the set A_2, we get

$$\mu\big(A \cap T^{-dn}A \cap \ldots \cap T^{-k(dn)}A\big) \geq \frac{1}{3}\mu(A)\nu(A_2) \geq \frac{1}{3P}\mu(A)\nu(A_1) = c_2.$$

We are done. □

4.2 An overview of multiple recurrence theorems

Furstenberg's ground-breaking proof of the ergodic Szemerédi theorem was the starting point of a new area: Ergodic Ramsey Theory. In this subsection we shall discuss various multiple recurrence results which are dynamical versions of corresponding Ramsey-theoretical density statements, and which have, so far, no conventional proof. While the proofs of these results are rather involved (which is manifested, in particular, by the length of papers such as [FuK2], [FuK4], [Le2], [BeM3], [BeLM]), they have a conspicuous commonality of the main structural features. One of our intentions in the following discussion is to stress the structural analogies between various proofs, while paying attention to new ideas whose introduction is necessary in the course of establishing new, stronger, and more refined results.

For a warm-up, let us start the discussion with the density version of Theorem 2.6.

Let V_F be a countably infinite vector space over a finite field F. As in Section 2, let us identify V_F with the direct sum F_∞ of countably many copies of F:

$$F_\infty = \big\{g = (a_1, a_2, \ldots) : a_i \in F \text{ and all but finitely many } a_i = 0\big\} = \bigcup_{n=1}^{\infty} F_n,$$

where $F_n = \big\{g = (a_1, a_2, \ldots), a_i \in F, a_i = 0 \text{ for } i > n\big\}$.

We shall say that a set $E \subset V_F \cong F_\infty$ *has positive upper density* if $\overline{d}_{F_\infty}(E) = \limsup_{N\to\infty} \frac{|E\cap F_n|}{|F_n|} > 0$.

Theorem 4.2.1. *Any set of positive upper density in the vector space V_F contains arbitrarily large affine subspaces.*

Note that since F is a finite field, saying "arbitrarily large" in the formulation above is tantamount to saying "of arbitrarily large dimension".

We now formulate an ergodic-theoretical theorem which is analogous to Theorem 4.1.2 and which implies Theorem 4.2.1.

Theorem 4.2.2. *For any measure preserving action $(T_g)_{g\in F_\infty}$ on a probability measure space $(X, \mathcal{B}, \mu)$ and for any $A \in \mathcal{B}$ with $\mu(A) > 0$, one has*

$$\liminf_{n\to\infty} \frac{1}{|F_n|} \sum_{g\in F_n} \mu\Big(\bigcap_{c\in F} T_{cg} A\Big) > 0.$$

For the derivation of Theorem 4.2.1 from Theorem 4.2.2, one can use the following version of Furstenberg's correspondence principle. Both the version below and the result stated above as Theorem 1.25 are special cases of Theorem 5.8, which will be proved in the next section.

Theorem 4.2.3. *For any set $E \subset F_\infty$ with $\overline{d}_{F_\infty}(E) > 0$, there exists a probability measure preserving system $\big(X, \mathcal{B}, \mu, (T_g)_{g \in F_\infty}\big)$ and a set $A \in \mathcal{B}$ with $\mu(A) = \overline{d}_{F_\infty}(E)$ such that for all $k \in \mathbb{N}$ and any $g_1, g_2, \ldots, g_k \in F_\infty$ one has*

$$\overline{d}_{F_\infty}\big(E \cap E - g_1 \cap \ldots \cap E - g_k\big) \geq \mu\big(A \cap T_{g_1}A \cap \ldots \cap T_{g_k}A\big).$$

Noting that the set $\{cg\}_{c \in F}$ forms a one-dimensional subspace of $V_F \cong F_\infty$, we see that Theorem 4.2.2 immediately implies, via Furstenberg's correspondence principle, that for some $g \neq \mathbf{0}$, $\overline{d}_{F_\infty}\big(\bigcap_{c \in F}(E - cg)\big) > 0$ and hence for any $x \in \bigcap_{c \in F}(E - cg)$, the one dimensional affine space $\{x + cg\}_{c \in F}$ is contained in E.

To obtain the full strength of Theorem 4.2.1, one can use the following "iterational" trick (which is very similar to that utilized in the proof of Theorem 1.12.) Namely, use Theorem 4.2.2 to find $g_1 = (b_1, b_2, \ldots, b_k, 0, 0, \ldots) \neq \mathbf{0}$ with the property that the set $A_1 = \bigcap_{c \in F} T_{cg_1}A$ has positive measure. Apply now Theorem 4.2.2 to the restriction of the action $(T_g)_{g \in F_\infty}$ to the subgroup $G_1 \subset F_\infty$ which is defined by

$$G_1 = \big\{g = (a_1, a_2, \ldots) \in F_\infty : a_1, a_2, \ldots, a_k = 0\big\}.$$

In other words, the supports of elements from G_1 are disjoint from the support of our $g_1 = (b_1, b_2, \ldots, b_k, 0, 0, \ldots)$. Note also that G_1 is isomorphic to the direct sum of countably many copies of F, and hence is isomorphic to F_∞. Find now $g_2 \in G_1$ with the property that $A_2 = \bigcap_{c \in F} T_{cg_2}A_1$ has positive measure, and continue in this fashion. After m steps of this iterational procedure, we will have found elements $g_1, g_2, \ldots, g_m$ such that the set $A_m = \bigcap_{c_1, c_2, \ldots, c_m \in F} T_{c_1g_1 + c_2g_2 + \ldots + c_mg_m}A$ has positive measure. It follows now from Theorem 4.2.3 that $\overline{d}_{F_\infty}\big(\bigcap_{c_1, c_2, \ldots, c_m \in F} E - (c_1g_1 + \ldots + c_mg_m)\big) > 0$, and this clearly implies that E contains (many) affine m-dimensional subspaces.

Let us now comment briefly on the proof of Theorem 4.2.2. It is not hard to check that Theorem 4.2.2 holds in the two "extremal" cases, namely the case when the action $(T_g)_{g \in F}$ is compact (which, as before, means that for any $f \in L^2(X, \mathcal{B}, \mu)$ the orbit $\{T_g f\}_{g \in F_\infty}$ is precompact), and the weakly mixing case (which can be defined, for example, by postulating the absence of compact functions). Moreover, the proof in each of these two cases is very similar to the analogous case of Theorem 4.1.2. Perhaps a few remarks are in order to clarify the situation with weak mixing for actions of $(T_g)_{g \in F_\infty}$. First, one can check that, like in the case of $\mathbb{Z}$-actions, weak mixing for F_∞-actions can be characterized in a variety of ways, all parallel to those occurring in the formulation of Theorem 4.1.4. (This remark actually applies to $-$ properly defined $-$ weak mixing actions of any countable or even locally compact group. See for example [BeR1], [Be6], [BeG].)

Second, one can check that an analogue of Theorem 4.1.6 also holds for more general groups (here the right generality is that of amenable groups; see more discussion in the next section.)

Now, one can define, in complete analogy to Definitions 4.1.11 and 4.1.15 the notions of relative weak mixing and relative compactness. The analogues of Theorems 4.1.12, 4.1.14, and 4.1.16 can also be established in a more or less similar fashion. So to finish the proof, one has to show that the multiple recurrence property lifts to compact extensions. As the perspicacious reader has probably guessed by now, one can use here the natural analogue of van der Waerden's theorem, namely Theorem 1.27.

Let us discuss now the multidimensional Szemerédi theorem or, rather, its measure-theoretical twin, Theorem 1.24. Here is the version which actually was proved by Furstenberg and Katznelson in [FuK1].

Theorem 4.2.4. *For any commuting measure preserving transformations $T_1, T_2, \ldots, T_k$ of a probability space $(X, \mathcal{B}, \mu)$ and for any $A \in \mathcal{B}$ with $\mu(A) > 0$ one has*

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

The main new difficulty which one faces when dealing with k general commuting transformations is that they generate a $\mathbb{Z}^k$-action, which may have different dynamical properties along the sub-actions of different subgroups. In other words, while Theorem 4.1.2 was about the joint behavior of k commuting transformations of a special form, namely $T, T^2, \ldots, T^k$, in Theorem 4.2.4 we have to study k commuting transformations which are in, so to say, general position. This complicates the underlying structure theory, which has to be "tuned up" to reflect the more complicated situation when different operators in the group generated by $T_1, \dots, T_k$ have different dynamical properties. What saves the day is Theorem 4.2.7 below, which is at the core of Furstenberg and Katznelson's proof of Theorem 4.2.4.

We need first to introduce some pertinent definitions. While these definitions make sense for any measure preserving group actions, (and are given below a general formulation for future reference), the reader should remember that in the discussion of the proof of Theorem 4.2.2, the group G which occurs in the next two definitions is meant to stand for $\mathbb{Z}^k$ (and hence the subgroups of G are themselves isomorphic to $\mathbb{Z}^l$ for some $0 \le l \le k$.)

Definition 4.2.5. (Cf. Def. 6.3 in [Fu3].) *An extension* $\big(X, \mathcal{B}, \mu, (T_g)_{g\in G}\big)$ *of* $\big(Y, \mathcal{D}, \nu, (S_g)_{g\in G}\big)$ *is a weakly mixing extension if for every* $g_0 \in G$, $g_0 \neq e$, *the system* $(X, \mathcal{B}, \mu, T_{g_0})$ *is a weakly mixing extension of* $(Y, \mathcal{D}, \nu, S_{g_0})$ *(in the sense of Definition 4.1.11).*

Definition 4.2.6. (Cf. Def. 6.5 in [Fu3].) *Assume that* $\big(X, \mathcal{B}, \mu, (T_g)_{g\in G}\big)$ *is an extension of* $\big(Y, \mathcal{D}, \nu, (S_g)_{g\in G}\big)$. *This extension is called primitive if* G *is a direct product of two subgroups,* $G_c \times G_{wm}$, *so that* $\big(X, \mathcal{B}, \mu, (T_g)_{g\in G_c}\big)$ *is a compact extension of* $\big(Y, \mathcal{D}, \nu, (S_g)_{g\in G_c}\big)$ *and* $\big(X, \mathcal{B}, \mu, (T_g)_{g\in G_{wm}}\big)$ *is a weakly mixing extension of* $\big(Y, \mathcal{D}, \nu, (S_g)_{g\in G_{wm}}\big)$.

Remark. We did not explicitly define the notion of compact extension for this more general situation because it is verbatim the same as Definition 4.1.15. (One just has to replace "for every $n \in \mathbb{Z}$" by "for every $g \in G$".) This should be juxtaposed with Definition 4.2.5 which, while still coinciding with Definition 4.1.11 when $G = \mathbb{Z}$, has the emphasis not on the weak mixing behavior of the group action $(T_g)_{g\in G}$, but on the behavior of $\mathbb{Z}$-actions generated by elements $g \in G$, $g \neq e$.

We are now ready to formulate the theorem which provides the main ingredient in the pertinent structure theory. The reader should keep in mind that in the theorem below G stands for $\mathbb{Z}^k$.

Theorem 4.2.7. *If* $\boldsymbol{X} = \big(X, \mathcal{B}, \mu, (T_g)_{g\in G}\big)$ *is an extension of* $\boldsymbol{Y} = \big(Y, \mathcal{D}, \nu, (S_g)_{g\in G}\big)$, *then there is an intermediate factor* $\boldsymbol{Z}$ *such that* $\boldsymbol{Z}$ *is a primitive extension of* $\boldsymbol{Y}$.

As was the case with Theorem 4.1.2, one can show that there is always a maximal factor for which Theorem 4.2.2 is valid. So, in view of Theorem 4.2.7, it remains only to make sure that the multiple recurrence property in question lifts to primitive extensions. This can be achieved by an argument which puts together the ideas behind the proofs of Theorems 4.1.14 and 4.1.17. The fact that primitive extensions utilize the appropriate splitting of $\mathbb{Z}^k$ plays a crucial role. In dealing with the compact part of this splitting, one uses this time the multidimensional van der Waerden theorem. For full details, see [FuK1] and [Fu3], Ch. 7.

Note that the coloring theorems used in the proofs of Theorems 4.1.2, 4.2.2, and 4.2.4 are all corollaries of the IP van der Waerden theorem (which was discussed in detail in Section 2.) This hints that there exists perhaps a more general theorem which bears the same relation to the IP van der Waerden theorem (Theorem 2.2 above) as, say, Theorem 4.1.1 to the (one-dimensional) van der Waerden theorem, and has Theorems 4.1.2, 4.2.2, and 4.2.4 as corollaries. Such a result, which is called the ergodic IP Szemerédi theorem, was established by Furstenberg and Katznelson in [FuK2] and will be briefly discussed below. But before turning our attention to the Furstenberg-Katznelson IP Szemerédi theorem, we want to discuss the polynomial extension of Szemerédi's theorem obtained in [BeL1]. While the paper [BeL1], which appeared in 1996, is more recent than the 1985 paper [FuK2], the structure theory which is utilized in [BeL1] is the same as that needed for the proof of Theorem 4.2.4, whereas in [FuK2] the authors deal with IP systems and develop the nontrivial and complicated IP version of structure theory.

Here then is the formulation of the polynomial Szemerédi theorem.

Theorem 4.2.8. ([BeL1], Thm. B′) *Let $r, l \in \mathbb{N}$ and let $P : \mathbb{Z}^r \to \mathbb{Z}^l$ be a polynomial mapping satisfying $p(0) = 0$. For any $S \subseteq \mathbb{Z}^l$ with $d^*(S) > 0$ and any finite set $F \subset \mathbb{Z}^r$, there is $n \in \mathbb{N}$ and $u \in \mathbb{Z}^l$ such that $u + P(nF) \subset S$.*

In order to formulate an ergodic result which would imply Theorem 4.2.8, let us first reformulate Theorem 4.2.8 in coordinate form.

Theorem 4.2.9. ([BeL1], Thm. B) *For $l \in \mathbb{N}$, let $S \subseteq \mathbb{Z}^l$ satisfy $d^*(E) > 0$. 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 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 $n \in \mathbb{N}$ and $v \in \mathbb{Z}^l$ such that $n + \sum_{j=1}^{t} p_{i,j}(n) v_j \in S$ for each $i \in \{1, 2, \dots, k\}$.*

To see that Theorem 4.2.8 implies Theorem 4.2.9, take $k = r$ and apply Theorem 4.2.8 to the polynomial mapping $P : \mathbb{Z}^r \to \mathbb{Z}^l$ defined by

$$P(n_1, n_2, \ldots, n_r) = \sum_{j=1}^{t} \sum_{i=1}^{r} p_{i,j}(n_i) v_j$$

and the finite set $F = \{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)\} \subset \mathbb{Z}^r$.

To see that Theorem 4.2.8 follows from Theorem 4.2.9, let $P : \mathbb{Z}^r \to \mathbb{Z}^l$ be a polynomial mapping satisfying $P(0) = 0$ and let $F = \{w_1, \ldots, w_k\}$ be an arbitrary finite set in $\mathbb{Z}^r$. Letting $t = l$ in Theorem 4.2.9, define polynomials $p_{i,j}(n)$ by

$$p_{i,j}(n) = P(nw_i)_j,\ n \in \mathbb{N},\ i = 1, 2, \ldots, k,\ j = 1, 2, \ldots, l.$$

Let $v_1, v_2, \ldots, v_l$ denote the unit vectors from the standard basis in $\mathbb{Z}^l$. Then, by Theorem 4.2.9 one has, for some $n \in \mathbb{N}$ and $u \in \mathbb{Z}^l$,

$$u + P(nw_i) = u + \sum_{j=1}^{t} P(nw_i)_j v_j \in S,\ i = 1, 2, \ldots k,$$

which is the same as $u + P(nF) \subset S$.

We formulate now an ergodic theoretic result which implies (via Furstenberg's correspondence principle) Theorems 4.2.8 and 4.2.9, and which may be viewed as a measure preserving analogue of the topological polynomial van der Waerden theorem, Theorem 2.9 above. (See also Theorem 4.2.26 below.)

Theorem 4.2.10. ([BeL1], Thm. A) *Let, for some $t, k \in \mathbb{N}$, $p_{1,1}(n), \ldots, p_{1,t}(n)$, $p_{2,1}(n), \ldots, p_{2,t}(n), \ldots, p_{k,1}(n), \ldots, p_{k,t}(n)$ be polynomials with rational coefficients taking integer values on the integers and satisfying $p_{i,j}(0) = 0$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, t$. Then, for any probability space $(X, \mathcal{B}, \mu)$, commuting invertible measure preserving transformations $T_1, T_2, \ldots, T_t$ of X and any $A \in \mathcal{B}$ with $\mu(A) > 0$, one has*

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

To get a feeling for how general Theorem 4.2.10 is (though this feeling, on a combinatorial level, should be provided by the formulation of Theorem 4.2.8), let us note that as a special case one has, for example, the following refinement of Theorem 4.2.4.

Theorem 4.2.11. *For any commuting invertible measure preserving transformations $T_1, \dots, T_k$ of a probability space $(X, \mathcal{B}, \mu)$, any polynomials $p_1(n), \dots, p_k(n)$ which have rational coefficients, take integer values on the integers, and satisfy $p_i(0) = 0$, $i = 1, 2, \dots, k$, any any $A \in \mathcal{B}$ with $\mu(A) > 0$, one has*

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

The proof of Theorem 4.2.10 in [BeL1] can be described as a "polynomialization" of the proof of Theorem 4.2.4. To ease the discussion, let us put in Theorem 4.2.4 $T_i = T^i$ and consider the expression

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

Assume first that $f = 1_A$ is a compact function. In this special case the positivity of the liminf above easily follows from the following simple fact.

Lemma 4.2.12. *Suppose that $p_1(n), p_2(n), \dots, p_k(n)$ are polynomials with rational coefficients which take integer values on integers and satisfy $p_i(0) = 0$, $i = 1, 2, \dots, k$. Let T be an isometry of a compact metric space (X, ρ). Then for any $\varepsilon > 0$ the set $\bigcap_{i=1}^{k} \Big\{n : \rho(T^{p_i(n)}x, x) < \varepsilon\Big\}$ is syndetic.*

To prove Lemma 4.2.12, one can, for example, invoke the fact that if T is an isometry then the dynamical system (X, T) is *semisimple*, i.e. is a disjoint union of minimal systems. Now, it is not hard to show that if the topological system (X, T), where T is an isometry, is minimal, then it is topologically isomorphic (conjugate) to a minimal translation on a compact abelian group. The desired result then can be deduced from Weyl's result on polynomial Diophantine approximation. Alternatively, one can observe that Lemma 4.2.12 is a corollary of the polynomial van der Waerden theorem (Theorem 2.9 above).

If $(X, \mathcal{B}, \mu, T)$ is weakly mixing, then the result in question also holds, due to the following refinement of Theorem 4.1.5.

Theorem 4.2.13. ([Be2]) *Let $(X, \mathcal{B}, \mu, T)$ be an invertible weakly mixing system. Assume that the polynomials $p_j(n), j = 1, 2, \dots, k$, take integer values on integers, have degree greater than or equal to one, and satisfy the* *condition* $p_i(n) - p_j(n) \not\equiv \text{const}$*,* $i \neq j$*. Then for any* $f_i \in L^\infty(X, \mathcal{B}, \mu)$ *one has*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{p_1(n)} f_1 T^{p_2(n)} f_2 \ldots T^{p_k(n)} f_k = \int f_1 d\mu \int f_2 d\mu \ldots \int f_k d\mu$$

in L^2*-norm.*

Putting $f_i = 1_A$, $i = 1, 2, \ldots, k$, multiplying by 1_A and integrating gives us

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\Big(A \cap T^{-p_1(n)} A \cap T^{-p_2(n)} A \cap \ldots \cap T^{-p_k(n)} A\Big) = \mu(A)^{k+1}.$$

The proof of Theorem 4.2.13 is achieved by an inductive procedure, based on Theorem 4.1.6, which is sometimes called PET-induction. See [Be2] for details.

The proof of Theorem 4.2.10 in its full generality takes, of course, some more work, but with the help of the polynomial van der Waerden theorem and the appropriately general form of Theorem 4.2.13, one is able to push the statement through the primitive extensions. See [BeL1] for the details.

We pass now to the discussion of the IP Szemerédi theorem which was obtained by Furstenberg and Katznelson in [FuK2]. The reader is encouraged to review the definition of an IP system introduced before the formulation of the IP van der Waerden theorem (Theorem 2.2) and to juxtapose the formulations of Theorem 2.2 and the following statement.

Theorem 4.2.14. (See [FuK2], Thm. A.) *Let* $(X, \mathcal{B}, \mu)$ *be a probability space and* G *an abelian group of measure-preserving transformations of* X*. For any* $k \in \mathbb{N}$*, any IP systems* $\big\{T_\alpha^{(1)}\big\}_{\alpha\in\mathcal{F}}, \big\{T_\alpha^{(2)}\big\}_{\alpha\in\mathcal{F}}, \ldots, \big\{T_\alpha^{(k)}\big\}_{\alpha\in\mathcal{F}}$ *in* G*, and any* $A \in \mathcal{B}$ *with* $\mu(A) > 0$ *there exists* $\alpha \in \mathcal{F}$ *such that*

$$\mu\big(A \cap T_\alpha^{(1)} A \cap T_\alpha^{(2)} A \cap \ldots \cap T_\alpha^{(k)} A\big) > 0.$$

The proof of the IP Szemerédi theorem is achieved via a sophisticated structure theory which could be viewed as an IP variation on the theme of primitive extensions discussed above. Curiously enough, it is not the IP van der Waerden theorem, but the more powerful Hales-Jewett theorem which has to be used when dealing with the IP version of compact extensions. We will give more details on the proof of Theorem 4.2.14 below, but first we want to discuss some of its corollaries.

Note that, since the notion of an IP set is a generalization of a (semi)group, the notion of an IP system of commuting invertible measure preserving transformations generalizes the notion of a measure preserving action of a countable abelian group. It follows that Theorems 4.1.2, 4.2.2, and 4.2.4 are immediate corollaries of Theorem 4.2.14. It also follows, via an appropriate version of Furstenberg's correspondence principle, that, on a combinatorial level, Theorem 4.2.4, the IP Szemerédi theorem, implies the multidimensional version of Szemerédi's theorem (Theorem 1.23 above) as well as Theorem 4.2.1. However, the IP Szemerédi theorem gives more! For example, it follows from it that the sets of configurations, always to be found in sets of positive density in $\mathbb{Z}^k$ or F_∞, are abundant in the sense that the set of parameters of these configurations form IP* sets. (See the discussion at the beginning of Section 2.)

One can derive these IP* versions of combinatorial results from the following corollary of Theorem 4.2.14. The IP and IP* sets in an abelian group are defined in complete (and obvious) analogy to the definitions in Sections 1 and 2 which were geared towards $\mathbb{N}$.

Theorem 4.2.15. *Let $(X, \mathcal{B}, \mu)$ be a probability space, and let G be a countable abelian group. For any k commuting measure preserving actions $\big(T_g^{(1)}\big)_{g\in G}$, $\big(T_g^{(2)}\big)_{g\in G}$, …, $\big(T_g^{(k)}\big)_{g\in G}$ of G on $(X, \mathcal{B}, \mu)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$, the set*

$$\big\{g \in G : \mu\big(A \cap T_g^{(1)}A \cap T_g^{(2)}A \cap \ldots \cap T_g^{(k)}A\big) > 0\big\}$$

is an IP* set in G.

Note that since any IP* set in $\mathbb{N}$ is obviously syndetic, Theorem 4.2.15 implies, for example, the following fact.

Corollary 4.2.16. *For any commuting transformations $T_1, T_2, \ldots, T_k$ of a probability space $(X, \mathcal{B}, \mu)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$, the set $\big\{n \in \mathbb{N} : \mu\big(A \cap T_1^{-n}A \cap T_2^{-n}A \cap \ldots \cap T_k^{-n}A\big) > 0\big\}$ is syndetic.*

Note that the conclusion of Corollary 4.2.16 would follow from Theorem 4.2.4 if one would be able to replace in its formulation the statement involving the regular Cesàro averages:

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

by a similar, but stronger, statement, involving "uniform" averages:

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

It is perhaps instructive to pinpoint the exact place in the proof of Theorem 4.2.4 (or its earlier version, Theorem 4.1.2) that does not work for the uniform Cesaro averages. This analysis will also allow the reader to get a better feeling for why one is forced to use the Hales-Jewett theorem. Careful examination of the proof reveals that it is actually only the case of compact extensions which causes the trouble.

Let us briefly review the main ingredients of the proof of Theorem 4.1.2. First, if the system $(X, \mathcal{B}, \mu, T)$ is compact, we saw that the set $\big\{n \in \mathbb{Z} : \big|\mu\big(A \cap T^{-n}A \cap \ldots \cap T^{-kn}A\big) - \mu(A)\big| < \epsilon\big\}$ is syndetic. Second, as was mentioned in Remark 4.1.7, in the case when $(X, \mathcal{B}, \mu, T)$ is weakly mixing, one has, for any $A \in \mathcal{B}$ with $\mu(A) > 0$,

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu\big(A \cap T^{-n}A \cap \ldots T^{-kn}A\big) = \big(\mu(A)\big)^{k+1},$$

and hence, in this case, the set

$$\big\{n \in \mathbb{Z} : \big|\mu\big(A \cap T^{-n}A \cap \ldots \cap T^{-kn}A\big) - \big(\mu(A)\big)^{k+1}\big| < \epsilon\big\}$$

is syndetic.

One can check that the case of relative weak mixing also works for uniform averages. However, it is the case of relatively compact extensions where the syndeticity property is lost in the passage to the extension. Indeed, in the proof of Theorem 4.1.17 we show that if $\boldsymbol{X} = (X, \mathcal{B}, \mu, T)$ is a relatively compact extension of $\boldsymbol{Y} = (Y, \mathcal{D}, \nu, S)$ and $A \in \mathcal{B}$ with $\mu(A) > 0$, then there is a set $A_1 \in \mathcal{D}$ with $\nu(A_1) > 0$ and a number $M \in \mathbb{N}$ such that for any n which is good for multiple recurrence of A, in $\boldsymbol{Y}$, there is a multiple dn with $d \leq M$, which is good for multiple recurrence of A in $\boldsymbol{X}$. So even if one would know in advance that the set $R_{A_1} = \{n_1, n_2, \ldots\}$ of multiple returns of A_1 is a syndetic set, the set of multiples $R_A = \{d_1 n_1, d_2 n_2, \ldots\}$ while being still of positive upper density (due to the fact that $d_i \leq M$ for all i), is no longer guaranteed to be syndetic, as it is not hard to see on some trivial examples.

A possible solution of this problem is to use a more powerful coloring theorem instead of van der Waerden's. It turns out that the Hales-Jewett theorem (see Theorems 1.26 and 1.28) which, as we saw in Section 2 (see Proposition 2.7) is very close to van der Waerden's, is strong enough to supply the missing link needed to assure that the syndeticity can be pushed through the transfinite induction. This added strength allows one to get the better result:

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

(See [McC], Section 5.2, for a presentation of the syndetic version of Theorem 4.1.2 via the Hales-Jewett theorem.)

We give now more details on the proof of the IP Szemerédi theorem. First, let us introduce, following [Fu3] and [FuK2], some pertinent terminology.

Let us recall that any sequence indexed by the set of nonempty subsets of $\mathbb{N}$ is called an $\mathcal{F}$-sequence. In particular, IP sets and IP systems that we have dealt with in earlier sections are examples of $\mathcal{F}$-sequences. As before, we will be writing, for $\alpha, \beta \in \mathcal{F}$, $\alpha < \beta$ (or $\beta > \alpha$) if $\max \alpha < \min \beta$. Assume that a collection of sets $\alpha_i \in \mathcal{F}$, $i = 1, 2, \ldots$ has the property $\alpha_i < \alpha_{i+1}$ for all $i \in \mathbb{N}$. The set $\mathcal{F}^{(1)} = \Big\{\bigcup_{i\in\beta} \alpha_i : \beta \in \mathcal{F}\Big\}$ is called an IP ring. Observe that $\mathcal{F}$ can be viewed as an IP set in the commutative semigroup $(\mathbb{N}, \cup)$, generated by the singletons $\{i\}$, $i \in \mathbb{N}$. By the same token, the IP ring $\mathcal{F}^{(1)}$ can be viewed as an IP set in $(\mathbb{N}, \cup)$ which is generated by the "atoms" α_i, $i \in \mathbb{N}$, and hence has the same structure as $\mathcal{F}$. More formally, let us define a mapping $\varphi : \mathcal{F} \to \mathcal{F}^{(1)}$ by $\varphi(\beta) = \bigcup_{i\in\beta} \alpha_i$. Clearly, φ is bijective and "structure preserving." It follows that any sequence indexed by the elements of an IP ring may itself be viewed as an $\mathcal{F}$-sequence.

The following is a version of Hindman's theorem, which will be needed below. The reader should have no problem establishing its equivalence to Theorem 1.10.

Theorem 4.2.17. *For any finite partition $\mathcal{F} = \bigcup_{i=1}^{r} C_i$, one of C_i contains an IP ring.*

Definition 4.2.18. *Assume that $(x_\alpha)_{\alpha\in\mathcal{F}}$ is an $\mathcal{F}$-sequence in a topological space X. Let $x \in X$ and let $\mathcal{F}^{(1)}$ be an IP ring. We shall write $\operatorname*{IP-lim}_{\alpha\in\mathcal{F}^{(1)}} x_\alpha = x$ if for any neighborhood U of x there exists $\alpha_0 = \alpha_0(U)$ such that, for any $\alpha \in \mathcal{F}^{(1)}$ with $\alpha > \alpha_0$, one has $x_\alpha \in U$.*

One has now the following IP version of the classical Bolzano-Weierstrass theorem.

Theorem 4.2.19. (cf. [Fu3], Thm. 8.14 and [FuK2], Thm. 1.3.) *If $(x_\alpha)_{\alpha\in\mathcal{F}}$ is an $\mathcal{F}$-sequence in a compact metric space X, then there exist an IP ring $\mathcal{F}^{(1)}$ and $x_0 \in X$ such that the $\mathcal{F}$-sequence $(x_\alpha)_{\alpha\in\mathcal{F}^{(1)}}$ has an IP-limit in X:*

$$\operatorname*{IP-lim}_{\alpha\in\mathcal{F}^{(1)}} x_\alpha = x_0.$$

Sketch of the proof. The proof goes along the lines of the classical "dichotomic" proof of the Bolzano-Weierstrass theorem, in which one replaces the pigeonhole principle by the (much more powerful) Hindman's theorem. For given $\epsilon > 0$, let $(B_i)_{i=1}^r$ be a finite family of open balls of radius $\frac{\epsilon}{2}$ which covers the compact space X. By Theorem 4.2.17 one can extract an IP ring $\mathcal{F}^{(1)}$ so that the $\mathcal{F}$-sequence $(x_\alpha)_{\alpha\in\mathcal{F}^{(1)}}$ has all of its elements within distance less than ϵ of one another. The proof is concluded by the diagonal procedure. $\square$

Remark. The notions and properties of IP convergence are very similar to those of the convergence along an idempotent ultrafilter, which was introduced and discussed in Section 3. One could advance this analogy even further by introducing $\beta\mathcal{F}$, the Stone-Čech compactification of $\mathcal{F}$. We have preferred to stick to IP convergence for two reasons. First, this allows us to follow more closely the work of Furstenberg and Katznelson in [FuK2]. Second, IP-limits seem, at least as of now, to be a more convenient tool for dealing with the polynomial extensions of the IP Szemerédi theorem. (See Theorems 4.2.23 and 4.2.26 below.)

The following result is an IP analogue of Proposition 3.17 above. For the (short) proof see [Fu3], Lemma 8.15, or [FuK2], p. 124.

Theorem 4.2.20. *Let $\{T_\alpha\}_{\alpha\in\mathcal{F}}$ be an IP system of continuous transformations of a metric space X. Assume that, for some $x, y \in X$, $\operatorname*{IP-lim}_{\alpha\in\mathcal{F}} T_\alpha x = y$. Then $\operatorname*{IP-lim}_{\alpha\in\mathcal{F}} T_\alpha y = y$.*

Assume now that $(U_\alpha)_{\alpha\in\mathcal{F}}$ is an IP system generated by commuting unitary operators acting on a separable Hilbert space $\mathcal{H}$. By using the fact that the closed ball $B_R = \{x \in \mathcal{H} : ||x|| \le R\}$ is a compact metrizable space in the weak topology, one has the following result.

Theorem 4.2.21. ([FuK2], Thm. 1.7) *If $\{U_\alpha\}_{\alpha\in\mathcal{F}}$ is an IP system of commuting unitary operators on a Hilbert space $\mathcal{H}$, then there is an IP ring $\mathcal{F}^{(1)}$ such that the IP subsystem $\{U_\alpha\}_{\alpha\in\mathcal{F}^{(1)}}$ converges weakly. Moreover, if one has $\operatorname*{IP-lim}_{\alpha\in\mathcal{F}} U_\alpha = P$ weakly, then P is an orthogonal projection.*

Sketch of the proof. Since, clearly, $||P|| \leq 1$, one needs only to show that $P^2 = P$. But this follows from Theorem 4.2.20.

□

The projection P occurring in the above theorem is an orthogonal projection on the space of *rigid* elements, i.e. elements R, satisfying $U_\alpha f \to f$. Note also that, by a classical exercise, $U_\alpha f \to f$ weakly if and only if $U_\alpha f \to f$ strongly. Assuming that $\operatorname*{IP-lim}_{\alpha\in\mathcal{F}} U_\alpha = P$ weakly, we have now the following decomposition of $\mathcal{H}$:

$$\mathcal{H} = \mathcal{H}_r \oplus \mathcal{H}_m, \text{ where}$$
$$\mathcal{H}_r = \big\{f \in \mathcal{H} : \operatorname*{IP-lim}_{\alpha\in\mathcal{F}} U_\alpha f = f\big\},$$
$$\mathcal{H}_m = \big\{f \in \mathcal{H} : \operatorname*{IP-lim}_{\alpha\in\mathcal{F}} U_\alpha f = 0 \text{ weakly}\big\}.$$

The reader should view this splitting as the IP analogue of the splitting $\mathcal{H} = \mathcal{H}_c \oplus \mathcal{H}_{wm}$ which was utilized in the proof of Theorem 4.1.2. This analogy is the starting point of the long list of facts about IP systems of commuting measure preserving transformations which parallel the familiar results pertaining to the structure theory of measure preserving systems and multiple recurrence. For example, when $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$ and the operators U_α are induced by measure preserving transformations T_α on the probability measure space, the space $\mathcal{H}_r$ of rigid functions can be represented, in complete analogy to Theorem 4.1.6, as $L^2(X, \mathcal{B}_1, \mu)$, where the σ-algebra $\mathcal{B}_1$ consists of sets A for which the indicator function 1_A is rigid. One can go even further and define the notions of relatively rigid and relatively mixing extensions. There is also an IP analogue of the van der Corput trick. (See for example Lemma 5.3 in [FuK2].) To handle relatively rigid (or relatively compact, as they are called in [FuK2]) extensions, one uses the Hales-Jewett theorem. Finally, (and mainly due to the fact that one deals with a finitely generated group of IP systems), one also has an analogue of primitive extensions and a theorem analogous to Theorem 4.2.7. (See Theorem 7.10 in [FuK2].)

While many details of the corresponding results demand much work and have to be worked out with care, it is shown in [FuK2] that all this can be glued together to obtain the proof of the IP Szemerédi theorem.

Being encouraged by the IP Szemerédi theorem, one can ask whether the polynomial Szemerédi theorem (Theorem 4.2.10) also admits an IP version. This question is already not trivial in the case of single recurrence, and we address it in this context first.

For an arbitrary invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, a set $A \in \mathcal{B}$ with $\mu(A) > 0$, and a polynomial $p(n)$ which takes integer values on the integers and satisfies $p(0) = 0$, consider the set

$$R_A = \{n : \mu(A \cap T^{p(n)}A) > 0\}.$$

As we have shown in the course of the proof of Theorem 1.31, one has $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \mu(A \cap T^{p(n)}A) > 0$, which clearly implies that R_A has positive upper density. Moreover, by using a modification of the van der Corput trick (Theorem 1.32) which deals with limits of the form $\lim_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M}^{N-1} x_n$, one can show that $\lim_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M}^{N-1} \mu(A \cap T^{p(n)}A) > 0$, which implies that, in fact, the set R_A is syndetic.

In order to obtain an IP version of Theorem 1.31, which would guarantee that the set R_A is an IP* set, one has to switch from Cesàro limits to IP limits. The following theorem, which is a special case of a more general result proved in [BeFM], not only implies that R_A is indeed an IP* set, but actually shows that for any $\epsilon > 0$, the set of returns with large intersections, $\{n : \mu(A \cap T^{p(n)}A) > \mu(A)^2 - \epsilon\}$ is also IP*.

Theorem 4.2.22. (See [Be3], Thm. 3.11.) *Assume that $p(t) \in \mathbb{Q}[t]$ satisfies $p(\mathbb{Z}) \subseteq \mathbb{Z}$ and $p(0) = 0$. Then for any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ with $\mu(A) > 0$, and any IP set $(n_\alpha)_{\alpha\in\mathcal{F}} \subset \mathbb{N}$, there exists an IP-ring $\mathcal{F}^{(1)} \subset \mathcal{F}$ such that*

$$\underset{\alpha\in\mathcal{F}^{(1)}}{\text{IP-lim}}\, \mu(A \cap T^{p(n_\alpha)}A) \geq \mu(A)^2.$$

The crucial role in the proof of Theorem 4.2.22 is played by the fact (obtained with the help of an IP version of van der Corput's trick) that there is an IP ring such that (denoting by U_T the unitary operator on $L^2(X, \mathcal{B}, \mu)$ which is induced by T) one has

$$\underset{\alpha\in\mathcal{F}^{(1)}}{\text{IP-lim}}\, U_T^{p(n_\alpha)} = P \text{ weakly},$$

where P is an orthogonal projection. In particular, it is the fact that P is an orthogonal projection which enables one to get large intersections along the sequence $\big(p(n_\alpha)\big)_{\alpha\in\mathcal{F}^{(1)}}$. Here is the proof:

$$\begin{aligned}\text{IP-}\lim_{\alpha\in\mathcal{F}^{(1)}}\mu\big(A\cap T^{p(n_\alpha)}A\big) &= \text{IP-}\lim_{\alpha\in\mathcal{F}^{(1)}}\langle U_T^{p(n_\alpha)}1_A,1_A\rangle\\ &= \langle P1_A,1_A\rangle=\langle P1_A,P1_A\rangle\langle 1,1\rangle\geq\langle P1_A,1\rangle^2=\langle 1_A,1\rangle^2=\mu(A)^2.\end{aligned}$$

While the proof of Theorem 4.2.22 is, in many respects, just an IP analogue of the proof of Theorem 1.31 above, there is one important distinction which we want to mention here. As we saw in the proof of Theorem 1.31, the splitting $\mathcal{H}=\mathcal{H}_{rat}\oplus\mathcal{H}_{tot.erg.}$ works *for all* $p(n)\in\mathbb{Z}[n]$. A novel feature encountered in the proof of the IP analogue of Theorem 1.31 is that the splitting of $\mathcal{H}=L^2(X,\mathcal{B},\mu)$ which enables one to distinguish between different kinds of asymptotic behavior of $U_T^{p(n)}$ along an IP set $(n_\alpha)_{\alpha\in\mathcal{F}}$ may depend on the polynomial $p(n)$.

However, a much more important novelty which is encountered when one deals with IP analogues of polynomial recurrence theorems is that one has now a bigger family of functions, namely the IP polynomials which form the IP analogue of the conventional polynomials, and for which the IP versions of familiar theorems make sense. Examples of IP polynomial recurrence results in topological dynamics were given in Section 2 (see for example Theorems 2.9 and 2.12). The results which we are going to formulate now can be characterized as polynomial IP extensions of Theorems 1.31 and 4.2.10, and involve a natural subclass of IP polynomials which can be obtained in the following way.

Let $q(t_1,\ldots,t_k)\in\mathbb{Z}[t_1,\ldots,t_k]$ and let $\big(n_\alpha^{(i)}\big)_{\alpha\in\mathcal{F}}$, $i=1,2,\ldots,k$ be IP sets. Then $q(\alpha)=q\big(n_\alpha^{(1)},n_\alpha^{(2)},\ldots,n_\alpha^{(k)}\big)$ is an example of an IP polynomial. For example, if $\deg q(t_1,\ldots,t_k)=2$, then $q(\alpha)$ will typically look like

$$g(\alpha)=\sum_{i=1}^{s}n_\alpha^{(i)}m_\alpha^{(i)}+\sum_{i=1}^{r}k_\alpha^{(i)}.$$

The following result, obtained in [BeFM], extends Theorem 4.2.22 to the case of several commuting transformations and to the the family of IP polynomials described above.

Theorem 4.2.23. ([BeFM], Cor. 2.1) *Suppose that $(X,\mathcal{B},\mu)$ is a probability space and that $\{T_1,T_2,\ldots,T_t\}$ is a collection of commuting invertible measure* *preserving transformations of X. Suppose that $\big(n_\alpha^{(i)}\big)_{\alpha\in\mathcal{F}} \subset \mathbb{N}$ are IP sets, $i = 1, 2, \ldots, k$, and that $p_j(x_1, \ldots, x_k) \in \mathbb{Z}[x_1, \ldots, x_k]$ satisfy $p_j(0, 0, \ldots, 0) = 0$ for $j = 1, 2, \ldots, t$. Then for any measurable $A \subseteq X$, there exists an IP-ring $\mathcal{F}^{(1)} \subset \mathcal{F}$ such that*

$$\operatorname*{IP-lim}_{\alpha\in\mathcal{F}^{(1)}} \mu\Big(A \cap \prod_{i=1}^{t} T_i^{p_i\left(n_\alpha^{(1)}, n_\alpha^{(2)}, \ldots, n_\alpha^{(k)}\right)} A\Big) \geq \mu(A)^2.$$

Theorem 4.2.23 is obtained in [BeFM] as a corollary of the following general fact about families of unitary operators, which can be viewed as a polynomial variation of Theorem 4.2.21. Note that the IP-ring $\mathcal{F}^{(1)}$ which occurs in the formulation, always exists due to the compactness of the weak topology.

Theorem 4.2.24. ([BeFM], Thm. 1.8) *Suppose that $\mathcal{H}$ is a Hilbert space, $\big(U_i\big)_{i=1}^t$ is a commuting family of unitary operators on $\mathcal{H}$, $\big(p_i(x_1, \ldots, x_k)\big)_{i=1}^t \subset \mathbb{Z}[x_1, \ldots, x_k]$ satisfy $p_i(0, 0, \ldots, 0) = 0$ for $1 \leq i \leq t$, and that $\big(n_\alpha^{(i)}\big)_{\alpha\in\mathcal{F}}$ are IP sets for $1 \leq j \leq t$. Suppose that $\mathcal{F}^{(1)}$ is an IP-ring such that for each $f \in \mathcal{H}$,*

$$\operatorname*{IP-lim}_{\alpha\in\mathcal{F}^{(1)}} \Big(\prod_{i=1}^{t} U_i^{p_i\left(n_\alpha^{(1)}, \ldots, n_\alpha^{(k)}\right)}\Big) f = P_{(p_1,\ldots,p_t)} f$$

exists in the weak topology. Then $P_{(p_1,\ldots,p_t)}$ is an orthogonal projection. Projections of this type commute, that is, if also $\big(q_i(x_1, \ldots, x_k)\big)_{i=1}^t \subset \mathbb{Z}[x_1, \ldots, x_k]$ satisfy $q_i(0, 0, \ldots, 0) = 0$ for $1 \leq i \leq t$, then

$$P_{(p_1,\ldots,p_t)} P_{(q_1,\ldots,q_t)} = P_{(q_1,\ldots,q_t)} P_{(p_1,\ldots,p_t)}.$$

An interesting feature of the proof of Theorem 4.2.23 is the usage of the following extension of Hindman's theorem, due independently to K. Milliken and A. Taylor. (See [M] and [T].)

Theorem 4.2.25. ([M], [T]) *Suppose that $\mathcal{F}^{(1)}$ is an IP-ring, $l, r \in \mathbb{N}$, and*

$$\big\{(\alpha_1, \ldots, \alpha_l) \in \big(\mathcal{F}^{(1)}\big)^l : \alpha_1 < \alpha_2 < \ldots < \alpha_l\big\} = \bigcup_{i=1}^{r} C_i.$$

Then there exists j, $1 \leq j \leq r$, and an IP-ring $\mathcal{F}^{(2)} \subset \mathcal{F}^{(1)}$ such that

$$\big\{(\alpha_1, \ldots, \alpha_l) \in \big(\mathcal{F}^{(2)}\big)^l : \alpha_1 < \alpha_2 < \ldots < \alpha_l\big\} \subset C_j.$$

The next natural step is to (try to) extend Theorem 4.2.23 to a multiple recurrence result. The following theorem obtained (as a corollary of a more general result) in [BeM2], which we will call the IP polynomial Szemerédi theorem, is an IP extension of Theorem 4.2.10. (Cf. Theorem 2.9 above.)

Theorem 4.2.26. ([BeM2], Thm. 0.9) *Suppose we are given r commuting invertible measure preserving transformations $T_1, \dots, T_r$ of a probability space $(X, \mathcal{B}, \mu)$. Let $k, t \in \mathbb{N}$ and suppose that $p_{i,j}(n_1, \dots, n_k) \in \mathbb{Q}[n_1, \dots, n_k]$ satisfy $p_{i,j}(\mathbb{Z}^k) \subseteq \mathbb{Z}$ and $p_{i,j}(0, 0, \dots, 0) = 0$ for $1 \le i \le r$, $1 \le j \le t$. Then for every $A \in \mathcal{B}$ with $\mu(A) > 0$, the set*

$$R_A = \Big\{(n_1, \dots, n_k) \in \mathbb{Z}^k : \mu\Big(\bigcap_{j=1}^{t}\Big(\prod_{i=1}^{r} T_i^{p_{i,j}(n_1,\dots,n_k)}\Big)A\Big) > 0\Big\}$$

is an IP* set in $\mathbb{Z}^k$.

We collect some of the corollaries of Theorem 4.2.26 in the following list.

(i) Already for $k = 1$, Theorem 4.2.26 gives a refinement of the polynomial Szemerédi theorem (Theorem 4.2.10). Indeed, it says that the set

$$\big\{n \in \mathbb{Z} : \mu\big(A \cap T_1^{p_{11}(n)} T_2^{p_{12}(n)} \dots T_r^{p_{1r}(n)} A \cap \dots \cap T_1^{p_{t1}(n)} T_2^{p_{t2}(n)} \dots T_r^{p_{tr}(n)} A\big) > 0\big\}$$

is IP*, hence syndetic, hence of positive lower density.

(ii) Theorem 4.2.26 also enlarges the family of configurations which can always be found in sets of positive upper Banach density in $\mathbb{Z}^n$. For example, using Furstenberg's correspondence principle, one obtains the following fact, in which the reader will recognize the density version of Theorem 2.12.

Theorem 4.2.27. *Let $P : \mathbb{Z}^r \to \mathbb{Z}^l$, $r, l \in \mathbb{N}$, be a polynomial mapping satisfying $P(0) = 0$, and let $F \subset \mathbb{Z}^r$ be a finite set. Then for any set $E \subset \mathbb{Z}^l$ with $d^*(E) > 0$ and any IP sets $\big(n_\alpha^{(i)}\big)_{\alpha \in \mathcal{F}}$, $i = 1, \dots, r$, there exist $u \in \mathbb{Z}^l$ and $\alpha \in \mathcal{F}$ such that*

$$\big\{u + P\big(n_\alpha^{(1)} x_1, n_\alpha^{(2)} x_2, \dots, n_\alpha^{(r)} x_r\big) : (x_1, \dots, x_r) \in F\big\} \subset S.$$

See [BeFM] for additional applications, both to combinatorics and to ergodic theory.

The proof of Theorem 4.2.26 that is given in [BeM2] is quite cumbersome (partly due to the fact that in order to push the statement through the transfinite induction over the factors with "manageable" behavior, one has to formulate and prove an even more general result.) In a way, it is a *polynomialization* of the proof of the IP Szemerédi theorem in [FuK2]. Not being able to go through the details of the proof here, we would like to mention the two combinatorial facts which play a decisive role in the proof. One of them is the Milliken-Taylor theorem, formulated above as Theorem 4.2.25. The other one is the polynomial Hales-Jewett theorem, Theorem 2.11.

We are going to discuss now the density versions (and their ergodic counterparts) of three more partition theorems which we encountered in Sections 1 and 2.

We start with Theorem 1.26, the Hales-Jewett theorem. As we saw in Section 2, some major corollaries of the Hales-Jewett theorem, such as the multidimensional van der Waerden theorem and the so-called geometric Ramsey theorem, Theorem 1.27, follow from the IP van der Waerden theorem. The streamlined measure theoretical extension of the IP van der Waerden theorem, the IP Szemerédi theorem, (Theorem 4.2.14) allows one to get the density versions of these corollaries. The IP Szemerédi theorem is, however, still not general enough to give the density version of the Hales-Jewett theorem. This density version, which we will refer to below as dHJ, was established by Furstenberg and Katznelson in [FuK4]. Here is one of a few equivalent formulations of dHJ.

Theorem 4.2.28. ([FuK4], Thm. E) *There is a function $R(\epsilon, k)$, defined for all $\epsilon > 0$ and $k \in \mathbb{N}$, so that if A is a set with k elements, $W_N(A)$ consists of words in A with length N, and if $N \geq R(\epsilon, k)$, then any subset $S \subset W_N(A)$ with $|S| \geq \epsilon k^N$ contains a combinatorial line.*

In order to formulate the ergodic counterpart of Theorem 4.2.28 which was proved in [FuK4], we shall need the following definition.

Definition 4.2.29. (See [FuK4], Def. 2.7.) *Let $W(k)$ denote the free semigroup over the k-element alphabet $\{1, 2, \ldots, k\}$. Given k sequences $\{T_n^{(1)}\}_{n=1}^{\infty}$, $\{T_n^{(2)}\}_{n=1}^{\infty}$, $\ldots$, $\{T_n^{(k)}\}_{n=1}^{\infty}$ of invertible measure preserving transformations of a probability space $(X, \mathcal{B}, \mu)$, define, for each $w = \big(w(1), w(2), \ldots, w(k)\big) \in W(k)$,*

$$T(w) = T_1^{w(1)} T_2^{w(2)} \ldots T_k^{w(k)}.$$

The family $\big(T(w), w \in W(k)\big)$ is called a $W(k)$-system.

Here is now the ergodic formulation of dHJ.

Theorem 4.2.30. (See [FuK4], Prop. 27.) *Let $\{T(w), w \in W(k)\}$ be a $W(k)$-system of invertible measure preserving transformations of a probability space $(X, \mathcal{B}, \mu)$. For any $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists a combinatorial line $\big(l(t)\big)_{t\in\{1,2,\ldots,k\}}$ in $W(k)$ such that*

$$\mu\Big(T\big(l(1)\big)^{-1}A \cap T\big(l(2)\big)^{-1}A \cap \ldots \cap T\big(l(k)\big)^{-1}A\Big) > 0.$$

The proof of Theorem 4.2.30, while following the general scheme of the other proofs discussed above, is significantly more involved, mainly due to the fact that the transformations forming the $W(k)$-system need not commute. (As a matter of fact, in the case where the $W(k)$-system is formed by commutative transformations, the situation is reduced to the IP Szemerédi theorem.) Despite the absence of commutativity, the proof of Theorem 4.2.30 has a strong IP flavor. In particular, the authors use the IP version of the van der Corput trick, Theorem 4.2.20, and a (noncommutative) version of Theorem 4.2.21. Much more importantly, the authors are using an *infinitary* combinatorial result which is a simultaneous extension of the Hindman, Milliken-Taylor, and Hales-Jewett theorems. This combinatorial fact was also obtained by Carlson. (See [FuK3], [Ca], and [BeBH].)

Before moving on with our discussion, we would like to stress that while Theorem 4.2.30 deals with an action of a free finitely generated semigroup, namely $W(k)$, it is a result about rather special configurations in $W(k)$.

Another multiple recurrence theorem involving a noncommutative group is Leibman's nil-Szemerédi theorem obtained in [Le2], which is a density version of his nil-van der Waerden theorem (Theorem 2.15), and, at the same time, is an extension of Theorem 4.2.10.

Theorem 4.2.31. (Cf. [Le2], Thm. NM) *Let $k, t, r \in \mathbb{N}$. Assume that G is a nilpotent group of measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$. Let $p_{i,j}(n_1, \ldots, n_k) \in \mathbb{Z}[n_1, \ldots, n_k]$ with $p_{i,j}(\mathbb{Z}^k) \subseteq \mathbb{Z}$ and $p_{i,j}(0, 0, \ldots, 0) = 0$, $1 \le i \le r$, $1 \le j \le t$. Then for every $A \in \mathcal{B}$ with $\mu(A) > 0$ and any $T_1, T_2, \ldots T_r \in G$, the set*

$$\Big\{(n_1, \ldots, n_k) \in \mathbb{Z}^k : \mu\Big(\bigcap_{j=1}^{t}\Big(\prod_{i=1}^{r} T_i^{p_{i,j}(n_1,\ldots,n_k)}\Big)A\Big) > 0\Big\}$$

is a syndetic set in $\mathbb{Z}^k$.

In his proof, Leibman builds a nilpotent version of primitive extensions similar to, but more sophisticated (due to the noncommutativity) than that which was introduced in [FuK1]. We will describe it now. Let Y be a measure space and let $\{X_i\}_{i\in I}$ be a system of measure spaces of the form $X_i = Y \times F_i$, $i \in I$; then the measure space $X = Y \times \prod_{i\in I} F_i$ is called *a relatively direct product of X_i, $i \in I$, over Y.*

Definition 4.2.32. (Cf. [Le2], Def. 11.10) *Let G be a finitely generated nilpotent group. An extension $\boldsymbol{X} = (X, \mathcal{B}, \mu, (T_g)_{g\in G})$ of a system $\boldsymbol{Y} = (Y, \mathcal{D}, \nu, (S_g)_{g\in G})$ is* primitive *if X is (isomorphic to) the relatively direct product over Y of a system $\{X_i\}_{i\in I}$ of measure spaces so that*

(i) *the transformations T_g, $g \in G$, on X permute the spaces X_i in the product: for any $g \in G$ and $i \in I$ one has $T_g(X_i) = X_j$ for some $j \in I$;*

(ii) *if $T = T_g$ preserves X_i, i.e. $T_g(X_i) = X_i$, then the action of T on X_i is either compact relative to Y or weak mixing relative to Y.*

Modulo this definition, the structure theorem for measure preserving actions of a finitely generated nilpotent group is the same as the structure theorem for $\mathbb{Z}^k$-actions, Theorem 4.2.7 above.

Theorem 4.2.33. ([Le2], Thm. 11.11) *If G is a finitely generated nilpotent group and $\boldsymbol{X} = \big(X, \mathcal{B}, \mu, (T_g)_{g\in G}\big)$ is an extension of $\boldsymbol{Y} = \big(Y, \mathcal{D}, \nu, (S_g)_{g\in G}\big)$, then there is an intermediate factor $\boldsymbol{Z}$ such that $\boldsymbol{Z}$ is a primitive extension of $\boldsymbol{Y}$.*

It is worth mentioning that the structure similar to that appearing in the case of measure preserving actions of nilpotent groups can already be observed on the unitary level.

Theorem 4.2.34. ([Le3], Thm. N) *Let $\{T_g\}$ be a unitary action of a finitely generated nilpotent group G on a Hilbert space $\mathcal{H}$. Then $\mathcal{H}$ is representable as the direct sum of a system $\{\mathcal{L}_i\}_{i\in I}$ of closed pairwise orthogonal subspaces so that:*

(i) *the operators T_g, $g \in G$, permute the subspaces $\mathcal{L}_i$: for any $g \in G$ and $i \in I$ one has $T_g(\mathcal{L}_i) = \mathcal{L}_j$ for some $j \in I$.*

(ii) *if $T = T_g$ preserves $\mathcal{L}_i$, i.e. if $T(\mathcal{L}_i) = \mathcal{L}_i$, then either T is scalar on $\mathcal{L}_i$ or T is weakly mixing on $\mathcal{L}_i$.*

Another interesting feature of Leibman's proof of Theorem 4.2.31 is that in order to lift the recurrence property in question to relatively compact extensions, a coloring theorem is employed which is close in spirit to Theorem 2.17.

Leibman's nil-Szemerédi theorem naturally leads to the question whether the assumptions can be further relaxed and whether, in particular, an analogue of Theorem 1.24 holds true if the measure preserving transformations $T_1, T_2, \ldots, T_k$ generate a solvable group. Note that any finitely generated solvable group is either of exponential growth or is virtually nilpotent, i.e. contains a nilpotent group of finite index. (See, for example, [Ros].) Since Theorem 4.2.31 easily extends to virtually nilpotent groups, the question boils down to solvable groups of exponential growth. The following result, proved in [BeL5], shows, in a strong way, that for solvable groups of exponential growth the answer to the above question is NO.

Theorem 4.2.35. ([BeL5], Thm. 1.1. (A)) *Assume that G is a finitely generated solvable group of exponential growth. There exist a measure preserving action $(T_g)_{g\in G}$ of G on a probability measure space $(X, \mathcal{B}, \mu)$, elements $g, h \in G$, and a set $A \in \mathcal{B}$ with $\mu(A) > 0$ such that $T_{g^n}A \cap T_{h^n}A = \emptyset$ for all $n \neq 0$.*

We conclude this section by formulating a conjecture about a density version of the polynomial Hales-Jewett theorem which, if true, extends both the partition polynomial Hales-Jewett theorem (Theorem 2.11) and the density version of the "linear" Hales-Jewett theorem (Theorem 4.2.28). For $q, d, N \in \mathbb{N}$, let $\mathcal{M}_{q,d,N}$ be the set of q-tuples of subsets of $\{1, 2, \ldots, N\}^d$:

$$\mathcal{M}_{q,d,N} = \big\{(\alpha_1, \alpha_2, \ldots, \alpha_q) : \alpha_i \subset \{1, 2, \ldots, N\}^d, i = 1, 2, \ldots, q\big\}.$$

Conjecture 4.2.36. *For any $q, d \in \mathbb{N}$ and $\epsilon > 0$, there exists $C = C(q, d, \epsilon)$ such that if $N > C$ and a set $S \subset \mathcal{M}_{q,d,n}$ satisfies $\frac{|S|}{|\mathcal{M}_{q,d,N}|} > \epsilon$ then S contains a "simplex" of the form:*

$$\Big\{(\alpha_1, \alpha_2, \ldots, \alpha_q),\ (\alpha_1 \cup \gamma^d, \alpha_2, \ldots, \alpha_q),\ (\alpha_1, \alpha_2 \cup \gamma^d, \ldots, \alpha_q),\ \ldots,\\ (\alpha_1, \alpha_2, \ldots, \alpha_q \cup \gamma^d)\Big\},$$

where $\gamma \subset \mathbb{N}$ is a nonempty set and $\alpha_i \cap \gamma^d = \emptyset$ for all $i = 1, 2, \ldots, q$.

5 Actions of amenable groups

One of the most striking theorems in mathematics, known as the Hausdorff-Banach-Tarski paradox, (see [Hau] and [BaT]), claims that given any two bounded sets A and B in $\mathbb{R}^n$, $n \geq 3$, each having nonempty interior, one can partition A into finitely many disjoint parts and rearrange them by rigid motions of $\mathbb{R}^n$ to form B. What makes this fact even more striking is that (as was shown by Banach in [Ba]) the analogous result does not take place in $\mathbb{R}$ or $\mathbb{R}^2$.

It was von Neumann who, in his fundamental work [Ne1], showed that the phenomenon of "paradoxicality" is related not so much to the structure of the space $\mathbb{R}^n$, but rather to the group of transformations which is used to rearrange the elements of the partition. In particular, von Neumann introduced and studied in [Ne1] a class of groups which he called "messbar" (measurable) and which do not allow for "paradoxical decompositions." These groups are called nowadays amenable, (see [OrW], p. 137, for the origin of the term amenable) and are known to have connections to many mathematical areas, including probability theory, geometry, theory of dynamical systems and representation theory. As we shall see in this section, countable amenable semigroups provide also a natural framework for Furstenberg's correspondence principle. (See [Gre], [Pier], [Pa] for comprehensive treatment of different aspects of amenability in the general framework of locally compact groups. See also [Wag] for a thorough and accessible discussion of amenability for discrete groups with the stress on connections to the Hausdorff-Banach-Tarski paradox.)

Definition 5.1. *Let G be a discrete semigroup. For $x \in G$ and $A \subset G$, let $x^{-1}A = \{y \in G : xy \in A\}$ and $Ax^{-1} = \{y \in G : yx \in A\}$. A semigroup G is called left-amenable (correspondingly, right-amenable) if there exists a finitely additive probability measure on the power set $\mathcal{P}(G)$ satisfying $\mu(A) = \mu(x^{-1}A)$ (correspondingly, $\mu(A) = \mu(Ax^{-1})$ for all $A \in \mathcal{P}(G)$ and $x \in G$). We say that G is amenable if it is both left- and right-amenable.*

It is easy to see (cf. [Wag], p. 147) that a semigroup G is amenable if and only if there exists an *invariant mean* on the space $B(G)$ of real-valued bounded functions on G, that is, a positive linear functional $L : B(G) \to \mathbb{R}$ satisfying
(i) $L(1_G) = 1$,
(ii) $L(f_g) = L({}_gf) = L(f)$ for all $f \in B(G)$ and $g \in G$, where $f_g(t) := f(tg)$ and ${}_gf(t) := f(gt)$.

The existence of an invariant mean is only one item from a long list of equivalent properties, (see, for example, [Wag], Thm. 10.11), some of which, such as the characterization of amenability given in the next theorem, are far from being obvious and, moreover, are valid for groups (or special classes of semigroups) only. One of the advantages of dealing with groups is that for groups, the notions of left and right amenability coincide. (For an easy proof of this fact see, for example, [HeR], Thm. 17.11.)

We will find the following characterization of amenability for discrete groups, which was established by Følner in [Fø], to be especially useful. (See also [Na] for a simplified proof.)

Theorem 5.2. *A countable group G is amenable if and only if it has a left Følner sequence, namely a sequence of finite sets $F_n \subset G$, $n \in \mathbb{N}$, with $|F_n| \to \infty$ and such that $\frac{|F_n \cap gF_n|}{|F_n|} \to 1$ for all $g \in G$.*

Remark 5.3.
1. A right Følner sequence is defined (in an obvious way) as a sequence of finite sets $F_n \subset G$, $n \in \mathbb{N}$, for which $|F_n| \to \infty$ and $\frac{|F_n \cap F_n g}{|F_n|} \to 1$ for all $g \in G$. While in non-commutative groups not every left Følner sequence is necessarily a right Følner sequence and vice versa, it is not hard to show that the existence of a sequence of either type in a semigroup implies the corresponding one-sided version of amenability. As was mentioned above, if G is a group, this is actually enough to get two-sided amenability. The hard part of Theorem 5.2 is establishing the existence of a Følner sequence.
2. Theorem 5.2 is also valid for semigroups possessing the cancellation law. See [Na] for details.
3. Theorem 5.2 can also be extended to general locally compact groups. See [Gre] for details.

It is not known how to construct Følner sequence in a general amenable group defined, say, by a finite set of generators and relations. On the other hand, in many concrete, especially abelian, situations, one has no problem finding a Følner sequence. For example, it is easy to see that the sets F_n which occurred in the proof of Theorem 2.6, form a Følner sequence in F_∞. The reader should also have no problem verifying that d-dimensional parallelepipeds $\Pi_n = [a_n^{(1)}, b_n^{(1)}] \times [a_n^{(2)}, b_n^{(2)}] \times \ldots \times [a_n^{(d)}, b_n^{(d)}]$, where $\min_{1\le i\le d} |a_i - b_i| \to \infty$ as $n \to \infty$, form a Følner sequence in $\mathbb{Z}^d$. Let us indicate how one can construct a Følner sequence in the cancellative abelian semigroup $(\mathbb{N}, \cdot)$.

Let $(a_n)_{n\in\mathbb{N}}$ be an arbitrary sequence in $\mathbb{N}$ and let

$$F_n = \{a_n p_1^{i_1} p_2^{i_2} \dots p_n^{i_n} : 0 \leq i_j \leq k_{j,n}, j = 1, 2, \dots, n\},$$

where $k_{j,n}$ is a doubly indexed sequence of positive integers such that, for every j, $k_{j,n} \to \infty$ as $n \to \infty$, and $\{p_n\}$ is the sequence of primes taken in arbitrary order. It is not hard to check that $\{F_n\}_{n \in \mathbb{N}}$ is a Følner sequence in $(\mathbb{N}, \cdot)$.

The following theorem summarizes some general facts about amenable (semi)groups which were established already in [Ne1]. (For accessible proofs, see [Gre], Ch. 1 and [Wag], Thm. 10.4.)

Theorem 5.4. (i) *Any abelian semigroup is amenable.*
(ii) *Homomorphic images and subgroups of amenable groups are amenable.*
(iii) *If N is a normal subgroup of an amenable group G, then G/N is amenable.*
(iv) *If N is a normal subgroup of a group G, and if both N and G/N are amenable, then G is amenable.*
(v) *If a group G is a union of a family of amenable subgroups $\{H_\alpha\}_{\alpha \in I}$ so that for any H_α, H_β there exists H_γ with $H_\gamma \supset H_\alpha \cup H_\beta$, then G is amenable.*

It follows that the class of amenable groups is quite rich. In particular, it contains all solvable groups, since they can be obtained (with the help of (iv)) from abelian groups by successive extensions with the help of abelian groups. It follows also that a group is amenable if and only if all of its finitely generated subgroups are amenable. This, in turn, implies that all locally finite groups (i.e. the groups in which every finite subset generates a finite subgroup) are amenable.

On the other hand, the group $F_2 = \langle a, b \rangle$, (the free group on two generators) and hence any group containing it as a subgroup, is not amenable.

To see that F_2 is not amenable, one can argue as follows. Let $F_2 = A^+ \cup A^- \cup B^+ \cup B^- \cup \{e\}$, where e is the unit of F_2 (the "empty" word) and the sets A^+, A^-, B^+, and B^- consist of the reduced words starting with a, a^{-1}, b, and b^{-1} respectively. Assume that μ is a finitely additive probability measure on $\mathcal{P}(F_2)$ satisfying $\mu(A) = \mu(gA)$ for any $A \in \mathcal{P}(F_2)$ and $g \in F_2$. Clearly, $\mu(\{e\}) = 0$. (If $\mu(\{e\}) = c > 0$, then, by translation invariance of μ, for any $n \in \mathbb{N}$, $\mu(\{b^n\}) = c$ and the set $\{e, b, b^2, \dots, b^N\}$, where $N \geq \frac{1}{c}$, would have measure bigger than one.) Assume that $\mu(A^+) = c > 0$. (The same proof will work for any other set of our partition which has positive measure.) Let $A_n = b^n A$, $n \geq 0$. Clearly, the sets A_n are disjoint and, by translation invariance, have the same measure $c > 0$. It follows that the set $\bigcup_{n=0}^{N} A_n$, where, as before, $N \geq \frac{1}{c}$, has measure bigger than one, which gives a contradiction. (Note that the simple argument used here is similar to that utilized in the proof of the abstract version of the Poincaré recurrence theorem, Theorem 1.4.)

It follows now that groups such as $SL(n,\mathbb{Z})$ with $n \geq 2$ or $SO(3,\mathbb{R})$ (with the discrete topology) are not amenable, since one can show that they contain a subgroup isomorphic to F_2. As was observed by von Neumann in [Ne1], it is the latter fact that is behind the Hausdorff-Banach-Tarski paradox. See [Wag] for a reader-friendly explanation of this fact. On the other hand, not every non-amenable group has to contain a subgroup isomorphic to F_2. Moreover, non-amenable groups can even be periodic. (See [Ol1], [Ol2], and [Pa], p. 182.)

As is well known to aficionados, many classical notions and results pertaining to 1-parameter group actions extend naturally to amenable groups. Here is, for example, a version of von Neumann's ergodic theorem for actions of countable amenable groups.

Theorem 5.5. *Let G be a countable amenable group. Assume that $(U_g)_{g\in G}$ is an antirepresentation of G as a group of unitary operators acting on a Hilbert space $\mathcal{H}$ (i.e. $U_{g_1}U_{g_2} = U_{g_2g_1}$ for all $g_1, g_2 \in G$). Let P be the orthogonal projection on the space $\mathcal{H}_{inv} = \{f \in \mathcal{H} : U_g f = f \ \forall g \in G\}$. Then for any left Følner sequence $(F_n)_{n\in\mathbb{N}}$ in G, one has*

$$\lim_{n\to\infty}\Big\|\frac{1}{|F_n|}\sum_{g\in F_n} U_g f - Pf\Big\| = 0.$$

Sketch of the proof. It is not hard to check that, in complete analogy to $\mathbb{Z}$-actions, the orthogonal complement of $\mathcal{H}_{inv}$ in $\mathcal{H}$, which we will denote by $\mathcal{H}_{erg}$, coincides with the space $\overline{\mathrm{Span}\{f - U_g f : f \in \mathcal{H}, g \in G\}}$. So it remains to verify that on $\mathcal{H}_{erg}$, the limit in question is zero. It is enough to check this for elements of the form $f - U_{g_0} f$. We have:

$$\begin{aligned}\Big\|\frac{1}{|F_n|}\sum_{g\in F_n} U_g(f - U_{g_0}f)\Big\| &= \Big\|\frac{1}{|F_n|}\sum_{g\in F_n} U_g f - \frac{1}{|F_n|}\sum_{g\in F_n} U_{g_0 g} f\Big\| \\ &= \Big\|\frac{1}{|F_n|}\sum_{g\in F_n} U_g f - \frac{1}{|F_n|}\sum_{g\in g_0 F_n} U_g f\Big\| \leq \frac{|F_n \triangle g_0 F_n|}{|F_n|}\|f\|.\end{aligned}$$

Since, by the definition of a left Følner sequence, $\frac{|F_n \triangle g_0 F_n|}{|F_n|} \underset{n\to\infty}{\longrightarrow} 0$, we are done.

□

Recall that a measure preserving action $(T_g)_{g\in G}$ of a group G on a probability space $(X, \mathcal{B}, \mu)$ is *ergodic* if any set $A \in \mathcal{B}$ which satisfies $\mu(T_g A \triangle A) = 0$ for all $g \in G$ has either measure zero or measure one. The reader should have no problem in verifying the following corollary of Theorem 5.5.

Theorem 5.6. *Assume that $(T_g)_{g\in G}$ is an ergodic measure preserving action of a countable amenable group G. Then for any (left or right) Følner sequence $(F_n)_{n\in\mathbb{N}}$ of G, and any $A_1, A_2 \in \mathcal{B}$, one has*

$$\frac{1}{|F_n|}\sum_{g\in F_n} \mu(A_1 \cap T_g A_2) \underset{n\to\infty}{\longrightarrow} \mu(A_1)\mu(A_2).$$

Here is another useful result, whose proof can be transferred almost verbatim from the proof of the classical Bogoliouboff-Kryloff theorem. (See, for example, [Wal], Thm. 6.9 and Cor. 6.9.1.)

Theorem 5.7. *Let $(T_g)_{g\in G}$ be an action of an amenable group G by homeomorphisms of a compact metric space X. Then there is a probability measure on the Borel σ-algebra $\mathcal{B}(X)$ such that for any $A \in \mathcal{B}(X)$ and any $g \in G$, one has $\mu(A) = \mu(T_g A)$.*

Remark. Unlike the von Neumann ergodic theorem, the pointwise theorem for actions of amenable groups is a much harder and more delicate result, which was proved in the right generality (that is, for any locally compact amenable group and for functions in L^1) only recently, in a remarkable paper of E. Lindenstrauss, [Li].

We are now going to discuss Ramsey-theoretical aspects of amenable groups.

Given a countable amenable group G and, say, a left Følner sequence $\{F_n\}_{n\in\mathbb{N}}$ in G, one can define the upper density with respect to $\{F_n\}_{n\in\mathbb{N}}$ by $\overline{d}_{\{F_n\}}(E) = \limsup_{n\to\infty} \frac{|E\cap F_n|}{|F_n|}$, $E \subset G$. Note that it immediately follows from the definition of a left Følner sequence that for all $g \in G$ and $E \subset G$, one has $\overline{d}_{\{F_n\}}(gE) = \overline{d}_{\{F_n\}}(E)$. By analogy with some known results about sets of positive density in abelian or nilpotent groups which were discussed in previous sections, one can expect that large sets in G, i.e. sets having positive upper density with respect to some Følner sequence, will contain some nontrivial configurations. The results which we will formulate below support this point of view and lead to a general conjecture, which will be formulated at the end of this section.

We start by formulating and proving a version of Furstenberg's correspondence principle for countable amenable groups.

Theorem 5.8. (See [Be5], Thm. 6.4.17.) *Let G be a countable amenable group and assume that a set $E \subset G$ has positive upper density with respect to some left Følner sequence $\{F_n\}_{n\in\mathbb{N}}$: $\overline{d}_{\{F_n\}}(E) = \limsup_{n\to\infty} \frac{|E\cap F_n|}{|F_n|} > 0$. Then there exists a probability measure preserving system $\big(X, \mathcal{B}, \mu, (T_g)_{g\in G}\big)$ and a set $A \in \mathcal{B}$ with $\mu(A) = \overline{d}_{\{F_n\}}(E)$ such that for any $k \in \mathbb{N}$ and $g_1, \ldots, g_k \in G$, one has*

$$\overline{d}_{\{F_n\}}\big(E \cap g_1^{-1}E \cap \ldots \cap g_k^{-1}E\big) \geq \mu\big(A \cap T_{g_1}^{-1}A \cap \ldots T_{g_k}^{-1}A\big).$$

Proof. We show first that there exists a left-invariant mean L on the space $B(G)$ of bounded real-valued functions on G such that

(i) $L(1_E) = \overline{d}_{\{F_n\}}(E)$
(ii) for any $k \in \mathbb{N}$ and any $g_1, \ldots, g_k \in G$, one has

$$\overline{d}_{\{F_n\}}\big(E \cap g_1^{-1}E \cap \ldots \cap g_k^{-1}E\big) \geq L\big(1_E \cdot 1_{g_1^{-1}E} \cdot \ldots \cdot 1_{g_k^{-1}E}\big).$$

Let $\mathcal{S}$ be the (countable) family of subsets of G of the form $\bigcap_{j=1}^k g_j^{-1}E$, where $k \in \mathbb{N}$ and $g_j \in G$, $j = 1, 2, \ldots, k$. By using the diagonal procedure, we can pass to a subsequence $\big\{F_{n_i}\big\}_{i=1}^\infty$ of our Følner sequence such that for our set E we have $\overline{d}_{\{F_n\}}(E) = \lim_{i\to\infty} \frac{|E\cap F_{n_i}|}{|F_{n_i}|}$ and for any $S \in \mathcal{S}$ the limit $L(S) = \lim_{i\to\infty} \frac{|S\cap F_{n_i}|}{|F_{n_i}|} = \lim_{i\to\infty} \frac{1}{|F_{n_i}|} \sum_{g\in F_{n_i}} 1_S(g)$ exists. Observe that for a typical set $S = \bigcap_{j=1}^k g_j^{-1}E \in \mathcal{S}$ this will give

$$\overline{d}_{\{F_n\}}\Big(\bigcap_{j=1}^k g_j^{-1}E\Big) = \limsup_{n\to\infty} \frac{\Big|\Big(\bigcap_{j=1}^k g_j^{-1}E\Big) \cap F_n\Big|}{|F_n|} \geq \lim_{i\to\infty} \frac{\Big|\Big(\bigcap_{j=1}^k g_j^{-1}E\Big) \cap F_{n_i}\Big|}{|F_{n_i}|}$$

$$= L\Big(\bigcap_{j=1}^k 1_{g_j^{-1}E}\Big).$$

Extending by linearity, we will get a positive linear functional L on the subspace $V \subset B(G)$ of finite linear combinations of characteristic functions of sets in $\mathcal{S}$. Note that it follows from the definition of a left Følner sequence that this functional L on V is left-invariant, i.e. for any $f \in V$ and $g \in G$, one has $L(f) = L({}_g f)$, where as before ${}_g f(t) = f(gt)$.

To extend L from V to $B(G)$, define the Minkowski functional $P(f)$ by $P(f) = \limsup_{i\to\infty} \frac{1}{|F_{n_i}|} \sum_{g\in F_{n_i}} f(g)$. Clearly, for any $f_1, f_2 \in B(G)$, one has $P(f_1 + f_2) \leq P(f_1) + P(f_2)$, and for any non-negative t, $P(tf) = tP(f)$. Note also that, on V, $L(f) = P(f)$. By the Hahn-Banach theorem, there is an extension of L (which we will denote by L as well) to $B(G)$ satisfying $L(f) \leq P(f)$ for all $f \in B(G)$. Clearly, L is a left-invariant mean satisfying conditions (i) and (ii) above.

Finally, in preparation for the next stage of the proof, let us note that L can be naturally extended to the space $B_{\mathbb{C}}(G)$ of complex valued bounded functions. We shall continue to denote this extension by L.

Let now $f(h) = 1_E(h)$ be the characteristic function of E and let $\mathcal{A}$ be the uniformly closed and closed under conjugation functional algebra generated by the function f and all of the functions of the form ${}_g f$, where $g \in G$. Then $\mathcal{A}$ is a separable (G is countable and linear combinations with rational coefficients are dense in $\mathcal{A}$) commutative C*-algebra with respect to the sup norm. By the Gelfand representation theorem, $\mathcal{A}$ is isomorphic to an algebra of the form $C(X)$, where X is a compact metric space. The linear functional L, which we constructed above, induces a positive linear functional $\tilde{L}$ on $C(X)$. By the Riesz representation theorem, there exists a regular measure μ on the Borel σ-algebra $\mathcal{B}$ of X such that for any $\varphi \in \mathcal{A}$,

$$L(\varphi) = \tilde{L}(\tilde{\varphi}) = \int_X \tilde{\varphi}\, d\mu,$$

where $\tilde{\varphi}$ denotes the image of φ in $C(X)$. Notice that since the Gelfand transform, establishing the isomorphism between $\mathcal{A}$ and $C(X)$, preserves the algebraic operations, and since the characteristic functions of sets are the only idempotents in $C(X)$, it follows that the image $\tilde{f}$ of our $f(h) = 1_E(h)$ is the characteristic function of some set $A \subset X$: $\tilde{f}(x) = 1_A(x)$. This gives

$$\overline{d}_{\{F_n\}}(E) = L(1_E) = \tilde{L}(1_A) = \int_X 1_A\, d\mu = \mu(A).$$

Notice also that the translation operators $\varphi(h) \to \varphi(gh)$, $\varphi \in \mathcal{A}$, $g \in G$, form an anti-action of G on $\mathcal{A}$, which induces an anti-action $(T_g)_{g\in G}$ on $C(X)$

defined for any $\varphi \in \mathcal{A}$ by $(T_g)\tilde{\varphi} = \tilde{_g\varphi}$. The transformations T_g, $g \in G$ are C*-isomorphisms of $C(X)$ (since they are induced by C*-isomorphisms $\varphi \to_g \varphi$ of $\mathcal{A}$). Now, it is known that algebra isomorphisms of $C(X)$ are induced by homeomorphisms of X, which we, by a slight abuse of notation, will also be denoting by T_g, $g \in G$. These homeomorphisms $T_g : X \to X$ form an action of G on X and preserve the measure μ. To see this, let $C \in \mathcal{B}$ and let $\varphi \in \mathcal{A}$ be the preimage of 1_C (so that $\tilde{\varphi} = 1_C$). For an arbitrary $g \in G$ we have:

$$\mu(C) = \int_X 1_C(x)d\mu(x) = \tilde{L}(\tilde{\varphi}) = L(\varphi) = L(_g\varphi) = \tilde{L}(\tilde{_g\varphi})$$

$$= \tilde{L}\big(\tilde{\varphi}(T_g x)\big) = \int_X 1_C(T_g x)d\mu(x) = \int 1_{T_g^{-1}C}(x)d\mu(x)$$

$$= \mu\big(T_g^{-1}C\big).$$

Notice also that since $L(\mathbf{1}) = 1$, $\mu(X) = \tilde{L}(1_X) = 1$. It follows that $\big(X, \mathcal{B}, \mu, (T_g)_{g\in G}\big)$ is a probability measure preserving system. (As a bonus, we have that, in this representation, the measure preserving transformations T_g are homeomorphisms of a compact metric space.) We finally have, for $f = 1_E$, $g_0 = e$, and any $g_1, \dots, g_k \in G$:

$$\overline{d}_{\{F_n\}}\Big(\bigcap_{j=0}^{k} g_j^{-1}E\Big) \geq L\Big(\prod_{j=0}^{k} {}_{g_j}f\Big) = \tilde{L}\Big(\prod_{j=0}^{k} \tilde{_{g_j}f}\Big)$$

$$= \tilde{L}\Big(\prod_{j=0}^{k} \big((T_{g_j})f\big)\Big) = \int_X \prod_{j=0}^{k} 1_{T_{g_j}^{-1}} A = \mu\Big(\bigcap_{j=0}^{k} T_{g_j}^{-1}A\Big).$$

We are done. □

Remark 5.9. It is not hard to modify the proof above to make Theorem 5.8 valid for any countable amenable semigroup possessing a left Følner sequence. As a matter of fact, Furstenberg's correspondence principle can be extended to general countable amenable semigroups if, instead of using Følner sequences, one defines a set $E \subset G$ to be large if for some left-invariant mean L on $B(G)$, one has $L(1_E) > 0$. (See [BeM3], Theorem 2.1.) The proof in [BeM3] is different also in that it avoids the usage of the Gelfand transform. See also Remark 6.4.21 in [Be5], which describes an approach to the proof of Theorem 5.8 which does not make use of C*-algebras.

The following useful lemma will be be used repeatedly in the sequel. (Note that while, in view of the pending applications, it is arranged in the amenable set-up, the lemma is actually completely general and has, in principle, very little to do with amenability.)

Lemma 5.10. (Cf. [Be1], Thm. 1.1.) *Let $\{F_n\}_{n\in\mathbb{N}}$ be a (left or right) Følner sequence in a countable amenable semigroup G, let $(X, \mathcal{B}, \mu)$ be a probability space, and let, for every $g \in G$, $A_g \in \mathcal{B}$ with $\mu(A_g) \geq a > 0$. Then there exists a set $S \subset G$ with $d_{\{F_n\}}(S) \geq A$, such that for any finite set $F \subset S$ one has $\mu\Big(\bigcap_{g\in F} A_g\Big) > 0$.*

Proof. For any finite set $F \subset G$, let $A_F = \bigcap_{g\in F} A_g$. Deleting, if needed, a set of measure zero from $\bigcup_{g\in G} A_g$, we may and will assume that if $A_F \neq 0$ then $\mu(A_F) > 0$. Let now

$$f_n(x) = \frac{1}{|F_n|}\sum_{g\in F_n} 1_{A_g}(x).$$

Note that $0 \leq f_n(x) \leq 1$ for all x and that $\int f_n d\mu \geq a > 0$ for all $n \in \mathbb{N}$. Let $f(x) = \limsup_{n\to\infty} f_n(x)$. By Fatou's lemma, we have

$$\int_X f d\mu = \int_X \limsup_{n\to\infty} f d\mu \geq \limsup_{n\to\infty} \int_X f_n d\mu \geq a.$$

Thus $\int_X f d\mu \geq a$ and, since $\mu(X) = 1$, there exists $x_0 \in X$ such that $\limsup_{n\to\infty} f_n(x_0) = f(x_0) \geq a$. It follows that there is a sequence $n_i \to \infty$ such that

$$f_{n_i}(x_0) = \frac{1}{|F_{n_i}|}\sum_{g\in F_{n_i}} 1_{A_g}(x) \underset{i\to\infty}{\longrightarrow} f(x_0) \geq a. \tag{5.1}$$

Let $P = \{g \in G : x_0 \in A_g\}$. It follows from (5.1) that $\overline{d}_{\{F_n\}}(P) \geq a$, and, since $x_0 \in A_g$ for all $g \in P$, we have that $\mu(A_F) > 0$ for every finite nonempty $F \subset P$.

□

Applying Theorem 5.8 (more precisely, the version of Theorem 5.8 for semigroups possessing a Følner sequence), we immediately obtain the following result.

Corollary 5.11. *Let $\{F_n\}_{n\in\mathbb{N}}$ be a left Følner sequence in an amenable semigroup G, and let $E \subset G$ satisfy $\overline{d}_{\{F_n\}}(E) = c > 0$. Then there exists a set $P \subset G$ with $\overline{d}_{\{F_n\}}(P) \geq c$ such that for any $g_1, \ldots, g_k \in P$, $\overline{d}_{\{F_n\}}\Big(\bigcap_{i=1}^k g_i^{-1}E\Big) > 0$.*

In order to formulate another application of Lemma 5.10, we need to introduce first the following definition.

Definition 5.12. *Let G be a countable semigroup. A set $R \subset G$ is called a set of measurable recurrence if for any measure preserving action $(T_g)_{g\in G}$ on a probability space $(X, \mathcal{B}, \mu)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists $g \in R$, $g \neq e$, such that $\mu(A \cap T_g^{-1}A) > 0$.*

Different semigroups have all kinds of peculiar sets of recurrence. For example, it follows from Theorem 1.31 that for any polynomial $p(n) \in \mathbb{Z}[n]$ with $p(0) = 0$, the set $\{p(n) : n \in \mathbb{Z}\}$ is a set of measurable recurrence for $\mathbb{Z}$-actions. Moreover, Theorem 4.2.22 tells us that for any IP set $(n_\alpha)_{\alpha\in\mathcal{F}} \subset \mathbb{N}$, the set $\{p(n_\alpha) : \alpha \in \mathcal{F}\}$ is a set of measurable recurrence. More generally, one can show (see [BeFM]) that, for any $p_1(n), \ldots, p_k(n) \in \mathbb{Z}[n]$ satisfying $p_i(0) = 0$, $i = 1, \ldots, k$, and any IP sets $\big(n_\alpha^{(1)}\big)_{\alpha\in\mathcal{F}}, \ldots, \big(n_\alpha^{(k)}\big)_{\alpha\in\mathcal{F}}$, the set

$$\big\{p_1\big(n_\alpha^{(1)}\big), \ldots, p_k\big(n_\alpha^{(k)}\big) : \alpha \in \mathcal{F}\big\} \subset \mathbb{Z}^k$$

is a set of measurable recurrence. Sets of the form $\{1 + \frac{1}{k} : k \in \mathbb{N}\}$ can be shown to be sets of measurable recurrence for the multiplicative group of positive rationals. This list can be continued indefinitely.

The following theorem shows that, for countable amenable semigroups possessing a Følner sequence, the notion of a set of measurable recurrence coincides with the notion of "density recurrence":

Theorem 5.13. *Let S be an amenable semigroup having a left Følner sequence. Then $R \subset S$ is a set of measurable recurrence if and only if for any left Følner sequence $\{F_n\}_{n\in\mathbb{N}}$ in G and any $E \subset G$ with $\overline{d}_{\{F_n\}}(E) > 0$ there exists $g \in R$, $g \neq e$, such that $E \cap g^{-1}E \neq \emptyset$.*

Proof. In one direction, the claim of the theorem immediately follows from Furstenberg's correspondence principle. So, it remains to show that if for any left Følner sequence $\{F_n\}_{n\in\mathbb{N}}$ and $E \subset G$ with $\overline{d}_{\{F_n\}}(E) > 0$ there exists $g \in R$, $g \neq e$, such that $E \cap g^{-1}E \neq \emptyset$, then R is a set of measurable recurrence. Let $(T_g)_{g\in G}$ be a measure preserving action on a probability space $(X, \mathcal{B}, \mu)$

and let $A \in \mathcal{B}$ with $\mu(A) > 0$. It follows from (the proof of) Lemma 5.10 that we may assume that, if $A \cap T_g^{-1}A \neq \emptyset$, then $\mu(A \cap T_g^{-1}A) > 0$, and that there exist a set $P \subset G$ with $\overline{d}_{\{F_n\}}(P) \geq \mu(A)$, and a point $x \in X$ such that, for any $g \in P$, one has $T_g x \in A$.

By our assumptions, there exists $g \in R$, $g \neq e$, such that $P \cap g^{-1}P \neq \emptyset$. Letting $h \in P \cap g^{-1}P$, we have $h, gh \in P$. It follows that $T_h x \in A$ and $T_{gh}x = T_g(T_h x) \in A$. This implies that, simultaneously, $T_h x \in A$ and $T_h x \in T_g^{-1}A$, which gives $A \cap T_g^{-1}A \neq \emptyset$ and, hence, $\mu\big(A \cap T_g^{-1}A\big) > 0$. We are done.

□

The following version of Theorem 5.13 is valid for any amenable semigroup. (See [BeM3], Thm. 2.2.)

Theorem 5.14. *Suppose that S is a countable left amenable semigroup. Then $R \subset S$ is a set of measurable recurrence if and only if for every left-invariant mean L and every $E \subset S$ with $L(1_E) > 0$, one has $E \cap g^{-1}E \neq \emptyset$ for some $g \in R$, $g \neq e$.*

Remark 5.15. One can show that both Furstenberg's correspondence principle and Theorem 5.13 fail for $\mathbb{R}$-actions. Indeed, it is proved in [BeBB] that

(i) For any $\alpha > 0$, $\{n^\alpha : n \in \mathbb{N}\}$ is a set of measurable recurrence for (continuous) measure preserving $\mathbb{R}$-actions.
(ii) For all but countably many $\alpha > 1$, one can find a measurable set $E \subset R$ such that

$$d(E) = \lim_{t\to\infty} \frac{m(E \cap [0,t])}{t} = \frac{1}{2}$$

and $E \cap (E - n^\alpha) = \emptyset$ for all $n \in \mathbb{N}$.

We shall use now Lemma 5.10 to obtain some new results about *multiplicatively large* sets in $\mathbb{N}$, namely sets E which, for some Følner sequence $\{F_n\}_{n\in\mathbb{N}}$ in $(\mathbb{N}, \cdot)$, satisfy $\overline{d}_{\{F_n\}}(E) > 0$.

We start by remarking that the notions of largeness for sets in $\mathbb{N}$, which are based on additive and multiplicative structures, are different. For example, the set O of odd natural numbers has (additive!) density $\frac{1}{2}$ with respect to any Følner sequence in $(\mathbb{N}, +)$. On the other hand, it is not hard to see that the set O will have zero density along any Følner sequence in $(\mathbb{N}, \cdot)$. Indeed, note that the sets $O_n = O/2^n$, $n \in \mathbb{N}$, are pairwise disjoint and, being multiplicative translates of O, have the same upper density with respect to any Følner sequence in $(\mathbb{N}, \cdot)$. It's not hard to see that these facts would lead to a contradiction if for some $\{F_n\}_{n\in\mathbb{N}}$ in $(\mathbb{N}, \cdot)$ one would have $\overline{d}_{\{F_n\}}(O) > 0$. In the other direction, consider, for example, a Følner sequence $\{a_n F_n\}_{n\in\mathbb{N}}$ in $(\mathbb{N}, \cdot)$, which is defined as follows. Let

$$F_n = \left\{p_1^{i_1} p_2^{i_2} \dots p_n^{i_n}, 0 \le i_j \le n, 1 \le j \le n\right\},$$

where p_i, $i = 1, 2, \dots$ are primes in arbitrary order, and let the integers a_n satisfy $a_n > |F_n|$, $n \in \mathbb{N}$ Let now $S = \bigcup_{n=1}^{\infty} a_n F_n$. It is easy to see that S has zero additive density with respect to any Følner sequence in $(\mathbb{N}, +)$. At the same time, S has multiplicative density one with respect to the Følner sequence $\{a_n F_n\}_{n\in\mathbb{N}}$.

As may be expected by mere analogy with additively large sets, multiplicatively large sets always contain (many) geometric progressions. (This can be derived, for example, with the help of the IP Szemerédi theorem, see Section 4.2.) It turns out, however, that multiplicatively large sets also contain arbitrarily long arithmetic progressions and some other, somewhat unexpected, configurations.

Theorem 5.16. (See [Be7], Thm. 3.2.) *Any multiplicatively large set $E \subseteq \mathbb{N}$ contains arbitrarily long arithmetic progressions.*

Proof. Invoking Furstenberg's correspondence principle, let $\big(X, \mathcal{B}, \mu, (T_n)_{n\in\mathbb{N}}\big)$ be the corresponding measure preserving system (where $(T_n)_{n\in\mathbb{N}}$ is a measure preserving action of $(\mathbb{N}, \cdot)$), and let $A \in \mathcal{B}$ be the set of positive measure corresponding to E. Let $A_n = T_n^{-1} A$. Clearly, $\mu(A) = \mu(A_n)$ for all $n \in \mathbb{N}$. By Lemma 5.10, there exists an additively large set S with the property that for any finite $F \subset S$, one has $\mu\Big(\bigcap_{n\in F} T_n^{-1} A\Big) > 0$. Using Szemerédi's theorem, we get, for arbitrary $k \in \mathbb{N}$, an arithmetic progression $P_k = \{n + id : i = 0, 1, \dots, k-1\} \subset S$ such that

$$\mu\Big(\bigcap_{n\in P_k} T_n^{-1} A\Big) > 0.$$

Applying again Furstenberg's correspondence principle, we see that the set $\bigcap_{n\in P_k} E/n$ is multiplicatively large and, in particular, nonempty. This implies that, for some $n \in \mathbb{N}$, $E \supset mP_k$.

$\square$

By working a little bit harder, one can show that any multiplicatively large set contains *geoarithmetic* progressions, namely configurations of the form $\{bq^j(a+id) : 0 \le i,j \le n\}$. (See [Be7], Thm. 3.11.) The following result describes yet another type of geoarithmetic configurations, which can always be found in multiplicatively large sets. (See [Be7] for more results on, and a discussion of, the combinatorial richness of multiplicatively large sets in $\mathbb{N}$.)

Theorem 5.17. (See [Be7], Thm. 3.15.) *Let $E \subset \mathbb{N}$ be a multiplicatively large set. For any $k \in \mathbb{N}$, there exist $a, b, d \in \mathbb{N}$ such that*

$$\left\{b(a+id)^j : 0 \le i,j \le k\right\} \subset E.$$

We shall address now the question about possible amenable extensions of the multiple recurrence results discussed in Section 4. While it is not clear at all how to even formulate an amenable generalization of the one-dimensional Szemerédi theorem (either ergodic or combinatorial), it is, curiously enough, not too hard to guess what should be an amenable version of the multidimensional Szemerédi theorem.

Conjecture 5.18. *Let G be a countable amenable group with a Følner sequence $\{F_n\}_{n\in\mathbb{N}}$. Let $\left(T_g^{(1)}\right)_{g\in G}, \dots, \left(T_g^{(k)}\right)_{g\in G}$ be k pairwise commuting measure preserving actions of G on a probability measure space $(X, \mathcal{B}, \mu)$. ("Pairwise commuting" means here that for any $1 \le i \ne j \le k$ and any $g, h \in G$, one has $T_g^{(i)}T_h^{(j)} = T_h^{(j)}T_g^{(i)}$.) Then for any $A \in \mathcal{B}$ with $\mu(A) > 0$ one has:*

$$\lim_{n\to\infty} \frac{1}{|F_n|} \sum_{g\in F_n} \mu\Big(A \cap T_g^{(1)}A \cap T_g^{(1)}T_g^{(2)}A \cap \ldots \cap T_g^{(1)}T_g^{(2)}\ldots T_g^{(k)}A\Big) > 0.$$

Remark 5.19. The "triangular" expressions

$$A \cap T_g^{(1)}A \cap T_g^{(1)}T_g^{(2)}A \cap \ldots \cap T_g^{(1)}T_g^{(2)}\ldots T_g^{(k)}A$$

appearing in the formulation above, seem to be the "right" configurations to consider. See the discussion and the counterexamples in [BeH2].

The following theorem lists the known instances of the validity of Conjecture 5.18.

Theorem 5.20. *Conjecture 5.18 holds true in the following situations:*
(i) ([BeMZ]) *For k =2.*
(ii) ([BeR2]) *For general k but under the additional assumption that each of the following actions is ergodic:*

$$\begin{aligned}
&\left(T_g^{(k)}\right)_{g\in G},\\
&\left(T_g^{(k-1)}\otimes T_g^{(k-1)}T_g^{(k)}\right)_{g\in G},\\
&\qquad\qquad\vdots\\
&\left(T_g^{(2)}\otimes T_g^{(2)}T_g^{(3)}\otimes\ldots\otimes T_g^{(2)}T_g^{(3)}\ldots T_g^{(k)}\right)_{g\in G},\\
&\left(T_g^{(1)}\otimes T_g^{(1)}T_g^{(2)}\otimes\ldots\otimes T_g^{(1)}T_g^{(2)}\ldots T_g^{(k)}\right)_{g\in G}.
\end{aligned}$$

(In this case, the limit in question equals $\big(\mu(A)\big)^{k+1}$).

While the case $k = 2$ corresponds to the intersection of three sets only, it allows one to derive some interesting combinatorial corollaries, some of which are brought together in the following theorem.

Theorem 5.21. (i) ([BeMZ], Thm. 6.1) *Suppose G is a countable amenable group and that $E \subset G\times G$ has positive upper density with respect to a left Følner sequence $\{F_n\}_{n\in\mathbb{N}}$ for $G\times G$. Then the set*

$$\Big\{g\in G:\ \textit{there exists } (a,b)\in G\times G \textit{ such that } \{(a,b),(ga,b),(ga,gb)\}\subset E\Big\}$$

is a syndetic set in G.

(ii) ([BeMZ], Cor. 7.2.) *Suppose that G is a countable amenable group, $r\in\mathbb{N}$, $G\times G\times G=\bigcup_{i=1}^r C_i$. Then the set*

$$\Big\{g\in G:\ \textit{there exist } i, 1\le i\le r,\ \textit{and } (a,b,c)\in G\times G\times G \textit{ such that}\\ \{(a,b,c),(ga,b,c),(ga,gb,c),(ga,gb,gc)\}\subset C_i\Big\}$$

is a syndetic set in G.

(iii) ([BeM3], Thm. 3.4.) *Suppose that G is a countable amenable group and that $G=\bigcup_{i=1}^r C_i$ is a finite partition. Let $A=\big\{g\in G : [G:C(g)]<\infty\big\}$, where $C(g)$ is the centralizer of g. If $[G:A]=\infty$, then there exist $x,y\in G$ and i, $1\le i\le r$, with $xy\ne yx$ and such that $\{x,y,xy,yx\}\subset C_i$.*

We conclude this section by fulfilling the promise made in Section 1:

Proof of Theorem 1.6. Let $A = \{a_1, a_2, \ldots\} \subset \Gamma$ be an infinite set. Denoting by $[A]^2$ the set of two-element subsets of A, let us define a finite coloring of $[A]^2$ by assigning to each $\{a_i, a_j\} \in [A]^2$, $i < j$, the coset $(a_i - a_j)\Gamma$. (This coloring is finite since Γ is of finite index in F^*.) We will apply now the infinite version of Ramsey's theorem (see [GraRS], Thm. 5, p.16), which says that for any finite coloring of the set of two-element subsets $[P]^2$ of an infinite set P there exists an infinite subset $P_1 \subset P$ such that the set of two-element subsets of P_1, $[P_1]^2$, is monochromatic. It follows that there is $c \in F^*$ and an infinite set $B = \{a_{n_1}, a_{n_2}, \ldots\} \subset A$ such that each member of $[B]^2$ has the same color, $c\Gamma$. This, in turn, says that $a_{n_i} - a_{n_j} \in c\Gamma$ for all $i < j$. Writing $b_i = c^{-1}a_{n_i}$, we see that Γ itself contains an infinite difference set $\{b_i - b_j\}_{i<j}$.

Using now the amenability of the abelian group F, let μ be a finitely additive translation-invariant probability measure on $\mathcal{P}(F)$. Since there are only finitely many disjoint cosets of Γ in F^* and since, clearly, $\mu(\{0\}) = 0$, one of the cosets, call it $c\Gamma$, has to satisfy $\mu(c\Gamma) > 0$. Let $x \in F^*$ be an arbitrary element and consider the sets $c\Gamma + xb_i$, $i \in \mathbb{N}$. Applying the familiar by now reasoning, we see that for some $i < j$, $\mu\big((c\Gamma + xb_i) \cap (c\Gamma + xb_j)\big) > 0$. This implies that $c\Gamma \cap \big(c\Gamma - x(b_i - b_j)\big) \neq \emptyset$, which, in turn, gives us $x(b_i - b_j) \in c\Gamma - c\Gamma$. Since $b_i - b_j \in \Gamma$, we get $x \in c\Gamma - c\Gamma$, and, since x was arbitrary, it gives us $F = c\Gamma - c\Gamma$, and, after the cancellation, $F = \Gamma - \Gamma$. We are done. $\square$

Remark 5.22. The methods used in the above proof can be applied to more general (and not necessarily commutative) rings. Some of the generalizations are given in [BeS]. By invoking stronger combinatorial theorems, such as the IP-Szemerédi theorem, one can actually show that if F is an infinite field and Γ is a multiplicative subgroup of finite index in F^*, then for any finite set $S \subset F$ there is $\gamma \in \Gamma$ such that $\gamma + S = \{\gamma + x : x \in S\} \subset \Gamma$. This, in turn, implies that there exists a finitely additive translation-invariant probability measure μ on F such that $\mu(\Gamma) = 1$.

6 Issues of convergence

This relatively short section is devoted to the discussion of ergodic theorems which are related to combinatorial and number-theoretic applications of ergodic theory.

Various convergence results and conjectures that we have already encountered in the previous sections typically emerged as a means of establishing various recurrence results. Yet, from a purely ergodic-theoretical point of view, these results are of significant interest on their own. While, in order to obtain combinatorial corollaries, one is perfectly satisfied with establishing the positivity of a liminf of Cesàro averages (see for example Theorem 4.1.2), the ideology and tradition of ergodic theory immediately leads to questions whether the limit of a pertinent Cesàro sum exists in norm or almost everywhere.

These questions usually lead to the development of new strong analytic techniques which, in turn, not only provide deeper knowledge about the structure of dynamical systems, but also enhance our understanding of the mutually perpetuating connections between ergodic theory, combinatorics, and number theory.

Consider, for example, Theorem 1.31. As we have seen in Section 1, a convenient way of showing that, for any measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ with $\mu(A) > 0$, and any polynomial $p(n) \in \mathbb{Z}[n]$ satisfying $p(0) = 0$, one has $\mu\big(A \cap T^{p(n)}A\big) > 0$, is to consider the averages

$$\frac{1}{N}\sum_{n=0}^{N-1} \mu\big(A \cap T^{p(n)}A\big)$$

and to show that the limit of these averages is positive. This implies that the set

$$\big\{n \in \mathbb{N} : \mu(A \cap T^{p(n)}A) > 0\big\} \tag{6.1}$$

has positive upper density, which, in turn, implies that the equation $x - y = p(n)$ has "many" integer solutions (x, y, n) with $x, y \in E$, $n \in \mathbb{N}$. At this point the interests of combinatorial number theory and conventional ergodic theory part. While the Cesàro averages are of little help if one wants to undertake the more refined study of the set (6.1) (see Theorem 4.2.22 and the discussion preceding it), it is the focus of the classical ergodic theory on the equidistribution of orbits, which makes the following question interesting.

Question 6.1. *Given an invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, a polynomial $p(n) \in \mathbb{Z}[n]$ and a function $f \in L^p(X, \mathcal{B}, \mu)$, where $p \geq 1$, is it true that*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f\big(T^{p(n)}x\big) \tag{6.2}$$

exists almost everywhere?

What makes this question especially peculiar is that while the norm convergence of the averages (6.2) is not hard to establish (we did it at the end of Section 1 for $p = 2$, which almost immediately implies the norm convergence in any L^p space for $p \geq 1$), the pointwise convergence is quite a bit harder. It was J. Bourgain who developed in the late eighties a powerful technique which allowed him to answer Question 6.1 in the affirmative first for $p = 2$ ([Bou1]) and soon after for any $p > 1$ ([Bou2]). The case $p = 1$ is still open and is perhaps one of the central open problems in that branch of ergodic theory which deals with almost everywhere convergence.

For an excellent survey of Bourgain's methods and a thorough discussion of various positive and negative results on pointwise ergodic theorems, the reader is referred to [RosW]. See also Appendix B, where A. Quas and M. Wierdl present a reader-friendly simplified proof of Bourgain's theorem on a.e. convergence along the set of squares (i.e. for $p(n) = n^2$) for functions in the L^2 space.

In view of the multiple recurrence results discussed in Section 4, the following question naturally suggests itself.

Question 6.2. *Let $T_1, T_2, \ldots, T_k$ be invertible measure preserving transformations which act on a probability space $(X, \mathcal{B}, \mu)$ and generate a nilpotent group. Is it true that for any polynomials $p_i(n) \in \mathbb{Z}[n]$ and $f_i \in L^\infty(X, \mathcal{B}, \mu)$, $i = 1, 2, \ldots, k$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f_1\big(T_1^{p_1(n)}x\big) f_2\big(T_2^{p_2(n)}x\big) \ldots f_k\big(T_k^{p_k(n)}x\big) \qquad (6.3)$$

exists in the L^2 norm? Almost everywhere?

We are going to describe the status of current knowledge in the following brief comments.

The only known result on almost everywhere convergence for $k > 1$ is due, again, to Bourgain, who showed in [Bou3] that for $k = 2$, $p_1(n) = an$, $p_2(n) = bn$, $a, b \in \mathbb{Z}$, the limit in (6.3) exists a.e. for any $f_1, f_2 \in L^\infty$. (It is not hard to show that this implies also the a.e. result for any $f_1, f_2 \in L^2$.)

Assume now that $T_1 = T_2 = \ldots = T_k = T$. As was already mentioned in Section 4, the convergence of the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) f_2(T^{2n} x) \ldots f_k(T^{kn} x) \qquad (6.4)$$

in L^2 norm was only recently established by Host and Kra ([HoK2]) and, independently, Ziegler ([Zie]). Appendix A, written by A. Leibman, gives, among other things, a glimpse into the structure of the proofs contained in the impressive papers [HoK2] and [Zie]. (While the main result proved in Appendix A is somewhat special and deals with the so-called *characteristic factors* for the averages $\frac{1}{N}\sum_{n=0}^{N-1} f_1\big(T^{a_1 n}x\big)f_2\big(T^{a_2 n}x\big)\dots f_k\big(T^{a_k n}x\big)$, the apparatus and techniques utilized there should provide the reader with a better understanding of the methods involved in the study of the averages (6.4).)

The following result, obtained very recently by A. Leibman ([Le5]), shows that the Host-Kra and Ziegler theorems can be extended to polynomial expressions.

Theorem 6.3. ([Le5]) *For $T_1 = T_2 = \dots = T_k = T$, the averages (6.3) converge in L^2.*

It was shown in [FuW2] that if T is totally ergodic (i.e. T^n is ergodic for any $n \neq 0$), then for any $f, g \in L^\infty(X, \mathcal{B}, \mu)$, one has

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f_1\big(T^n x\big) f_2\big(T^{n^2}x\big) = \int f_1 d\mu \int f_2 d\mu$$

in the L^2 norm.

The following theorem, proved in [FrK], gives a nice generalization of this fact.

Theorem 6.4. ([FrK]) *Assume that $(X, \mathcal{B}, \mu, T)$ is an invertible totally ergodic system. Then for any rationally independent polynomials $p_1(n), p_2(n), \dots, p_k(n) \in \mathbb{Z}[n]$ and any $f_i \in L^\infty(X, \mathcal{B}, \mu)$, $i = 1, 2, \dots, k$, one has*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f_1\big(T^{p_1(n)}x\big) f_2\big(T^{p_2(n)}x\big)\dots f_k\big(T^{p_k(n)}x\big) = \int f_1 d\mu \int f_2 d\mu \dots \int f_k d\mu.$$

The L^2-convergence of the averages $\frac{1}{N}\sum_{n=0}^{N-1} f_1(T_1^n x) f_2(T_2^n x)$ for commuting T_1, T_2 (where $f_1, f_2 \in L^\infty(X, \mathcal{B}, \mu)$) was established in [CL1]. The following result, obtained in [BeL3], provides a nilpotent extension of this fact.

Theorem 6.5. ([BeL3]) *Let T_1, T_2 be measure preserving transformations of a probability space $(X, \mathcal{B}, \mu)$ generating a nilpotent group. Then for any $f_1, f_2 \in L^\infty(X, \mathcal{B}, \mu)$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f_1(T_1^n x) f_2(T_2^n x) \tag{6.5}$$

exists in the L^2 norm.

Remark 6.6. Similarly to the situation with recurrence (See Theorem 4.2.35), one can show that if T_1, T_2 generate a solvable group of exponential growth, then the averages (6.5) do not always converge. See Theorem 1.1.(B) in [BeL5].

Due to our specific interest in ergodic theorems related to the material surveyed in the previous sections, we have focused here only on rather special (but important) convergence issues. For more information on pointwise convergence, the reader is referred to [Kre2] and [Ga] as well as to the survey [RosW] mentioned above and the article of A. Nevo in this volume.

A Appendix: Host-Kra and Ziegler factors, and convergence of $\frac{1}{N}\sum_{n=1}^{N} T^{a_1 n} f_1 \cdot \ldots \cdot T^{a_k n} f_k$

by **A. Leibman**[1]
The Ohio State University
Columbus, OH 43210, USA
E-mail: *leibman@math.ohio-state.edu*

The *nonconventional*, or *multiple* ergodic averages

$$\frac{1}{N}\sum_{n=1}^{N} T^n f_1 \cdot \ldots \cdot T^{kn} f_k, \tag{A.1}$$

where T is a measure preserving transformation of a probability measure space X and $f_1, \ldots, f_k$ are (bounded) measurable functions on X, were introduced by H. Furstenberg in his ergodic-theoretical proof of Szemerédi's theorem ([Fu2]). In order to prove Szemerédi's theorem, it was sufficent to show that, in the case $f_1 = \ldots = f_k \geq 0, \not\equiv 0$, the liminf of the averages (A.1) is nonzero, and Furstenberg had confined himself to proving this fact. The question whether the limit of the multiple ergodic averages exists in L^1-sense was an open problem for more than twenty years, until it was answered positively by Host and Kra ([HoK2]) and, independently, by Ziegler ([Zie]). The way of solving this problem was suggested in [Fu2]: one has to determine a factor Z of X which is *characteristic* for the averages (A.1), which means that the limiting behavior of (A.1) only depends on the conditional expectation of f_i with respect to Z:

$$\Big\|\frac{1}{N}\sum_{n=1}^{N}\big(T^n f_1 \cdot \ldots \cdot T^{kn} f_k - T^n E(f_1|Z) \cdot \ldots \cdot T^{kn} E(f_k|Z)\big)\Big\|_{L^1(X)} \underset{N\to\infty}{\longrightarrow} 0$$

for any $f_1, \ldots, f_k \in L^\infty(X)$, or equivalently, that $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} T^n f_1 \cdot \ldots \cdot T^{kn} f_k = 0$ whenever one of $E(f_i|Z)$, $i = 1, \ldots, k$, is equal to 0. Once a characteristic factor Z has been found, the problem is restricted to the system (Z,T); one therefore succeeds if he/she manages to show that every system (X,T) possesses a characteristic factor with a relatively simple structure, so that the convergence of averages (A.1) can be easily established

[1]Supported by NSF, grant DMS-0345350.

for it. For example, under the assumption that T is ergodic, one can show that the Kronecker factor K of X is characteristic for the two-term averages $\frac{1}{N}\sum_{n=1}^{N} T^n f_1 \cdot T^{2n} f_2$ (see [Fu2]; see also Section 4.1 of this survey). Since K has a structure of a compact abelian group on which T acts as a translation, it is not hard to see that the averages above converge for $f_1, f_2 \in L^\infty(K)$.

A k-step nilsystem is a pair (N, T) where N is a compact homogeneous space of a k-step nilpotent group G and T is a translation of N defined by an element of G. When G is a nilpotent Lie group, N is called *a k-step nilmanifold*; if G is an inverse limit of nilpotent Lie groups, N is called *a k-step pro-nilmanifold*. After Conze and Lesigne had shown ([CL1], [CL2], [CL3]) that the characteristic factor for the three-term multiple ergodic averages is a two-step nilsystem, it was natural to conjecture that the characteristic factor for the averages (A.1) with arbitrary k is a $(k-1)$-step nilsystem. Host-Kra and Ziegler have confirmed this conjecture by constructing such factors.

Ziegler's factors $Y_{k-1}(X, T)$, $k = 2, 3, \ldots$, are characteristic for the averages of the form

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{a_1 n} f_1 \cdot \ldots \cdot T^{a_k n} f_k \tag{A.2}$$

for any $a_1, \ldots, a_k \in \mathbb{Z}$. Ziegler's construction is a (very complicated) extension of that of Conze-Lesigne: she obtains the factor $Y_k(X, T)$ as a product of $Y_{k-1}(X, T)$ and a compact abelian group H so that T acts as a skew-product transformation on $Y_k(X, T) = Y_{k-1}(X, T) \times H$, $T(y, h) = (Ty, h + \rho(y))$, with ρ satisfying certain conditions that allow one to impose on $Y_k(X, T)$ the structure of a k-step pro-nilmanifold with T being a translation on it. She also shows that $Y_{k-1}(X, T)$ is the minimal factor of X which is characteristic for all averages of the form (A.2), and the maximal factor of X having the structure of a $(k-1)$-step pro-nilmanifold.

Host and Kra used another, very elegant construction. They first describe the characteristic factor for the (numerical) averages of the form

$$\lim_{N_k\to\infty} \frac{1}{N_k} \sum_{n_k=1}^{N_k} \cdots \lim_{N_1\to\infty} \frac{1}{N_1} \sum_{n_1=1}^{N_1} \int_X \prod_{\epsilon_1,\ldots,\epsilon_k \in \{0,1\}} T^{\epsilon_1 n_1 + \ldots + \epsilon_k n_k} f_{\epsilon_1,\ldots,\epsilon_k}. \tag{A.3}$$

(These averages are not introduced in [HoK2] explicitly, but can be clearly observed in the very construction of the Host-Kra factors; see Proposition A.9 below.) While the expression (A.3) looks scary, it is quite natural. (For instance, when $k=2$ it is just $\lim\limits_{N_2\to\infty}\frac{1}{N_2}\sum\limits_{n_2=1}^{N_2}\lim\limits_{N_1\to\infty}\frac{1}{N_1}\sum\limits_{n_1=1}^{N_1}\int_X f_{0,0}\cdot T^{n_1}f_{1,0}\cdot T^{n_2}f_{0,1}\cdot T^{n_1+n_2}f_{1,1}$.) The corresponding characteristic factor, which will be denoted by $Z_{k-1}(X,T)$, can be easily constructed inductively (we will describe this construction below), and Host and Kra prove that, for each k, the factor $Z_{k-1}(X,T)$ possesses a structure of a $(k-1)$-step pro-nilmanifold. The averages (A.3) turn out to be "universal": successive applications of the van der Corput lemma (see [Be2]; see also Theorems 1.32 and 4.1.6 of the main text) allow one to majorize by the averages (A.3), with suitable $k=k(l)$, all averages of the form

$$\lim_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_l(u)}f_l, \tag{A.4}$$

where $\varphi_1,\ldots,\varphi_l$ are linear functions $\mathbb{Z}^d\longrightarrow\mathbb{Z}$ and $\{\Phi_N\}_{N=1}^{\infty}$ is any Følner sequence in $\mathbb{Z}^d$. (The averages (A.3) also majorize the "polynomial" averages of the form $\lim_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{p_1(u)}f_1\cdot\ldots\cdot T^{p_l(u)}f_l$, where $p_1,\ldots,p_l$ are polynomials $\mathbb{Z}^d\longrightarrow\mathbb{Z}$; see [HoK3] and [Le5].) It follows that the factors $Z_k(X,T)$, with $k=k(l)$, are characteristic for the averages (A.4). In particular, it is shown in [HoK2] that $Z_{k-1}(X,T)$ is characteristic for the averages (A.1), and in [HoK3] that $Z_k(X,T)$ is characteristic for the averages of the form (A.2).

In this note we first describe the Host-Kra construction. We then show that the Host-Kra factors associated with a nontrivial power T^l of a transformation T are the same as the factors associated with T itself. (In [HoK3] this was done for the case of a totally ergodic T only; we give a different proof of this fact.) Next, we prove that, actually, for $k\geq 2$ already $Z_{k-1}(X,T)$ is characteristic for the averages (A.4) and, in particular, (A.2). (The existence of the limit (A.4) will now follow from two facts: (i) $(Z_{k-1}(X,T),T)$ is isomorphic to a nil-system on a pro-nilmanifold, and (ii) the averages (A.4) converge for such a nilsystem; for a (quite nontrivial) proof of the first fact see [HoK2], for a proof of the second fact see [Les] and [Le4].) As a corollary, we obtain that the Host-Kra factors $Z_{k-1}(X,T)$ coincide with the corresponding Ziegler factors $Y_{k-1}(X,T)$. Indeed, being a $(k-1)$-step pro-nilmanifold, $Z_{k-1}(X,T)$ is a factor of $Y_{k-1}(X,T)$; on the other hand, since $Y_{k-1}(X,T)$ is the minimal characteristic factor for the averages (A.2), it is a factor of $Z_{k-1}(X,T)$.

We will now set up some terminology and notation. We will assume that the measure spaces we deal with are regular, that is, are metric spaces endowed with a probability Borel measure. (Any separable measure preserving system has a regular model; see, for example, [Fu3], Chapter 5.) Let $\pi\colon X \longrightarrow V$ be a measurable mapping from a measure space $(X, \mathcal{B}, \mu)$ to a measure space $(V, \mathcal{D}, \nu)$. If π is measure preserving, that is, $\mu(\pi^{-1}(A)) = \nu(A)$ for all $A \in \mathcal{D}$, V is called *a factor* of X. (Note that here V is a factor of a space, not of a dynamical system.) We will denote by X_v the fiber $\pi^{-1}(v)$, $v \in V$. Let $f \in L^1(X)$; then $\mu_f(D) = \int_{\pi^{-1}(D)} f\, d\mu$ is a (signed) measure on V absolutely continuous with respect to ν. By the Radon-Nikodym theorem $d\mu_f/d\nu$ is an integrable function on V; it is denoted by $E(f|V)$ and is called the *conditional expectation* of f with respect to V. The fibers X_v, $v \in V$, may be given a structure of measure spaces with probability measures μ_v, $v \in V$, such that $\int_{X_v} f\, d\mu_v = E(f|V)(v)$ for all $f \in L^1(X)$. (See [Fu3], Chapter 5). We will refer to the partition $X = \bigcup_{v\in V} X_v$ as to *the decomposition of* X *with respect to* V.

If $(V, \mathcal{D}, \nu)$ is a factor of $(X, \mathcal{B}, \mu)$, with $\pi\colon X \longrightarrow V$ being the factorization mapping, then $\pi^{-1}(\mathcal{D})$ is a sub-σ-algebra of $\mathcal{B}$, which we will identify with $\mathcal{D}$. Conversely, with any sub-σ-algebra $\mathcal{D}$ of $\mathcal{B}$ a factor V of X is associated; roughly speaking, V is the partition of X induced by $\mathcal{D}$. (One can construct V in the following way. Choose a countable system $\{D_n\}_{n\in\mathbb{N}}$ generating $\mathcal{D}$. (Such a system may not, actually, exist; we then take a countable system that generates the subsets from $\mathcal{D}$ up to measure zero.) For each n, let $D_n^0 = D_n$ and $D_n^1 = X \setminus D_n$. Put $V = \{0,1\}^{\mathbb{N}}$, and for each $v = (e_1, e_2, \ldots) \in V$ put $X_v = \bigcap_{n=1}^{\infty} D_n^{e_n}$. This defines a mapping $X \longrightarrow V$, $X_v \mapsto \{v\}$; a measure on V is inherited from X.)

Let $(V, \mathcal{D}, \nu)$ be a factor of $(X, \mathcal{B}, \mu)$ and $\pi\colon X \longrightarrow V$ be the factorization mapping. *The relative square* $X \times_V X$ is the subspace $\big\{(x_1, x_2) : \pi(x_1) = \pi(x_2)\big\} = \bigcup_{v\in V} X_v \times X_v$ of $X \times X$, with the measure $\mu \times_V \mu = \int_V \mu_v \times \mu_v d\nu$ thereon. A mapping $X \times_V X \longrightarrow V$ is naturally defined by $(x_1, x_2) \mapsto \pi(x_1) (= \pi(x_2))$, and turns V into a factor of $X \times_V X$ with fibers $(X \times_V X)_v = X_v \times X_v$, $v \in V$, so that $\bigcup_{v\in V} X_v \times X_v$ is the decomposition of $X \times_V X$ with respect to V. (Note that, if we start with a σ-subalgebra $\mathcal{D}$ of $\mathcal{B}$, the space representing the corresponding factor V is not defined canonically, and $X \times_V X$ is only defined up to measure zero. Usually, no new underlying space is introduced for V, and V is simply taken to be the non-regular measure space $(X, \mathcal{D}, \mu)$. The relative square $X \times_V X$ is then defined as X^2 with the measure given by $(\mu \times_V \mu)(A \times B) = \int_V \mu_v(A)\mu_v(B)\, d\nu$, $A, B \in \mathcal{B}$. We use the "set-theoretical" approach to make the geometric picture more transparent. This however leads to some delicate problems related to the fact that our constructions are only defined up to measure zero. The reader is referred to [Fu3], Chapter 5 for a detailed treatment of measure-theoretical issues.) To simplify notation, starting from this moment we will not designate measures; for each space appearing below it will be clear from the context what measure is assumed thereon.

Now let T be a measure preserving transformation of X. We will denote by $\mathcal{I}(X,T)$ the σ-algebra of T-invariant measurable subsets of X and by $I(X,T)$ the factor of X associated with $\mathcal{I}(X,T)$. The decomposition $X = \bigcup_{v\in I(X,T)} X_v$ of X with respect to $I(X,T)$ is then the ergodic decomposition of X. To simplify notation, we will write $X \times_T X$ for $X \times_{I(X,T)} X$.

The Host-Kra factors of X with respect to T are constructed in the following way. One puts $X_T^{[0]} = X$, $T^{[0]} = T$, and when $X_T^{[k]}$ and $T^{[k]}$ have been defined for certain k, let $X_T^{[k+1]} = X_T^{[k]} \times_{T^{[k]}} X_T^{[k]}$ and let $T^{[k+1]}$ be the restriction of $T^{[k]} \times T^{[k]}$ on $X_T^{[k+1]}$. For any $k = 0, 1, \ldots$, $X_T^{[k]}$ is a measurable subspace of X^{2^k}; let $\mathcal{Z}_k(X,T)$ be the minimal σ-algebra on X such that $\mathcal{I}(X_T^{[k]}, T^{[k]}) \subseteq \mathcal{Z}_k(X,T)^{\otimes 2^k}$. *The k-th Host-Kra factor $Z_k(X,T)$ of X with respect to T* is the factor of X associated with $\mathcal{Z}_k(X,T)$.

Assume that $(V,\mathcal{D})$ is a factor of X such that the fibers X_v, $v \in V$, are T-invariant, $T(X_v) = X_v$. Since "life is independent" in distinct fibers X_v, we have:

Lemma A.1. *For any k, the spaces $X_T^{[k]}$ and $I(X_T^{[k]}, T^{[k]})$ decompose with respect to V to, respectively, $\bigcup_{v\in V} (X_v)_T^{[k]}$ and $\bigcup_{v\in V} I((X_v)_T^{[k]}, T^{[k]})$.*

Proof. Let $X = \bigcup_{\alpha\in I(X,T)} X_\alpha$ be the decomposition of X with respect to $I(X,T)$, that is, the ergodic decomposition of X. Elements of the σ-algebra $\mathcal{D} \subseteq \mathcal{B}$ are preserved by T, thus $\mathcal{D} \subseteq \mathcal{I}(X,T)$, and V is a factor of $I(X,T)$. Let $I(X,T) = \bigcup_{v\in V} I_v$ be the decomposition of $I(X,T)$ with respect to V. For (almost) every $v \in V$ the decomposition $X_v = \bigcup_{\alpha\in I_v} X_\alpha$ is the ergodic decomposition of X_v, and thus $I(X_v,T) = I_v$. Hence, $\bigcup_{v\in V} I(X_v,T)$ is the decomposition of $I(X,T)$ with respect to V. Now,

$$X_T^{[1]} = \bigcup_{\alpha\in I(X,T)} X_\alpha \times X_\alpha = \bigcup_{v\in V} \bigcup_{\alpha\in I_{X_v,T}} X_\alpha \times X_\alpha = \bigcup_{v\in V} (X_v)_T^{[1]}.$$

(To be accurate, we also have to check that the measure on $X_T^{[1]}$ agrees with this decomposition. It does:

$$\mu \times_{I(X,T)} \mu = \int_{I(X,T)} \mu_\alpha \times \mu_\alpha d\alpha = \int_V \int_{I_v} \mu_\alpha \times \mu_\alpha d\alpha\, dv = \int_V \mu_v dv.)$$

We then proceed by induction on k. $\square$

Let $\bigcup_{i\in V} X_i$ be a finite measurable partition of X. The finite set V can then be considered as a factor of X (with measure defined by $\nu(\{i\}) = \mu(X_i)$, and the fibers X_i having measures $\mu_i = \mu/\mu(X_i)$, $i \in V$). For this case, Lemma A.1 says that $X_T^{[k]}$ and $I(X_T^{[k]}, T^{[k]})$ partition to, respectively, $\bigcup_{i\in V}(X_i)_T^{[k]}$ and $\bigcup_{i\in V} I((X_i)_T^{[k]}, T^{[k]})$. It follows that $\mathcal{I}(X_T^{[k]}, T^{[k]}) = \prod_{i\in V} \mathcal{I}((X_i)_T^{[k]}, T^{[k]})$, and that $\mathcal{Z}_k(X,T) = \prod_{i\in V} \mathcal{Z}_k(X_i,T)$ and $Z_k(X,T) = \bigcup_{i\in V} Z_k(X_i,T)$.

Our first goal is to investigate the Host-Kra factors associated with T^l, $l \neq 0$.

Theorem A.2. *For any $l \neq 0$ and $k \geq 1$ the k-th Host-Kra factor $Z_k(X,T^l)$ of X with respect to T^l coincides with the k-th Host-Kra factor $Z_k(X,T)$ of X with respect to T.*

Proof. We fix a nonzero integer l. It follows from Lemma A.1 that it suffices to prove Theorem A.2 for an ergodic T only. We first assume that T^l is also ergodic. Given a measure preserving transformation S of a measure space Y, let us denote by $\mathcal{E}_\lambda(Y,S)$ the eigenspace of S in $L^1(Y)$ corresponding to the eigenvalue λ, $\mathcal{E}_\lambda(Y,S) = \big\{f \in L^1(Y) : Sf = \lambda f\big\}$. In particular, $\mathcal{E}_1(Y,S)$ is the space of S-invariant integrable functions on Y, which we will denote by $\mathcal{L}(Y,S)$.

Lemma A.3. (Cf. [HoK3]) *Let S be a measure preserving transformation of a measure space Y. If S^l is ergodic, then $I(Y \times Y, S^l \times S^l) = I(Y \times Y, S \times S)$.*

Proof. S^l is ergodic means that $\mathcal{E}_\lambda(Y,S) = \{0\}$ for all $\lambda \neq 1$ with $\lambda^l = 1$. We have $\mathcal{L}(Y \times Y, (S \times S)^l) \subseteq \operatorname{span}\big\{\mathcal{E}_\lambda(Y \times Y, S \times S) : \lambda^l = 1\big\}$. For any $\lambda \in \mathbb{C}$, $|\lambda| = 1$, the space $\mathcal{E}_\lambda(Y \times Y, S \times S)$ is spanned by the the functions of the form $f \otimes g$ where $f \in \mathcal{E}_{\lambda_1}(Y,S)$ and $g \in \mathcal{E}_{\lambda_2}(Y,S)$ with $\lambda_1\lambda_2 = \lambda$. For such a function, $fg \in \mathcal{E}_\lambda(Y,S)$. If $\lambda \neq 1$ and $\lambda^l = 1$, we have $fg = 0$; since S is ergodic, $|f| = \text{const}$ and $|g| = \text{const}$, so either $f = 0$ or $g = 0$. Thus, for any $\lambda \neq 1$ with $\lambda^l = 1$ we have $\mathcal{E}_\lambda(Y \times Y, S \times S) = \{0\}$.

Hence, $\mathcal{L}(Y\times Y,S^l\times S^l)\subseteq\mathcal{E}_1(Y\times Y,S\times S)=\mathcal{L}(Y\times Y,S\times S)$. With the evident opposite inclusion $\mathcal{L}(Y\times Y,S\times S)\subseteq\mathcal{L}(Y\times Y,S^l\times S^l)$ this implies $\mathcal{I}(Y\times Y,S^l\times S^l)=\mathcal{I}(Y\times Y,S\times S)$. □

Lemma A.4. (Cf. [HoK3]) *Let T be a measure preserving transformation of a measure space X. If T^l is ergodic then $X^{[k]}_{T^l}=X^{[k]}_T$ and $I\big(X^{[k]}_{T^l},(T^l)^{[k]}\big)=I\big(X^{[k]}_T,T^{[k]}\big)$ for all $k\geq 0$.*

Proof. For $k=0$ the statement is trivial. Assume by induction that, for some $k\geq 0$, $Y=X^{[k]}_{T^l}=X^{[k]}_T$ and $I=I(Y,(T^l)^{[k]})=I(Y,T^{[k]})$. Then $X^{[k+1]}_{T^l}=X^{[k+1]}_T=Y\times_I Y$. Let $Y=\bigcup_{\alpha\in I}Y_\alpha$ be the decomposition of Y with respect to I and for each $\alpha\in I$ let $S_\alpha=T^{[k]}\big|_{Y_\alpha}$. By the induction assumption S^l_α is ergodic on Y_α for every $\alpha\in I$, thus by Lemma A.1 and Lemma A.3 applied to the systems (Y_α,S_α),

$$\begin{aligned}I(Y\times_I Y,(T^l)^{[k]}\times(T^l)^{[k]})&=\textstyle\bigcup_{\alpha\in I}I\big(Y_\alpha\times Y_\alpha,S^l_\alpha\times S^l_\alpha\big)\\&=\textstyle\bigcup_{\alpha\in I}I\big(Y_\alpha\times Y_\alpha,S_\alpha\times S_\alpha\big)=I(Y\times_I Y,T^{[k]}\times T^{[k]}).\end{aligned}$$

□

It follows that $Z_k(X,T^l)=Z_k(X,T)$ for all $k\geq 0$, which proves Theorem A.2 in the case T^l is ergodic.

Now assume that T is ergodic whereas T^l is not. We may assume that l is a prime integer. In this case X is partitioned, up to a subset of measure 0, to measurable subsets $X_0,\ldots,X_{l-1}$ such that $T(X_i)=X_{i+1}$ for all $i\in\mathbb{Z}_l$. (We identify $\{0,\ldots,l-1\}$ with $\mathbb{Z}_l=\mathbb{Z}/(l\mathbb{Z})$ in order to have $(l-1)+1=0$.)

Lemma A.5. *Let X be a disjoint union of measure spaces $X_0,\ldots,X_{l-1}$ and let T be an invertible measure preserving transformation of X such that $T(X_i)=X_{i+1}$, $i\in\mathbb{Z}_l$. Then $X_0,\ldots,X_{l-1}\in\mathcal{Z}_1(X,T)$.*

Proof. We may assume that T is ergodic; otherwise we pass to the ergodic components of X with respect to T. Then $X^{[1]}_T=X^2$ and $T^{[1]}=T\times T$. The “diagonal” $W=X_0^2\cup\ldots\cup X_{l-1}^2\subseteq X^{[1]}_T$ is $T^{[1]}$-invariant and therefore W is $\mathcal{Z}_1(X,T)\otimes\mathcal{Z}_1(X,T)$-measurable. By Fubini's theorem the “fibers” $X_0,\ldots,X_{l-1}$ of W are $\mathcal{Z}_1(X,T)$-measurable. □

Lemma A.6. *Let Y be a disjoint union of measure spaces $Y_0,\ldots,Y_{l-1}$ and let S be an invertible measure preserving transformation of Y such that $S(Y_i)=Y_{i+1}$, $i\in\mathbb{Z}_l$. Then $Y\times_S Y$ is partitioned to $\bigcup_{i,j\in\mathbb{Z}_l}Y_{i,j}$ where $Y_{i,i}=Y_i\times_{S^l}Y_i$*

for all $i \in \mathbb{Z}_l$, and for all $i, j, s, t \in \mathbb{Z}_l$, $(S^s \times S^t)\big|_{Y_{j,j}}$ is an isomorphism between $Y_{i,j}$ and $Y_{i+s,j+t}$. In particular, $(S \times S)(Y_{i,j}) = Y_{i+1,j+1}$ for all i, j, thus the subsets $V_i = \bigcup_{j \in \mathbb{Z}_l} Y_{j,j+i}$, $i \in \mathbb{Z}_l$ are $S \times S$-invariant and partition $Y \times_S Y$, and $\mathrm{Id}_{Y_0} \times S^i$ is an isomorphism between V_0 and V_i.

Proof. We first determine $I(Y, S)$. Let A be a measurable S-invariant subset of Y. Let $A_i = A \cap Y_i$, $i \in \mathbb{Z}_l$. Then A_0 is S^l-invariant, and $A_i = S^i(A_0)$ for $i \in \mathbb{Z}_l$. So, the mapping $A \mapsto A \cap Y_0$ is an isomorphism between $\mathcal{I}(Y, S)$ and $\mathcal{I}(Y_0, S^l)$, which induces an isomorphism between $I(Y, S)$ and $I(Y_0, S^l)$.

Let $Y_0 = \bigcup_{\alpha \in I} Y_{0,\alpha}$ be the decomposition of Y_0 with respect to $I = I(Y_0, S^l)$. For every $\alpha \in I$ and $i \in \mathbb{Z}_l \setminus \{0\}$ define $Y_{i,\alpha} = S^i(Y_{0,\alpha})$ and $Y_\alpha = \bigcup_{i \in \mathbb{Z}_l} Y_{i,\alpha}$. Then $Y = \bigcup_{\alpha \in I} Y_\alpha$ is the decomposition of Y with respect to I. We have

$$Y_S^{[1]} = \bigcup_{\alpha \in I} Y_\alpha \times_S Y_\alpha = \bigcup_{\alpha \in I} \bigcup_{i,j \in \mathbb{Z}_l} Y_{i,\alpha} \times Y_{j,\alpha} = \bigcup_{i,j \in \mathbb{Z}_l} \bigcup_{\alpha \in I} Y_{i,\alpha} \times Y_{j,\alpha} = \bigcup_{i,j \in \mathbb{Z}_l} Y_{i,j},$$

where $Y_{i,j} = \bigcup_{\alpha \in I} Y_{i,\alpha} \times Y_{j,\alpha}$. In particular, $Y_{i,i} = \bigcup_{\alpha \in I} Y_{i,\alpha} \times Y_{i,\alpha} = Y_i \times_{S^l} Y_i$ for all $i \in \mathbb{Z}_l$. $\square$

Lemma A.7. *Let X be a disjoint union of measure spaces $X_0, \ldots, X_{l-1}$ and let T be an invertible measure preserving transformation of X such that $T(X_i) = X_{i+1}$, $i \in \mathbb{Z}_l$. Then for any $k \geq 0$, $X_T^{[k]}$ can be partitioned, $X_T^{[k]} = \bigcup_{j=1}^{l^k} W_j$, into $T^{[k]}$-invariant measurable subsets $W_1, \ldots, W_{l^k}$, such that $W_1 = \bigcup_{i \in \mathbb{Z}_l} (X_i)_{T^l}^{[k]}$ with $T^{[k]}\big((X_i)_{T^l}^{[k]}\big) = (X_{i+1})_{T^l}^{[k]}$ for each i, and for each $j = 2, \ldots, l^k$ there exists an isomorphism $\tau_j \colon W_1 \longrightarrow W_j$, which in each coordinate is given by a power of T (that is, if $\pi_n \colon X^{[k]} \longrightarrow X$, $n = 1, \ldots, 2^k$, are the projection mappings, for each n there exists $m \in \mathbb{Z}$ such that $\pi_n \circ \tau_j = T^m \circ \pi_n\big|_{W_1}$).*

Proof. We use induction on k; for $k = 0$ the statement is trivial. Assume that it holds for some $k \geq 0$. Then by Lemma A.1, $X_T^{[k+1]} = \bigcup_{j=1}^{l^k} W_j \times_{T^{[k]}} W_j$. The isomorphisms τ_j between W_1 and W_j, commuting with $T^{[k]}$, induce isomorphisms $\tau_j \times \tau_j$ between $W_1 \times_{T^{[k]}} W_1$ and $W_j \times_{T^{[k]}} W_j$, $j = 1, \ldots, l^k$, and $\tau_j \times \tau_j$ acts on coordinates as powers of T if τ_j does. Thus, we may focus on $W_1 \times_{T^{[k]}} W_1$ only.

By Lemma A.6 applied to $W_1 = \bigcup_{i \in \mathbb{Z}_l} (X_i)_{T^l}^{[k]}$ and $T^{[k]}\big|_{W_1}$, $W_1 \times_{T^{[k]}} W_1$ is partitioned into $T^{[k]} \times T^{[k]} = T^{[k+1]}$-invariant subsets $V_0, \ldots, V_{l-1}$ such that

$$V_0 = \bigcup_{i \in \mathbb{Z}_l} (X_i)_{T^l}^{[k]} \times_{(T^{[k]})^l} (X_i)_{T^l}^{[k]} = \bigcup_{i \in \mathbb{Z}_l} (X_i)_{T^l}^{[k+1]}$$

and $V_1,\dots,V_{l-1}$ are isomorphic to V_0 by isomorphisms whose projections on the factors $(X_i)^{[k]}_{T^l}$ coincide with some powers of $T^{[k]}$. $\square$

End of the proof of Theorem A.2. Assume that T is ergodic on X, l is a prime integer and T^l is not ergodic on X. Let $k\geq 1$. Ignoring a subset of measure 0 in X, partition X to measurable subsets $X_0,\dots,X_{l-1}$ such that, for each i, $T(X_i)=X_{i+1}$. Let $k\geq 1$ and let $W_1,\dots,W_{l^k}$ be as in Lemma A.7. Since $X_0,\dots,X_{l-1}$ are T^l-invariant, by Lemma A.1 we have $\mathcal{I}(X^{[k]},(T^l)^{[k]})=\prod_{i\in\mathbb{Z}_l}\mathcal{I}(X_i^{[k]},(T^l)^{[k]})$ and $\mathcal{Z}_k(X,T^l)=\prod_{i\in\mathbb{Z}_l}\mathcal{Z}_k(X_i,T^l)$. Any $T^{[k]}$-invariant measurable subset A of $W_1=\bigcup_{i\in\mathbb{Z}_l}(X_i)^{[k]}_{T^l}$ has form $A=\bigcup_{i\in\mathbb{Z}_l}A_i$ where $A_i\in\mathcal{I}(X_i,(T^l)^{[k]})$ and $T^{[k]}(A_i)=A_{i+1}$, $i\in\mathbb{Z}_l$. Thus, $\mathcal{I}(W_1,T^{[k]})\subseteq\mathcal{I}(X^{[k]},(T^l)^{[k]})\subseteq\mathcal{Z}_k(X,T^l)^{\otimes 2^k}$. Since $\mathcal{Z}_k(X,T^l)$ is T-invariant and $W_n=\tau_n(W_1)$ where τ_n is an isomorphism acting on each coordinate as a power of T, $\mathcal{I}(W_n,T^{[k]})\subseteq\mathcal{Z}_k(X,T^l)^{\otimes 2^k}$ for any n. Hence, $\mathcal{Z}_k(X,T)\subseteq\mathcal{Z}_k(X,T^l)$.

We will now show that for any $i\in\mathbb{Z}_l$ and any $B\in\mathcal{I}(X_i^{[k]},(T^l)^{[k]})$ one has $B\in\mathcal{Z}_k(X,T)^{\otimes 2^k}$; this will imply that $\mathcal{Z}_k(X,T^l)\subseteq\mathcal{Z}_k(X,T)$. Put $A_j=(T^{[k]})^{j-i}(B)$, $j\in\mathbb{Z}_l$, and $A=\bigcup_{j\in\mathbb{Z}_l}A_j$. Then $A\in\mathcal{I}(W_1,T^{[k]})\subseteq\mathcal{Z}_k(X,T)^{\otimes 2^k}$. By Lemma A.5, $X_i\in\mathcal{Z}_1(X,T)\subseteq\mathcal{Z}_k(X,T)$, thus $(X_i)^{[k]}_{T^l}\in\mathcal{Z}_k(X,T)^{\otimes 2^k}$, and therefore $B=A_i=A\cap(X_i)^{[k]}_{T^l}\in\mathcal{Z}_k(X,T)^{\otimes 2^k}$.

$\square$

We now pass to our second result:

Theorem A.8. *For any $k\geq 2$, any $d\in\mathbb{N}$, any linear functions $\varphi_1,\dots,\varphi_k\colon\mathbb{Z}^d\longrightarrow\mathbb{Z}$ and any Følner sequence $\{\Phi_N\}_{N=1}^\infty$ in $\mathbb{Z}^d$, $Z_{k-1}(X,T)$ is a characteristic factor for the averages $\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k$ in $L^1(X)$, that is,*

$$\lim_{N\to\infty}\Bigl\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\Bigl(T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k -T^{\varphi_1(u)}E\bigl(f_1|Z_{k-1}(X,T)\bigr)\cdot\ldots\cdot T^{\varphi_k(u)}E\bigl(f_k|Z_{k-1}(X,T)\bigr)\Bigr)\Bigr\|_{L^1(X)}=0 \tag{A.5}$$

for any $f_1,\dots,f_k\in L^\infty(X)$.

In order to prove Theorem A.8 we will first show that $Z_{k-1}(X,T)$ is a characteristic factor for averages of a very special form. Let us bring more facts from [HoK2]. Starting from this moment, we will only be considering real-valued functions on X. Given $f_0, f_1 \in L^\infty(X)$, by the ergodic theorem we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X f_0 \cdot T^n f_1 = \int_{I(X,T)} E\big(f_0|I(X,T)\big) \cdot E\big(f_1|I(X,T)\big) = \int_{X_T^{[1]}} f_0 \otimes f_1.$$

Applying this twice we get, for $f_{0,0}, f_{0,1}, f_{1,0}, f_{1,1} \in L^\infty(X)$,

$$\begin{aligned}
&\lim_{N_2\to\infty} \frac{1}{N_2} \sum_{n_2=1}^{N_2} \lim_{N_1\to\infty} \frac{1}{N_1} \sum_{n_1=1}^{N_2} \int_X f_{0,0} \cdot T^{n_1} f_{1,0} \cdot T^{n_2} f_{0,1} \cdot T^{n_1+n_2} f_{1,1} \\
&\quad = \lim_{N_2\to\infty} \frac{1}{N_2} \sum_{n_2=1}^{N_2} \lim_{N_1\to\infty} \frac{1}{N_1} \sum_{n_1=1}^{N_2} \int_X (f_{0,0} \cdot T^{n_2} f_{0,1}) \cdot T^{n_1} (f_{1,0} \cdot T^{n_2} f_{1,1}) \\
&\quad = \lim_{N_2\to\infty} \frac{1}{N_2} \sum_{n_2=1}^{N_2} \int_{X^{[1]}} (f_{0,0} \otimes f_{1,0}) \cdot T^{n_2} (f_{0,1} \otimes f_{1,1}) \\
&\quad = \int_{X^{[2]}} (f_{0,0} \otimes f_{1,0}) \otimes (f_{0,1} \otimes f_{1,1}).
\end{aligned}$$

By induction, for any k and any collection $f_{\epsilon_1,\ldots,\epsilon_k} \in L^\infty(X)$, $\epsilon_1, \ldots, \epsilon_k \in \{0,1\}$,

$$\begin{aligned}
\lim_{N_k\to\infty} \frac{1}{N_k} \sum_{n_k=1}^{N_k} \cdots \lim_{N_1\to\infty} \frac{1}{N_1} \sum_{n_1=1}^{N_1} \int_X \prod_{\epsilon_1,\ldots,\epsilon_k \in \{0,1\}} & T^{\epsilon_1 n_1 + \ldots + \epsilon_k n_k} f_{\epsilon_1,\ldots,\epsilon_k} \\
&= \int_{X^{[k]}} \bigotimes_{\epsilon_1,\ldots,\epsilon_k \in \{0,1\}} f_{\epsilon_1,\ldots,\epsilon_k}
\end{aligned}$$

(where the tensor product is taken in a certain order, which we do not specify here).

For $k \in \mathbb{N}$ and $f \in L^\infty(X)$ the seminorm $|||f|||_{T,k}$ associated with T is defined by $|||f|||_{T,k} = \big(\int_{X_T^{[k]}} f^{\otimes 2^k}\big)^{1/2^k}$. Equivalently,

$$|||f|||_{T,k}^{2^k} = \lim_{N_k\to\infty} \frac{1}{N_k} \sum_{n_k=1}^{N_k} \cdots \lim_{N_1\to\infty} \frac{1}{N_1} \sum_{n_1=1}^{N_1} \int_X \prod_{\epsilon_1,\ldots,\epsilon_k \in \{0,1\}} T^{\epsilon_1 n_1 + \ldots + \epsilon_k n_k} f.$$

It is proved in [HoK2] that for any $f_1, \ldots, f_{2^k} \in L^\infty(X)$ one has

$$\Big| \int_{X_T^{[k]}} \bigotimes_{j=1}^{2^k} f_j \Big| \le \prod_{j=1}^{2^k} |||f_j|||_{T,k}.$$

For any $k \in \mathbb{N}$ and $f \in L^\infty(X)$ we have

$$|||f|||_{T,k}^{2^k} = \int_{X_T^{[k]}} f^{\otimes 2^k} = \int_{I(X_T^{[k-1]},T^{[k-1]})} E\big(f^{\otimes 2^{k-1}} | I(X_T^{[k-1]},T^{[k-1]})\big)^2.$$

Since $\mathcal{I}(X_T^{[k-1]},T^{[k-1]}) \subseteq \mathcal{Z}_{k-1}(X,T)^{\otimes 2^{k-1}}$, one has $|||f|||_{T,k} = 0$ whenever $E\big(f|Z_{k-1}(X,T)\big) = 0$.

Proposition A.9. *For any $k \geq 2$, nonzero integers $l_1,\dots,l_k$ and a collection $f_{\epsilon_1,\dots,\epsilon_k} \in L^\infty(X)$, $\epsilon_1,\dots,\epsilon_k \in \{0,1\}$, if $E\big(f_{\epsilon_1,\dots,\epsilon_k}|Z_{k-1}(X,T)\big) = 0$ for some $\epsilon_1,\dots,\epsilon_k$ then*

$$\lim_{N_k\to\infty} \frac{1}{N_k}\sum_{n_k=1}^{N_k} \cdots \lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1} \int_X \prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{\epsilon_1 l_1 n_1+\dots+\epsilon_k l_k n_k} f_{\epsilon_1,\dots,\epsilon_k} = 0.$$

Proof. Let l be a common multiple of $l_1,\dots,l_k$. Since, by Theorem A.2, $Z_{k-1}(X,T^l) = Z_{k-1}(X,T)$, $E\big(f_{\epsilon_1,\dots,\epsilon_k}|Z_{k-1}(X,T)\big) = 0$ implies $|||f_{\epsilon_1,\dots,\epsilon_k}|||_{T^l,k} = 0$.

Let $r_i = l/l_i$, $i = 1,\dots,k$. We have

$$\begin{aligned}
&\lim_{N_k\to\infty} \frac{1}{N_k}\sum_{n_k=1}^{N_k} \cdots \lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1} \int_X \prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{\epsilon_1 l_1 n_1+\dots+\epsilon_k l_k n_k} f_{\epsilon_1,\dots,\epsilon_k}\\
&= \frac{1}{r_1\dots r_k}\sum_{m_k=0}^{r_k-1}\cdots\sum_{m_1=0}^{r_1-1} \lim_{N_k\to\infty} \frac{1}{N_k}\sum_{n_k=1}^{N_k} \cdots \lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1}\\
&\qquad \int_X \prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{\epsilon_1 l n_1+\dots+\epsilon_k l n_k}\big(T^{\epsilon_1 l_1 m_1+\dots+\epsilon_k l_k m_k} f_{\epsilon_1,\dots,\epsilon_k}\big)\\
&= \frac{1}{r_1\dots r_k}\sum_{m_k=0}^{r_k-1}\cdots\sum_{m_1=0}^{r_1-1} \int_{X_{T^l}^{[k]}} \bigotimes_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{\epsilon_1 l_1 m_1+\dots+\epsilon_k l_k m_k} f_{\epsilon_1,\dots,\epsilon_k}.
\end{aligned}$$

And for any $m_{\epsilon_1,\dots,\epsilon_k} \in \mathbb{Z}$, $\epsilon_1,\dots,\epsilon_k \in \{0,1\}$,

$$\begin{aligned}
\Big|\int_{X_{T^l}^{[k]}} \bigotimes_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{m_{\epsilon_1,\dots,\epsilon_k}} f_{\epsilon_1,\dots,\epsilon_k}\Big| &\leq \prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} |||T^{m_{\epsilon_1,\dots,\epsilon_k}} f_{\epsilon_1,\dots,\epsilon_k}|||_{T^l,k}\\
&= \prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} |||f_{\epsilon_1,\dots,\epsilon_k}|||_{T^l,k} = 0.
\end{aligned}$$

$\square$

Let $\varphi\colon \mathbb{Z}^d \longrightarrow \mathbb{Z}$ be a nonzero linear function, that is, a function of the form
$\varphi(n_1,\ldots,n_d) = a_1n_1+\ldots+a_dn_d$ with $a_1,\ldots,a_d \in \mathbb{Z}$ not all zero. Then for any measure preserving system (Y,S), any $f\in L^1(Y)$ and any Følner sequence $\{\Phi_N\}_{N=1}^\infty$ in $\mathbb{Z}^d$ one has $\lim_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}S^{\varphi(u)}f = E(f|I(Y,S^l)) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N S^{ln}f$, where $l=\gcd(a_1,\ldots,a_d)$. Applying this fact k times, we come to the following generalization of Proposition A.9:

Proposition A.10. *For any $k\geq 2$, positive integers $d_i\in\mathbb{N}$, nonzero linear functions $\varphi_i\colon \mathbb{Z}^{d_i}\longrightarrow\mathbb{Z}$, Følner sequences $\{\Phi_{i,N}\}_{N=1}^\infty$ in $\mathbb{Z}^{d_i}$, $i=1,\ldots,k$, and a collection of functions $f_{\epsilon_1,\ldots,\epsilon_k}\in L^\infty(X)$, $\epsilon_1,\ldots,\epsilon_k\in\{0,1\}$, if $E\big(f_{\epsilon_1,\ldots,\epsilon_k}|Z_{k-1}(X,T)\big)=0$ for some $\epsilon_1,\ldots,\epsilon_k$ then*

$$\lim_{N_k\to\infty}\frac{1}{|\Phi_{k,N_k}|}\sum_{u_k\in\Phi_{k,N_k}}\cdots\lim_{N_1\to\infty}\frac{1}{|\Phi_{1,N_1}|}\sum_{u_1\in\Phi_{1,N_1}}\int_X\prod_{\epsilon_1,\ldots,\epsilon_k\in\{0,1\}}T^{\epsilon_1\varphi_1(u_1)+\ldots+\epsilon_k\varphi_k(u_k)}f_{\epsilon_1,\ldots,\epsilon_k}=0.$$

The proof of Theorem A.8 will be based on the following lemma:

Lemma A.11. *For any linear functions $\varphi_1,\ldots,\varphi_k\colon\mathbb{Z}^d\longrightarrow\mathbb{Z}$ and any $f_1,\ldots,f_k\in L^\infty(X)$,*

$$\begin{aligned}&\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k\Big\|_{L^2(X)}\\ &\leq\Big(\lim_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{(v_1,w_1)\in\Phi_{N_1}^2}\lim_{N_k\to\infty}\frac{1}{|\Phi_{N_k}|^2}\sum_{(v_k,w_k)\in\Phi_{N_k}^2}\cdots\lim_{N_2\to\infty}\frac{1}{|\Phi_{N_2}|^2}\sum_{(v_2,w_2)\in\Phi_{N_2}^2}\\ &\qquad\int_X\prod_{\epsilon_1,\epsilon_2,\ldots,\epsilon_k\in\{0,1\}}T^{\epsilon_1\varphi_1(v_1-w_1)+\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)}f_1\Big)^{1/2^k}\\ &\qquad\cdot\prod_{i=2}^k\|f_i\|_{L^\infty(X)}.\end{aligned}$$

Proof. Let $\{\Phi_N\}_{N=1}^\infty$ be a Følner sequence in $\mathbb{Z}^d$. We will use the van der Corput lemma in the following form: if $\{f_u\}_{u\in\mathbb{Z}^d}$ is a bounded family of elements of a Hilbert space, then

$$\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}f_u\Big\|^2\leq\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\langle f_u,f_{u+v-w}\rangle.$$

We may assume that $|f_2|,\dots,|f_k|\le 1$. By the van der Corput lemma we have:

$$\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k\Big\|^2_{L^2(X)}$$
$$\le\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\int_X T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k$$
$$\cdot T^{\varphi_1(u+v-w)}f_1\cdot\ldots\cdot T^{\varphi_k(u+v-w)}f_k$$
$$=\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\int_X T^{\varphi_1(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{\varphi_k(u)}(f_k\cdot T^{\varphi_k(v-w)}f_k)$$
$$=\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\int_X T^{\varphi_1(u)-\varphi_k(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{\varphi_{k-1}(u)-\varphi_k(u)}(f_{k-1}\cdot T^{\varphi_{k-1}(v-w)}f_{k-1})\cdot(f_k\cdot T^{\varphi_k(v-w)}f_k)$$
$$=\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\int_X\Big(\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{(\varphi_1-\varphi_k)(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{(\varphi_{k-1}-\varphi_k)(u)}(f_{k-1}\cdot T^{\varphi_{k-1}(v-w)}f_{k-1})\Big)\cdot(f_k\cdot T^{\varphi_k(v-w)}f_k)$$

$$\le\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{(v,w)\in\Phi^2_{N_1}}\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{(\varphi_1-\varphi_k)(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{(\varphi_{k-1}-\varphi_k)(u)}(f_{k-1}\cdot T^{\varphi_{k-1}(v-w)}f_{k-1})\Big\|_{L^2(X)}.$$

By the induction hypothesis, applied to the linear functions $\varphi_i-\varphi_k\colon\mathbb{Z}^d\longrightarrow\mathbb{Z}$ and to the functions $f_i\cdot T^{\varphi_i(v-w)}f_i\in L^\infty(X)$, $i=1,\dots,k-1$, for any $(v,w)\in\mathbb{Z}^{2d}$ we have

$$\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{(\varphi_1-\varphi_k)(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{(\varphi_{k-1}-\varphi_k)(u)}(f_{k-1}\cdot T^{\varphi_{k-1}(v-w)}f_{k-1})\Big\|_{L^2(X)}$$
$$\le\Big(\lim_{N_k\to\infty}\frac{1}{|\Phi_{N_k}|^2}\sum_{(v_k,w_k)\in\Phi^2_{N_k}}\cdots\lim_{N_2\to\infty}\frac{1}{|\Phi_{N_2}|^2}\sum_{(v_2,w_2)\in\Phi^2_{N_2}}$$
$$\int_X\prod_{\epsilon_2,\dots,\epsilon_k\in\{0,1\}}T^{\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\Big)^{1/2^{k-1}}$$

$$= \Big(\lim_{N_k\to\infty} \frac{1}{|\Phi_{N_k}|^2} \sum_{(v_k,w_k)\in\Phi^2_{N_k}} \cdots \lim_{N_2\to\infty} \frac{1}{|\Phi_{N_2}|^2} \sum_{(v_2,w_2)\in\Phi^2_{N_2}}$$
$$\int_X \prod_{\epsilon_1,\epsilon_2,\ldots,\epsilon_k\in\{0,1\}} T^{\epsilon_1\varphi_1(v-w)+\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)} f_1\Big)^{1/2^{k-1}}.$$

Thus,

$$\limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} T^{\varphi_1(u)} f_1 \cdot \ldots \cdot T^{\varphi_k(u)} f_k \Big\|_{L^2(X)}$$
$$\leq \Big(\limsup_{N_1\to\infty} \frac{1}{|\Phi_{N_1}|^2} \sum_{(v,w)\in\Phi^2_{N_1}} \Big(\lim_{N_k\to\infty} \frac{1}{|\Phi_{N_k}|^2} \sum_{(v_k,w_k)\in\Phi^2_{N_k}} \cdots \lim_{N_2\to\infty} \frac{1}{|\Phi_{N_2}|^2} \sum_{(v_2,w_2)\in\Phi^2_{N_2}}$$
$$\int_X \prod_{\epsilon_1,\epsilon_2,\ldots,\epsilon_k\in\{0,1\}} T^{\epsilon_1\varphi_1(v-w)+\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)} f_1\Big)^{1/2^{k-1}}\Big)^{1/2}$$
$$\leq \Big(\lim_{N_1\to\infty} \frac{1}{|\Phi_{N_1}|^2} \sum_{(v,w)\in\Phi^2_{N_1}} \lim_{N_k\to\infty} \frac{1}{|\Phi_{N_k}|^2} \sum_{(v_k,w_k)\in\Phi^2_{N_k}} \cdots \lim_{N_2\to\infty} \frac{1}{|\Phi_{N_2}|^2} \sum_{(v_2,w_2)\in\Phi^2_{N_2}}$$
$$\int_X \prod_{\epsilon_1,\epsilon_2,\ldots,\epsilon_k\in\{0,1\}} T^{\epsilon_1\varphi_1(v-w)+\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)} f_1\Big)^{1/2^k}.$$

□

Proof of Theorem A.8. Because of the multilinearity of (A.5), it suffices to show that $\lim_{N\to\infty} \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} T^{\varphi_1(u)} f_1 \cdot \ldots \cdot T^{\varphi_k(u)} f_k = 0$ in $L^1(X)$ whenever $E\big(f_1|Z_{k-1}(X,T)\big) = 0$. We may assume that the functions $\varphi_1, \ldots, \varphi_k$ are all nonzero and distinct. Then, combining Lemma A.11 and Proposition A.10, applied to the nonzero linear functions $\varphi_1(v-w)$, $(\varphi_1-\varphi_2)(v-w)$, $\ldots$, $(\varphi_1-\varphi_k)(v-w)$ on $\mathbb{Z}^{2d}$ and the Følner sequence $\{\Phi^2_N\}_{N=1}^\infty$ in $\mathbb{Z}^{2d}$, we get $\lim_{N\to\infty} \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} T^{\varphi_1(u)} f_1 \cdot \ldots \cdot T^{\varphi_k(u)} f_k = 0$ in $L^2(X)$ and so, in $L^1(X)$. □

Acknowledgment. I thank V. Bergelson and E. Lesigne for valuable comments on the draft of this note.

B Appendix: Ergodic averages along the squares

by **Anthony Quas**[1] and **Máté Wierdl**[2]
University of Memphis
Department of Mathematical Sciences
373 Dunn Hall
Memphis, TN 38152-3240, USA
E-mail: *quasa@msci.memphis.edu*, *mw@csi.hu*

B.1 Enunciation of the result

In this note we want to present a proof of the almost everywhere convergence of the ergodic averages along the sequence of squares.

Theorem B.1. *Let τ be a measurable, measure preserving transformation of the σ-finite measure space (X, Σ, μ).*

Then, for $f \in L^2$, the averages

$$S_t f(x) = \frac{1}{t} \sum_{n \le t} f(\tau^{n^2} x)$$

converge for almost every $x \in X$.

The theorem is due to J. Bourgain. To keep our presentation as continuous as possible, we present historical remarks, and cite references in the last section, Section B.6.

B.2 Subsequence lemma

The main idea of the proof is to analyze the Fourier transform $\widehat{S}_t(\alpha) = 1/t \sum_{n \le t} e^{2\pi i n^2 \alpha}$ of the averages. This analysis permits us to replace the averages S_t by other operators that are easier to handle. The replace-ability of the sequence (S_t) by another sequence (A_t) means that we have an inequality of the form

$$\int \sum_t |S_t f - A_t f|^2 < c \int |f|^2. \tag{B.1}$$

[1]A. Quas' research is partially supported by NSF grant #DMS–0200703
[2]M. Wierdl's research is partially supported by NSF grant #DMS–0100577

Now, if somehow we prove that the sequence $(A_t f(x))$ converges for a.e. x, then the above inequality implies, since its left hand side is finite for $f \in L^2$, that the sequence $(S_t f(x))$ converges a.e. as well.

Well, we will not be able to prove an inequality of the type (B.1) exactly. In the real inequality, we will be able to have an inequality where the t runs through a lacunary sequence. But this is quite all right since it is enough to prove the a.e. convergence of the $(S_t f)$ along a lacunary sequence:

Lemma B.2. *For $\sigma > 1$ denote*

$$I = I_\sigma = \{t \mid t = \sigma^n \text{ for some positive integer } n\}.$$

Suppose that for each fixed $\sigma > 1$, the sequence $(S_t f)_{t \in I}$ converges a.e.

Then the full (S_t) sequence converges a.e.

Proof. We can assume that the function f is nonnegative. For a given t, choose k so that $\sigma^k \leq t < \sigma^{k+1}$. We can then estimate as

$$S_t f(x) \leq \frac{1}{\sigma^k} \sum_{n \leq \sigma^{k+1}} f(\tau^{n^2} x) = \sigma \cdot S_{\sigma^{k+1}} f(x),$$

and similarly, we have $\sigma^{-1} \cdot S_{\sigma^k} f(x) \leq S_t f(x)$. This means that

$$\sigma^{-1} \cdot \lim_k S_{\sigma^k} f(x) \leq \liminf_t S_t f(x) \leq \limsup_t S_t f(x) \leq \sigma \cdot \lim_k S_{\sigma^k} f(x).$$

Choosing now $\sigma_p = 2^{2^{-p}}$, we get that $\lim_k S_{\sigma_p^k} f(x)$ is independent of p for a.e. x, and, by the above estimates, it is equal to $\lim_t S_t f(x)$. $\square$

> *For the rest of the proof, we fix $\sigma > 1$, and unless we say otherwise, we always assume that $t \in I = I_\sigma$.*

Definition B.3. *If two sequences (A_t) and (B_t) of $L^2 \to L^2$ operators satisfy*

$$\int \sum_t |A_t f - B_t f|^2 < c \int |f|^2; \quad f \in L^2,$$

then we say that (A_t) and (B_t) are equivalent.

B.3 Oscillation and an instructive example

One standard way of proving a.e. convergence for the usual ergodic averages $1/t\sum_{n\leq t} f(\tau^n x)$ is to first prove a maximal inequality, and then note that there is a natural dense class for which a.e. convergence holds.

Unfortunately, the second part of this scheme does not work for the averages along the squares, since there is no known class of functions for which it would be easy to prove a.e. convergence of the averages.

Instead, for the squares, we will prove a so called *oscillation inequality*: for any $t(1) < t(2) < \dots$ with $t(k) \in I$, there is a constant c so that we have

$$\int \sum_k \sup_{t(k)<t<t(k+1)} |S_t f - S_{t(k+1)} f|^2 \leq c \int f^2. \tag{B.2}$$

We leave it to the reader to verify why an oscillation inequality implies a.e. convergence of the sequence $(S_t f)$. We also leave it to the reader to verify that if two operator sequences (A_t) and (B_t) are equivalent and (A_t) satisfies an oscillation inequality, then so does (B_t).

An important remark is that by the so called *transference principle* of Calderón, it is enough to prove the inequality in (B.2) on the integers $\mathbb{Z}$ which we consider equipped with the counting measure and the right shift. In this case, we have $S_t f(x) = 1/t\sum_{n\leq t} f(x+n^2)$.

To see how Fourier analysis can help in proving an oscillation inequality, let us look at a simpler example first: the case of the usual ergodic averages $U_t f(x) = 1/t\sum_{n\leq t} f(x+n)$ (by the transference principle, we only need to prove the oscillation inequality on the integers).

Let us assume that we already know the maximal inequality

$$\int_{\mathbb{Z}} \sup_t |U_t f|^2 \leq c \cdot \int_{\mathbb{Z}} |f|^2$$

For the Fourier transform $\widehat{U}_t(\alpha) = 1/t\sum_{n\leq t} e^{2\pi i n\alpha}$, $\alpha \in (-1/2, 1/2)$ we easily obtain the estimates

$$|\widehat{U}_t(\alpha) - 1| \leq c \cdot t \cdot |\alpha|; \tag{B.3}$$

$$|\widehat{U}_t(\alpha)| \leq \frac{c}{t \cdot |\alpha|}. \tag{B.4}$$

The first estimate is effective (nontrivial) when $|\alpha| < 1/t$ and it says that $\widehat{U}_t(\alpha)$ is close to 1. The second estimate is effective when $|\alpha| > 1/t$, and it says that then $|\widehat{U}_t(\alpha)|$ is small. In other words, the estimates in (B.3) and (B.4) say that the function $\mathbb{1}_{(-1/t,1/t)}(\alpha)$ captures the "essence" of $\widehat{U}_t(\alpha)$. How? Let us define the operator A_t via its Fourier transform as $\widehat{A}_t(\alpha) = \mathbb{1}_{(-1/t,1/t)}(\alpha)$. The great advantage of the (A_t) is that it is a monotone sequence of projections. We'll see in a minute how this can help. First we claim that the sequences (U_t) and (A_t) are equivalent. To prove this claim, start by observing that

$$\widehat{U_t f}(\alpha) = \widehat{U}_t(\alpha) \cdot \widehat{f}(\alpha); \quad \widehat{A_t f}(\alpha) = \widehat{A}_t(\alpha) \cdot \widehat{f}(\alpha),$$

and then estimate, using Parseval's formula, as

$$\begin{aligned}
\int_{\mathbb{Z}} \sum_{t \in I} |A_t f - U_t f|^2 &= \int_{-1/2}^{1/2} \sum_{t \in I} |\widehat{A}_t(\alpha) - \widehat{U}_t(\alpha)|^2 \cdot |\widehat{f}(\alpha)|^2 \; d\alpha \\
&\leq \int_{-1/2}^{1/2} |\widehat{f}(\alpha)|^2 \; d\alpha \cdot \sup_{\alpha} \sum_{t \in I} |\widehat{U}_t(\alpha) - \widehat{A}_t(\alpha)|^2 \\
&= \int_{\mathbb{Z}} f^2 \cdot \sup_{\alpha} \sum_{t \in I} |\widehat{U}_t(\alpha) - \widehat{A}_t(\alpha)|^2
\end{aligned}$$

It follows that it is enough to prove the inequality

$$\sup_{\alpha} \sum_{t \in I} |\widehat{U}_t(\alpha) - \widehat{A}_t(\alpha)|^2 < \infty.$$

To see this, for a fixed α, divide the summation on t into two parts, $t < |\alpha|^{-1}$ and $t > |\alpha|^{-1}$. For the case $t < |\alpha|^{-1}$, use the estimate in (B.3) and in case $t > |\alpha|^{-1}$ use the estimate in (B.4). In both cases, we end up with a geometric progression with quotient $1/\sigma$.

Since (U_t) and (A_t) are equivalent and (U_t) satisfies a maximal inequality, the operators A_t also satisfy a maximal inequality. But then the sequence $(A_t f(x))$ satisfies an oscillation inequality. To see this, first note that if $t(k) \leq t \leq t(k+1)$ then $A_t f(x) - A_{t(k+1)} f(x) = A_t \left(A_{t(k)} f(x) - A_{t(k+1)} f(x)\right)$. It follows, that

$$\begin{aligned}
\int_{\mathbb{Z}} \sup_{t(k)<t<t(k+1)} |A_t f - A_{t(k+1)} f|^2 &= \int_{\mathbb{Z}} \sup_{t} \left| A_t \left(A_{t(k)} f - A_{t(k+1)} f\right) \right|^2 \\
&\leq c \cdot \int_{\mathbb{Z}} \left| A_{t(k)} f - A_{t(k+1)} f \right|^2,
\end{aligned}$$

since the sequence (A_t) satisfies a maximal inequality. But now the oscillation inequality follows from the inequality

$$\int_{\mathbb{Z}} \sum_k |A_{t(k)}f - A_{t(k+1)}f|^2 \le \int_{\mathbb{Z}} f^2.$$

This inequality, in turn, follows by examining the Fourier transform of the left hand side.

Now the punchline is that the ergodic averages (U_t) also satisfy the oscillation inequality since (U_t) and (A_t) are equivalent.

Let us summarize the scheme above: the maximal inequality for (U_t) implies a maximal inequality for the (A_t) since the two sequences are equivalent. But the (A_t), being a monotone sequence of projections, satisfy an oscillation inequality. But then, again appealing to the equivalence of the two sequences, the (U_t) satisfies an oscillation inequality.

What we have learned is that if a sequence of operators $(B_t f)$ satisfies a maximal inequality, and it is equivalent to a monotone sequence (A_t) of projections, then $(B_t f)$ satisfies an oscillation inequality..

In the remaining sections we will see that the scheme of proving an oscillation inequality for the averages along the squares (S_t) is similar, and ultimately it will be reduced to proving a maximal inequality for a monotone sequence of projections.

B.4 Periodic systems and the circle method

The difference between the usual ergodic averages and the averages along squares is that the squares are not uniformly distributed in residue classes. Indeed, for example no number of the form $3n-1$ is a square. This property of the squares is captured well in the behavior of the Fourier transform, $\widehat{S}_t(\alpha) = 1/t \sum_{n \le t} e^{2\pi i n^2 \alpha}$: for a typical rational $\alpha = b/q$, $\lim_{t\to\infty} \widehat{S}_t(\alpha)$ is nonzero (while it would be 0 if the squares were uniformly distributed $\mod q$).

We need some estimates on the Fourier transform $\widehat{S}_t(\alpha)$. Since we will often deal with the function $e^{2\pi i\beta}$, we introduce the notation $e(\beta) = e^{2\pi i \beta}$. Also, the estimates for the Fourier transform $\widehat{S}_t(\alpha)$ are simpler if instead of the averages $\frac{1}{t}\sum_{n\le t} \tau^{n^2} f(x)$ we consider the weighted averages $1/t \sum_{n^2 \le t}(2n - 1)\tau^{n^2} f(x)$. The weight $2n-1$ is motivated by $n^2 - (n-1)^2 = 2n-1$. Everything we said about the averages along the squares applies equally well to these new weighted averages. Furthermore, it is an exercise in summation by parts to show that the a.e. convergence of the weighted and non weighted averages is equivalent.

So from now on, we use the notation

$$S_t f(x) = \frac{1}{t} \sum_{n^2 \le t} (2n-1) \tau^{n^2} f(x).$$

Let $\widehat{\Lambda}(\alpha) = \lim_t \widehat{S}_t(\alpha)$. By Weyl's theorem, $\widehat{\Lambda}(\alpha) = 0$ for irrational α and for rational $\alpha = b/q$, if b/q is in reduced terms, we have the estimate

$$|\widehat{\Lambda}(b/q)| \le \frac{c}{q^{1/2}} \tag{B.5}$$

This inequality tells us that while the squares are not uniformly distributed in residue classes $\mod q$, at least they try to be: $\widehat{\Lambda}_t(b/q) \to 0$ as $q \to \infty$.

Now the so called *circle method* of Hardy and Littlewood tells us about the structure of $\widehat{S}_t(\alpha)$. Let us introduce the notations $P(t) = t^{1/3}$, $Q(t) = 2t/P(t) = 2t^{2/3}$. According to the circle method, we have the following estimates

$$\left| \widehat{S}_t(\alpha) - \widehat{\Lambda}(b/q) \cdot \widehat{U}_t(\alpha - b/q) \right| \le c \cdot t^{-1/6}; \quad q \le P(t), \quad |\alpha - b/q| < 1/Q(t) \tag{B.6}$$

$$\left| \widehat{S}_t(\alpha) \right| < c \cdot t^{-1/6}, \quad \text{otherwise,} \tag{B.7}$$

where recall that U_t denotes the usual ergodic averages so $\widehat{U}_t(\beta) = 1/t \sum_{n \le t} e(n\alpha)$. In other words, the estimate above tells us that $\widehat{S}_t(\alpha)$ is close to $\widehat{\Lambda}(b/q) \cdot \widehat{U}_t(\alpha - b/q)$ if α is close to a rational point b/q with small denominator, and otherwise $|\widehat{S}_t(\alpha)|$ is small.

Given these estimates, it is easy to see that the sequence (S_t) is equivalent to the sequence (A_t) defined by its Fourier transform as

$$\widehat{A}_t(\alpha) = \sum_{\substack{b/q \\ q \le P(t)}} \widehat{\Lambda}(b/q) \cdot \widehat{U}_t(\alpha - b/q) \cdot \mathbb{1}_{(-1/Q(t), 1/Q(t))}(\alpha - b/q)$$

It remains to prove an oscillation inequality for the A_t. To do this, first we group those b/q for which q is of similar size:

$$E_p = \{b/q \mid 2^p \le q < 2^{p+1}\}$$

By the estimates in (B.5), we have

$$\sup_{b/q \in E_p} |\widehat{\Lambda}(b/q)| \leq c \cdot 2^{-p/2}. \tag{B.8}$$

Note also that if $b/q \in E_p$ then the term $\widehat{\Lambda}(b/q)$ occurs in the definition of A_t only when $t > 2^{3p}$. Define the operator $A_{p,t}$ by its Fourier transform as

$$\widehat{A}_{p,t}(\alpha) = \sum_{b/q \in E_p} \widehat{\Lambda}(b/q) \cdot \widehat{U}_t(\alpha - b/q) \cdot \mathbb{1}_{(-1/Q(t),1/Q(t))}(\alpha - b/q), \quad t > 2^{3p}.$$

Using the triangle inequality for the summation in p, we see that an oscillation inequality for (A_t) would follow from the inequality

$$\int_{\mathbb{Z}} \sum_k \sup_{t(k)<t<t(k+1)} |A_{p,t}f - A_{p,t(k+1)}f|^2 \leq c \cdot \frac{p^2}{2^p} \cdot \int_{\mathbb{Z}} f^2 \tag{B.9}$$

We have learned in the previous section, Section B.3, that it is useful to try work with projections. As a step, we introduce the operators $B_{p,t}$ defined via

$$\widehat{B}_{p,t}(\alpha) = \sum_{b/q \in E_p} \widehat{\Lambda}(b/q) \cdot \mathbb{1}_{(-1/t,1/t)}(\alpha - b/q), \quad t > 2^{3p}.$$

Note that for each α there is at most one $b/q \in E_p$ so that $\mathbb{1}_{(-1/t,1/t)}(\alpha - b/q) \neq 0$ or $\mathbb{1}_{(-1/Q(t),1/Q(t))}(\alpha - b/q) \neq 0$ for some $t > 2^{3p}$. Hence, using the estimates in (B.3), (B.4), and (B.8), we get

$$|\widehat{A}_{p,t}(\alpha) - \widehat{B}_{p,t}(\alpha)| \leq c \cdot 2^{-p/2} \cdot \min\{t|\alpha|, (t|\alpha|)^{-1}\}; \quad t > 2^{3p}.$$

It follows that we can replace the $(A_{p,t})$ by the $(B_{p,t})$:

$$\int_{\mathbb{Z}} \sum_{t > 2^{3p}} |A_{p,t}f - B_{p,t}f|^2 \leq c \cdot 2^{-p} \cdot \int_{\mathbb{Z}} f^2$$

In order to prove the required oscillation inequality for the $B_{p,t}$, we make one more reduction. Namely, we claim that defining $C_{p,t}$ by

$$\widehat{C}_{p,t}(\alpha) = \sum_{b/q \in E_p} \mathbb{1}_{(-1/t,1/t)}(\alpha - b/q), \quad t > 2^{3p}.$$

(so $\widehat{C}_{p,t}$ is just $B_{p,t}$ without the multipliers $\widehat{\Lambda}(b/q)$), we need to prove

$$\int_{\mathbb{Z}} \sum_k \sup_{t(k)<t<t(k+1)} |C_{p,t} - C_{p,t(k+1)}|^2 \leq c \cdot p^2 \cdot \int_{\mathbb{Z}} f^2. \tag{B.10}$$

We leave the proof of this implication to the reader with the hint to replace the function f by g defined by its Fourier transform as

$$\widehat{g}(\alpha) = \sum_{b/q \in E_p} \widehat{\Lambda}(b/q) \cdot \mathbb{1}_{(-2^{-3p}, 2^{-3p})}(\alpha - b/q) \cdot \widehat{f}(\alpha).$$

Indeed, then $B_{p,t}f(x) = C_{p,t}g(x)$ and $\int_{\mathbb{Z}} g^2 \leq c \cdot 2^{-p} \int_{\mathbb{Z}} f^2$ by (B.8).

Now, the $C_{p,t}$ form a monotone (in t) sequence of projections, and hence they will satisfy the oscillation inequality in (B.10) once they satisfy the maximal inequality

$$\int_{\mathbb{Z}} \sup_{t>2^{3p}} |C_{p,t}|^2 \leq c \cdot p^2 \cdot \int_{\mathbb{Z}} f^2. \tag{B.11}$$

To encourage the reader, we emphasize that our only remaining task is to prove the inequality in (B.11) above.

B.5 The main inequality

Since the least common multiple of the denominators of rational numbers in the set E_p is not greater than 2^{cp2^p} and the distance between two elements of E_p is at least 2^{-2p}, the estimate in (B.11) follows from the following result

Theorem B.4. *Let $0 < \delta < 1/2$ and $e(\alpha_1), e(\alpha_2), \ldots, e(\alpha_J)$ be distinct complex Q-th roots of unities with $|\alpha_i - \alpha_j| > \delta$ for $i \neq j$. We assume that $\delta^{-1} \leq Q$. Define the projections R_t by*

$$\widehat{R}_t(\alpha) = \sum_{j \leq J} \mathbb{1}_{(-1/t, 1/t)}(\alpha - \alpha_j).$$

Then we have, with an absolute constant c,

$$\int_{\mathbb{Z}} \sup_{t \geq \delta^{-1}} |R_t f|^2 \leq c \cdot (\log \log Q)^2 \cdot \int_{\mathbb{Z}} |f|^2.$$

We restrict the the range on t to $t \geq \delta^{-1}$, because then the sum making up R_t contains pairwise orthogonal elements—as a result on the separation hypothesis $|\alpha_i - \alpha_j| > \delta$.

Proof. Two essentially different techniques will be used to handle the supremum. The first technique will handle the range $\delta^{-1} \leq t < Q^4$, and the other technique will handle the remaining $t > Q^4$ range.

Let us start with proving the inequality

$$\int_{\mathbb{Z}} \sup_{\delta^{-1} \leq t \leq Q^4} |R_t f|^2 \leq c \cdot (\log \log Q)^2 \cdot \int_{\mathbb{Z}} |f|^2. \tag{B.12}$$

We can assume that Q^4 is a power of σ, say $Q^4 = \sigma^S$, and then the range $\delta^{-1} \leq t \leq Q^4$ can be rewritten as $c \log \delta^{-1} \leq s \leq S$, where we take log with base σ. Introduce the monotone sequence of projections $P_s = R_{\sigma^{S-s}}$, $s \leq S - c \log \delta^{-1}$. All follows from

$$\int_{\mathbb{Z}} \sup_{s \leq S - c \log \delta^{-1}} |P_s f|^2 \leq c \cdot \log^2 S \cdot \int_{\mathbb{Z}} |f|^2.$$

It is clearly enough to show the inequality for dyadic $S - c \log \delta^{-1}$:

$$\int_{\mathbb{Z}} \sup_{s \leq 2^M} |P_s f|^2 \leq c \cdot M^2 \cdot \int_{\mathbb{Z}} |f|^2.$$

For each integer $m \leq M$ consider the sets

$$H_m = \left\{ P_{(d+1) \cdot 2^m} - P_{d \cdot 2^m} \mid d = 0, 1, \ldots, 2^{M-m} - 1 \right\}.$$

If the dyadic expansion of s is $s = \sum_{m \leq M} \epsilon_m \cdot 2^m$, where ϵ_m is 0 or 1, then for some $X_m \in H_m$, $P_s = \sum_{m \leq M} \epsilon_m \cdot X_m$. It follows that

$$|P_s f(x)|^2 \leq M \cdot \sum_{m \leq M} |X_m f(x)|^2.$$

For each m, we have

$$|X_m f(x)|^2 \leq \sum_{d \leq 2^{M-m}} |P_{(d+1) \cdot 2^m} f(x) - P_{d \cdot 2^m} f(x)|^2,$$

hence

$$
\begin{aligned}
\int_{\mathbb{Z}} \sup_{s\le 2^M} |P_s f|^2 &\le M \cdot \int_{\mathbb{Z}} \sum_{m\le M} \sum_{d\le 2^{M-m}} |P_{(d+1)\cdot 2^m} f(x) - P_{d\cdot 2^m} f(x)|^2 \\
&\le M \cdot \sum_{m\le M} \sum_{s\le 2^M} \int_{\mathbb{Z}} |P_{s+1} f - P_s f|^2 \\
&\le M^2 \cdot 2 \cdot \int_{\mathbb{Z}} |f|^2.
\end{aligned}
$$

Let us now handle the remaining range for t. We want to prove

$$\int_{\mathbb{Z}} \sup_{t>Q^4} |R_t f|^2 \le c \cdot \int_{\mathbb{Z}} |f|^2. \tag{B.13}$$

It seems best if we replace the operators R_t by the operators

$$A_t f(x) = \frac{1}{t} \sum_{n\le t} \sum_{j\le J} e(n\alpha_j) f(x+n).$$

This replacement is possible if we prove the following two inequalities

$$\int_{\mathbb{Z}} \sum_{t>\delta^{-2}} |A_t f - R_t f|^2 \le c \cdot \int_{\mathbb{Z}} |f|^2. \tag{B.14}$$

and

$$\int_{\mathbb{Z}} \sup_{t>Q^4} |A_t f|^2 \le c \cdot \int_{\mathbb{Z}} |f|^2. \tag{B.15}$$

Let us start with proving (B.14). By Parseval's formula, we need to prove

$$\sup_{\alpha} \sum_{t} |\widehat{A}_t(\alpha) - \widehat{R}_t(\alpha)|^2 < \infty.$$

Fix α. Without loss of generality we can assume that of the α_j, the point α_1 is closest to α. Possibly dividing the sum on j into two and reindexing them, we also assume that $\alpha_1 < \alpha_2 < \cdots < \alpha_J$. Using the separation hypothesis $|\alpha_i - \alpha_j| > \delta$ for $i \ne j$, we have that $|\alpha - \alpha_j| > (j-1)\delta$ for $j > 1$.

For $t \le 1/|\alpha-\alpha_1|$ we can thus estimate (recall that $\widehat{U}_t(\beta) = 1/t \sum_{n\le t} e(n\beta))$ as

$$\begin{aligned}
|\widehat{A}_t(\alpha) - \widehat{R}_t(\alpha)| &\le |\widehat{U}_t(\alpha-\alpha_1)| + \sum_{2\le j\le J} |\widehat{U}_t(\alpha-\alpha_j)| \text{ by (B.3) and (B.4)}\\
&\le c\cdot\left(t|\alpha-\alpha_1| + \sum_{2\le j\le J}\frac{1}{t(j-1)\delta}\right)\\
&\le c\cdot (t|\alpha-\alpha_1| + \log J/(\delta t))\\
&\le c\cdot \big(t|\alpha-\alpha_1| + \delta^{-2}/t\big)
\end{aligned}$$

where we used in the last estimate that $J \le \delta^{-1}$. Summing this estimate over $t\in I$ with $\delta^{-2}\le t \le 1/|\alpha-\alpha_1|$ we get a finite bound independent of α.

For $t > 1/|\alpha-\alpha_1|$, we have

$$|\widehat{A}_t(\alpha) - \widehat{R}_t(\alpha)| \le \sum_{1\le j\le J} |\widehat{U}_t(\alpha-\alpha_j)| \le c\cdot\frac{\delta^{-2}}{t},$$

which, upon summing over the full range $\delta^{-2} < t$, again gives a finite bound independent of α.

Let us single out a consequence of inequality (B.14): there is a constant c so that

$$\int_{\mathbb{Z}} |A_t f|^2 \le c\cdot \int_{\mathbb{Z}} |f|^2; \quad t > \delta^{-2}. \tag{B.16}$$

Our only remaining task is to prove inequality (B.15).

For a given t, let q be the largest integer so that $qQ^2 \le t$. Note that $q \ge Q^2$ since $t > Q^4$. We can estimate as

$$\begin{aligned}
&\Big|\sum_{n\le t}\sum_{j\le J} e(n\alpha_j) f(x+n)\Big|\\
&\quad\le \Big|\sum_{n\le qQ^2}\sum_{j\le J} e(n\alpha_j) f(x+n)\Big| + \Big|\sum_{qQ^2<n\le t}\sum_{j\le J} e(n\alpha_j) f(x+n)\Big|.
\end{aligned}\tag{B.17}$$

We estimate the second term on the right trivially as

$$\Big|\sum_{qQ^2<n\le t}\sum_{j\le J} e(n\alpha_j) f(x+n)\Big| \le J\cdot \sum_{qQ^2<n\le (q+1)Q^2} \big|f(x+n)\big|.$$

With this, we have

$$\begin{aligned}
&\sup_{t>Q^4}\Big(\frac{1}{t}\Big|\sum_{qQ^2<n\le t}\sum_{j\le J} e(n\alpha_j)f(x+n)\Big|\Big)^2\\
&\quad\le \sup_{q\ge Q^2}\Big(\frac{J}{qQ^2}\cdot\sum_{qQ^2<n\le(q+1)Q^2}|f(x+n)|\Big)^2 \text{ by Cauchy's inequality}\\
&\quad\le \sup_{q\ge Q^2}\frac{J^2\cdot Q^2}{qQ^2}\cdot\frac{\sum_{qQ^2<n\le(q+1)Q^2}|f(x+n)|^2}{qQ^2}\\
&\quad\le \sum_{q\ge Q^2}\frac{J^2}{q^2}\cdot\frac{1}{Q^2}\sum_{qQ^2<n\le(q+1)Q^2}|f(x+n)|^2
\end{aligned}$$

Integrating the last line, we obtain the bound

$$\sum_{q\ge Q^2}\frac{J^2}{q^2}\cdot\int_{\mathbb{Z}}|f|^2\le c\cdot\frac{J^2}{Q^2}\int_{\mathbb{Z}}|f|^2\le c\cdot\int_{\mathbb{Z}}|f|^2.$$

since $J\le Q$.

Let us now handle the first term on the right of (B.17). Since $e(\alpha_j)$ satisfies $e((mQ^2+h)\alpha_j)=e(h\alpha_j)$ (this is the first and last time we use that the $e(\alpha_j)$ are Q-th roots of unities), we can write, defining $Tg(x)=g(x+Q^2)$,

$$\Big|\frac{1}{t}\sum_{n\le qQ^2}\sum_{j\le J}e(n\alpha_j)f(x+n)\Big|\le\Big|\frac{1}{q}\sum_{m\le q}T^m\frac{1}{Q^2}\sum_{h\le Q^2}\sum_{j\le J}e(h\alpha_j)f(x+h)\Big|.$$

By the ergodic maximal inequality, applied to T, the ℓ^2 norm of our maximal operator is bounded by the ℓ^2 norm of

$$\frac{1}{Q^2}\sum_{h\le Q^2}\sum_{j\le J}e(h\alpha_j)f(x+h).$$

But the estimate in (B.14) says, the ℓ^2 norm of the above is bounded independently of Q since $Q^2>\delta^{-2}$ by assumption. □

B.6 Notes

More details More details and references can be found in [RosW]. In particular, the circle method and the transference principle are described in complete details—though no proof of the main inequality of Bourgain, Theorem B.4, is given. The inequalities (B.6) and (B.7) appear as (4.23) and (4.24) in [RosW].

Theorem B.1 The result is due to Bourgain ([Bou1]). He later extended the result to $f \in L^p$, $p > 1$; cf [Bou2]. The case $p = 1$ is the most outstanding unsolved problem in this subject.

Idea of proof The basic structure of the proof is that of Bourgain's ([Bou2]) but we used ideas from Lacey's paper [La] as well—not to mention some personal communication with M. Lacey.

Other sequences The sequence of primes is discussed in [Wi]. But we'd like to emphasize that the L^2 theory of the primes is identical to the case of the squares. The only difference is in the estimates in (B.6) and (B.7).

A characterization of sequences which are good for the pointwise and mean ergodic theorems can be found in [BoQW].

Acknowledgment. The authors thank E. Lesigne for useful comments.

References

[A] J. Auslander, On the proximal relation in topological dynamics, *Proc. Amer. Math. Soc.* **11** (1960), 890–895.

[Ba] S. Banach, Sur le problème de la mesure, *Fund. Math.* **4** (1923), 7–33.

[BaT] S. Banach and A. Tarski, Sur la décomposition des ensembles de points en parties respectivement congruentes, *Fund. Math.* **6** (1924) 244–277.

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

[Be1] V. Bergelson, Sets of recurrence of $\mathbb{Z}^m$-actions, and properties of sets of differences in $\mathbb{Z}^m$, *J. London Math Soc. (2)* **31** (1985), 295–304.

[Be2] V. Bergelson, Weakly mixing PET, *Erg. Th. and Dynam. Sys.* **7** (1987), 337–349.

[Be3] V. Bergelson, Ergodic Ramsey theory – an update, *Ergodic theory of $\mathbb{Z}^d$-actions (Warwick, 1993-1994)*, 273–296, London Math. Soc. Lecture Note Ser., 228, Cambridge Univ. Press, Cambridge, 1996.

[Be4] V. Bergelson, The multifarious Poincare recurrence theorem, *Descriptive set theory and dynamical systems (Marseille-Luminy, 1996)*, 31–57, London Math Soc. Lecture Note Ser. 277, Cambridge Univ. Press, Cambridge, 2000.

[Be5] V. Bergelson, Ergodic theory and diophantine problems, *Topics in symbolic dynamics and applications (Temuco, 1997)*, 167-205, London Math Soc. Lecture Note Ser. 279, Cambridge Univ. Press, Cambridge, 2000.

[Be6] V. Bergelson, Minimal idempotents and ergodic Ramsey theory, *Topics in dynamics and ergodic theory, 8–39*, London Math. Soc. Lecture Note Ser., 310, Cambridge Univ. Press, Cambridge, 2003.

[Be7] V. Bergelson, Multiplicatively large sets and ergodic Ramsey theory, to appear in *Israel J. Math.*

[BeBB] V. Bergelson, M. Boshernitzan, and J. Bourgain, Some results on non-linear recurrence, *J. Anal. Math.* **62** (1994), 29–46.

[BeBH] V. Bergelson, A. Blass and N. Hindman, Partition theorems for spaces of variable words, *Proc. London Math. Soc. (3)* **68** (1994), no. 3, 449–476.

[BeFHK] V. Bergelson, H. Furstenberg, N. Hindman, and Y. Katznelson, An algebraic proof of van der Waerden's theorem, *Enseign. Math. (2)* **35** (1989), no. 3-4, 209–215.

[BeFM] V. Bergelson, H. Furstenberg, and R. McCutcheon, IP-sets and polynomial recurrence, *Ergod. Th. and Dynam. Sys.* **16** (1996), no. 5, 963–974.

[BeG] V. Bergelson and A. Gorodnik, Weakly mixing group actions: a brief survey and an example, *Modern Dynamical Systems and Applications, B. Hasselblatt, M. Brin, Y. Pesin, eds.*, 3–25, Cambridge University Press, New York, 2004.

[BeH1] V. Bergelson and N. Hindman, Nonmetrizable topological dynamics and Ramsey theory, *Trans. Amer. Math. Soc.* **320** (1990), 293–320.

[BeH2] V. Bergelson and N. Hindman, Some topological semicommutative van der Waerden type theorems and their combinatorial consequences, *J. London Math. Soc. (2)* **45** (1992), no. 3, 385–403.

[BeH3] V. Bergelson and N. Hindman, Additive and multiplicative Ramsey theorems in N – some elementary results, *Combin. Probab. Comput.* **2** (1993), no. 3, 221–241.

[BeH4] V. Bergelson and N. Hindman, On IP* sets and central sets, *Combinatorica* **14** (1994), no. 3, 269–277.

[BeL1] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems, *J. Amer. Math. Soc.* **9** (1996), 725 –753.

[BeL2] V. Bergelson and A. Leibman, Set-polynomials and polynomial extension of the Hales-Jewett theorem. *Ann. of Math. (2)* **150** (1999), no. 1, 33–75.

[BeL3] V. Bergelson and A. Leibman, A nilpotent Roth theorem. *Invent. Math.* **147** (2002), no. 2, 429–470.

[BeL4] V. Bergelson and A. Leibman, Topological multiple recurrence for polynomial configurations in nilpotent groups. *Adv. Math.* **175** (2003), no. 2, 271–296.

[BeL5] V. Bergelson and A. Leibman, Failure of the Roth theorem for solvable groups of exponential growth, *Ergod. Th. and Dynam. Sys.* **24** (2004), 45–53.

[BeLM] V. Bergelson, A. Leibman and R. McCutcheon, Polynomial Szemerédi theorems for countable modules over integral domains and finite fields, to appear in *J. d'Analyse Math.*

[BeM1] V. Bergelson and R. McCutcheon, Uniformity in the polynomial Szemerédi theorem, *Ergodic theory of $\mathbb{Z}^d$-actions (Warwick, 1993-1994)*, 273–296, London Math. Soc. Lecture Note Ser., 228, Cambridge Univ. Press, Cambridge, 1996.

[BeM2] V. Bergelson and R. McCutcheon, An ergodic IP Szemerédi theorem, *Mem. Amer. Math. Soc.* **146** (2000), no. 695, viii+106 pp.

[BeM3] V. Bergelson and R. McCutcheon, Recurrence for semigroup actions and a non-commutative Schur theorem, *Topological dynamics and applications (Minneapolis, MN, 1995)*, 205–222, Contemp. Math., 215, Amer. Math. Soc., Providence, RI, 1998.

[BeMZ] V. Bergelson, R. McCutcheon and Q. Zhang, A Roth theorem for amenable groups, *Amer. J. Math.* **119** (1997), no. 6, 1173–1211.

[BeR1] V. Bergelson and J. Rosenblatt, Mixing actions of groups, *Illinois J. Math.* **32** (1988), no. 1, 65–80.

[BeR2] V. Bergelson and J. Rosenblatt, Joint ergodicity for group actions, *Ergod. Th. and Dynam. Sys.* **8** (1988), no. 3, 351–364.

[BeS] V. Bergelson and D. Shapiro, Multiplicative subgroups of finite index in a ring. *Proc. Amer. Math. Soc.* **116** (1992), no. 4, 885–896.

[Bi] P. Billingsley, *Probability and Measure*, Wiley, New York, 1986.

[BPT] A. Blaszczyk, S. Plewik, S. Turek, Topological multidimensional van der Waerden theorem. *Comment. Math. Univ. Carolin.* **30** (1989), no. 4, 783–787.

[BoQW] M. Boshernitzan, G. Kolesnik, A. Quas and M. Wierdl, Ergodic averaging sequences, 33 pages, to appear in *J. d'Analyse Math.*, available at `http://www.csi.hu/mw/hardy_ergav.pdf` or `http://www.csi.hu/mw/hardy_ergav.dvi`

[Bou1] J. Bourgain, On the maximal ergodic theorem for certain subsets of the integers, *Israel J. Math.* **61** (1988), 39–72.

[Bou2] J. Bourgain, Pointwise ergodic theorems for arithmetic sets (appendix: The return time theorem), *Publ. Math. I.H.E.S.* **69** (1989), 5–45.

[Bou3] J. Bourgain, Double recurrence and almost sure convergence, *J. Reine Angew. Math.* **404** (1990), 140–161.

[Br] T. C. Brown, An interesting combinatorial method in the theory of locally finite semigroups. *Pacific J. Math.* **36** (1971) 285–289.

[Ca] T.J. Carlson, Some unifying principles in Ramsey theory, *Discrete Math.* **68** (1988), 117–169.

[CL1] J.P. Conze and E. Lesigne, Théorèmes ergodiques pour des mesures diagonales, *Bull. Soc. Math. France* **112** (1984) no. 2, 143–175.

[CL2] J.P. Conze and E. Lesigne, Sun un théorème ergodique pour des mesures diagonales, *Probabilités*, 1–31, Publ. Inst. Rech. Math. Rennes, 1987–1, Univ. Rennes I, Rennes, 1988.

[CL3] J.P. Conze and E. Lesigne, Sun un théorème ergodique pour des mesures diagonales, *C. R. Acad. Sci. Paris Sér. I Math.* **306** (1988) no. 12, 491–493.

[CN] W. Comfort and S. Negrepontis, *The Theory of Ultrafilters*, Springer-Verlag, Berlin and New York, 1974.

[CY] P. Civin and B. Yood, The second conjugate space of a Banach algebra as an algebra, *Pac. J. Math.* **11** (1961), 847–870.

[D1] L. E. Dickson, On the congruence $x^n + y^n = z^n \equiv 0 \pmod p$, *J. Reine Angew. Math.* **135** (1909) p.134–141.

[D2] L. E. Dickson, *History of the Theory of Numbers, vol. II (Diophantine analysis)*, Chelsea, New York, 1971.

[E1] R. Ellis, Distal transformation groups, *Pac. J. Math.* **8** (1958), 401–405.

[E2] R. Ellis, A semigroup associated with a transformation group, *Trans. Amer. Math. Soc.* **94** (1960), 272–281.

[ET] P. Erdős and P. Turán, On some sequences of integers, *J. London Math. Soc.* **11** (1936), 261–264.

[Fø] E. Følner, On groups with full Banach mean values, *Math. Scand.* **3** (1955), 243–254.

[FrK] N. Frantzikinakis and B. Kra, Polynomial averages converge to the product of integrals, to appear in *Israel J. Math.*

[Fu1] H. Furstenberg, The structure of distal flows, *Amer. J. Math.* **85** (1963), 477–515.

[Fu2] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. d'Analyse Math.* **31** (1977), 204–256.

[Fu3] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, 1981.

[Fu4] H. Furstenberg, Poincaré recurrence and number theory, *Bull. Amer. Math. Soc. (New Series)* **5** (1981), 211–234.

[FuK1] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, *J. d'Analyse Math.* **34** (1978), 275–291.

[FuK2] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for IP-systems and combinatorial theory, *J. d'Analyse Math.* **45** (1985), 117–168.

[FuK3] H. Furstenberg and Y. Katznelson, Idempotents in compact semigroups and Ramsey theory, *Israel J. Math.* **68** (1990), 257–270.

[FuK4] H. Furstenberg and Y. Katznelson, A density version of the Hales-Jewett theorem, *J. d'Analyse Math.* **57** (1991), 64–119.

[FuKO] H. Furstenberg, Y. Katznelson, D. Ornstein, The ergodic theoretical proof of Szemerédi's theorem, *Bull. Amer. Math. Soc. (N.S.)* **7** (1982), no. 3, 527–552.

[FuW1] H. Furstenberg and B. Weiss, Topological dynamics and combinatorial number theory, *J. d'Analyse Math.* **34** (1978), 61–85.

[FuW2] H. Furstenberg and B. Weiss, A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, *Convergence in ergodic theory and probability (Columbus, OH, 1993)*, 193–227, Ohio State Univ. Math. Res. Inst. Publ., 5, de Gruyter, Berlin, 1996.

[Ga] A. Garcia, *Topics in Almost Everywhere Convergence*, Markham, Chicago, 1970.

[GiJ] L. Gillman and M. Jerison, *Rings of Continuous Functions*, Springer-Verlag, New York, 1976.

[GraRS] R. Graham, B. Rothschild and J. Spencer, *Ramsey Theory*, Wiley, New York, 1980.

[Gre] F. Greenleaf, *Invariant Means on Topological Groups*, von Nostrand, New York, 1969.

[HaJ] A.W. Hales and R.I. Jewett, Regularity and positional games, *Trans. Amer. Math. Soc.* **106** (1963), 222–229.

[Hal] P. Halmos, *Lectures on Ergodic Theory*, Mathematical Society of Japan, Tokyo, 1956.

[HalN] P. Halmos and J. von Neumann, Operator methods in classical mechanics, II, *Ann. of Math.* **43** (1942), 332–350.

[HaL] G. Hardy and J. Littlewood, Some problems of Diophantine approximation, *Acta Mathematica* **37** (1914), 155–191.

[Hau] F. Hausdorff, *Grundzüge der Mengenlehre*, Verlag von Veit, Leipzig, 1914. Reprinted by Chelsea, New York, 1949.

[Hel] H. Helson, *Harmonic Analysis*, Addison-Wesley, Reading, MA, 1983.

[HeR] E. Hewitt and K. Ross, *Abstract Harmonic Analysis I*, Springer-Verlag, 1963.

[Hil] D. Hilbert, Über die Irreduzibilität ganzer rationaler Funktionen mit ganzzahligen Koeffizienten, *J. Math.* **110** (1892), 104–129.

[Hin1] N. Hindman, Finite sums from sequences within cells of a partition of $\mathbb{N}$, *J. Combin. Theory* (Ser. A) **17** (1974), 1–11.

[Hin2] N. Hindman, Partitions and sums and products of integers, *Trans. Amer. Math. Soc.* **247** (1979), 227–245.

[HinS] N. Hindman and D. Strauss, *Algebra in the Stone-Čech Compactification. Theory and Applications*, de Gruyter Expositions in Mathematics, 27. Walter de Gruyter & Co., Berlin, 1998. xiv+485 pp.

[Hopf] E. Hopf, *Ergodentheorie,* Chelsea, New York, 1948.

[HoK1] B. Host and B. Kra, Convergence of Conze-Lesigne averages, *Ergod. Th. and Dynam. Sys.* **21** (2001), no. 2, 493–509.

[HoK2] B. Host and B. Kra, Nonconventional ergodic averages and nilmanifolds, to appear in *Ann. of Math.*

[HoK3] B. Host and B. Kra, Convergence of polynomial ergodic averages, to appear in *Israel J. Math.*

[KM] T. Kamae and M. Mendès-France, Van der Corput's difference theorem, *Israel J. Math.* **31** (1978), no. 3-4, 335–342.

[Kh] A. Y. Khintchine, *Three Pearls of Number Theory*, Graylock Press, Rochester, N.Y., 1952.

[KoN] B.O. Koopman and J. von Neumann, Dynamical systems of continuous spectra, *Proc. Nat. Acad. Sci. U.S.A.* **18** (1932), 255–263.

[Kre1] U. Krengel, Weakly wandering vectors and weakly independent partitions, *Trans. Amer. Math. Soc.* **164** (1972), 199–226.

[Kre2] U. Krengel, *Ergodic Theorems*, de Gruyter, Berlin, 1985.

[Kro] L. Kronecker, Die Periodensysteme von Funktionen Reeller Variablen, *Berliner Sitzungsberichte* (1884), 1071–1080.

[La] M. Lacey, On an inequality due to Bourgain, *Illinois J. Math.* **41** (1997), 231–236.

[Le1] A. Leibman, Multiple recurrence theorem for nilpotent group actions, *Geom. Funct. Anal.* **4** (1994), no. 6, 648–659.

[Le2] A. Leibman, Multiple recurrence theorem for measure preserving actions of a nilpotent group *Geom. and Funct. Anal.* **8** (1998), 853–931.

[Le3] A. Leibman, The structure of unitary actions of finitely generated nilpotent groups, *Ergod. Th. and Dynam. Sys.* **20** (2000), 809–820.

[Le4] A. Leibman, Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold, to appear in *Ergod. Th. and Dynam. Sys.*

[Le5] A. Leibman, Convergence of multiple ergodic averages along polynomials of several variables, to appear in *Israel J. Math.*

[Les] E. Lesigne, Sur une nil-variété, les parties minimales assocées à une translation sont uniquement ergodiques, *Ergod. Th. and Dynam. Sys.* **11** (1991), 379-391.

[Li] E. Lindenstrauss, Pointwise theorems for amenable groups, *Invent. Math.* **146** (2001), 259–295.

[McC] R. McCutcheon, *Elemental Methods in Ergodic Ramsey Theory*, Springer Lecture Notes in Mathematics 1722, Springer-Verlag, Berlin, 1999.

[M] K. Milliken, Ramsey's theorem with sums or unions, *J. Combin. Theory* (Ser. A) **18** (1975), 276–290.

[Na] I. Namioka, Følner's condition for amenable semi-groups, *Math. Scand.* **15** (1964), 18–28.

[Ne1] J. von Neumann, Zur algemeinen Theorie des Masses, *Fund. Math.* **13** 1929, 73–116.

[Ne2] J. von Neumann, Zur Operatorenmethode in der klassischen Mechanik, *Ann. of Math.* **33** (1932), 587–642.

[Nu] K. Numakura, On bicompact semigroups, *Math. J. of Okayama University* **1** (1952), 99–108.

[Ol1] A. Olshanskii, On the question of the existence of an invariant mean on a group, *Uspekhi Mat. Nauk.* **35** (1980), no. 4 (214), 199–200.

[Ol2] A. Olshanskii, An infinite group with subgroups of prime orders, *Izv. Akad. Nauk SSSR Ser. Mat.* **44** (1980), no. 2, 309–321.

[OrW] D. Ornstein, B. Weiss, Entropy and isomorphism theorems for actions of amenable groups, *J. d' Analyse Math.* **48** (1987), 1–141.

[Pa] A. Patterson, *Amenability*, Amer. Math. Soc., Providence, 1988.

[Pe] K. Petersen, *Ergodic Theory*, Cambridge Univ. Press, Cambridge, 1981.

[Pier] J.-P. Pier, *Amenable Locally Compact Groups*, John Wiley and Sons, 1984.

[Po] H. Poincaré, Sur le problème des trois corps et les équations de la Dynamique, Acta Mathematica **13** (1890), 1–270.

[Rado] R. Rado, Note on combinatorial analysis, *Proc. London Math. Soc.* **48** (1993), 122–160.

[Ram] F. P. Ramsey, On a problem of formal logic, *Proc. London Math. Soc.* **30** (1930), 264–286.

[Ri] P. Ribenboim, *13 Lectures on Fermat's Last Theorem*, Springer-Verlag, New York, 1979.

[Ros] J. Rosenblatt, Invariant measures and growth conditions, *Trans. Amer. Math. Soc.* **193** (1974), 33–53.

[RosW] J. Rosenblatt and M. Wierdl, Pointwise ergodic theorems via harmonic analysis, *Ergodic Theory and Its Connection with Harmonic Analysis (K. Petersen and I. Salama, eds.)*, 3–151, London Math. Soc. Lecture Note Ser. 205, Cambridge University Press, Cambridge, 1995.

[Roy] H. L. Royden, *Real Analysis*, Macmillan, New York, 1968.

[Sa] A. Sárközy, On difference sets of integers III, *Acta. Math. Acad. Sci. Hungar.* **31** (1978) 125–149.

[Schm] W. Schmidt, *Equations Over Finite Fields: An Elementary Approach*, Springer-Verlag, Berlin ; New York, 1976.

[Schur] I. Schur, Über die Kongruenz $x^m + y^m \equiv z^m \pmod p$, *Jahresbericht der Deutschen Math.-Ver.* **25** (1916), 114–117.

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

[T] A. Taylor, A canonical partition relation for finite subsets of ω, *J. Combin. Theory* (Ser. A) **17** (1974), 1–11.

[vdW1] B. van der Waerden, Beweis einer Baudetschen Vermutung, *Nieuw. Arch. Wisk.* **15** (1927), 212–216.

[vdW2] B. van der Waerden, How the proof of Baudet's conjecture was found, *Studies in Pure Mathematics presented to Richard Rado (1971),* L. Mirsky, ed., Academic Press, London, 251–260.

[Wag] S. Wagon, *The Banach-Tarski Paradox*, Cambridge University Press, Cambridge, 1985.

[Wal] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, New York, 1982.

[Weyl] H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, *Math. Ann.* **77** (1916), 313–352.

[Wi] M. Wierdl, Pointwise ergodic theorem along the prime numbers, *Israel J. Math.* **64** (1988), 315–336.

[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.

[Zie] T. Ziegler, Universal characteristic factors and Furstenberg averages, preprint.

[Zim1] R. Zimmer, Extensions of ergodic group actions, *Illinois J. Math* **20** (1976), no. 3, 373–409.

[Zim2] R. Zimmer, Ergodic actions with generalized discrete spectrum, *Illinois J. Math* **20** (1976), no. 4, 555–588.