Ergodic Theory of $\mathbb{Z}^d$-actions
London Math. Soc. Lecture Note Series 228, 1996, 1–61

Ergodic Ramsey Theory–an Update

VITALY BERGELSON
The Ohio State University
Columbus, OH 43210 U.S.A.

0. Introduction.

This survey is an expanded version and elaboration of the material presented by the author at the *Workshop on Algebraic and Number Theoretic Aspects of Ergodic Theory* which was held in April 1994 as part of the 1993/1994 *Warwick Symposium on Dynamics of* $\mathbf{Z}^n$*-actions and their connections with Commutative Algebra, Number Theory and Statistical Mechanics.* The leitmotif of this paper is: Ramsey theory and ergodic theory of multiple recurrence are two beautiful, tightly intertwined and mutually perpetuating disciplines. The scope of the survey is mostly limited to Ramsey-theoretical and ergodic questions about $\mathbf{Z}^n$–partly because of the proclaimed goals of the Warwick Symposium and partly because of the author's hope that $\mathbf{Z}^n$-related combinatorics, number theory and ergodic theory can serve as an ideal lure through which the author's missionary zeal will reach as wide an audience of potential adherents to the subject as possible.

To compensate for the selective neglect of details and for the lack of full generality in some of the proofs, which were imposed by natural time and space limitations, a significant effort was spent on accentuation and motivation of ideas which lead to conjectures and techniques on which the proofs of conjectures hinge.

Here now is a brief description of the content of the five sections constituting the body of this survey. In Section 1 three main principles of Ramsey theory are introduced and their connection with the ergodic theory of multiple recurrence is emphasized. This section contains a lot of discussion and very few proofs. The goal in this section is to help create in the reader a feeling of what Ergodic Ramsey Theory is all about.

Section 2 is devoted to a multifaceted treatment of a special case of the polynomial ergodic Szemerédi theorem recently obtained in [BL1] (Theorem 1.19 of Section 1). Different approaches are discussed and brought to (hopefully) a convincing level of detail.

In Section 3 the somewhat esoteric, but fascinating and very useful object $\beta\mathbf{N}$, the Stone-Čech compactification of $\mathbf{N}$, is introduced and discussed in some detail. An ultrafilter proof of the celebrated Hindman's theorem is given and some applications of $\beta\mathbf{N}$ and Hindman's theorem to topological dynamics, especially to distal systems, are discussed. This section concludes with a formulation and discussion of an ultrafilter refinement of the

Furstenberg-Sárközy theorem on recurrence along polynomials, and a proof of a special case of this refinement.

Section 4 is devoted to ramifications of results brought in previous sections. Most of the discussion is devoted to polynomial ergodic theorems along IP-sets. In addition, the role of a polynomial refinement of the combinatorial Hales-Jewett theorem is emphasized. The flow of this discussion naturally leads to some open problems which are collected and commented on in Section 5.

I was fortunate to be a graduate student at the Hebrew University of Jerusalem at the time of the inception and early development of Ergodic Ramsey Theory. It is both my duty and pleasure to acknowledge the influence of and express my gratitude to Izzy Katznelson, Benji Weiss, and especially my Ph.D. thesis advisor, Hillel Furstenberg.

I wish to express my indebtedness to my friend and co-author Neil Hindman for many useful discussions of ultrafilter lore.

I also owe a large debt to Randall McCutcheon, whose numerous and most pertinent suggestions for improvements of presentation greatly facilitated the preparation of this survey.

In addition, my thanks go to Boris Begun and Paul Larick for their useful remarks on a preliminary version of this paper.

Finally, I would like to thank the organizers and hosts of the Symposium and the editors of these Proceedings, Bill Parry, Mark Pollicott, Klaus Schmidt, Caroline Series, and Peter Walters for creating a great atmosphere and for their efforts to promote and advance the beautiful science of ergodic theory.

1. Three main principles of Ramsey theory and its connection with the ergodic theory of multiple recurrence.

> A mathematician, like a painter
> or a poet, is a maker of patterns.
>
> –G.H. Hardy, [Ha], p. 84.

Van der Waerden's Theorem, one of Khintchine's "Three Pearls of Number Theory" ([K1]), states that whenever the natural numbers are finitely partitioned (or, as it is customary to say, finitely colored), one of the cells of the partition contains arbitrarily long arithmetic progressions. One can reformulate van der Waerden's theorem in the following, "finitistic" form:

Theorem 1.1. For any natural numbers k and r there exists $N = N(k,r)$ such that whenever $m \geq N$ and $\{1,\cdots,m\} = \bigcup_{i=1}^{r} C_i$, one of C_i, $i = 1,\cdots,r$ contains a k-term arithmetic progression.

Exercise 1. Show the equivalence of van der Waerden's theorem and

Theorem 1.1.

Van der Waerden's theorem belongs to the vast variety of results which form the body of *Ramsey theory* and which have the following general form: If V is an infinite, "highly organized" structure (a semi-group, a vector space, a complete graph, etc.) then for any finite coloring of V there exist arbitrarily large (and sometimes even infinite) highly organized monochromatic substructures. In other words, the high level of organization cannot be destroyed by partitioning into finitely many pieces–one of these pieces will still be highly organized. To fit van der Waerden's theorem into this framework, let us call a subset of $\mathbf{Z}$ *a.p.-rich* if it contains arbitrarily long arithmetic progressions. Then van der Waerden's theorem can be reformulated in the following way:

Theorem 1.2. If $S \subset \mathbf{Z}$ is an a.p.-rich set and, for some $r \in \mathbf{N}$, $S = \bigcup_{i=1}^{r} C_i$, then one of C_i, $i = 1, \cdots, r$ is a.p.-rich.

(Since $\mathbf{Z}$ is a.p.-rich, van der Waerden's theorem is obviously a special case of Theorem 1.2. On the other hand, it is not hard to derive Theorem 1.2 from Theorem 1.1.)

We cannot resist the temptation to bring here two more equivalent forms of van der Waerden's theorem, each revealing still another of its facets.

Theorem 1.3. For any finite partition of $\mathbf{Z}$, one of the cells of the partition contains an affine image of any finite set. (An affine image of a set $F \subset \mathbf{Z}$ is any set of the form $a + bF = \{a + bx : x \in F\}$.)

Exercise 2. Show the equivalence of Theorems 1.3 and 1.1.

Theorem 1.4 (A special case of a theorem due to Furstenberg and Weiss, [FW1]). Suppose $k \in \mathbf{N}$ and $\epsilon > 0$. For any continuous self-mapping of a compact metric space (X, ρ), there exists $x \in X$ and $n \in \mathbf{N}$ such that $\rho(T^{in}x, x) < \epsilon$, $i = 1, \cdots, k$.

Theorem 1.3 shows that van der Waerden's theorem is actually a *geometric* rather than number theoretic fact. On the other hand, Theorem 1.4 establishes the seminal connection between partition theorems of van der Waerden type with *topological dynamics*–the link which proved to be extremely useful.

Another example of "unbreakable" structure is given by Hindman's finite sums theorem ([H2]). To formulate Hindman's theorem let us (following notation in [FW1]) call a set $S \subset \mathbf{N}$ an *IP-set* if it consists of an infinite sequence $(x_n)_{n=1}^{\infty} \subset \mathbf{N}$ together with all finite sums of the form $x_\alpha = \sum_{n \in \alpha} x_n$, where α ranges over the finite non-empty subsets of $\mathbf{N}$.

Theorem 1.5 (Hindman). If $E \subset \mathbf{N}$ is an IP-set, then for any finite coloring $E = \bigcup_{i=1}^{r} C_i$, one of C_i, $i = 1, \cdots, r$ contains an IP-set.

We shall return to Hindman's theorem in the discussions of Section 3. We refer the reader to [GRS] for many more examples illustrating the *first principle of Ramsey theory*–the *preservation of structure under finite partitions.*

After being convinced of the validity of this first principle of Ramsey theory, one is led to the next natural question: why is this so? What exactly is responsible for this stubborn tendency of highly organized infinite structures to preserve their (rightly interpreted) replicas in at least one cell of an arbitrary finite partition? The answer is, and this is the *second principle of Ramsey theory*, that there is always an appropriate notion of *largeness* which is behind the scenes and such that *any* large set contains the sought-after highly organized sub-structures. The only other requirement that the notion of largeness should satisfy is that if A is large and $A = \bigcup_{i=1}^{r} C_i$, then one of C_i, $1 = 1, \cdots, r$ is also large. It is the mathematician's task when dealing with this or that result of partition Ramsey theory to guess what the appropriate notion of largeness responsible for the truth of the proposition is. It is the almost intentional vagueness of the approach which allows one to obtain stronger and stronger theorems by modifying and playing with different notions of largeness. To illustrate this second principle of Ramsey theory we shall now give some examples.

Given a set $A \subset \mathbf{N}$, define its *upper density* $\overline{d}(A)$ by

$$\overline{d}(A) = \limsup_{N\to\infty} \frac{|A \cap \{1, \cdots, N\}|}{N}.$$

If the limit (rather than lim sup) exists, we say that A has density, and denote it by $d(A)$. Being of positive upper density is obviously a notion of largeness and it is natural to ask (as P. Erdös and P. Turán did in [ET]) whether this notion of largeness is responsible for the validity of van der Waerden's theorem. Namely, is it true that any set $A \subset \mathbf{N}$ of positive upper density is a.p.-rich?

The question turned out to be very hard. After some partial results were obtained in [Ro] and [Sz1], Szemerédi [Sz2] settled the Erdös-Turán conjecture affirmatively, thus providing a convenient sufficient condition for a set to be a.p.-rich.

Theorem 1.6 (Szemerédi, [Sz2]). Any set $E \subset \mathbf{N}$ having positive upper density is a.p.-rich.

Exercise 3. Derive from Theorem 1.6 the following finitistic version of it:

For any $\epsilon > 0$ and any $k \in \mathbf{N}$ there exists $N = N(\epsilon, k)$ such that if $m > N$ and $A \subset \{1, 2, \cdots, m\}$ satisfies $\frac{|A|}{m} > \epsilon$, then A contains a k-term arithmetic progression.

It follows from Exercise 3 that in the formulation of Theorem 1.6 a somewhat weaker notion of largeness would do, namely, the notion of *upper Banach density.* Given a set $E \subset \mathbf{Z}$ define its upper Banach density $d^*(E)$ by

$$d^*(E) = \limsup_{N-M\to\infty} \frac{|E \cap \{M, M+1, \cdots, N\}|}{N-M+1}.$$

It is the notion of positive upper Banach density and its natural extensions to $\mathbf{Z}^d$ and, indeed, to any countable *amenable* group which naturally participate in many questions and results of *density Ramsey theory.*

It is easy to check that for any $E \subset \mathbf{Z}$ and any $t \in \mathbf{Z}$ the set $E - t := \{x-t : x \in E\}$ satisfies $d^*(E-t) = d^*(E)$. This shift-invariance of the upper Banach density hints that there is a genuine measure preserving system behind any set $E \subset \mathbf{Z}$ with $d^*(E) > 0$. This is indeed so (see below). On the other hand, the notion of upper Banach density does not provide the right notion of largeness for results like Hindman's theorem. For example, the set $E = 2\mathbf{N}+1$ is large in the sense that $d^*(E) = \frac{1}{2}$ but obviously cannot contain any IP-set, or even any triple of the form $\{x, y, x+y\}$ (see also Exercise 8 in Section 3). We shall see in Section 3 that a notion of largeness relevant for Hindman's theorem is provided by idempotent ultrafilters in $\beta\mathbf{N}$, the Stone-Čech compactification of $\mathbf{N}$. This notion of largeness will also have a mild form of shift invariance which will allow us to prove Hindman's theorem by repeated utilization of a kind of Poincaré recurrence theorem adapted to the situation at hand.

Ergodic Ramsey Theory started with the publication of [F1], in which Furstenberg derived Szemerédi's theorem from a beautiful, far reaching extension of the classical Poincaré recurrence theorem, which corresponds to the case $k = 1$ in the following:

Theorem 1.7 (Furstenberg, [F1]). Let $(X, \mathcal{B}, \mu, T)$ be a probability measure preserving system. For any $k \in \mathbf{N}$ and for any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $n \in \mathbf{N}$ such that

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > 0.$$

In order to derive Szemerédi's Theorem 1.6, Furstenberg introduced a *correspondence principle*, which provides the link between density Ramsey theory and ergodic theory.

Theorem 1.8 *Furstenberg's correspondence principle.* Given a set $E \subset \mathbf{Z}$ with $d^*(E) > 0$ there exists a probability measure preserving system $(X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{B}$, $\mu(A) = d^*(E)$, such that for any $k \in \mathbf{N}$ and any $n_1, \cdots n_k \in \mathbf{Z}$ one has:

$$d^*\big(E \cap (E-n_1) \cap \cdots \cap (E-n_k)\big) \geq \mu(A \cap T^{-n_1}A \cap \cdots \cap T^{-n_k}A).$$

Since the set E contains a progression $\{x, x+n, \cdots, x+kn\}$ if and only if $E \cap (E-n) \cap \cdots \cap (E-kn) \neq \emptyset$, it is clear that Furstenberg's multiple recurrence Theorem 1.7 together with the correspondence principle imply Szemerédi's theorem. We remark that as a matter of fact, Theorem 1.7 follows from Szemerédi's theorem using fairly elementary arguments. Alternatively one can utilize the following refinement of the Poincaré recurrence theorem.

Theorem 1.9 ([B1]). For any probability measure preserving system $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists a sequence $E = (n_m)_{m=1}^{\infty}$ whose density exists and satisfies $d(E) \geq \mu(A)$ such that for any $m \in \mathbf{N}$

$$\mu(A \cap T^{-n_1}A \cap \cdots \cap T^{-n_m}A) > 0.$$

Van der Waerden's theorem has a natural multidimensional extension which is hinted at by the geometric formulation (Theorem 1.3).

Theorem 1.10 *Multidimensional van der Waerden theorem* (Gallai-Grünwald). For any finite coloring of $\mathbf{Z}^d$, $\mathbf{Z}^d = \bigcup_{i=1}^r C_i$, one of C_i, $i = 1, \cdots, r$ contains an affine image of any finite subset $F \subset \mathbf{Z}^d$. In other words, there exists i, $1 \leq i \leq r$, such that for any finite $F \subset \mathbf{Z}^d$, there exists $u \in \mathbf{Z}^d$ and $a \in \mathbf{N}$ such that $u + aF = \{u + ax : x \in F\} \subset C_i$.

Remark. An attribution of Theorem 1.10 to G. Grünwald is made in [Ra], p. 123. As far as we know, Grünwald never published his proof. He later changed his name to Gallai, to whom the result is attributed in [GRS].

In accordance with the second principle of Ramsey theory one should expect that Theorem 1.10 has a density version. This is indeed so and was proved in [FK1]. The multidimensional Szemerédi theorem established by Furstenberg and Katznelson there was the first in a chain of strong combinatorial results ([FK2], [FK4], [BL1]) which were achieved by means of ergodic theory and which so far have no conventional combinatorial proof.

Let us say that a set $S \subset \mathbf{Z}^k$ has positive upper Banach density if for some sequence of parallelepipeds $\Pi_n = [a_n^{(1)}, b_n^{(1)}] \times \cdots \times [a_n^{(k)}, b_n^{(k)}] \subset \mathbf{Z}^k$, $n \in \mathbf{N}$, with $b_n^{(i)} - a_n^{(i)} \to \infty$, $i = 1, \cdots, k$ one has:

$$\frac{|S \cap \Pi_n|}{|\Pi_n|} > \epsilon$$

for some $\epsilon > 0$.

The natural question now is whether it is true that any set of positive upper Banach density in $\mathbf{Z}^k$ contains an affine image of any finite set $F \subset \mathbf{Z}^k$. Furstenberg and Katznelson answered this question affirmatively by deducing the answer from the following generalization of Furstenberg's multiple recurrence theorem.

Theorem 1.11 ([FK1], Theorem A). Let $(X, \mathcal{B}, \mu)$ be a measure space with $\mu(X) < \infty$, let $T_1, \cdots, T_k$ be commuting measure preserving transformations of X and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then

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

Corollary 1.12 ([FK1], Theorem B). Let $S \subset \mathbf{Z}^k$ be a subset with positive upper Banach density and let $F \subset \mathbf{Z}^k$ be a finite configuration. Then there exists a positive integer n and a vector $u \in \mathbf{Z}^k$ such that $u+nF \subset S$.

For the derivation of Corollary 1.12 from Theorem 1.11, the reader is referred to [F2], where the correspondence principle is spelled out for $\mathbf{Z}^k$.

The *third* and last *principle of Ramsey theory* which we want to discuss in this section is the following: *the sought-after configurations always to be found in large sets are abundant.* Abundance in our context means not only that the parameters describing the configurations form large sets in the space of parameters, but also that these parameters are nicely spread in all kinds of families of subsets of integers. Let us consider some examples. Take, for instance, Szemerédi's theorem. Let $E \subset \mathbf{Z}$ with $d^*(E) > 0$. For fixed k the progressions $\{x, x+d, \cdots, x+kd\} \subset E$ are naturally parametrized by pairs (x, d). Let us call a point $x \in E$ a (d, k)-starter if $\{x, x+d, \cdots, x+kd\} \subset E$ and a non-(d, k)-starter otherwise. One can show that for any k, "almost every" point of E is a (d, k)-starter for some d. In other words, for any fixed k, the set of (d, k)-starters in E has upper Banach density equal to $d^*(E)$.

Exercise 4. Show that the set of non-$(d, 2)$-starters of a set $E \subset \mathbf{Z}$ with $d^*(E) > 0$ may be infinite.

Let us turn now to a much more interesting set of those d which appear as differences of arithmetic progressions in E. One of the ways of measuring how well "spread" a subset of integers is, would be to see whether it has a nonempty intersection with different families of subsets of integers (analogy: a set $S \subset [0, 1]$ is dense if for any $0 \le a < b \le 1$, $S \cap (a, b) \ne \emptyset$). We shall need a few definitions. Given a countable abelian group G, a set $S \subset G$ is called *syndetic* if for some finite set $F \subset G$ one has: $S + F = \{x + y : x \in S, y \in F\} = G$. It is easy to see that a set $S \subset \mathbf{Z}$ is syndetic if and only if it has bounded gaps, namely intersects non-trivially any big enough interval. We note that any syndetic set $S \subset \mathbf{Z}$ is a.p.-rich. Indeed, as finitely many shifts of S cover $\mathbf{Z}$ completely, by van der Waerden's theorem one of these shifts is a.p.-rich, and as the property of a.p.-richness is clearly shift invariant, we see that S itself must have this property.

Following the terminology introduced in [F2], given a family $\mathcal{S}$ of subsets of $\mathbf{Z}$ let us call a set $E \subset \mathbf{Z}$ an $\mathcal{S}^*$-set if for any $S \in \mathcal{S}$, $E \cap S \neq \emptyset$. In particular, a set $E \subset \mathbf{Z}$ is an IP*-set if E has non-trivial intersection with any IP-set. It is not hard to see that any IP* set is syndetic. Indeed, if a set E was an IP*-set but not syndetic, its complement would contain a union of intervals $[a_n, b_n]$ with $b_n - a_n \to \infty$. One can easily show that any such union of intervals contains an IP-set which leads to contradiction with the assumed IP*-ness of E.

Exercise 5. Show that for any finite measure preserving system $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$, the set $\{n : \mu(A \cap T^{-n}A) > 0\}$ is an IP*-set.

On the other hand, it is easy to see that not every syndetic set is an IP*-set: take, for example, the odd integers.

Now, IP*-sets are large in a few different senses. Besides having positive lower density (the lower density of a set S is defined as $\liminf\limits_{N\to\infty} \frac{|S\cap[1,\cdots,N]|}{N}$), they, for example, have a finite intersection property.

Lemma 1.13. If $S_1, S_2, \cdots, S_k$ are IP*-sets then $\bigcap_{i=1}^k S_i$ is also an IP*-set.

Proof. It is enough to prove the result for $k = 2$. Let E be an IP-set. Consider the following partition of E: $E = (E \cap S_1) \cup (E \cap S_1^c)$. By Hindman's theorem at least one of $E \cap S_1$, $E \cap S_1^c$ contains an IP-set E_1. Since S_1 is an IP*-set, $E_1 \cap S_1 \neq \emptyset$, hence $E_1 \subset E \cap S_1$. Also S_2 is an IP*-set, hence $E_1 \cap S_2 \neq \emptyset$, which implies that $E \cap (S_1 \cap S_2) \neq \emptyset$. As E was an arbitrary IP-set, the lemma is proved.

□

Given $E \subset \mathbf{Z}$ with $d^*(E) > 0$ let

$$R_k(E) = \big\{d \in \mathbf{Z} : \{x, x+d, \cdots, x+kd\} \subset E \text{ for some } x \in \mathbf{Z}\big\}.$$

The question of how well spread the sets $R_k(E)$ are in $\mathbf{Z}$ is interesting already for $k = 1$. The illustrative results about sets $R_1(E)$ which we collect here are special cases of sometimes very far reaching generalizations. Notice that $R_1(E) = E - E = \{x - y : x, y \in E\}$. It follows immediately from Exercise 5 via Furstenberg's correspondence principle that $R_1(E)$ is an IP*-set. We remark that this result has also a simple completely elementary proof: given an IP-set, generated, say, by $n_1, n_2, \cdots$, one considers the sets

$E_i = E - (n_1 + \cdots + n_i)$, $i = 1, 2, \cdots$ and observes that since $d^*(E) > 0$ we have, for some $1 \leq i < j \leq \frac{1}{d^*(E)} + 1$,

$$d^*(E_i \cap E_j) = d^*\Big(E \cap \big(E - (n_{i+1} + \cdots + n_j)\big)\Big) > 0.$$

This implies that the set of differences $E - E$ contains the element $n_{i+1} + \cdots + n_j$ from our IP-set.

We sketch now a curious application of this circle of ideas to the theory of almost periodic functions. For the sake of simplicity we shall deal only with functions on $\mathbf{Z}$, remarking that easy modifications of these arguments would apply to almost periodic functions on an arbitrary topological group. Recall that, according to H. Bohr ([Bo1]), a function $f\colon \mathbf{Z} \to \mathbf{C}$ is called almost periodic if for any $\epsilon > 0$ the set of "ϵ-periods",

$$E(\epsilon, f) = \{h \in \mathbf{Z} : |f(x+h) - f(x)| < \epsilon \text{ for all } x\}$$

is syndetic. Later Bogolioùboff, [Bo2], and Følner, [Fø] showed that the condition of syndeticity in the definition may be replaced by the weaker condition of positive upper Banach density. This result is contained in the following proposition:

Theorem 1.14. For a function $f\colon \mathbf{Z} \to \mathbf{C}$ the following conditions are equivalent:

(i) For any $\epsilon > 0$ the set $E(\epsilon, f)$ has positive upper Banach density.

(ii) For any $\epsilon > 0$ the set $E(\epsilon, f)$ is syndetic.

(iii) For any $\epsilon > 0$ the set $E(\epsilon, f)$ is an IP*-set.

Proof. It is enough to show that (i)$\rightarrow$(iii). But this follows immediately from two facts:

(1) $E(\frac{\epsilon}{2}, f) - E(\frac{\epsilon}{2}, f) \subset E(\epsilon, f)$.

(2) If $d^*(E) > 0$ then $E - E$ is an IP*-set.

$\square$

As a byproduct one obtains the following fact, which is not obvious from Bohr's definition (but is obvious from some other equivalent definitions of almost periodicity).

Corollary 1.15. If f, g are almost periodic functions then $f + g$ is also an almost periodic function.

Proof. Observe that $E(\frac{\epsilon}{2}, f) \cap E(\frac{\epsilon}{2}, g) \subset E(\epsilon, f + g)$. The result follows from Lemma 1.13.

$\square$

Following Furstenberg ([F3]) let us call a set $R \subset \mathbf{Z}$ a *set of recurrence* if for any invertible finite measure preserving system $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $n \in R$, $n \neq 0$, with $\mu(A \cap T^n A) > 0$.

Exercise 6. Show that the following are sets of recurrence:

(i) Any $E \subset \mathbf{Z}$ with $d^*(E) = 1$.

(ii) $a\mathbf{N} = \{an : n \in \mathbf{N}\}$, for any $0 \neq a \in \mathbf{Z}$.

(iii) $E - E$, for any infinite set $E \subset \mathbf{Z}$.

(iv) Any IP-set.

(v) Any set of the form $\bigcup_{n=1}^{\infty} \{a_n, 2a_n, \cdots, na_n\}$, $a_n \in \mathbf{N}$.

Denote by $\mathcal{R}$ the family of all sets of recurrence in $\mathbf{Z}$. According to our adopted conventions, a set E is $\mathcal{R}^*$ if it intersects nontrivially any set of recurrence. Similarly to IP-sets, sets of recurrence possess the Ramsey property: if a set of recurrence R is finitely partitioned, $R = \bigcup_{i=1}^{r} C_i$, then one of C_i, $i = 1, \cdots, r$ is itself a set of recurrence. To see this, assume that this is not true. So for a set of recurrence R and some partition $R = \bigcup_{i=1}^{r} C_i$ one can find measure preserving systems $(X_i, \mathcal{B}_i, \mu_i, T_i)$ and sets $A_i \in \mathcal{B}_i$, $i = 1, \cdots, r$, with $\mu_i(A_i) > 0$, such that $\mu_i(A_i \cap T_i^n A_i) = 0$ for all $n \in C_i$. Let $(X, \mathcal{B}, \mu, T)$ be the product system of $(X_i, \mathcal{B}_i, \mu_i, T_i)$, $i = 1, \cdots, r$ and take $A = A_1 \times \cdots \times A_r \in \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_r$. Since R is a set of recurrence, there exists $n \in R$ such that $\mu(A \cap T^n A) > 0$, where $\mu = \mu_1 \times \cdots \times \mu_r$, $T = T_1 \times \cdots \times T_r$. This implies that for $i = 1, \cdots, r$, $\mu_i(A_i \cap T_i^n A_i) > 0$ which is a contradiction. The discussion above together with the fact that for any $E \subset \mathbf{Z}$ with $d^*(E) > 0$ the set $E - E$ is an $\mathcal{R}^*$-set imply the following combinatorial fact (cf. [F2], p. 75).

Theorem 1.16. Given sets $E_i \subset \mathbf{Z}$ with $d^*(E_i) > 0$, $i = 1, \cdots, k$, the set $D = (E_1 - E_1) \cap (E_2 - E_2) \cap \cdots \cap (E_k - E_k)$ is $\mathcal{R}^*$. In particular, D is IP* and hence syndetic.

The following fact, due independently to Furstenberg ([F2]) and Sárközy ([S]), provides an example of a set of recurrence of a quite different nature than those of Exercise 6.

Theorem 1.17. Assume that $p(t) \in \mathbf{Q}[t]$ with $p(\mathbf{Z}) \subset \mathbf{Z}$, $\deg p(t) > 0$, and $p(0) = 0$. The the set $\{p(n) : n \in \mathbf{N}\}$ is a set of recurrence.

For more examples and further discussion of sets of recurrence the reader is referred to [F3], [B1], [B2], [BHå], [Fo], and [M]. We comment now on some extensions of these results to multiple recurrence.

Given a finite invertible measure preserving system $(X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{B}$ with $\mu(A) > 0$ consider the set

$$R_k(A) = \{n \in \mathbf{Z} : \mu(A \cap T^n A \cap \cdots \cap T^{kn} A) > 0\}.$$

According to the third principle of Ramsey theory, the sets $R_k(A)$ should be "well spread" in $\mathbf{Z}$. They are. The fact that $R_k(A)$ is syndetic was contained already in the pioneering paper of Furstenberg, [F1]: he established Theorem 1.7 by actually showing that

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

The next question to ask about the sets $R_k(A)$ is whether they are always IP*-sets. This turned out to be true and is a special case of a deep and highly nontrivial "IP-Szemerédi theorem" due to Furstenberg and Katznelson, [FK2]. We give a formulation of a more general fact than IP*-ness of $R_k(A)$ which still is quite a special case of the main theorem in [FK2].

Theorem 1.18 ([FK2]). For any finite measure space $(X, \mathcal{B}, \mu)$, any $k \in \mathbf{N}$, any commuting invertible measure preserving transformations $T_1, \cdots, T_k$ of $(X, \mathcal{B}, \mu)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$ the set

$$\{n : \mu(A \cap T_1^n A \cap \cdots \cap T_k^n A) > 0\}$$

is IP*.

Another desirable refinement of Furstenberg's Szemerédi theorem is hinted upon by Theorem 1.17, an equivalent form of which is that for any invertible measure preserving system $(X, \mathcal{B}, \mu, T)$, any $A \in \mathcal{B}$ with $\mu(A) > 0$, and any polynomial $p(t) \in \mathbf{Q}[t]$ with $p(\mathbf{Z}) \subset \mathbf{Z}$ and $p(0) = 0$, the set $R_1(A) = \{n \in \mathbf{Z} : \mu(A \cap T^n A) > 0\}$ intersects non-trivially the set $p(\mathbf{N}) = \{p(n) : n \in \mathbf{N}\}$. Is it true, for example, that the sets

$$R_k(A) = \{n \in \mathbf{Z} : \mu(A \cap T^n A \cap \cdots \cap T^{kn} A) > 0\}$$

also have such an intersection property? Or, even more ambitiously, does a joint extension of Theorems 1.11 and 1.17 hold? Namely, given any k polynomials $p_i(t) \in \mathbf{Q}[t]$ with $p_i(\mathbf{Z}) \subset \mathbf{Z}$ and $p_i(0) = 0$, $i = 1, \cdots, k$ and any k commuting invertible measure preserving transformations $T_1, \cdots, T_k$ of a probability measure space $(X, \mathcal{B}, \mu)$, is it true that for any $A \in \mathcal{B}$ with $\mu(A) > 0$ one has

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T_1^{p_1(n)} A \cap \cdots \cap T_k^{p_k(n)} A) > 0?$$

It turns out that the answers to these questions are positive and that a stronger, more general result holds.

Theorem 1.19 (Polynomial Szemerédi theorem, [BL1]). Let $(X, \mathcal{B}, \mu)$ be a probability space, let $T_1, \cdots, T_t$ be commuting invertible measure preserving transformations of X, let $p_{i,j}(n)$ be polynomials with rational coefficients, $1 \le i \le k$, $1 \le j \le t$, satisfying $p_{i,j}(0) = 0$ and $p_{i,j}(\mathbf{Z}) \subset \mathbf{Z}$. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then

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

As a corollary we have:

Theorem 1.20. Let $P: \mathbf{Z}^r \to \mathbf{Z}^l$, $r, l \in \mathbf{N}$ be a polynomial mapping satisfying $P(0) = 0$, let $F \subset \mathbf{Z}^r$ be a finite set and let $S \subset \mathbf{Z}^l$ be a set of positive upper Banach density. Then for some $n \in \mathbf{N}$ and $u \in \mathbf{Z}^l$ one has

$$u + P(nF) = \big\{u + P(nx_1, nx_2, \cdots, nx_r) : (x_1, \cdots, x_r) \in F\big\} \subset S.$$

The proof of Theorem 1.19 in [BL1] is, in a sense, a *polynomialization* of the proof of Furstenberg's and Katznelson's multidimensional Szemerédi theorem (Theorem 1.11). We shall try to convey some of the flavor of the proof of Theorem 1.19 in the next section, where we shall treat a special case of it. See also [BM1] where a concise proof of the following refinement of a special case of Theorem 1.19 is given.

Theorem 1.21. Suppose that $(X, \mathcal{B}, \mu, T)$ is an invertible measure preserving system, $k \in \mathbf{N}$, $A \in \mathbf{B}$ with $\mu(A) > 0$ and $p_i(x) \in \mathbf{Q}[x]$ with $p_i(\mathbf{Z}) \subset \mathbf{Z}$ and $p_i(0) = 0$, $1 \le i \le k$. Then

$$\liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T^{p_1(n)} A \cap \cdots \cap T^{p_k(n)} A) > 0.$$

Motivated by the third principle of Ramsey theory, (and by Theorem 1.21), one should expect that Theorem 1.19 has an IP-refinement, similar to the way in which Theorem 1.18 refines Theorem 1.11. This again turns out to be true ([BM2]) and will be discussed in more detail in Section 4. Notice that even in the case of single recurrence along polynomials, it is not obvious at all whether, say, the set $\{n : \mu(A \cap T^{n^2} A) > 0\}$ is an IP*-set. (It is. See Theorem 3.11.)

2. Special case of polynomial Szemerédi theorem: single recurrence.

In this section we shall discuss in detail the following special case of Theorem 1.19, which corresponds to the case $k = 1$.

Theorem 2.1. Suppose $t \in \mathbf{N}$. For any t invertible commuting measure preserving transformations $T_1, T_2, \cdots T_t$ of a probability space $(X, \mathcal{B}, \mu)$, for any $A \in \mathcal{B}$ with $\mu(A) > 0$, and for any polynomials $p_i(x) \in \mathbf{Q}[x]$, $i = 1, \cdots, t$ satisfying $p_i(\mathbf{Z}) \subset \mathbf{Z}$ and $p_i(0) = 0$ one has:

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T_1^{p_1(n)} T_2^{p_2(n)} \cdots T_t^{p_t(n)} A) > 0.$$

Before embarking on the proof we wish to make some remarks and formulate the facts that will be instrumental to the proof.

The first remark that we want to make is that Theorem 2.1, being a result about *single recurrence*, is *Hilbertian* in nature in the sense that it follows from (more general) facts shared by all unitary $\mathbf{Z}^t$-actions, rather than specifically those induced by measure preserving transformations. One, and from some point of view natural way of proving Theorem 2.1 would be to employ the spectral theorem for unitary $\mathbf{Z}^t$-actions. This is done for $t = 1$ in [F2] and [F3] and the extension for general $t \in \mathbf{N}$ goes through with no problems. Unfortunately, the spectral theorem is of no use when one has to deal with multiple recurrence. That is why we prefer to use a "softer", spectral theorem-free approach, which, on the one hand, looks more involved, but on the other hand is more easily susceptible to generalization and refinement.

The main idea which will govern our approach is that in order to prove this or that sort of ergodic theorem, one looks for a suitable splitting of the underlying Hilbert space into orthogonal invariant subspaces on which the behavior of the studied unitary action along, say, polynomials can be well understood and controlled. Consider, for example, the classical mean ergodic theorem of von Neumann for a unitary operator U on a Hilbert space $\mathcal{H}$. The (almost trivial) splitting in this case is $\mathcal{H} = \mathcal{H}_{inv} \oplus \mathcal{H}_{erg}$, where

$$\mathcal{H}_{inv} = \{f \in \mathcal{H} : Uf = f\}, \text{ and}$$

$$\mathcal{H}_{erg} = \overline{\{f \in \mathcal{H} : \text{ there exists } g \in \mathcal{H} \text{ with } f = g - Ug\}}$$

$$= \{f \in \mathcal{H} : \left\| \frac{1}{N} \sum_{n=0}^{N-1} U^n f \right\| \to 0\}.$$

The $\mathbf{Z}$-action generated by U is trivial on the subspace $\mathcal{H}_{inv}$ of invariant elements, whereas it is easily manageable on the *ergodic* subspace $\mathcal{H}_{erg}$.

This splitting is too trivial to help with ergodic theorems along polynomials, but it hints that if one enlarges $\mathcal{H}_{inv}$ just a little bit (and at the same time appropriately shrinks $\mathcal{H}_{erg}$) the situation will be suitable for a polynomial ergodic theorem. Indeed, consider the splitting $\mathcal{H}_{rat} \oplus \mathcal{H}_{tot.erg}$, where

$$\mathcal{H}_{rat} = \overline{\{f \in \mathcal{H}: \text{ there exists } i \in \mathbf{N} \text{ with } U^i f = f\}}, \text{ and}$$

$$\mathcal{H}_{tot.erg} = \{f \in \mathcal{H}: \text{ for all } i \in \mathbf{N}, \|\frac{1}{N}\sum_{n=0}^{N-1} U^{in} f\| \to 0\}.$$

Exercise 7. Show that $\mathcal{H} = \mathcal{H}_{rat} \oplus \mathcal{H}_{tot.erg}$. Attempt not to use the spectral theorem. Show that

$$\mathcal{H}_{tot.erg} = \Big\{f \in \mathcal{H}: \text{ for all } i \in \mathbf{N}, \Big\|\frac{1}{N}\sum_{n=0}^{N-1} U^{-in} f\Big\| \to 0\Big\}.$$

(This fact is related to N. Wiener's speculations about "observable past" and "unattainable future". See [W], Section 1.4.)

Let $p(x) \in \mathbf{Q}[x]$ with $p(\mathbf{Z}) \subset \mathbf{Z}$ and $\deg p(x) > 0$. Let us show how the splitting $\mathcal{H} = \mathcal{H}_{rat} \oplus \mathcal{H}_{tot.erg}$ allows one to establish the existence of the limit

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} U^{p(n)} f.$$

If f belongs to the *rational spectrum* subspace $\mathcal{H}_{rat}$, there is almost nothing to prove: indeed, it is enough to check the case when for some i, $U^i f = f$. If, say, $p(n) = n^2$ and $i = 6$, then

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} U^{n^2} f = \Big(\frac{I + 2U + U^3 + 2U^4}{6}\Big) f.$$

On the other hand, on the *totally ergodic* subspace $\mathcal{H}_{tot.erg}$, the following theorem applies and does the job.

Theorem 2.2 (van der Corput trick). If $(u_n)_{n\in\mathbf{N}}$ is a bounded sequence in a Hilbert space $\mathcal{H}$ and if for any $h \in \mathbf{N}$

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

then $\left\|\frac{1}{N}\sum_{n=1}^{N} u_n\right\| \to 0$.

Proof. Notice that for any $\epsilon > 0$ and any $H \in \mathbf{N}$, if N is large enough one has

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

But

$$\limsup_{N\to\infty} \left\| \frac{1}{N}\frac{1}{H}\sum_{n=1}^{N}\sum_{h=0}^{H-1} u_{n+h} \right\|^2 \le \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \left\| \frac{1}{H}\sum_{h=0}^{H-1} u_{n+h} \right\|^2$$

$$= \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \frac{1}{H^2}\sum_{h_1,h_2=0}^{H-1} \langle u_{n+h_1}, u_{n+h_2}\rangle \le \frac{B}{H},$$

where $B = \sup_n \|u_n\|^2$. Since H was arbitrary, we are done.

□

Remark. The classical van der Corput difference theorem in the theory of uniform distribution says that if $(x_n)_{n\in\mathbf{N}}$ is a sequence of real numbers such that for any $h \in \mathbf{N}$ the sequence $(x_{n+h} - x_n)_{n\in\mathbf{N}}$ is uniformly distributed mod 1 then the sequence $(x_n)_{n\in\mathbf{N}}$ is also uniformly distributed mod 1. It is easy to see that this result follows from the Hilbertian van der Corput trick, applied to the 1-dimensional Hilbert space $\mathbf{C}$. Indeed, by Weyl's criterion, a sequence $(x_n)_{n\in\mathbf{N}}$ is uniformly distributed mod 1 if and only if for any $m \in \mathbf{Z}\setminus\{0\}$,

$$\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i m x_n} \to 0.$$

Writing $u_n^{(m)} = e^{2\pi i m x_n}$, we see that the assumption of van der Corput's difference theorem is that for any $m \in \mathbf{Z}\setminus\{0\}$,

$$\frac{1}{N}\sum_{n=1}^{N} \langle u_{n+h}^{(m)}, u_n^{(m)}\rangle \to 0.$$

Hence, $\frac{1}{N}\sum_{n=1}^{N} u_n^{(m)} \to 0$, which gives the uniform distribution of $(x_n)_{n\in\mathbf{N}}$.

Now, if $f \in \mathcal{H}_{tot.erg}$, then an induction on the degree of the polynomial $p(x)$ reduces the situation to the classical von Neumann theorem. Indeed, if $d_p = \deg p(x) > 1$, then writing $u_n = U^{p(n)}f$, we have:

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

Notice that for any $h \in \mathbf{N}$ the degree of $p(n+h) - p(n)$ equals $d_p - 1$. Since weak convergence follows from strong convergence, we have by the induction hypothesis:

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

and Theorem 2.2 implies that $\left\| \frac{1}{N} \sum_{n=1}^{N} U^{p(n)} f \right\| \to 0$.

Let us explain now how the splitting $\mathcal{H} = \mathcal{H}_{rat} \oplus \mathcal{H}_{tot.erg}$ can be used to establish the existence of the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_1^{p_1(n)} U_2^{p_2(n)} \cdots U_t^{p_t(n)} f$$

for $t > 1$. Take for simplicity $t = 2$ (it will be clear from the discussion that the proof for general t is completely analogous). Let U_1, U_2 be commuting unitary operators generating a unitary $\mathbf{Z}^2$-action on a Hilbert space $\mathcal{H}$. Consider the splittings $\mathcal{H} = \mathcal{H}_{rat}^{(i)} \oplus \mathcal{H}_{tot.erg}^{(i)}$, $i = 1, 2$ which correspond to U_1 and U_2. We have the decomposition into invariant subspaces $\mathcal{H} = \mathcal{H}_{rr} \oplus \mathcal{H}_{rt} \oplus \mathcal{H}_{tr} \oplus \mathcal{H}_{tt}$, where $\mathcal{H}_{rr} = \mathcal{H}_{rat}^{(1)} \cap \mathcal{H}_{rat}^{(2)}$, $\mathcal{H}_{rt} = \mathcal{H}_{rat}^{(1)} \cap \mathcal{H}_{tot.erg}^{(2)}$, $\mathcal{H}_{tr} = \mathcal{H}_{tot.erg}^{(1)} \cap \mathcal{H}_{rat}^{(2)}$, and $\mathcal{H}_{tt} = \mathcal{H}_{tot.erg}^{(1)} \cap \mathcal{H}_{tot.erg}^{(2)}$. For $f \in \mathcal{H}$ we have the corresponding decomposition $f = f_{rr} + f_{rt} + f_{tr} + f_{tt}$, where $f_{\alpha\beta} \in \mathcal{H}_{\alpha\beta}$, $\alpha, \beta \in \{r, t\}$. Now, in each of the subspaces $\mathcal{H}_{rr}$, $\mathcal{H}_{rt}$, and $\mathcal{H}_{tr}$, the problem of establishing the existence of $\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_1^{p_1(n)} U_2^{p(n)} f$ is reducible to the case of $\mathbf{Z}$-actions, already discussed above. Consider, for example, $f \in \mathcal{H}_{tr}$. Since $f \in \mathcal{H}_{rat}^{(2)}$ we can assume without loss of generality that for some $a \in \mathbf{N}$, $U_2^a f = f$. Notice that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_1^{p_1(n)} U_2^{p_2(n)} f$$

exists if each of

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U_1^{p_1(an+r)} U_2^{p_2(an+r)} f; \quad r = 0, 1, \cdots, a-1$$

exists. As $U_2^{p(an+r)} f$ does not depend on n (since we assumed that $U_2^a f = f$), we see that in this case (i.e. on the invariant subspace $\mathcal{H}_{tr}$) the problem is reduced to that of the $\mathbf{Z}$-action generated by U_1.

It remains to show the existence of the limit in question on $\mathcal{H}_{tt}$. Let $f \in \mathcal{H}_{tt}$. First of all, assume without loss of generality that there do not exist non-zero $a, b \in \mathbf{Z}$ and $g \in \mathcal{H}_{tt}$, $g \neq 0$, with $U_1^a U_2^b g = g$ (the set of such $g \in \mathcal{H}_{tt}$ comprises a U_1 and U_2-invariant subspace on which the situation again reduces to $\mathbf{Z}$-actions). Under this assumption, one shows that the limit is zero. The result follows by induction on $d = \max\{\deg p_1(x), \deg p_2(x)\}$. For $d = 1$ one has $p_1(n) = c_1 n$, $p_2(n) = c_2 n$ where at least one of c_1, c_2 is non-zero. In this case we are done by von Neumann's ergodic theorem. If $d > 1$ then, as before, van der Corput's trick reduces the case to $d - 1$.

The case of general $t > 1$ is treated in a similar fashion. We have 2^t invariant subspaces; on all but one of them at least one of the generators $U_1, \cdots, U_t$ has rational spectrum. On these $2^t - 1$ spaces the situation is naturally reduced to that of a $\mathbf{Z}^n$-action, $n < t$. On the remaining subspace, call it $\mathcal{H}_t$, on which all of $U_1, \cdots, U_t$ are totally ergodic, one disposes first with the potential degeneration caused by "linear dependence" between $U_1, \cdots, U_t$; again, on any subspace of $\mathcal{H}_t$ where such a dependence exists, the situation reduces to that of a lower-dimensional unitary action. Finally, if no degeneration occurs, one shows that the limit is zero by induction on $\max_{1 \leq i \leq t} \deg p_i(x)$ with the help of the van der Corput trick.

To establish Theorem 2.1, it suffices to show that for the unitary $\mathbf{Z}^t$-action induced by measure preserving transformations T_i, $i = 1, \cdots, t$ on $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$ and for the characteristic function $f = 1_A$, where $\mu(A) > 0$, the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T_1^{p_1(n)} T_2^{p_2(n)} \cdots T_t^{p_t(n)} A)$$
$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle U_1^{p_1(n)} U_2^{p_2(n)} \cdots U_t^{p_t(n)} f, f\rangle$$

is positive.

Here is one possible way to see this (we will offer below another proof of the positivity of the limit). Let $L^2(X, \mathcal{B}, \mu) = \mathcal{H} = \mathcal{H}_{rat}^{(i)} \oplus \mathcal{H}_{tot.erg}^{(i)}$, $i = 1, 2, \cdots t$, be the splittings corresponding to the unitary operators U_i, $i = 1, 2, \cdots, t$, defined by $(U_i f)(x) := f(T_i x)$, $f \in L^2(X, \mathcal{B}, \mu)$. We have then a natural splitting of $\mathcal{H}$ into 2^t invariant subspaces each having the form

$$\mathcal{H}_C = \Big(\bigcap_{i\in C} \mathcal{H}_{rat}^{(i)}\Big) \cap \Big(\bigcap_{i\in\{1,2,\cdots,t\}\setminus C} \mathcal{H}_{tot.erg}^{(i)}\Big),$$

where C is any of the 2^t subsets of $\{1, 2, \cdots, t\}$. Let $f = \sum_{C\subset\{1,2,\cdots,t\}} f_C$ be the corresponding orthonormal decomposition of $f = 1_A$.

We claim that without loss of generality one can assume that the polynomials $p_1(n), \cdots, p_t(n)$ all have distinct (and positive) degrees. Indeed one can always arrive at such a situation by regrouping and, possibly, collapsing some of the U_i. (Example: $U_1^{2n^2+3n}U_2^{5n^2-n}U_3^{-n^2} = \tilde{U}_1^{n^2}\tilde{U}_2^{n}$, where $\tilde{U}_1 = U_1^2U_2^5U_3^{-1}$, $\tilde{U}_2 = U_1^3U_2^{-1}$.)

Now, if the polynomials $p_1(n), \cdots, p_t(n)$ have pairwise distinct positive degrees, then one can check, by carefully examining the effect of the van der Corput trick, that on each of the subspaces $\mathcal{H}_C$, C a proper subset of $\{1, 2, \cdots, t\}$, the limit in question is zero, so that

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \mu\big(A \cap T_1^{p_1(n)}T_2^{p_2(n)}\cdots T_t^{p_t(n)}A\big)\\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \langle U_1^{p_1(n)}U_2^{p_2(n)}\cdots U_t^{p_t(n)}f, f\rangle\\
&= \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \langle U_1^{p_1(n)}U_2^{p_2(n)}\cdots U_t^{p_t(n)}\overline{f}, \overline{f}\rangle,
\end{aligned}$$

where $\overline{f} = f_{\{1,2,\cdots,t\}} = Pf$ and P is the projection onto the subspace $\bigcap_{i=1}^t \mathcal{H}_{rat}^{(i)} = \overline{\bigcup_{a\in\mathbf{Z}^t}\mathcal{H}_a}$, where the (potentially containing constants only) subspaces $\mathcal{H}_a$, $a = (a_1, \cdots, a_t) \in \mathbf{Z}^t$ are defined by

$$\mathcal{H}_a = \{f \in \mathcal{H} : U_i^{a_i}f = f,\ i = 1, \cdots, t\}.$$

Now, since $f = 1_A \geq 0$, and since $f \not\equiv 0$, one has $\overline{f} \geq 0$, $\overline{f} \not\equiv 0$. Indeed, $\overline{f}$ minimizes the distance from $\bigcap_{i=1}^t \mathcal{H}_{rat}^{(i)}$ to f and $\max\{\overline{f}, 0\}$ would do at least as well in minimizing this distance (cf. [F2], Lemma 4.23). By the same token, for any $a \in \mathbf{Z}^t$, the projection f_a of $f = 1_A$ onto $\mathcal{H}_a$ satisfies $f_a \geq 0$, $f_a \not\equiv 0$. Since the limit in question is strictly positive for any such f_a (and since the subspaces $\mathcal{H}_a$ span $\bigcap_{i=1}^t \mathcal{H}_{rat}^{(i)}$) we are done.

We shall consider now still another, and from the author's point of view, most important splitting theorem. An appropriate generalization of this splitting plays a significant role in the proofs of Theorems 1.11 and 1.19. Again, the splitting which we are going to introduce follows easily from the spectral theorem, but may be proved without resorting to it. A form of it appears for the first time in [KN].

Theorem 2.3. For any unitary $\mathbf{Z}$-action $(U^n)_{n\in\mathbf{Z}}$ on a Hilbert space $\mathcal{H}$ one has a decomposition $\mathcal{H} = \mathcal{H}_c \oplus \mathcal{H}_{wm}$, where

$$\mathcal{H}_c = \{f \in \mathcal{H} :\ \text{the orbit } (U^nf)_{n\in\mathbf{Z}} \text{ is precompact in norm topology}\},$$

$$\mathcal{H}_{wm} = \{f \in \mathcal{H} : \text{ for all } g \in \mathcal{H}, \frac{1}{N}\sum_{n=0}^{N-1} |\langle U^n f, g\rangle| \to 0\}.$$

Remark. One can show that the space $\mathcal{H}_c$ of *compact elements* coincides with the span of the eigenvectors of U:

$$\mathcal{H}_c = \overline{\text{span}\{f \in \mathcal{H} : \text{ there exists } \lambda \in \mathbf{C} \text{ with } Uf = \lambda f\}}.$$

The orthocomplement of $\mathcal{H}_c$, the space $\mathcal{H}_{wm}$ on which U acts in a *weakly mixing* manner can be characterized as follows:

$$\begin{aligned}\mathcal{H}_{wm} = \{f \in \mathcal{H} : \text{ there exists } S \subset \mathbf{N}, d(S) = 1 \\ \text{such that for all } g \in \mathcal{H}, \lim_{n\to\infty, n\in S} \langle U^n f, g\rangle = 0\}.\end{aligned}$$

(This terminology comes from measurable ergodic theory where one normally deals with operators induced by measure preserving transformations.)

Let us show now how the splitting $\mathcal{H} = \mathcal{H}_c \oplus \mathcal{H}_{wm}$ allows one to prove the existence of the limit in Theorem 2.1. Consider first of all the case $t = 1$. Since $\mathcal{H}_{wm} \subset \mathcal{H}_{tot.erg}$ and since on $\mathcal{H}_{tot.erg}$, as we saw above, the van der Corput trick, together with von Neumann's ergodic theorem, does the job, we have to care only about the space $\mathcal{H}_c$. One possibility is to use the characterization of $\mathcal{H}_c$ given in the remark above. Upon momentary reflection the reader will agree that without loss of generality one has to consider only

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} U^{p(n)} f,$$

where f satisfies $Uf = \lambda f$, $\lambda \in \mathbf{C}$, $|\lambda| = 1$. Since the situation on $\mathcal{H}_{rat}$ was already discussed above, assume additionally that $f \in \mathcal{H}_c \setminus \mathcal{H}_{rat}$, i.e., assume that for no integer $k \neq 0$ is $\lambda^k = 1$. Then, by Weyl's theorem on equidistribution of polynomials mod 1 one gets $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \lambda^{p(n)} = 0$ and we are done. The extension of the proof to the case of general $t > 0$ is done in complete similarity to the proof above in which the splitting $\mathcal{H} = \mathcal{H}_{rat} \oplus \mathcal{H}_{tot.erg}$ was utilized.

Having in mind generalizations in the direction of multiple recurrence we want to indicate now still another approach which, while utilizing the same splitting $\mathcal{H} = \mathcal{H}_c \oplus \mathcal{H}_{wm}$, is "soft" enough to be susceptible to generalization. This is the gain; the loss is that, unlike the approaches discussed above, this approach leads only to results like

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A \cap T_1^{p_1(n)} T_2^{p_2(n)} \cdots T_t^{p_t(n)} A) > 0$$

instead of

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mu(A\cap T_1^{p_1(n)}T_2^{p_2(n)}\cdots T_t^{p_t(n)}A)>0.$$

Theorem 2.4. Assume that $U_1,\cdots,U_t$ are commuting unitary operators on a Hilbert space $\mathcal{H}$. Let $p_i(x)\in\mathbf{Q}[x]$ with $p_i(\mathbf{Z})\subset\mathbf{Z}$ and $p_i(0)=0$, $i=1,\cdots,t$. If for any $f\in\mathcal{H}$ and for any i, $i=1,\cdots,t$ the orbit $(U_i^n f)_{n\in\mathbf{Z}}$ is precompact in the strong topology, then for any $\epsilon>0$ the set

$$\{n\in\mathbf{Z}:\|U_1^{p_1(n)}U_2^{p_2(n)}\cdots U_t^{p_t(n)}f-f\|\le\epsilon\}$$

is IP* and hence syndetic.

In the proof of Theorem 2.4 we shall utilize the Gallai-Grünwald theorem (Theorem 1.10) in the following refined form, which can be derived from Theorem 3.2 in [FW1] as well as from the Hales-Jewett theorem (see Exercise 16 in Section 4). See also theorem 2.18 in [F2].

Theorem 2.5. If $\mathbf{Z}^t=\bigcup_{i=1}^r C_i$ is an r-coloring of $\mathbf{Z}^t$, then one of C_i, $i=1,\cdots,r$ contains a "t-cube" of the form

$$\begin{aligned}&K(n_1,n_2,\cdots,n_t;h)\\&=\big\{(n_1+\epsilon_1 h,n_2+\epsilon_2 h,\cdots,n_t+\epsilon_t h):\epsilon_i\in\{0,1\},i=1,2,\cdots,t\big\}.\end{aligned}$$

The set of $h\in\mathbf{Z}$ such that for some $(n_1,n_2,\cdots,n_t)$ the t-cube $K(n_1,n_2,\cdots,n_t;h)$ is contained in one of the C_i is an IP*-set.

Proof of Theorem 2.4. For $f\in\mathcal{H}$, let X be the closure in the strong topology of $\mathcal{H}$ of the orbit $(U_1^{n_1}U_2^{n_2}\cdots U_t^{n_t}f)_{(n_1,\cdots,n_t)\in\mathbf{Z}^t}$. It is easy to see that one can assume without loss of generality that $p_i(x)\in\mathbf{Z}[x]$, $i=1,2,\cdots,t$. By the increasing if needed the number of commuting operators involved, we may and will assume that the polynomials $p_i(n)$, $i=1,\cdots,t$, have the form $p_i(n)=n^{b_i}$, $b_i\ge 1$. (For example, if $p_1(n)=5n^2-17n^3$, we would rewrite $U_1^{p_1(n)}$ as $T^{n^2}S^{n^3}$ where $T=U_1^5$, $S=U_1^{-17}$.) Given $\epsilon>0$, let

$$\epsilon_0=\frac{\epsilon}{\sum_{i=1}^t 2^{b_i+1}}$$

and let $(g_j)_{j=1}^r$ be an ϵ_0-net in the (compact metric) space X. For each i, $i=1,2,\cdots,t$ consider the r-coloring of $\mathbf{Z}^{b_i}$ defined by

$$\chi_i(n_1,n_2,\cdots,n_{b_i})=\min\{j:\|U_i^{n_1n_2\cdots n_{b_i}}f-g_j\|\le\epsilon_0\}.$$

By Theorem 2.5, for each i there exist χ_i-monochrome b_i-cubes $K(n_1^{(i)}, n_2^{(i)}, \cdots, n_{b_i}^{(i)}; h_i)$. Let S_i, $i = 1, 2, \cdots, t$ be the sets of "sizes" h of such monochrome cubes:

$$\begin{aligned} S_i = \{h : &\text{ there exists } (n_1, \cdots, n_{b_i}) \in \mathbf{Z}^{b_i} \\ &\text{with } K(n_1, \cdots, n_{b_i}; h)\ \chi_i - \text{monochrome}\}. \end{aligned}$$

Again by Theorem 2.5 the sets S_i are IP* and by Lemma 1.13 the set $S = \bigcap_{i=1}^t S_i$ is also an IP*-set.

We shall utilize the following identity:

$$h^b = \sum_{a=0}^{b} \sum_{A \subset \{1,2,\cdots,b\}, |A|=a} (-1)^a \prod_{i \in A} n_i \prod_{i \notin A} (n_i + h).$$

(For example, for $b = 3$ one has:

$$\begin{aligned} h^3 = &(n_1 + h)(n_2 + h)(n_3 + h) - n_1(n_2 + h)(n_3 + h) \\ &- (n_1 + h)n_2(n_3 + h) - (n_1 + h)(n_2 + h)n_3 + n_1 n_2 (n_3 + h) \\ &+ n_1(n_2 + h)n_3 + (n_1 + h)n_2 n_3 - n_1 n_2 n_3.) \end{aligned}$$

Notice that for any b the sum of coefficients

$$\sum_{a=0}^{b} \sum_{\substack{A \subset \{1,2,\cdots,b\} \\ |A|=a}} (-1)^a$$

equals zero and the number of them equals 2^b.

Since for every two vertices $v' = (v'_1, \cdots, v'_{b_i})$, $v'' = (v''_1, \cdots, v''_{b_i})$ of a monochrome b_i-cube $K(n_1^{(i)}, \cdots, n_{b_i}^{(i)}; h)$ one has

$$\|U_i^{v'_1 v'_2 \cdots v'_{b_i}} f - U_i^{v''_1 v''_2 \cdots v''_{b_i}} f\| \leq 2\epsilon_0,$$

it follows from the identity above that for any $h \in S_i$, $i = 1, 2, \cdots, t$

$$\|U_i^{h^{b_i}} f - f\| \leq 2^{b_i+1}\epsilon_0.$$

It is clear then that for any $h \in \bigcap_{i=1}^t S_i$ one has:

$$\|U_1^{h^{b_1}} U_2^{h^{b_2}} \cdots U_t^{h^{b_t}} f - f\| \leq \sum_{i=1}^{t} 2^{b_i+1}\epsilon_0 = \epsilon.$$

This finishes the proof of Theorem 2.4.

$\square$

Let us show now how Theorem 2.4 can be used for a proof of Theorem 2.1. As before, let U_i, $i = 1, 2, \cdots, t$ be unitary operators on $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$ defined by $(U_i f)(x) = f(T_i x)$.

Let $\mathcal{H} = \mathcal{H}_c^{(i)} \oplus \mathcal{H}_{wm}^{(i)}$ be the corresponding splittings. For each $B \subset \{1, 2, \cdots, t\}$ we have the $U_1, \cdots, U_t$-invariant subspace

$$\mathcal{H}_B = \Big(\bigcap_{i \notin B} \mathcal{H}_c^{(i)} \Big) \cap \Big(\bigcap_{i \in B} \mathcal{H}_{wm}^{(i)} \Big).$$

Then $\mathcal{H} = \bigoplus_{B \subset \{1,2,\cdots,t\}} \mathcal{H}_B$. For the same reasons as before we shall assume without loss of generality that on each $\mathcal{H}_B$ no degeneration caused by "linear dependence" between $U_1, U_2, \cdots, U_t$ occurs.

Let $f = 1_A$ and consider the corresponding decomposition

$$f = \sum_{B \subset \{1,2,\cdots,t\}} f_B,$$

where $f_B \in \mathcal{H}_B$.

Taking into account that for distinct B the spaces $\mathcal{H}_B$ are mutually orthogonal, we have:

$$\begin{aligned} \mu(A \cap T_1^{p_1(n)} T_2^{p_2(n)} \cdots T_t^{p_t(n)} A) &= \langle f, U_1^{p_1(n)} U_2^{p_2(n)} \cdots U_t^{p_t(n)} f \rangle \\ &= \sum_{B \subset \{1,2,\cdots,t\}} \langle f_B, U_1^{p_1(n)} U_2^{p_2(n)} \cdots U_t^{p_t(n)} f_B \rangle. \end{aligned}$$

One can show with the help of the van der Corput trick that under our assumptions, for any $B \neq \emptyset$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle f_B, U_1^{p_1(n)} U_2^{p_2(n)} \cdots U_t^{p_t(n)} f_B \rangle = 0,$$

so that

$$\begin{aligned} &\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T_1^{p_1(n)} T_2^{p_2(n)} \cdots T_t^{p_t(n)} A) \\ = &\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle \tilde{f}, U_1^{p_1(n)} U_2^{p_2(n)} \cdots U_t^{p_t(n)} \tilde{f} \rangle, \end{aligned}$$

where $\tilde{f}$ is the component of f in the subspace $\bigcap_{i=1}^t \mathcal{H}_c^{(i)}$.

As with the component $\overline{f} \in \bigcap_{i=1}^{t} \mathcal{H}_{rat}^{(i)}$ above, and for the same reasons, the function $\tilde{f}$ satisfies $\tilde{f} \geq 0$, $\tilde{f} \not\equiv 0$. Also, by Theorem 2.4, for any $\epsilon > 0$ the set

$$S_\epsilon = \{n \in \mathbf{Z} : \|U_1^{p_1(n)} U_2^{p_2(n)} \cdots U_t^{p_t(n)} \tilde{f} - \tilde{f}\| < \epsilon\}$$

is syndetic. Therefore, if ϵ is small enough, we shall have

$$\begin{aligned}
&\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T_1^{p_1(n)} T_2^{p_2(n)} \cdots T_t^{p_t(n)} A) \\
=&\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle \tilde{f}, U_1^{p_1(n)} U_2^{p_2(n)} \cdots U_t^{p_t(n)} \tilde{f}\rangle \\
\geq&\liminf_{N\to\infty} \frac{1}{N} \sum_{n\in S_\epsilon \cap [0,N-1]} \langle \tilde{f}, U_1^{p_1(n)} U_2^{p_2(n)} \cdots U_t^{p_t(n)} \tilde{f}\rangle > 0.
\end{aligned}$$

3. Discourse on βN and some of its applications.

In this section we shall address among other things the question of what the density version of Hindman's theorem (Theorem 1.5) is. It will turn out, somewhat surprisingly, that, appropriately formulated, Hindman's theorem is its own density version! First of all, recall from Section 1 that positive upper Banach density does not provide us with the right notion of largeness for Hindman's theorem. One of the reasons that upper Banach density seems to have nothing to do with Hindman's theorem is that it, unlike Szemerédi's theorem, which deals with *shift-invariant* configurations, deals with configurations which are not shift-invariant. On the other hand, any set of positive upper Banach density contains plenty of configurations of the form

$$t + FS(x_j)_{j=1}^n = t + \Big\{ \sum_{j\in\alpha} x_j : \ \emptyset \neq \alpha \subset \{1, 2, \cdots, n\}\Big\}.$$

Here "plenty" means, in particular, that if $E \subset \mathbf{Z}$ with $d^*(E) > 0$, then for any n, there are $x_1, \cdots, x_n$ such that

$$d^*\{t : t + FS(x_j)_{j=1}^n \subset E\} > 0.$$

This result has a very simple proof which we shall describe now. Let $E \subset \mathbf{Z}$, $d^*(E) > 0$. As observed above (see the speculations following the proof of Lemma 1.13), there exists x_1 with $1 \leq x_1 \leq \frac{1}{d^*(E)} + 1$ satisfying $d^*\big(E \cap (E - x_1)\big) > 0$. Repeating this argument with the set $E \cap (E - x_1)$ leads to finding $x_2 > x_1$ with

$$\begin{aligned}
&d^*\Big(\big(E \cap (E - x_1)\big) \cap \Big(\big(E \cap (E - x_1)\big) - x_2\Big)\Big) \\
&\qquad = d^*\big(E \cap (E - x_1) \cap (E - x_2) \cap (E - (x_1 + x_2))\big) > 0.
\end{aligned}$$

It is clear that after $n-2$ additional steps we will arrive at

$$d^*\Big(E \cap \bigcap_{x \in FS(x_j)_{j=1}^n} (E-x)\Big) > 0.$$

Any t belonging to this intersection will satisfy $t + FS(x_j)_{j=1}^n \subset E$ and we are done.

Being as simple as it is, the result proved just now is a strengthening of a key lemma needed by D. Hilbert in his famous paper [H1] where he showed among other things that if $f(x,y) \in \mathbf{Z}[x,y]$ is an irreducible polynomial then for some $x_0 \in \mathbf{Z}$ the polynomial of one variable $f(x_0, y)$ is also irreducible. The interesting thing is that although Hilbert's lemma is weaker than the density version of it just proved, namely he shows that for any finite coloring $\mathbf{N} = \bigcup_{i=1}^r C_i$ and for any n, one of C_i, $i = 1, \cdots, r$ has the property that for some x_j, $j = 1, 2, \cdots, n$ and for infinitely many t one has $t + FS(x_j)_{j=1}^n \subset C_i$, his proof is two pages long. Another interesting thing is that the proof we have indicated contains in embryo the main idea behind the proof of the much stronger Hindman's theorem that we are going to pass to now. Before starting, we want to formulate an Exercise which shows that it is hopeless to look for infinite configurations of the form $t + FS(x_j)_{j=1}^\infty$ in arbitrary sets of positive upper density.

Exercise 8. Show that for any $\epsilon > 0$ there exists a set $E \subset \mathbf{N}$ with $d(E) > 1 - \epsilon$ and such that E does not contain a subset of the form $t + FS(x_j)_{j=1}^\infty$. (The reader may try first to produce E with $d^*(E) > 1 - \epsilon$. This is much simpler.)

Though the original proof of Hindman's theorem in [H2] (as well as that in [Ba]) was combinatorial in nature, the real key to understanding Hindman's theorem lies in $\beta\mathbf{N}$–the Stone-Čech compactification of $\mathbf{N}$. More precisely, it is the algebraic structure of $\beta\mathbf{N}$, naturally inherited from addition in $\mathbf{N}$, that is behind a proof of Hindman's theorem which we want to present in this section. We shall also indicate in this section some other applications of $\beta\mathbf{N}$.

Since according to the author's experience most mathematicians (unless they are logicians or set-theoretical topologists) are not overly knowledgeable about or thrilled by $\beta\mathbf{N}$ in general and its algebraic structure in particular, we shall start with some generalities. For more details the reader is encouraged to consult, say [GJ], [C], or Sections 6-9 in [H3]. It was the paper [H3] which convinced the author of the effectiveness of ultrafilters in solving Ramsey-theoretical questions and indeed in the ergodic theory of multiple recurrence.

We take $\beta\mathbf{N}$, the Stone-Čech compactification of $\mathbf{N}$ to be the set of *ultrafilters* on $\mathbf{N}$. Recall that a filter p on $\mathbf{N}$ is a set of subsets of $\mathbf{N}$ satisfying

(i) $\emptyset \notin p$,
(ii) $A \in p$ and $A \subset B$ implies $B \in p$, and
(iii) $A \in p$ and $B \in p$ implies $A \cap B \in p$.

A filter is an ultrafilter if, in addition, it satisfies

(iv) if $r \in \mathbf{N}$ and $\mathbf{N} = A_1 \cup A_2 \cup \cdots \cup A_r$ then $A_i \in p$ for some i, $1 \leq i \leq r$.

In other words, an ultrafilter is a maximal filter. An alternative way of looking at ultrafilters (and actually the one that we shall adopt) is to identify each ultrafilter $p \in \beta\mathbf{N}$ with a finitely additive, $\{0,1\}$-valued probability measure μ_p on the power set $\mathcal{P}(\mathbf{N})$. The measure μ_p is naturally defined by the condition $\mu_p(A) = 1$ if and only if $A \in p$. Without saying so explicitly, we will always think of ultrafilters as such measures, but will prefer to write $A \in p$ instead of $\mu_p(A) = 1$. It is their $\{0,1\}$-valuedness and finite additivity (as well as abundance of ultrafilters with diversified properties) which makes ultrafilters so suitable for Ramsey theory: the property of being a member of an ultrafilter fits nicely into the notion of largeness which the second principle of Ramsey theory encourages us to look for.

Any element $n \in \mathbf{N}$ is naturally identified with an ultrafilter $\{A \subset \mathbf{N} : n \in A\}$. Such (and only such) ultrafilters are called *principal*. A natural question which the shrewd reader may ask at this point is: are there any less dull examples of ultrafilters? The answer is yes ... modulo Zorn's Lemma, which the reader is kindly encouraged to accept. The reader is also asked not to attempt to produce a non-principal ultrafilter without the use of Zorn's Lemma (it will not work). See the discussion on pp. 161-162 of [CN].

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

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

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 $\beta\mathbf{N}$. To see that non-principal ultrafilters exist take for example

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

Clearly $\mathcal{C}$ has the finite intersection property, so there is an ultrafilter $p \in \beta\mathbf{N}$ 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 $\mathcal{D} = \{A \subset \mathbf{N} : d(A) = 1\}$. 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 $A^c = (\mathbf{N} \setminus A) \in \mathcal{D}$.) These examples hint that the space $\beta\mathbf{N}$ is quite large. It is indeed: the cardinality of $\beta\mathbf{N}$ equals that of $\mathcal{P}\big(\mathcal{P}(\mathbf{N})\big)$ ([GJ], 6.10 (a)).

Now a few words about topology in $\beta\mathbf{N}$. Given $A \subset \mathbf{N}$, let $\overline{A} = \{p \in \beta\mathbf{N} : A \in p\}$. The set $\mathcal{G} = \{\overline{A} : A \subset \mathbf{N}\}$ 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 $\beta\mathbf{N}$ observe that if $A, B \subset \mathbf{N}$, then $\overline{A} \cap \overline{B} = \overline{A \cap B}$. Also, $\overline{\mathbf{N}} = \beta\mathbf{N}$ and hence $\bigcup_{\overline{A} \in \mathcal{G}} \overline{A} = \beta\mathbf{N}$. (Notice also that $\overline{A} \cup \overline{B} = \overline{A \cup B}$.) The crucial fact for us is that, with this topology, $\beta\mathbf{N}$ satisfies the following.

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

Proof. Let $\mathcal{K}$ be a cover of $\beta\mathbf{N}$ by sets belonging to the base $\mathcal{G} = \{\overline{A} : A \subset \mathbf{N}\}$. Let $\mathcal{C} \subset \mathcal{P}(\mathbf{N})$ 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 $\beta\mathbf{N}$, 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 $\bigcup_{i=1}^r A_i = \mathbf{N}$, which implies that $\bigcup_{i=1}^r \overline{A}_i = \beta\mathbf{N}$. Again, this is a contradiction, as we assumed that $\mathcal{K}$ has no finite subcover.

As for the Hausdorff property, notice that if $p, q \in \beta\mathbf{N}$ 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 $A \in p \setminus q$, then $A^c \in q \setminus p$, which means that $\overline{A}$ and $\overline{A^c}$ are disjoint neighborhoods of p and q.

□

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

Since $\overline{\mathbf{N}} = \beta\mathbf{N}$, it is natural to attempt to extend the operation of addition from (the densely embedded) $\mathbf{N}$ to $\beta\mathbf{N}$. Since ultrafilters are measures (principal ultrafilters being just the *point measures* corresponding to the elements of $\mathbf{N}$), 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 $\beta\mathbf{N}$ was explicitly introduced only about 35 years ago (in [CY]).

In the following definition, $A - n$ (where $A \subset \mathbf{N}$, $n \in \mathbf{N}$) is the set of all m for which $m + n \in A$. For $p, q \in \mathbf{N}$, define

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

Exercise 9. Check that for principal ultrafilters the operation + corresponds to addition in $\mathbf{N}$.

Remarks. 1. Despite the somewhat repelling 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. [HR], 19.11):

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

2. 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 $(\mathbf{N}, +)$ 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 $\{n \in \mathbf{N} : (A - n) \in p\} \in q$ and $\{n \in \mathbf{N} : (B - n) \in p\} \in q$. Since p and q are ultrafilters, we have:

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

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

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

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

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

Proof. Let $q \in \beta\mathbf{N}$ 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 $A \subset \mathbf{N}$ 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

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

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 $B = \{n \in \mathbf{N} : (A - n) \in p\} \in r$. This means that $A \in p + r = \lambda_p(r)$, or $\lambda_p(r) \in \overline{A} \in \mathcal{U}$.

□

With the operation $+$, $\beta\mathbf{N}$ becomes, in view of Theorem 3.2, a compact *left topological* semigroup. Such semigroups are known to have idempotents, which is the last preliminary result that we need for the proof of Hindman's theorem. For compact *topological semigroups* (i.e. with an operation which is continuous in both variables), this result is due to Numakura, [N]; for left topological semigroups the result is due to Ellis, [E1].

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.

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 non-empty. 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 $x * x = x$.

□

For a fixed $p \in \beta\mathbf{N}$ we shall call a set $C \subset \mathbf{N}$ *p-big* if $C \in p$. Clearly, being a member of an ultrafilter is a notion of largeness in the sense discussed in Section 1. The notion of largeness induced by idempotent ultrafilters is special (and promising) in that it inherently has a shift-invariance property. Indeed, if $p \in \beta\mathbf{N}$ with $p+p=p$ then

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

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 $n \in \mathbf{N}$ the shifted set $(A-n)$ is p-big. Or, still somewhat differently: $A \subset \mathbf{N}$ is p-big if for p-almost all $n \in \mathbf{N}$ 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).

Exercise 10. Let $p \in \beta\mathbf{N}$ with $p+p=p$.
(i) Show that p cannot be a principal ultrafilter.
(ii) Show that for any $a \in \mathbf{N}$, $a\mathbf{N} \in p$.

Each idempotent ultrafilter $p \in \beta\mathbf{N}$ induces a "measure preserving dynamical system" with the phase space $\mathbf{N}$, σ-algebra $\mathcal{P}(\mathbf{N})$, measure p, and "time" being the "p-preserving" $\mathbf{N}$-action induced by the shift. The two peculiarites 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, $A \cap (A-n) \in p$. We are now in position to prove Hindman's theorem which we rephrase slightly as follows.

Theorem 3.4. Let $p \in \beta\mathbf{N}$ be an idempotent and let $\mathbf{N} = \bigcup_{i=1}^{r} C_i$. If $C = C_i$ is p-big, then there exists an infinite sequence $(x_j)_{j=1}^{\infty} \subset C$ such that

$$FS\big((x_j)_{j=1}^{\infty}\big) = \Big\{ \sum_{j\in\alpha} x_j : \emptyset \neq \alpha \subset \mathbf{N},\ |\alpha| < \infty \Big\} \subset C.$$

Proof. Let $p \in \beta\mathbf{N}$ be an idempotent ultrafilter. Assume that $\mathbf{N} = \bigcup_{i=1}^{r} C_i$ and choose $i \in \{1, \cdots, r\}$ such that $C = C_i \in p$. Since $C \in p$ and $p+p=p$, one has $R_1 = \{n \in \mathbf{N} : (C-n) \in p\} \in p$. Choose any $x_1 \in R_1 \cap C \in p$. Then $x_1 \in C$ and $A_1 = C \cap (C-x_1) \in p$. Consider

$$R_2 = \{n \in \mathbf{N} : A_1 - n = \big((C \cap (C-x_1)-n\big) \in p\} \in p.$$

Choose any $x_2 \in R_2 \cap A_1 \in p$. Since p is an idempotent, it is a non-principal ultrafilter and its members are infinite sets. This allows us always to assume in the course of this proof that an element chosen from a member of p lies outside of any chosen finite set. In particular, we may assume that $x_2 > x_1$. Notice that

$$x_2 \in A_1 = C \cap (C - x_1) \Rightarrow \{x_1, x_2, x_1 + x_2\} = FS(x_j)_{j=1}^2 \subset C.$$

Now, since $x_2 \in R_2$,

$$A_2 = A_1 \cap (A_1 - x_2) = C \cap (C - x_1) \cap \big(C \cap (C - x_1) - x_2\big)$$

$$= C \cap (C - x_1) \cap (C - x_2) \cap (C - x_1 - x_2) \in p.$$

And so on! At the nth step we will have

$$A_k = A_{k-1} \cap (A_{k-1} - x_k) = C \cap \Bigg(\bigcap_{t \in FS(x_j)_{j=1}^k} (C - t) \Bigg) \in p,$$

and defining

$$R_k = \{n \in \mathbf{N} : (A_k - n) \in p\} \in p$$

and choosing $x_{k+1} \in R_{k+1} \cap A_k$, $x_{k+1} > x_1 + \cdots + x_k$, we will have:

$$x_{k+1} \in \bigcap_{t \in FS(x_j)_{j=1}^k} (C - t),$$

$$A_{k+1} = A_k \cap (A_k - x_{k+1}) = C \cap \Bigg(\bigcap_{t \in FS(x_j)_{j=1}^{k+1}} (C - t) \Bigg) \in p.$$

The sequence $(x_j)_{j=1}^\infty$ thus created has the property that for any $n \in \mathbf{N}$, $FS(x_j)_{j=1}^n \subset C$. Hence $FS(x_j)_{j=1}^\infty \subset C$ and we are done.

$\square$

Remarks. 1. The proof of Theorem 3.4 which we have presented here is essentially due to S. Glaser, who never published it. For an account of the story behind the discovery of Glaser's proof, see [H5].

2. Theorem 3.4 tells us that a notion of largeness which is behind Hindman's theorem is that of being a member of an idempotent ultrafilter. In a sense, it is *the* notion. Indeed, though the IP-set $FS(x_j)_{j=1}^\infty$ which in the course of the proof was found inside a p-big set C is not necessarily itself p-big, one can show that for any IP-set S there exists an idempotent p such that S is p-big. One can say that Hindman's theorem is the density version of itself!

One can show that $\beta\mathbf{N}$ has 2^c idempotents, where c is the power of continuum (this follows from the fact that $\beta\mathbf{N}$ has 2^c disjoint closed subsemigroups. See for example [D]). One can expect that some of these idempotents are good not only for Hindman's theorem but, say, for van der Waerden's theorem too. As we remarked above, any IP-set is a member of an idempotent. Since there are "thin" IP-sets that do not contain length 3 arithmetic progressions (take for example $FS(5^n)_{n=1}^{\infty}$), it is clear that not every idempotent may reveal something about van der Waerden's theorem. But, there is an important and natural class of idempotents, namely *minimal* ones, which do (this became clear after Furstenberg and Katznelson put to use in [FK3] the minimal idempotents in Ellis' enveloping semigroup–an object similar in many ways to $\beta\mathbf{N}$). An idempotent $p \in \beta\mathbf{N}$ is called minimal if it belongs to a minimal right ideal in $\beta\mathbf{N}$. One can show that any minimal right ideal in $\beta\mathbf{N}$ has the form $q + \beta\mathbf{N}$ (warning: not every right ideal has this form).

This terminology fits well with the usual definition of minimality in topological dynamics. Recall that a topological dynamical system (X, T), where X is a compact space and T is a continuous self-map of X, is called minimal if for any $x \in X$ one has $\overline{(T^n x)_{n\in\mathbf{N}}} = X$. Notice that for each $p \in \beta\mathbf{N}$ the right ideal $p + \beta\mathbf{N}$ is compact (since $p + \beta\mathbf{N}$ is the continuous image of $\beta\mathbf{N}$ under the continuous function λ_p–see Theorem 3.2). Let $X_p = p + \beta\mathbf{N}$ and define $T: X_p \to X_p$ by $Tx = x + 1$, $x \in X_p$ (here 1 is the principal ultrafilter of sets containing the integer 1). It is not hard to show that the ideal $p+\beta\mathbf{N}$ is minimal if and only if the dynamical system (X_p, T) is minimal.

One can also show that any minimal idempotent $p \in \beta\mathbf{N}$ has the property that if $C \in p$ then C is *piecewise syndetic*, namely the intersection of a syndetic set with a union of intervals $[a_n, b_n]$, where $b_n - a_n \to \infty$.

Exercise 11. (i) Show that a set $S \subset \mathbf{N}$ is piecewise syndetic if and only if there exist $t_1, \cdots, t_k \in \mathbf{N}$ such that $d^*\big(\bigcup_{i=1}^{k}(S - t_i)\big) = 1$.

(ii) Derive from (i) that any piecewise syndetic set is a.p.-rich.

Theorem 3.5. If $\mathbf{N} = \bigcup_{1=1}^{r} C_i$ then one of C_i, $i = 1, \cdots, r$ is a.p.-rich and contains an IP-set.

Proof. Take any minimal idempotent $p \in \beta\mathbf{N}$. If $C_i \in p$ then by Theorem 3.4, C_i contains an IP-set. Also, since p is minimal, C_i is piecewise syndetic and hence a.p.-rich.

□

The combinatorial results mentioned in this section so far can be obtained also by methods of topological dynamics introduced in [FW1] and further developed in [F2] and [FK3]. As a matter of fact, one can show

([BH1]) that so-called *central sets*, which are defined dynamically and which are shown in [F2] to be combinatorially rich, are exactly the members of minimal idempotents in $\beta\mathbf{N}$. The following is an example of a result for whose proof the space of ultrafilters seems better suited.

Theorem 3.6 ([BH1]). For any finite coloring of $\mathbf{N}$, one of the cells contains arbitrarily long arithmetic progressions, arbitrarily long geometric progressions, an additive IP-set, and a multiplicative IP-set.

A *multiplicative* IP-set is, in analogy with the additive case, any infinite sequence together with all finite products of its distinct elements. The reason that $\beta\mathbf{N}$ is an appropriate tool for proving the foregoing result lies with the fact that it has also another semigroup structure–the one inherited from the usual multiplication in $\mathbf{N}$, and is a left topological compact semigroup with respect to this structure, too. In particular, there are (many) *multiplicative* idempotents, and in complete analogy to the proof of Theorem 3.4 one can show that any member of a multiplicative idempotent contains a multiplicative IP-set. It follows that for any finite coloring $\mathbf{N} = \bigcup_{i=1}^{r} C_i$, there exists a monochromatic additive IP-set and a monochromatic multiplicative IP-set, but not necessarily of the same color. It was shown by Hindman in [H4] that a single C_i, $i = 1, \cdots, r$ always contains both an additive and a multiplicative IP-set. This as well as the other assertions of Theorem 3.6 are consequences of the following result combining the two structures in $\mathbf{N}$.

Theorem 3.7 ([BH1], Corollary 5.5). For any finite partition $\mathbf{N} = \bigcup_{i=1}^{r} C_i$, one of C_i, $i = 1, \cdots, r$, is a member of a minimal additive idempotent and also a member of a minimal multiplicative idempotent.

It would be interesting to find a dynamical proof of Theorem 3.6.

The discussion above shows that ultrafilters are a convenient tool in partition Ramsey theory. We shall try to explain now why idempotent ultrafilters (and IP-sets which are intrinsically related to them) are helpful in topological dynamics and ergodic theory, and through this, in density Ramsey theory. This discussion will continue into the next section where we shall touch upon issues of multiple recurrence and further refinements of the polynomial Szemerédi theorem. Throughout this section we shall be concerned with single recurrence (both topological and measure theoretical). Since we are mainly concerned with $\mathbf{Z}$ and $\mathbf{Z}^t$ actions we shall confine ourselves to dealing exclusively with the additive structure of $\beta\mathbf{N}$.

Let X be a topological space and let $p \in \beta\mathbf{N}$. Given a sequence $(x_n)_{n\in\mathbf{N}}$ we shall write $p\text{-}\lim_{n\in\mathbf{N}} x_n = y$ if for every neighborhood U of y one has $\{n : x_n \in U\} \in p$. It is easy to see that $p\text{-}\lim_{n\in\mathbf{N}} x_n$ exists and is unique in any compact Hausdorff space.

The following is a special case of 6.10 in [BH1].

Theorem 3.8. Let X be a compact Hausdorff space, let $p, q \in \beta\mathbf{N}$ and let $(x_n)_{n\in\mathbf{N}}$ be a sequence in X. Then

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

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

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

Proof. Recall that

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

Let $x = q+p\text{-}\lim_{r\in\mathbf{N}} x_r$. It will suffice for us to show that for any neighborhood U of x, we have that for p-many t, $q\text{-}\lim_{s\in\mathbf{N}} x_{s+t} \in U$. Fix such a U. We have $\{r : x_r \in U\} \in q + p$, so that

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

This implies, in particular, that for p-many t, $q\text{-}\lim_{s\in\mathbf{N}} x_{s+t} \in U$.

□

Here is a simple application of Theorem 3.8. Recall that a continuous self-map of a compact metric space X is called *distal* if

$$\inf_{n\geq 1} d(T^n x, T^n y) = 0 \Rightarrow x = y.$$

The innocent looking fact that distal transformations are invertible is not transparent from the definition. Using Theorem 3.8 we can prove this as follows.

Let $T\colon X \to X$ be a distal transformation of a compact metric space X. Since a distal map is clearly injective we need only show that it is onto. Fix an idempotent $p \in \beta\mathbf{N}$. We shall show something stonger, namely that for any $x \in X$ $p\text{-}\lim_{n\in\mathbf{N}} T^n x = x$ and hence $x = T(p\text{-}\lim_{n\in\mathbf{N}} T^{n-1}x)$. Let $y = p\text{-}\lim_{n\in\mathbf{N}} T^n x$. By Theorem 3.8 one has

$$\begin{aligned} p\text{-}\lim_{n\in\mathbf{N}} T^n y &= p\text{-}\lim_{n\in\mathbf{N}} T^n (p\text{-}\lim_{k\in\mathbf{N}} T^k x) \\ &= p\text{-}\lim_{n\in\mathbf{N}} p\text{-}\lim_{k\in\mathbf{N}} T^{n+k} x = p\text{-}\lim_{n\in\mathbf{N}} T^n x = y. \end{aligned}$$

Since T is distal, the relations $p\text{-}\lim_{n\in\mathbf{N}} T^n x = y = p\text{-}\lim_{n\in\mathbf{N}} T^n y$ imply $x = y$. We are done.

The conventional proof of the fact that distal transformations are invertible goes by way of the Ellis enveloping semigroup (see for example [Br], p. 60, or [E2]), which is typically non-metrizable. Our proof uses $\beta\mathbf{N}$, which is also non-metrizable, but a modification of the proof using the notion of IP-convergence (which will be explained below) in place of p-limits can serve to place this fact entirely within the scope of metric spaces.

Recall that a point $x \in X$ is uniformly recurrent with respect to $T: X \to X$ if for any neighborhood U of x the set $\{n : T^n x \in U\}$ is syndetic. It is well known that if x is a uniformly recurrent point then the orbit closure $\overline{(T^n x)_{n\in\mathbf{N}}}$ is a minimal T-invariant subset of X. It is not hard to show (see 6.9 in [BH1]) that if $p \in \beta\mathbf{N}$ is a minimal idempotent, and $p\text{-}\lim_{n\in\mathbf{N}} T^n y = y$ then y is a uniformly recurrent point. It follows that in a distal system every point is uniformly recurrent. This in its turn implies by a routine argument (see for instance Thm. 1.17 in [F2]) that any distal system is *semisimple*, namely, the disjoint union of minimal subsets.

Let us indicate now how one can modify proofs involving limits along idempotent ultrafilters so that there will be no reference to non-metrizable spaces. The notation that we will introduce will also be used in the next section.

Let $\mathcal{F}$ denote the family of non-empty finite subsets of $\mathbf{N}$. Following the notation in [F2] and [FK2], let us call any sequence indexed by $\mathcal{F}$ an $\mathcal{F}$-*sequence*. Notice that an IP-set in $\mathbf{Z}$ is nothing but an $\mathcal{F}$-sequence $(n_\alpha)_{\alpha\in\mathcal{F}}$ with the property that $n_\alpha + n_\beta = n_{\alpha\cup\beta}$ whenever $\alpha \cap \beta = \emptyset$. For $\alpha, \beta \in \mathcal{F}$ we shall write $\alpha < \beta$ if $\max\alpha < \min\beta$.

If a collection of sets $\alpha_i \in \mathcal{F}$, $i \in \mathbf{N}$ has the property $\alpha_i < \alpha_{i+1}$, $i = 1, 2, \cdots$, then the set

$$\mathcal{F}^{(1)} = \{\bigcup_{i\in\beta} \alpha_i : \beta \in \mathcal{F}\}$$

is called an *IP-ring*. Note that the mapping $\varphi: \mathcal{F} \to \mathcal{F}^{(1)}$, $\varphi(\beta) = \bigcup_{i\in\beta} \alpha_i$ is bijective and structure preserving.

Since elements of $\mathcal{F}^{(1)}$ are naturally indexed by elements of $\mathcal{F}$, any sequence indexed by $\mathcal{F}^{(1)}$ may itself be viewed as an $\mathcal{F}$-sequence. The following exercise is an equivalent form of Hindman's theorem.

Exercise 12. If $\mathcal{F} = \bigcup_{i=1}^{r} C_i$, then one of C_i, $i = 1, \cdots, r$ contains an IP-ring.

Let $(x_\alpha)_{\alpha\in\mathcal{F}}$ be an $\mathcal{F}$-sequence in a topological space X, let $x \in X$, and let $\mathcal{F}^{(1)}$ be an IP-ring. One writes

$$\underset{\alpha\in\mathcal{F}^{(1)}}{\text{IP-lim}}\ x_\alpha = x$$

if for any neighborhood U of x there exists $\alpha_U \in \mathcal{F}^{(1)}$ such that for any $\alpha \in \mathcal{F}^{(1)}$ with $\alpha > \alpha_U$ one has $x_\alpha \in U$.

The following theorem is Theorem 8.14 in [F2] and is proved by a diagonal procedure with the help of Hindman's theorem as formulated in Exercise 12.

Theorem 3.9. If $(x_\alpha)_{\alpha\in\mathcal{F}}$ is an $\mathcal{F}$-sequence with values in a compact space X, then there exists $x \in X$ and an IP-ring $\mathcal{F}^{(1)}$ such that

$$\underset{\alpha\in\mathcal{F}^{(1)}}{\text{IP-lim}}\ x_\alpha = x.$$

The following result is a special case of Lemma 8.15 in [F2].

Theorem 3.10. Let T be a continuous self map of a compact metric space X and let $(n_\alpha)_{\alpha\in\mathcal{F}} \subset \mathbf{N}$ be an IP-set. Then for any $x \in X$ there exists an IP-ring $\mathcal{F}^{(1)}$ and a point $y \in X$ such that

$$\underset{\alpha\in\mathcal{F}^{(1)}}{\text{IP-lim}}\ T^{n_\alpha}x = \underset{\alpha\in\mathcal{F}^{(1)}}{\text{IP-lim}}\ T^{n_\alpha}y = y.$$

We leave it now safely to the reader to verify that the proof of the invertibility of distal transformations given above can be rewritten in the language of IP-limits with no significant changes.

As we shall see in the next section, besides being more constructive, IP-limits allow refined formulations and non-trivial strengthenings of polynomial ergodic theorems. On the other hand, limits along idempotent ultrafilters, when applicable, have the advantage of there being no need to constantly be passing to convergent subsequences, thereby making the statements and proofs cleaner and more albebraic.

We shall conclude this section by presenting an ultrafilter proof of a special case of the following refinement of the Furstenberg-Sárközy theorem, Theorem 1.17. (At the same time this refinement is a special case of a more general theorem proved in [BFM] which will be discussed in the next section.)

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

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

In particular, if $\deg p(n) > 0$ then $\{p(n) : n \in (n_\alpha)_{\alpha\in\mathcal{F}}\}$ is a set of recurrence for any IP-set $(n_\alpha)_{\alpha\in\mathcal{F}}$.

In a situation similar to that of Section 2, Theorem 3.11 follows from a Hilbertian result which we now formulate in the language of ultrafilters:

Theorem 3.12. Let $q(t) \in \mathbf{Q}[t]$ with $q(\mathbf{Z}) \subset \mathbf{Z}$ and $q(0) = 0$. Let U be a unitary operator on a Hilbert space $\mathcal{H}$. Let $p \in \beta\mathbf{N}$ be an idempotent. Then, letting $p\text{-}\lim_{n\in\mathbf{N}} U^{q(n)} f = P_p f$, where the limit is in the weak topology, P_p is an orthogonal projection onto a subspace of $\mathcal{H}$.

To see that Theorem 3.11 follows from Theorem 3.12, take $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$, and take U to be the unitary operator induced by T, that is, $(Ug)(x) := g(Tx)$, and let $f = 1_A$. One then has

$$p\text{-}\lim_{n\in\mathbf{N}} \mu(A \cap T^{q(n)} A) = p\text{-}\lim_{n\in\mathbf{N}} \langle U^{q(n)} f, f\rangle = \langle P_p f, f\rangle$$

$$= \langle P_p f, P_p f\rangle\langle 1, 1\rangle \geq \langle P_p f, 1\rangle^2 = \langle f, 1\rangle^2 = \langle 1_A, 1\rangle^2 = (\mu(A))^2.$$

We leave to the reader to justify the replacement of $p\text{-}\lim_{n\in\mathbf{N}} \mu(A \cap T^{q(n)} A)$ by

$$\text{IP-}\lim_{\alpha\in\mathcal{F}^{(1)}} \mu(A \cap T^{q(n)} A).$$

(Hint: any member of any idempotent contains an IP-set, and any IP-set is a member of some idempotent.)

Proof of Theorem 3.12 (for $q(n) = n^2$). As the reader (after reading Section 2) may guess, the proof boils down to finding an appropriate splitting of $\mathcal{H}$. This guess is true but the situation is more complicated when compared with the one which was encountered in Section 2. As we saw in Section 2, when one is interested in studying limits of the form

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U^{q(n)} f,$$

the splitting $\mathcal{H} = \mathcal{H}_c \oplus \mathcal{H}_{wm}$ works *for all* $q(t) \in \mathbf{Z}[t]$. Our case here is more delicate. It may occur, for example, that for some $f \in \mathcal{H}$, $p\text{-}\lim_{n\in\mathbf{N}} U^{an} f = f$ weakly for all $a \in \mathbf{N}$ but $p\text{-}\lim_{n\in\mathbf{N}} U^{n^2} f = 0$ weakly.

It follows that the splittings which enable one to distinguish between different kinds of asymptotic behaviour of $U^{q(n)}$ along $p \in \beta\mathbf{N}$ may depend on the polynomial $q(n)$. To convey the gist of the proof we shall utilize a splitting which allows one to prove the theorem for $q(n) = n^2$ (and indeed for any other quadratic polynomial). Recall that any ball of fixed radius is compact in the weak topology, which will assure the existence of the weak p-limits that we shall deal with in the course of the proof.

Let $P_p f = p\text{-}\lim_{n\in\mathbf{N}} U^{n^2} f$, $f \in \mathcal{H}$. Put

$$\mathcal{H}_r = \overline{\{f \in \mathcal{H} : \text{ there exists } a \in \mathbf{N} \text{ with } p\text{-}\lim_{n\in\mathbf{N}} U^{an} f = f\}}.$$

Notice that since $\|U^{an}f\| = \|f\|$, the relation $p\text{-}\lim_{n\in\mathbf{N}} U^{an}f = f$ may be thought of as valid in the sense of both strong and weak convergence. It is not hard to see that the orthocomplement of $\mathcal{H}_r$, which we shall denote by $\mathcal{H}_m$, may be described as follows:

$$\mathcal{H}_m = \{f \in \mathcal{H} : \text{ for all } a \in \mathbf{N},\ \underset{n\in\mathbf{N}}{p\text{-}\lim}\, U^{an}f = 0 \text{ weakly }\}.$$

Let us start by showing that $P_pf = 0$ for $f \in \mathcal{H}_m$. To this end we shall need the following IP-version of the van der Corput trick:

If $(x_n)_{n\in\mathbf{N}} \subset \mathcal{H}$ is a bounded sequence and if for any $h \in \mathbf{N}$ we have

$$\underset{n\in\mathbf{N}}{p\text{-}\lim}\langle x_{n+h}, x_n\rangle = 0,$$

then $p\text{-}\lim_{n\in\mathbf{N}}\langle x_n, g\rangle = 0$ for any $g \in \mathcal{H}$.

The proof of this is almost trivial: without loss of generality consider $g \in \operatorname{span}(x_n)_{n\in\mathbf{N}}$ and use the hypotheses. The fact that for $f \in \mathcal{H}_m$, $p\text{-}\lim_{n\in\mathbf{N}} U^{n^2}f = 0$ weakly follows routinely now with the help of the IP van der Corput trick.

Let us consider now the space $\mathcal{H}_r$. As $\|P_pf\| \le \|f\|$ for all $f \in \mathcal{H}$, in order to show that P_p is a projection, all we have left to establish is that $P_p^2f = P_pf$ for all $f \in \mathcal{H}_r$. Since $\mathcal{H}_r = \overline{\bigcup_{a\in\mathbf{N}}\mathcal{H}_a}$, where the potentially trivial spaces $\mathcal{H}_a$, $a \in \mathbf{N}$, are defined by

$$\mathcal{H}_a = \{f \in \mathcal{H} : \underset{n\in\mathbf{N}}{p\text{-}\lim}\, U^{an}f = f\},$$

we may assume without loss of generality that for some $a \in \mathbf{N}$,

$$\underset{n\in\mathbf{N}}{p\text{-}\lim}\, U^{an}f = f.$$

Notice that this implies that for any b dividing a one also has

$$\underset{n\in\mathbf{N}}{p\text{-}\lim}\, U^{bn}f = f$$

(since the convergence in this case is strong as well as weak). We shall need to make use of Exercise 10 (ii), which enables us to take p-limits along those $n \in \mathbf{N}$ which are divisible by any prescribed integer. This will be expressed by the notation $p\text{-}\lim_{n\in\mathbf{N},a|n}$. We also remark that since for any m with $a|m$, $p\text{-}\lim_{n\in\mathbf{N}} U^{mn}f = f$, for any such m we have $p\text{-}\lim_{n\in\mathbf{N}} U^{n^2}f = p\text{-}\lim_{n\in\mathbf{N}} U^{n^2+mn}f$.

We now may write

$$\begin{aligned}
P_p^2 f &= \underset{m\in\mathbf{N}}{p\text{-lim}}\, U^{m^2}\big(\underset{n\in\mathbf{N}}{p\text{-lim}}\, U^{n^2} f\big) = \underset{m\in\mathbf{N},a|m}{p\text{-lim}}\, U^{m^2}\big(\underset{n\in\mathbf{N}}{p\text{-lim}}\, U^{n^2} f\big) \\
&= \underset{m\in\mathbf{N},a|m}{p\text{-lim}}\, U^{m^2}\big(\underset{n\in\mathbf{N}}{p\text{-lim}}\, U^{n^2+2mn} f\big)\, \underset{m\in\mathbf{N}}{p\text{-lim}}\,\underset{n\in\mathbf{N}}{p\text{-lim}}\, U^{(n+m)^2} f \\
&= \underset{n\in\mathbf{N}}{p\text{-lim}}\, U^{n^2} f = P_p f.
\end{aligned}$$

(We have used at one stage weak continuity of the operator U^{m^2}, and at another stage Theorem 3.8.) This finishes the proof in the case $q(n) = n^2$.

4. IP-polynomials, recurrence, and polynomial Hales-Jewett theorem.

The results discussed in the previous sections, especially Theorems 1.18, 1.19, 1.21, and 3.11 indicate that there are two types of subsets of $\mathbf{Z}$ which stand out as being related to particularly nice refinements of ergodic theorems pertaining to single and multiple recurrence: IP-sets and *polynomial sets*, namely sets of the form $p(\mathbf{Z}) = \{p(n) : n \in \mathbf{Z}\}$, where $p(n) \in \mathbf{Q}[n]$, $p(\mathbf{Z}) \subset \mathbf{Z}$, and $p(0) = 0$. As regards IP-sets, these may be viewed as kinds of generalized additive subsemigroups of $\mathbf{Z}$. It is so since given an IP-set $(x_\alpha)_{\alpha\in\mathcal{F}}$ generated by the sequence $(x_j)_{j=1}^{\infty} \subset \mathbf{Z}$ (where $\mathcal{F}$ is the set of non-empty finite subsets of $\mathbf{N}$ and $x_\alpha := \sum_{j\in\alpha} x_j$, $\alpha \in \mathcal{F}$), the commutative and associative partial operation defined by the formula $x_\alpha + x_\beta = x_{\alpha\cup\beta}$ has only one flaw: it is only valid when $\alpha\cap\beta = \emptyset$. However, since one generally deals in treatments such as ours with limits of expressions in which the parameter (in this case $\alpha \in \mathcal{F}$) goes to infinity, this limitation turns out not to hinder us, particularly in light of the fact that in *IP ergodic theory* one deals with IP-convergence rather than with the Cesaro convergence typical of classical ergodic theory.

On the other hand, a fundamental property of polynomials is that after finitely many applications of the difference operator they become linear. This often allows one to apply an inductive procedure deducing results for polynomials of a certain degree from similar results for polynomials of lesser degree.

The following weakly mixing PET (Polynomial Ergodic Theorem) is obtained by an application of this procedure. (It is at the same time an important special case of Theorem 1.19. See also [BM1].) Recall that a measure preserving system $(X, \mathcal{B}, \mu, T)$ is weakly mixing if the only eigenfunctions are the constants, that is, if $f(Tx) = \lambda f(x)$ for $f \in L^2(X, \mathcal{B}, \mu)$, $\lambda \in \mathbf{C}$ implies that f =const a.e.

Theorem 4.1 ([B4]). Suppose that $(X, \mathcal{B}, \mu, T)$ is a weakly mixing system and let $p_i(t) \in \mathbf{Q}[t]$, $i = 1, 2, \cdots, k$ be polynomials satisfying $p_i(\mathbf{Z}) \subset$

$\mathbf{Z}$, $\deg p_i(t) > 0$, and $\deg\big(p_i(t) - p_j(t)\big) > 0$, $1 \le i \ne j \le k$. Then for any $f_1, \cdots, f_k \in L^\infty(X, \mathcal{B}, \mu)$ one has

$$\lim_{N\to\infty} \Big\| \frac{1}{N} \sum_{n=1}^{N} f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) \cdots f_k(T^{p_k(n)}x) - \prod_{i=1}^{k} \int f_i \, d\mu \Big\|_{L^2(X)} = 0.$$

Returning to IP-sets, notice that whereas $\mathbf{Z}$ has only countably many additive subsemigroups (and only countably many polynomial subsets of the form $p(\mathbf{Z})$ with $p(t) \in \mathbf{Q}[t]$), an IP-set has uncountably many subsets which are themselves IP-sets. This fact lends additional peculiarity to and hints at greater diversity of the ergodic phenomena inherent in IP-ergodic theory.

We want now to address the question as to whether or not there exists a viable joint framework for ergodic theory along polynomial sets and along IP-sets. Theorem 3.11, for example, would be in this vein. It turns out, in fact, that Theorem 3.11 admits of significant strengthening even in the case of single recurrence. The goal of our present discussion is to provide some motivation for such a generalization. As we shall see, polynomial images of IP-sets (namely sets of the form $\{p(x_\alpha) : \alpha \in \mathcal{F}\}$), which are dealt with in Theorem 3.11, are only a special case of a much wider family of subsets of $\mathbf{Z}$ which deserve to be called *polynomial IP-sets.*

Let us return for a moment to usual polynomials belonging to $\mathbf{Z}[t]$ (we leave to the reader the verification of the fact that the whole discussion is extendable to polynomials from $\mathbf{Q}[t]$ taking on integer values on integers). More specifically, let us examine some of the (numerous) ways of arriving at quadratic polynomials with zero constant term. One possibility is exemplified by formulas like $n^2 = 1 + 3 + \cdots + (2n-1)$, $n \in \mathbf{N}$. Another approach is to seek solutions to functional equations such as

$$p(a+b+c) - p(a+b) - p(b+c) - p(a+c) + p(a) + p(b) + p(c) = 0, \ a, b, c \in \mathbf{Z}.$$

Still another possibility is to take the "diagonal" of a bilinear form. For example, if $g(n,m) = anm + bn + cm$, then putting $n = m$ one gets $p(n) = an^2 + (b+c)n$.

The latter two approaches make perfect sense for $\mathbf{Z}$-valued functions of $\mathcal{F}$-variables. Call a function $f\colon \mathcal{F} \times \mathcal{F} \to \mathbf{Z}$ bilinear if the $\mathcal{F}$-sequence obtained by fixing one of the arguments of g satisfies the IP-equation $x_\alpha + x_\beta = x_{\alpha\cup\beta}$, $\alpha \cap \beta = \emptyset$. Then the "diagonal" $p(\alpha) = g(\alpha, \alpha)$ is a natural analog of a quadratic polynomial. One will arrive at the same family of functions $g: \mathcal{F} \to \mathbf{Z}$ by solving functional equations of the form

$$g(\alpha \cup \beta \cup \gamma) + g(\alpha) + g(\beta) + g(\gamma) = g(\alpha \cup \beta) + g(\beta \cup \gamma) + g(\alpha \cup \gamma),$$
$$\alpha, \beta, \gamma \in \mathcal{F}, \alpha \cap \beta = \emptyset, \alpha \cap \gamma = \emptyset, \beta \cap \gamma = \emptyset.$$

The *IP-quadratic* functions which one obtains this way include (but are not limited to) expressions like $g(\alpha) = n_\alpha k_\alpha + l_\alpha m_\alpha$, $\alpha \in \mathcal{F}$, where $(n_\alpha)_{\alpha\in\mathcal{F}}$, $(k_\alpha)_{\alpha\in\mathcal{F}}$, $(l_\alpha)_{\alpha\in\mathcal{F}}$ and $(m_\alpha)_{\alpha\in\mathcal{F}}$ are IP-sets. A natural subclass of IP-polynomials may be obtained in the following way. Let $q(t_1, t_2, \cdots, t_k) \in \mathbf{Z}[t_1, \cdots, t_k]$ and let $(n_\alpha^{(i)})_{\alpha\in\mathcal{F}}$, $i = 1, 2, \cdots k$ be IP-sets. Then $g(\alpha) = q(n_\alpha^{(1)}, n_\alpha^{(2)}, \cdots, n_\alpha^{(k)})$ is an example of an IP-polynomial. If, say, $\deg q(t_1, \cdots, t_k) = 2$, then $g(\alpha)$ will typically look like

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

For IP-polynomials of the type just described we have the following joint refinement of Theorems 1.17 and 3.11.

Theorem 4.2 ([BFM]). For any polynomial $p(x_1, \cdots, x_k) \in \mathbf{Z}[x_1, \cdots, x_k]$ satisfying $p(0, \cdots, 0) = 0$, and for any k IP-sets $(n_\alpha^{(1)})_{\alpha\in\mathcal{F}}, \cdots, (n_\alpha^{(k)})_{\alpha\in\mathcal{F}}$, the set $\{p(n_\alpha^{(1)}, \cdots, n_\alpha^{(k)}) : \alpha \in \mathcal{F}\}$ is a set of recurrence.

While generalizing Theorem 1.17, Theorem 4.2 does not contain as a special case the more general Theorem 2.1. To formulate a result which would contain Theorem 2.1 as well, we need the following definition. Suppose that $\mathbf{T} = \{T_w : w \in W\}$ is an indexed family of measure preserving transformations of a probability space $(X, \mathcal{B}, \mu)$. One says that $\mathbf{T}$ has the R-property if for any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $w \in W$ such that $\mu(A \cap T_w^{-1} A) > 0$. Theorem 4.2 tells us that the family $\{T^{p(n_\alpha^{(1)}, \cdots, n_\alpha^{(k)})} : \alpha \in \mathcal{F}\}$ has the R-property. This is a special case of the following.

Theorem 4.3 ([BFM]). Suppose that $p_i(x_1, \cdots, x_k) \in \mathbf{Z}[x_1, \cdots, x_k]$ satisfy $p_i(0, \cdots, 0) = 0$, $1 \leq i \leq m$ and that $(n_\alpha^{(i)})_{\alpha\in\mathcal{F}}$ are IP-sets, $1 \leq i \leq k$. Let $p_\alpha^{(j)} := p_j(n_\alpha^{(1)}, \cdots, n_\alpha^{(k)})$, $1 \leq j \leq m$. Then for any commuting invertible measure preserving transformations $T_1, \cdots, T_m$, the family $\{\prod_{i=1}^m T_i^{p_\alpha^{(i)}} : \alpha \in \mathcal{F}\}$ has the R-property.

Exercise 13. Derive Theorem 2.1 from Theorem 4.3.

Theorem 4.3 has the following combinatorial corollary, which can be obtained by applying Furstenberg's correspondence principle for $\mathbf{Z}^t$-actions.

Theorem 4.4. Suppose that a set $E \subset \mathbf{Z}^t$ has positive upper Banach density and let $p_i(x_1, \cdots, x_k) \in \mathbf{Z}[x_1, \cdots, x_k]$ satisfy $p_i(0, \cdots, 0) = 0$, $1 \leq i \leq t$. Let $(n_\alpha^{(i)})_{\alpha \in \mathcal{F}}$ be IP-sets in $\mathbf{N}$, $1 \leq i \leq k$. Then for some $(x_1, \cdots, x_t)$, $(y_1, \cdots, y_t) \in E$ and $\alpha \in \mathcal{F}$ one has

$$\begin{aligned} x_1 - y_1 &= p_1(n_\alpha^{(1)}, \cdots, n_\alpha^{(k)}) \\ x_2 - y_2 &= p_2(n_\alpha^{(1)}, \cdots, n_\alpha^{(k)}) \\ &\vdots \\ x_t - y_t &= p_t(n_\alpha^{(1)}, \cdots, n_\alpha^{(k)}). \end{aligned}$$

The next natural question is whether Theorem 4.4 is a special case of a general *IP-polynomial* Szemerédi theorem, which would bear to Theorem 4.4 the same relation as Theorem 1.19 bears to Theorem 2.1. The answer is yes and is provided by the following theorem.

To formulate it we need the notion of an IP*-set in $\mathbf{Z}^k$. This notion (which actually can be defined in any semigroup) is introduced in a way completely analogous to that for $\mathbf{Z}$ in Section 1. Given any infinite sequence $G = \{g_i : i \in \mathbf{N}\} \subset \mathbf{Z}^k$, the IP-set generated by G is the set $\Gamma = \{g_\alpha\}_{\alpha \in \mathcal{F}}$, where $g_\alpha := \sum_{i \in \alpha} g_i$. A subset $S \subset \mathbf{Z}^k$ is called IP* if for any IP-set $\Gamma \subset \mathbf{Z}^k$ one has $S \cap \Gamma \neq \emptyset$.

Exercise 14. Check that Lemma 1.13 holds for IP*-sets in $\mathbf{Z}^t$.

Theorem 4.5 ([BM2]). Suppose that $T_1, \cdots, T_r$ are commuting invertible measure preserving transformations of a probability space $(X, \mathcal{B}, \mu)$. Suppose that $k, t \in \mathbf{N}$ and that we have polynomials $p_{ij}(n_1, \cdots, n_k) \in \mathbf{Z}[n_1, \cdots, n_k]$, $1 \leq i \leq r$, $1 \leq j \leq t$ having zero constant term. Then for every $A \in \mathcal{B}$ with $\mu(A) > 0$ the set

$$R_A = \left\{ (n_1, \cdots, n_k) \in \mathbf{Z}^k : \mu\Big(\bigcap_{i=1}^{t} \Big(\prod_{j=1}^{r} T_j^{p_{ij}(n_1, \cdots, n_k)} \Big) A \Big) > 0 \right\}$$

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

Some of the corollaries of Theorem 4.5 are collected in the following list.

(i) Already the case $k = 1$ of Theorem 4.5 gives a refinement of the polynomial Szemerédi theorem (Theorem 1.19 above) as well as a strengthened form of its special case, Theorem 1.21. Indeed, Theorem 4.5 says that, when $k = 1$, the set

$$\begin{aligned} R_A = \big\{ n \in \mathbf{Z} : \mu(A \cap T_1^{p_{11}(n)} T_2^{p_{12}(n)} \cdots T_r^{p_{1r}(n)} A \cap \cdots \\ \cap T_1^{p_{t1}(n)} T_2^{p_{t2}(n)} \cdots T_r^{p_{tr}(n)} A) > 0 \big\} \end{aligned}$$

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

(ii) In addition, Theorem 4.5 enlarges the family of configurations which can always be found in sets of positive upper Banach density in $\mathbf{Z}^n$. For example, since for any IP-sets $(n_\alpha^{(i)})_{\alpha\in\mathcal{F}}$, $i = 1, 2, \cdots, k$, any measure preserving system $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $\alpha \in \mathcal{F}$ such that

$$\mu(A \cap T^{n_\alpha^{(1)}} A \cap T^{n_\alpha^{(1)} n_\alpha^{(2)}} A \cap \cdots \cap T^{n_\alpha^{(1)} n_\alpha^{(2)} \cdots n_\alpha^{(k)}} A) > 0,$$

one obtains (via Furstenberg's correspondence principle) the fact that for any $E \subset \mathbf{Z}$ with $d^*(E) > 0$ there exists $x \in \mathbf{Z}$ and $\alpha \in \mathcal{F}$ such that

$$\{x, x + n_\alpha^{(1)}, x + n_\alpha^{(1)} n_\alpha^{(2)}, \cdots, x + n_\alpha^{(1)} n_\alpha^{(2)} \cdots n_\alpha^{(k)}\} \subset E.$$

The following theorem gives a more general corollary of Theorem 4.5 (which includes Theorem 1.20 as a special case).

Theorem 4.6. Let $P\colon \mathbf{Z}^r \to \mathbf{Z}^l$, $r, l \in \mathbf{N}$ be a polynomial mapping satisfying $P(0) = 0$, let $F \subset \mathbf{Z}^r$ be a finite set, let $S \subset \mathbf{Z}^l$ be a set of positive upper Banach density and let $(n_\alpha^{(i)})_{\alpha\in\mathcal{F}}$, $i = 1, 2, \cdots, r$ be arbitrary IP-sets. Then for some $u \in \mathbf{Z}^l$ and $\alpha \in \mathcal{F}$ one has:

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

(iii) The following exercise serves as a good reaffirmation of the third principle of Ramsey theory.

Exercise 15. Given k commuting invertible measure preserving transformations $T_1, \cdots, T_k$ of a probability space $(X, \mathcal{B}, \mu)$, and polynomials $p_i(n, m) \in \mathbf{Z}[n, m]$ with $p_i(0, 0) = 0$, $i = 1, 2, \cdots, k$, and a set $A \in \mathcal{B}$ with $\mu(A) > 0$, let

$$R_A = \{(n, m) : \mu(A \cap T_1^{p_1(n,m)} A \cap \cdots \cap T_k^{p_k(n,m)} A) > 0\}.$$

Show that for any polynomials $q_1, q_2 \in \mathbf{Z}[n]$ satisfying $q_1(0) = q_2(0) = 0$, the set $\{n : (q_1(n), q_2(n)) \in R_A\}$ is an IP*-set in $\mathbf{Z}$.

We want to conclude this section with a brief discussion of the combinatorial tool which is instrumental in the proof of Theorem 4.5, namely the polynomial extension of Hales-Jewett theorem ([HJ]), which was recently obtained in [BL2]. The by now classical Hales-Jewett theorem which deals with finite sequences formed from a finite alphabet rather than with integers, may be regarded as an abstract extension of van der Waerden's theorem. To formulate the Hales-Jewett theorem we introduce some definitions.

Let A be a finite alphabet, $A = \{a_1, \cdots, a_k\}$. The set $W_n(A)$ of words of length n over A can be viewed as an abstract n-dimensional vector space (over A). Having agreed on such a geometric terminology, we look now for an appropriate notion of a (combinatorial) line. Since A is not supposed to have any algebraic structure, the only candidates for combinatorial lines in $W_n(A)$ are sets of n-tuples of elements from A which obey the equations $x_i = x_j$ or $x_j = a$, $a \in A$. For example, if $A = \{1, 2, 3\}$ and $n = 2$, there are 4 lines: $\{11, 22, 33\}$, $\{11, 12, 13\}$, $\{21, 22, 23\}$ and $\{31, 32, 33\}$. Another way of introducing lines is the following. Let $W_n(A, t)$ be the set of words of length n from the alphabet $A \cup \{t\}$, where t is a letter not belonging to A which will serve as a variable. If $w(t) \in W_n(A, t)$ is a word in which the variable t actually occurs, then the set $\{w(t)\}_{t \in A} = \{w(a_1), \cdots, w(a_k)\}$ is a combinatorial line.

Theorem 4.7 ([HJ]). Given any alphabet $A = \{a_1, \cdots, a_k\}$ and $r \in \mathbf{N}$ there exists $N = N(k, r) \in \mathbf{N}$ such that if $n > N$ and $W_n(A)$ is partitioned into r classes then at least one of these classes contains a combinatorial line.

Remarks. (i) Taking $A = \{0, 1, 2, \cdots, s-1\}$ and interpreting $W_n(A)$ as integers to base s having n or less digits in their s-expansion one sees that in this situation the elements of a combinatorial line form an arithmetic progression (with difference of the form $d = \sum_{i=0}^{n-1} a_i s^i$ where $a_i = 0$ or 1). Thus van der Waerden's theorem is a corollary of the Hales-Jewett theorem.

(ii) If one takes A to be a finite field F, then $W_n(F) = F^n$ has a natural structure of an n-dimensional vector space over F. In this case a combinatorial line is an affine linear one-dimensional subspace of F^n.

An interesting feature of the Hales-Jewett theorem (and the one showing that it is, in a sense, the "right" result) is that one can easily derive from it its multidimensional version. Let $t_1, \cdots, t_m$ be m variables and let $W_n(A; t_1, \cdots, t_m)$ be the set of words of length n over the alphabet $A \cup \{t_1, \cdots, t_m\}$. If for some n $w(t_1, \cdots, t_m) \in W_n(A; t_1, \cdots t_m)$ is a word in which all the variables appear, the result of the substitution

$$\{w(t_1, \cdots, t_m)\}_{(t_1, \cdots, t_m) \in A^m} = \{w(a_{i_1}, \cdots, a_{i_m}) : a_{i_j} \in A, j = 1, 2, \cdots m\}$$

is called a combinatorial m-space. It easily follows from Theorem 4.7 that for any $r, m \in \mathbf{N}$ there exists $N = N(k, r, m)$ such that if $n > N$ and $A = \{a_1, \cdots, a_k\}$ is partitioned into r classes then at least one of these classes contains a combinatorial m-space.

Exercise 16. Derive from the multidimensional (or, rather, multiparameter) version of the Hales-Jewett theorem just described the Gallai-Grünwald theorem (Theorem 1.10 above). (With a little extra effort one should also be able to get Theorem 2.5.)

We are going to formulate now one more version of the Hales-Jewett theorem. Given an infinite set M and $k \in \mathbf{N}$, denote by $P_f^{(k)}(M)$ the set of k-tuples of finite (potentially empty) subsets of M and let us call any $(k+1)$-element subset of $P_f^{(k)}(M)$ of the form

$$\big\{(\alpha_1, \alpha_2, \cdots, \alpha_k), (\alpha_1 \cup \gamma, \alpha_2, \cdots, \alpha_k), (\alpha_1, \alpha_2 \cup \gamma, \cdots, \alpha_k), \\ \cdots, (\alpha_1, \alpha_2, \cdots, \alpha_k \cup \gamma)\big\},$$

where γ is non-empty and disjoint from $\alpha_1, \cdots, \alpha_k$, a *simplex* and denote it by $S(\alpha_1, \alpha_2, \cdots, \alpha_k; \gamma)$ (a familiar Euclidian simplex with vertices

$$(s_1, \cdots, s_k), (s_1 + h, s_2, \cdots, s_k), (s_1, s_2 + h, \cdots, s_k), \cdots, (s_1, s_2, \cdots, s_k + h)$$

should come to the reader's mind).

Theorem 4.8. For any $k \in \mathbf{N}$ and any finite coloring of $P_f^{(k)}(\mathbf{N})$, there exists a monochome simplex.

Let us show that Theorem 4.8 follows from Theorem 4.7. Let

$$\chi: P_f^{(k)}(\mathbf{N}) \to \{1, 2, \cdots, r\}$$

be a finite coloring. Let $W = \bigcup_{i=1}^{\infty} W_i(A)$ denote the set of all finite words over the alphabet $A = \{0, 1, \cdots, k\}$. Let $\varphi: W \to P_f^{(k)}(\mathbf{N})$ be the mapping which corresponds to any word $w = w_1 w_2 \cdots w_m \in \mathcal{W}$ a k-tuple $(\alpha_1, \cdots, \alpha_k) \in P_f^{(k)}(\mathbf{N})$ by the rule:

$$\alpha_i = \{j : w_j = i\}, \ \ i = 1, 2, \cdots, k.$$

Notice that this induces a coloring $\tilde{\chi}$ of W defined by $\tilde{\chi} = \chi \circ \varphi$. By Theorem 4.7 there exists a $\tilde{\chi}$-monochrome combinatorial line $\{w(t)\}_{t \in A}$. One easily checks that the χ-monochrome image of this line under φ forms a simplex. The following example should make it completely clear. Let $A = \{0, 1, 2, 3, 4\}$ and assume that

$$\begin{aligned} l &= \{2t1t213\}_{t \in A} \\ &= \big\{(2010213), (2111213), (2212213), (2313213), (2414213)\big\} \end{aligned}$$

is a $\tilde{\chi}$-monochrome combinatorial line. Letting $\gamma = \{2, 4\}$, one observes that

$$\begin{aligned} \varphi(2010213) &= (\{3, 6\}, \{1, 5\}, \{7\}, \emptyset) \\ \varphi(2111213) &= (\{3, 6\} \cup \gamma, \{1, 5\}, \{7\}, \emptyset) \\ \varphi(2212213) &= (\{3, 6\}, \{1, 5\} \cup \gamma, \{7\}, \emptyset)\,. \\ \varphi(2313213) &= (\{3, 6\}, \{1, 5\}, \{7\} \cup \gamma, \emptyset) \\ \varphi(2414213) &= (\{3, 6\}, \{1, 5\}, \{7\}, \gamma) \end{aligned}$$

This gives us a monochrome simplex $S(\{3,6\},\{1,5\},\{7\},\emptyset;\gamma)$.

Exercise 17. Show that Theorem 4.8 implies Theorem 4.7.

To give the reader a feeling of what the polynomial Hales-Jewett theorem is about we shall bring now two equivalent formulations of one of its simplest cases, the "quadratic" nature of which is self-evident. The first formulation is a natural refinement of Theorem 4.8. The second one shows the connection between the polynomial Hales-Jewett theorem and topological dynamics, by means of which the polynomial Hales-Jewett theorem is proved in [BL2].

Theorem 4.9. For any $k \in \mathbf{N}$ and any finite coloring of $P_f^{(k)}(\mathbf{N}\times\mathbf{N})$ there exists a monochrome simplex of the form

$$\big\{(\alpha_1,\alpha_2,\cdots,\alpha_k),(\alpha_1\cup(\gamma\times\gamma),\alpha_2,\cdots,\alpha_k),(\alpha_1,\alpha_2\cup(\gamma\times\gamma),\cdots,\alpha_k),$$

$$\cdots,(\alpha_1,\alpha_2,\cdots,\alpha_k\cup(\gamma\times\gamma))\big\},$$

where γ is a finite non-empty subset of $\mathbf{N}$ and the Cartesian square $\gamma\times\gamma$ is disjoint from $\alpha_1,\cdots,\alpha_n$.

Theorem 4.10. Let (X,ρ) be a compact metric space. For some $k\in\mathbf{N}$, let

$$T(\alpha_1,\alpha_2,\cdots,\alpha_k)=T^{(\alpha_1,\alpha_2,\cdots,\alpha_k)},\ \ (\alpha_1,\alpha_2,\cdots,\alpha_k)\in P_f^{(k)}(\mathbf{N}\times\mathbf{N})$$

be a family of self-mappings of X satisfying the condition

$$(*)\qquad \begin{array}{l}\text{for any finite sets } \alpha_i,\beta_i\subset\mathbf{N}\times\mathbf{N} \text{ satisfying } (\alpha_i\cap\beta_i)=\emptyset,\\ i=1,\cdots,k,\ \ T^{(\alpha_1\cup\beta_1,\cdots,\alpha_k\cup\beta_k)}=T^{(\alpha_1,\cdots,\alpha_k)}T^{(\beta_1,\cdots,\beta_k)}.\end{array}$$

Then for any $\epsilon>0$ and for any $x\in X$ there exist a non-empty set $\gamma\subset\mathbf{N}$ and finite sets $\alpha_1,\cdots,\alpha_k\subset\mathbf{N}\times\mathbf{N}$ such that $\alpha_i\cap(\gamma\times\gamma)=\emptyset$, $i=1,2,\cdots,k$ and

$$\operatorname{diam}\big\{T^{(\alpha_1,\alpha_2,\cdots,\alpha_k)}x,T^{(\alpha_1\cup(\gamma\times\gamma),\alpha_2,\cdots,\alpha_k)}x,T^{(\alpha_1,\alpha_2\cup(\gamma\times\gamma),\cdots,\alpha_k)}x,$$

$$\cdots,T^{(\alpha_1,\alpha_2,\cdots,\alpha_k\cup(\gamma\times\gamma))}x\big\}<\epsilon.$$

Before discussing some applications of Theorems 4.9 and 4.10, let us show their equivalence.

(4.9) $\to$ (4.10). Given the compact space (X, ρ) together with the family of its self-mappings

$$T^{(\alpha_1,\alpha_2,\cdots,\alpha_k)}, \quad (\alpha_1,\alpha_2,\cdots,\alpha_k) \in P_f^{(k)}(\mathbf{N}\times\mathbf{N}),$$

a point $x \in X$ and $\epsilon > 0$, let $\{x_1, \cdots, x_r\}$ be an $\frac{\epsilon}{2}$-net in X. This net naturally defines a coloring $\chi\colon P_f^{(k)}(\mathbf{N}\times\mathbf{N}) \to \{1, 2, \cdots, r\}$ by the rule:

$$\chi\big((\alpha_1,\alpha_2,\cdots,\alpha_k)\big) = \min\big\{i : \rho(T^{(\alpha_1,\alpha_2,\cdots,\alpha_k)}x, x_i) < \frac{\epsilon}{2}\big\}.$$

Let $S(\alpha_1, \alpha_2, \cdots, \alpha_k; \gamma\times\gamma)$ be a monochrome simplex as guaranteed by Theorem 4.9. Then clearly,

$$\operatorname{diam}\big\{T^{(\alpha_1,\alpha_2,\cdots,\alpha_k)}x, T^{(\alpha_1\cup(\gamma\times\gamma),\alpha_2,\cdots,\alpha_k)}x, T^{(\alpha_1,\alpha_2\cup(\gamma\times\gamma),\cdots,\alpha_k)}x,$$
$$\cdots T^{(\alpha_1,\alpha_2,\cdots,\alpha_k\cup(\gamma\times\gamma))}x\big\} < \epsilon.$$

(4.10) $\to$ (4.9) For fixed $r, k \in \mathbf{N}$ let $\Omega_r^{(k)}$ be the space of all r-colorings of $P_f^{(k)}(\mathbf{N}\times\mathbf{N})$ (namely, the set of mappings $\chi\colon P_f^{(k)}(\mathbf{N}\times\mathbf{N}) \to \{1, 2, \cdots, r\}$ equipped with the metric

$$\rho(\chi_1,\chi_2) = \inf\big\{\frac{1}{N+1} : \chi\big((\alpha_1,\cdots,\alpha_k)\big) = \chi_2\big((\alpha_1,\cdots,\alpha_k)\big) \text{ for any}$$
$$(\alpha_1,\cdots,\alpha_k) \in \{1,2,\cdots,N\}\times\{1,2,\cdots,N\}\big\}.$$

Clearly, $(\Omega_r^{(k)}, \rho)$ is a compact metric space. Note that $\rho(\chi_1,\chi_2) < 1$ if and only if $\chi_1(\emptyset) = \chi_2(\emptyset)$. Define mappings

$$T^{(\alpha_1,\alpha_2,\cdots,\alpha_k)}\colon \Omega_r^{(k)} \to \Omega_r^{(k)}, \quad (\alpha_1,\alpha_2,\cdots,\alpha_k) \in P_f^{(k)}(\mathbf{N}\times\mathbf{N})$$

by

$$\big(T^{(\alpha_1,\alpha_2,\cdots,\alpha_k)}\chi\big)(\beta_1,\beta_2,\cdots,\beta_k) = \chi\big((\alpha_1\cup\beta_1, \alpha_2\cup\beta_2, \cdots, \alpha_k\cup\beta_k)\big).$$

Applying Theorem 4.10 for $\epsilon = 1$ and given r-coloring χ, we get $(\alpha_1, \cdots, \alpha_k) \in P_f^{(k)}(\mathbf{N}\times\mathbf{N})$ and a finite non-empty set $\gamma \subset \mathbf{N}$, such that $\alpha_i \cap (\gamma\times\gamma) = \emptyset$, $i = 1, 2, \cdots, k$ and such that

$$\operatorname{diam}\big\{T^{(\alpha_1,\alpha_2,\cdots,\alpha_k)}\chi, T^{(\alpha_1\cup(\gamma\times\gamma),\alpha_2,\cdots,\alpha_k)}\chi, T^{(\alpha_1,\alpha_2\cup(\gamma\times\gamma),\cdots,\alpha_k)}\chi,$$
$$\cdots T^{(\alpha_1,\alpha_2,\cdots,\alpha_k\cup(\gamma\times\gamma))}\chi\big\} < 1.$$

By the remark above we get

$$\begin{aligned}T^{(\alpha_1,\alpha_2,\cdots,\alpha_k)}\chi(\emptyset) &= T^{(\alpha_1\cup(\gamma\times\gamma),\alpha_2,\cdots,\alpha_k)}\chi(\emptyset)\\ &= T^{(\alpha_1,\alpha_2\cup(\gamma\times\gamma),\cdots,\alpha_k)}\chi(\emptyset)\\ &\vdots\\ &= T^{(\alpha_1,\alpha_2,\cdots,\alpha_k\cup(\gamma\times\gamma))}\chi(\emptyset),\end{aligned}$$

or, equivalently,

$$\begin{aligned}\chi\big((\alpha_1,\alpha_2,\cdots,\alpha_k)\big) &= \chi\big((\alpha_1\cup(\gamma\times\gamma),\alpha_2,\cdots,\alpha_k)\big)\\ &= \chi\big((\alpha_1,\alpha_2\cup(\gamma\times\gamma),\cdots,\alpha_k)\big)\\ &\vdots\\ &= \chi\big((\alpha_1,\alpha_2,\cdots,\alpha_k\cup(\gamma\times\gamma))\big).\end{aligned}$$

Let us derive now some combinatorial consequences from the "quadratic" Hales-Jewett theorem. Let $k,r\in\mathbf{N}$ be given and let $\chi_{\mathbf{N}}\colon\mathbf{N}\to\{1,2,\cdots,r\}$ be a coloring of $\mathbf{N}$. Induce a coloring $\chi\colon P_f^{(k)}(\mathbf{N})\to\{1,2,\cdots,r\}$ in the following way. First of all, for any finite non-empty set $\alpha\subset\mathbf{N}\times\mathbf{N}$ define

$$\chi(\alpha)=\chi_{\mathbf{N}}\Big(\sum_{(t,s)\in\alpha}ts\Big).$$

Notice that if $\alpha=\gamma\times\gamma$, then $\sum_{(t,s)\in\alpha}ts$ is a perfect square. If $\alpha=\emptyset$, let $\chi(\alpha)=1$. For any $(\alpha_1,\alpha_2,\cdots,\alpha_k)\in P_f^{(k)}(\mathbf{N}\times\mathbf{N})$ let

$$\chi\big((\alpha_1,\alpha_2,\cdots,\alpha_k)\big):=\chi_{\mathbf{N}}\Big(\sum_{(t,s)\in\alpha_1}ts+2\sum_{(t,s)\in\alpha_2}ts+\cdots+k\sum_{(t,s)\in\alpha_k}ts\Big).$$

Let $S(\alpha_1,\cdots,\alpha_k;\gamma\times\gamma)$ be a χ-monochrome simplex whose existence is guaranteed by Theorem 4.10. Then, letting $v=\sum_{(t,s)\in\alpha_1}ts$ and $c^2=\sum_{(t,s)\in\gamma\times\gamma}ts$, we see that

$$\chi_{\mathbf{N}}(v)=\chi_{\mathbf{N}}(v+c^2)=\chi_{\mathbf{N}}(v+2c^2)=\cdots=\chi_{\mathbf{N}}(v+kc^2).$$

We have obtained the following "quadratic" van der Waerden theorem.

Theorem 4.11. For any $k,r\in\mathbf{N}$, if $\mathbf{N}=\bigcup_{i=1}^r C_i$ then one of C_i, $i=1,2,\cdots,r$ contains a configuration of the form $\{v,v+c^2,\cdots,v+kc^2\}$.

Exercise 18. Derive from Theorem 4.9 the following combinatorial result. For any finite coloring of $\mathbf{Z}^n$ there exist $(a_1, \cdots, a_n) \in \mathbf{Z}^n$ and $c \in \mathbf{N}$ so that the configuration

$$\big\{(a_1, a_2, \cdots, a_n), (a_1 + c^2, a_2, \cdots, a_n), \\ (a_1, a_2 + c^2, \cdots, a_n), \cdots, (a_1, a_2, \cdots, a_n + c^2)\big\}$$

is monochromatic.

The general theorem proved in [BL2] allows one to derive many more combinatorial results as well as results belonging to the realm of topological dynamics, all of which have intrinsic polynomial features. To get a feeling how the polynomial Hales-Jewett theorem may be utilized in the course of a proof of a result pertaining to measurable multiple recurrence, the reader is referred to [BM1].

We conclude this section by formulating a topological recurrence theorem which is, in a sense, the most general "commutative topological quadratic" recurrence result.

Theorem 4.12. Let (X, ρ) be a compact metric space. For a fixed $k \in \mathbf{N}$, let $(T_{ij}^{(l)})_{(i,j)\in\mathbf{N}\times\mathbf{N}}$, $l = 1, 2, \cdots, k$ be commuting continuous self-mappings of X. For any finite nonempty $\alpha \in \mathbf{N} \times \mathbf{N}$ define

$$T_\alpha^{(l)} = \prod_{(i,j)\in\alpha} T_{ij}^{(l)}, \ \ l = 1, 2, \cdots, k.$$

Then for any $\epsilon > 0$ there exists $x \in X$ and non-empty finite $\gamma \subset \mathbf{N}$, such that for $l = 1, 2, \cdots, k$ one has

$$\rho(T_{\gamma\times\gamma}^{(l)} x, x) < \epsilon.$$

5. Some open problems and conjectures.

> Our achievements on the theoretical front will be very poor indeed if...we close our eyes to problems and can only memorize isolated conclusions or principles...
>
> –Mao Tsetung, "Rectify the Party's style of work", [Mao], p. 212.

A mathematical discipline is alive and well if it has many exciting open problems of different levels of difficulty. This section's goal is to show that this is the case with Ergodic Ramsey Theory.

To warm up we shall start with some results and problems related to single recurrence. The following result ([K2]) is usually called Khintchine's recurrence theorem (cf. [Pa], p. 22; [Pe], p. 37).

Theorem 5.1. For any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$, $\epsilon > 0$, and any $A \in \mathcal{B}$ the set $\{n \in \mathbf{Z} : \mu(A \cap T^n A) \geq \mu(A)^2 - \epsilon\}$ is syndetic.

One possible way of proving Theorem 5.1 is to use the uniform version of von Neumann's ergodic theorem: if U is a unitary operator acting on a Hilbert space $\mathcal{H}$, then for any $f \in \mathcal{H}$

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} U^n f = P_{inv} f = f^*,$$

where the convergence is in norm and P_{inv} is the orthogonal projection onto the subspace of U-invariant elements.

Noting that $\langle f^*, f\rangle = \langle f^*, f^*\rangle$ and taking $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$, $(Ug)(x) := g(Tx)$, $g \in L^2(X, \mathcal{B}, \mu)$, and $f = 1_A$ one has

$$\begin{aligned}
&\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T^n A) \\
=&\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \langle U^n f, f\rangle = \langle f^*, f\rangle \\
=&\langle f^*, f^*\rangle = \langle f^*, f^*\rangle\langle 1, 1\rangle \geq \big(\langle f^*, 1\rangle\big)^2 = \big(\langle f, 1\rangle\big)^2 = \mu(A)^2
\end{aligned}$$

(cf. [Ho]).

The following alternative way of proving Theorem 5.1 is more elementary and has two additional advantages: it enables one to prove a stronger fact, namely the IP*-ness of the set $\{n \in \mathbf{Z} : \mu(A \cap T^n A) \geq \mu(A)^2 - \epsilon\}$ and is easily adjustable to measure preserving actions of arbitrary (semi)groups.

Note first that if A_k, $k = 1, 2, \cdots$ are sets in a probability measure space such that $\mu(A_k) \geq a > 0$ for all $k \in \mathbf{N}$ then for any $\epsilon > 0$ there exist $i < j$ such that $\mu(A_i \cap A_j) \geq a^2 - \epsilon$. Indeed, if this would not be the case, the following inequality would be contradictive for sufficiently large n:

$$n^2 a^2 \leq \Big(\int \sum_{i=1}^n 1_{A_i}\Big)^2 \leq \int \Big(\sum_{i=1}^n 1_{A_i}\Big)^2 = \sum_{i=1}^n \mu(A_i) + 2 \sum_{1\leq i<j\leq n} \mu(A_i \cap A_j)$$

(cf. [G]).

To show that $\{n \in \mathbf{Z} : \mu(A \cap T^n A) \geq \mu(A)^2 - \epsilon\}$ is an IP*-set, let $(n_i)_{i=1}^{\infty}$ be an arbitrary sequence of integers and let $A_k = T^{n_1+\cdots+n_k} A$, $k \in \mathbf{N}$. By the above remark, there exist $i < j$ such that

$$a^2 - \epsilon \leq \mu(A_i \cap A_j) = \mu(T^{n_1+\cdots+n_i} A \cap T^{n_1+\cdots+n_j} A) = \mu(A \cap T^{n_{i+1}+\cdots+n_j} A).$$

This shows that

$$\{n \in \mathbf{Z} : \mu(A \cap T^n A) \geq \mu(A)^2 - \epsilon\} \cap FS(n_i)_{i=1}^{\infty} \neq \emptyset$$

and we are done. We remark also that the IP*-ness of the set $\{n \in \mathbf{Z} : \mu(A \cap T^n A) \geq \mu(A)^2 - \epsilon\}$ is equivalent to the "linear" case of Theorem 3.11.

Since in mixing measure preserving systems for any $A \in \mathcal{B}$ one has $\lim_{n\to\infty} \mu(A \cap T^n A) = \mu(A)^2$, we see that in a sense, Khintchine's recurrence theorem is the best possible. We have however the following.

Question 1. Is it true that for any invertible mixing measure preserving system $(X, \mathcal{B}, \mu, T)$ there exists $A \in \mathcal{B}$ with $\mu(A) > 0$ such that for all $n \neq 0$, $\mu(A \cap T^n A) < \mu(A)^2$? How about the reverse inequality $\mu(A \cap T^n A) > \mu(A)^2$?

Definition 5.1. A set $R \subset \mathbf{Z}$ is called a set of *nice recurrence* if for any invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$ one has $\limsup_{n\to\infty, n\in R} \mu(A \cap T^n A) \geq \mu(A)^2$.

Exercise 19. Check that all the sets of recurrence mentioned in Sections 1 through 4 are sets of nice recurrence.

A natural question arises whether any set of recurrence at all is actually a set of nice recurrence. Forrest showed in [Fo] that this is not always so. See also [M] for a shorter proof.

We saw in Section 1 that sets of recurrence have the Ramsey property: if R is a set of recurrence and $R = \bigcup_{i=1}^{r} C_i$ then at least one of C_i, $i = 1, \cdots, r$ is itself a set of recurrence.

Question 2. Do sets of nice recurrence possess the Ramsey property?

A natural necessary condition for a set $R \subset \mathbf{Z} \setminus \{0\}$ to be a set of recurrence is that for any $a \in \mathbf{Z}$, $a \neq 0$, $R \cap a\mathbf{Z} \neq \emptyset$. In particular, the set $\{2^n 3^k : n, k \in \mathbf{N}\}$ is not a set of recurrence. But what if one restricts oneself to some special classes of systems?

Question 3. Is it true that for any invertible weakly mixing system $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exist $n, k \in \mathbf{N}$ such that $\mu(A \cap T^{2^n 3^k} A) > 0$?

Some sets of recurrence have an additional property that the ergodic averages along these sets exhibit regular behavior. For example, we saw in

Section 2 that for any $q(t) \in \mathbf{Z}[t]$ and for any unitary operator $U: \mathcal{H} \to \mathcal{H}$ the norm limit $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} U^{q(n)}f$ exists for every $f \in \mathcal{H}$. The following theorem, due to Bourgain, shows that much more delicate pointwise convergence also holds along the polynomial sets.

Theorem 5.2 ([Bo3]). For any measure preserving system $(X, \mathcal{B}, \mu, T)$, for any polynomial $q(t) \in \mathbf{Z}[t]$ and for any $f \in L^p(X, \mathcal{B}, \mu)$, where $p > 1$, $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(T^{q(n)}x)$ exists almost everywhere.

Question 4. Does Theorem 5.2 hold true for any $f \in L^1(X, \mathcal{B}, \mu)$?

Another interesting question related to ergodic averages along polynomials is concerned with uniquely ergodic systems. A topological dynamical system (X, T), where X is a compact metric space and T is a continuous self mapping of X is called uniquely ergodic if there is a unique T-invariant probability measure on the σ-algebra of Borel sets in X. The following well known result appeared for the first time in [KB]:

Theorem 5.3. A topological system (X, T) is uniquely ergodic if and only if for any $f \in C(X)$ and any $x \in X$ one has

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) = \int f \, d\mu,$$

where μ is the unique T-invariant Borel measure.

Question 5. Assume that a topological dynamical system (X, T) is uniquely ergodic and let $p(t) \in \mathbf{Z}[t]$ and $f \in C(X)$. Is it true that for all but a first category set of points $x \in X$ $\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^{p(n)}x)$ exists?

The next question that we would like to pose is concerned with the possibility of extending results like Theorem 4.2 and 4.3 to polynomial expressions involving infinitely many commuting operators. We shall formulate it for a special "quadratic" case which is a measure theoretic analogue of Theorem 4.12 for $k = 1$. Recall that an indexed family $\{T_w : w \in \mathcal{W}\}$ of measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$ is said to have the R-property if for any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $w \in \mathcal{W}$ such that $\mu(A \cap T_w^{-1}A) > 0$.

Question 6. Let $(T_{ij})_{(i,j)\in\mathbf{N}\times\mathbf{N}}$ be commuting measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$. For any finite non-empty set $\alpha \subset \mathbf{N} \times \mathbf{N}$ let $T_\alpha = \prod_{(i,j)\in\alpha} T_{ij}$. Is it true that the family of measure preserving transformations

$$\{T_{\gamma\times\gamma} : \emptyset \neq \gamma \subset \mathbf{N}, \gamma \text{ finite}\}$$

has the R-property?

We move on now to questions related to multiple recurrence.

Definition 5.2. Let $k \in \mathbf{N}$. A set $R \subset \mathbf{Z}$ is a set of *k-recurrence* if for every invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$ and any $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $n \in R$, $n \neq 0$, such that $\mu(A \cap T^n A \cap T^{2n} A \cap \cdots \cap T^{kn} A) > 0$.

One can show that items (i), (ii), (iv) and (v) of Exercise 6 are examples of sets of k-recurrence for any k. On the other hand, an example due to Furstenberg ([F], p. 178) shows that not every infinite set of differences (item (iii) of Exercise 6) is a set of 2-recurrence (although every such is a set of 1-recurrence).

Question 7. Given $k \in \mathbf{N}$, $k \geq 2$, what is an example of a set of k-recurrence which is not a set of $(k+1)$-recurrence?

Question 8. Given a set of 2-recurrence S, is it true that for any pair T_1, T_2 of invertible commuting measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$ and a set $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $n \in S$ such that $\mu(A \cap T_1^n A \cap T_2^n A) > 0$? (The answer is very likely *no*.) Same question for S a set of k-recurrence for any k.

Question 9. Let $k \in \mathbf{N}$, let $T_1, T_2, \cdots, T_k$ be commuting invertible measure preserving transformations of a probability measure space $(X, \mathcal{B}, \mu)$ and let $p_1(t), p_2(t), \cdots, p_k(t) \in \mathbf{Z}[t]$. Is it true that for any $f_1, \cdots, f_k \in L^\infty(X, \mathcal{B}, \mu)$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T_1^{p_1(n)} x) f_2(T_2^{p_2(n)} x) \cdots f_k(T_k^{p_k(n)} x)$$

exists in L^2-norm? Almost everywhere?

Remark. The following results describe the status of current knowledge: The answer to the question about L^2-convergence is *yes* in the following cases:

(i) $k = 2$, $p_1(t) = p_2(t) = t$ ([CL1]).
(ii) $k = 2$, $T_1 = T_2$, $p_1(t) = t$, $p_2(t) = t^2$ ([FW2]).
(iii) $k = 3$, $T_1 = T_2 = T_3$, $p_1(t) = at$, $p_2(t) = bt$, $p_3(t) = ct$, $a, b, c \in \mathbf{Z}$ ([CL2], [FW2]).

The answer to the question about almost everywhere convergence is *yes* for $k = 2$, $T_1 = T_2$, $p_1(t) = at$, $p_2(t) = bt$, $a, b \in \mathbf{Z}$ ([Bo4]).

Question 10. Let $k \in \mathbf{N}$. Assume that $(X, \mathcal{B}, \mu, T)$ is a totally ergodic system (i.e. $(X, \mathcal{B}, \mu, T^k)$ is ergodic for any $k \neq 0$). Is it true that for any set of polynomials $p_i(t) \in \mathbf{Z}[t]$, $i = 1, 2, \cdots, k$ having pairwise distinct (non-zero) degrees, and any $f_1, \cdots, f_k \in L^\infty(X, \mathcal{B}, \mu)$ one has:

$$\lim_{N\to\infty} \Big\| \frac{1}{N} \sum_{n=1}^{N} f_1(T^{p_1(n)}x) f_2(T^{p_2(n)}x) \cdots f_k(T^{p_k(n)}x) - \prod_{i=1}^{k} \int f_i \, d\mu \Big\|_{L^2(X,\mathcal{B},\mu)} = 0?$$

Remark. It is shown in [FW2] that the answer is *yes* when $k = 2$, $p_1(t) = t$, $p_2(t) = t^2$. See also Theorem 4.1.

We now formulate a few problems related to partition Ramsey theory. A unifying property that many configurations of interest (such as arithmetic progressions or sets of the form $FS(x_j)_{j=1}^n$) have is that they constitute sets of solutions of (not necessarily linear) diophantine equations or systems thereof. A system of diophantine equations is called *partition regular* if for any finite coloring of $\mathbf{Z} \setminus \{0\}$ (or of $\mathbf{N}$) there is a monochromatic solution. For example, the following systems of equations are partition regular:

$$\begin{aligned} x_1 + x_3 &= 2x_2 \\ x_2 + x_4 &= 2x_3 \\ x_3 + x_5 &= 2x_4 \\ x_4 + x_6 &= 2x_5 \end{aligned} \qquad\qquad \begin{aligned} x + y &= t \\ x + z &= u \\ z + y &= v \\ x + y + z &= w \end{aligned}$$

A general theorem due to Rado gives necessary and sufficient conditions for a system of *linear* equations to be partition regular (cf. [Ra], [GRS] or [F2]). The results involving polynomials brought forth in Sections 1-4 hint that there are some nonlinear equations that are partition regular too. For example, the equation $x - y = p(z)$ is partition regular for any $p(t) \in \mathbf{Z}[t]$, $p(0) = 0$. To see this, fix $p(t)$ and let $\mathbf{N} = \bigcup_{i=1}^r C_i$ be an arbitrary partition. Arguing as in [B2] one can show that one of the cells C_i, call it C, has the property that it contains an IP-set and has positive upper density. Let $\{n_\alpha\}_{\alpha\in\mathcal{F}}$ be an IP-set in C. According to Theorem 3.11, $\{p(n_\alpha)\}_{\alpha\in\mathcal{F}}$ is a set of recurrence. This together with Furstenberg's correspondence principle gives that for some $\alpha \in \mathcal{F}$,

$$\overline{d}\Big(C_i \cap \big(C_i - p(n_\alpha)\big)\Big) > 0.$$

If $y \in \big(C_i \cap \big(C_i - p(n_\alpha)\big)\big)$ then $x = y + p(n_\alpha) \in C_i$. This establishes the partition regularity of $x - y = p(z)$. In accordance with the third principle of Ramsey theory one should expect that there are actually many x, y, z having the same color and satisfying $x - y = p(z)$. This is indeed so: using the fact that $\{p(n_\alpha)\}_{\alpha\in\mathcal{F}}$ is a set of nice recurrence one can show, for example, that for any $\epsilon > 0$ and any partition $\mathbf{N} = \bigcup_{i=1}^r C_i$ one of C_i, $i = 1, 2, \cdots, r$ satisfies

$$\overline{d}\Big(\big\{z \in C_i : \overline{d}\big(C_i \cap \big(C_i - p(z)\big)\big) \geq \big(\overline{d}(C_i)\big)^2 - \epsilon\big\}\Big) > 0$$

(cf. [B2], see also Theorem 0.4 in [BM1]).

Question 11. Are the following systems of equations partition regular?
(i) $x^2 + y^2 = z^2$.
(ii) $xy = u$, $x + y = w$.
(iii) $x - 2y = p(z)$, $p(t) \in \mathbf{Z}[t]$, $p(0) = 0$.

The discussion in this survey so far has concentrated mainly on topological and measure preserving $\mathbf{Z}^n$-actions. Ergodic Ramsey theory of actions of more general, especially non-abelian groups is much less developed and offers many interesting problems.

In complete analogy with the case of the group $\mathbf{Z}$, given a semigroup G call a set $R \subset G$ a set of recurrence if for any measure preserving action $(T_g)_{g\in G}$ of G on a finite measure space $(X, \mathcal{B}, \mu)$ and for 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, one can show that the set $\{1 + \frac{1}{k} : \ k \in \mathbf{N}\}$ is a set of recurrence for the multiplicative group of positive rationals. Sets of the form $\{n^\alpha : n \in \mathbf{N}\}$, where $\alpha > 0$, are sets of recurrence for $(\mathbf{R}, +)$. As a matter of fact, one can show (see [BBB]) that for any measure preserving $\mathbf{R}$-action $(S^t)_{t\in\mathbf{R}}$ on a probability space $(X, \mathcal{B}, \mu)$ one has for every $A \in \mathcal{B}$ that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap S^{n^\alpha} A) \geq \mu(A)^2.$$

On the other hand one has the following negative result.

Theorem 5.4 ([BBB]). Let $(S^t)_{t\in\mathbf{R}}$ be an ergodic measure preserving flow acting on a probability Lebesgue space $(X, \mathcal{B}, \mu)$. For all but countably many $\alpha > 0$ (in particular for all positive $\alpha \in (\mathbf{Q} \setminus \mathbf{Z})$) one can find an L^∞-function f for which the averages $\frac{1}{N}\sum_{n=1}^N f(S^{n^\alpha}x)$ fail to converge for a set of x of positive measure.

It is possible that the countable set of "good" α coincides with $\mathbf{N}$. Such a result would follow from a positive answer to the following number-theoretical question which we believe is of independent interest.

Question 12. Let us call an increasing sequence $\{a_n : n \in \mathbf{N}\} \subset \mathbf{R}$ *weakly independent* over $\mathbf{Q}$ if there exists an increasing sequence $(n_i)_{i=1}^\infty \subset \mathbf{N}$

having positive upper density such that the sequence $\{a_{n_i} : i \in \mathbf{N}\}$ is linearly independent over $\mathbf{Q}$. Is it true that for every $\alpha > 0$, $\alpha \notin \mathbf{N}$, the sequence $\{n^\alpha : n \in \mathbf{N}\}$ is weakly independent over $\mathbf{Q}$? (It is known that the answer is yes for all but countably many α.)

Definition 5.3. Given a (semi)group G, a set $R \subset G$ is called a set of topological recurrence if for any minimal action $(T_g)_{g\in G}$ of G on a compact metric space X and for any open, non-empty set $U \subset X$ there exists $g \in R$, $g \neq e$, such that $(U \cap T_g^{-1}U) \neq \emptyset$.

Exercise 20. Prove that in an amenable group any set of (measurable) recurrence is a set of topological recurrence.

An interesting result due to Kriz ([Kr], see also [Fo], [M]) says that in $\mathbf{Z}$ there are sets of topological recurrence which are not sets of measurable recurrence. While the same kind of result ought to hold in any abelian group, and while for any amenable group sets of measurable recurrence are, according to Exercise 20, sets of topological recurrence, the situation for more general groups is far from clear. We make the following

Conjecture. A group G is amenable if and only if any set of measurable recurrence $R \subset G$ is a set of topological recurrence.

An intriguing question is, what is the right formulation of the Szemerédi (or van der Waerden) theorem for general group actions. In this connection we want to mention a very nice noncommutative extension of Theorem 1.19 which was recently obtained by Leibman in [L2]: he was able to show that the conclusion of Theorem 1.19 holds if one replaces the assumption about the commutativity of the measure preserving transformations T_i by the demand that they generate a nilpotent group. He also proved earlier in [L1] a topological van der Waerden-type theorem of a similar kind. This should be contrasted with an example due to Furstenberg of a pair of homeomorphisms T_1, T_2 of a compact metric space X generating a metabelian group such that no point of X is simultaneously recurrent for T_1, T_2 (this implies that for metabelian groups one should look for another formulation of a Szemerédi-type theorem).

A possible way of extending multiple recurrence theorems to a situation involving non-commutative groups is to consider a finite family of pairwise commuting actions of a given group. Results obtained within such framework ought to be called *semicommutative*. We have the following

Conjecture. Assume that G is an amenable group with a Følner sequence $(F_n)_{n=1}^{\infty}$. Let $(T_g^{(1)})_{g\in G}, \cdots, (T_g^{(k)})_{g\in G}$ be k pairwise commuting measure preserving actions of G on a measure space $(X, \mathcal{B}, \mu)$ ("pairwise commuting" means here that for any $1 \leq i \neq j \leq k$ 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:

$$\liminf_{n\to\infty} \frac{1}{|F_n|} \sum_{g\in F_n} \mu(A \cap T_g^{(1)}A \cap T_g^{(1)}T_g^{(2)}A \cap \cdots \cap T_g^{(1)}T_g^{(2)}\cdots T_g^{(k)}A) > 0.$$

Remarks. (i) We have formulated the conjecture for amenable groups for two major reasons. First of all, the conjecture is known to hold true for $k = 2$ ([BMZ], see also [BeR]). Second, in case the group G is countable, a natural analogue of Furstenberg's correspondence principle, which was formulated in Section 1, holds and allows one to obtain combinatorial corollaries which, should the conjecture turn out to be true for any k, contain Szemerédi's theorem as quite a special case.

(ii) The "triangular" expressions

$$A \cap T_g^{(1)}A \cap T_g^{(1)}T_g^{(2)}A \cap \cdots \cap T_g^{(1)}T_g^{(2)}\cdots T_g^{(k)}A$$

appearing in the formulation of the conjecture seem to be the "right" configurations to consider. See the discussion and counterexamples in [BH2] where a topological analogue of the conjecture is treated (but not fully resolved). We suspect that the answer to the following question is, in general, negative.

Question 13. Given an amenable group G and a Følner sequence $(F_n)_{n=1}^{\infty}$ for G, let $(T_g)_{g\in G}$ and $(S_g)_{g\in G}$ be two commuting measure preserving actions on a probability space $(X, \mathcal{B}, \mu)$. Is it true that for any $A \in \mathcal{B}$ the following limit exists:

$$\lim_{n\to\infty} \frac{1}{|F_n|} \sum_{g\in F_n} \mu(T_gA \cap S_gA)?$$

We want to conclude by formulating a conjecture about a density version of the polynomial Hales-Jewett theorem which would extend both the partition results from [BL2] and the density version of the ("linear") Hales-Jewett theorem proved in [FK4]. For $q, d, N \in \mathbf{N}$ let $\mathcal{M}_{q,d,N}$ be the set of q-tuples of subsets of $\{1, 2, \cdots, N\}^d$:

$$\mathcal{M}_{q,d,N} = \big\{(\alpha_1, \cdots, \alpha_q) : \alpha_i \subset \{1, 2, \cdots, N\}^d,\ i = 1, 2, \cdots, q\big\}.$$

Conjecture. For any $q, d \in \mathbf{N}$ 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:

$$\big\{(\alpha_1, \alpha_2, \cdots, \alpha_q), (\alpha_1 \cup \gamma^d, \alpha_2, \cdots, \alpha_q), (\alpha_1, \alpha_2 \cup \gamma^d, \cdots, \alpha_q), \\ \cdots, (\alpha_1, \alpha_2, \cdots, \alpha_q \cup \gamma^d)\big\},$$

where $\gamma \subset \mathbf{N}$ is a non-empty set and $\alpha_i \cap \gamma^d = \emptyset$ for all $i = 1, 2, \cdots, q$.

Remark. For $d = 1$ the conjecture follows from [FK4]. This paper contains a wealth of related material and is strongly recommended for rewarding reading.

REFERENCES

[Ba] J.E. Baumgartner, A short proof of Hindman's theorem, *J. Combinatorial Theory (A)* **17** (1974), 384-386.

[B1] V. Bergelson, Sets of recurrence of $\mathbf{Z}^m$-actions and properties of sets of differences in $\mathbf{Z}^m$, *J. London Math. Soc.* (2) **31** (1985), 295-304.

[B2] V. Bergelson, A density statement generalizing Schur's theorem, *J. Combinatorial Theory (A)* **43** (1986), 338-343.

[B3] V. Bergelson, Ergodic Ramsey Theory, *Logic and Combinatorics* (edited by S. Simpson), *Contemporary Mathematics* **65** (1987), 63-87.

[B4] V. Bergelson, Weakly mixing PET, *Ergodic Theory and Dynamical Systems*, **7** (1987), 337-349.

[BBB] V. Bergelson, M. Boshernitzan, and J. Bourgain, Some results on non-linear recurrence, *Journal d'Analyse Mathématique*, **62** (1994), 29-46.

[BeR] V. Bergelson and J. Rosenblatt, Joint ergodicity for group actions, *Ergodic Theory and Dynamical Systems* (1988) **8**, 351-364.

[BFM] V. Bergelson, H. Furstenberg and R. McCutcheon, IP-sets and polynomial recurrence, *Ergodic Theory and Dynamical Systems* (to appear).

[BHå] V. Bergelson and I.J. Håland, Sets of recurrence and generalized polynomials, to appear in *The Proceedings of the Conference on Ergodic Theory and Probability, June 1993*, Ohio State University Math. Research Institute Publications, de Gruyter, 1995.

[BH1] V. Bergelson and N. Hindman, Nonmetrizable topological dynamics and Ramsey theory, *Trans. AMS* **320** (1990), 293-320.

[BH2] V. Bergelson and N. Hindman, Some topological semicommutative van der Waerden type theorems and their combinatorial consequences, *J. London Math. Soc.* (2) **45** (1992) 385-403.

[BL1] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems, *Journal of AMS* (to appear).

[BL2] V. Bergelson and A. Leibman, Set polynomials and a polynomial extension of Hales-Jewett theorem. Preprint.

[BM1] V. Bergelson and R. McCutcheon, Uniformity in polynomial Szemerédi theorem, these Proceedings.

[BM2] V. Bergelson and R. McCutcheon, An IP polynomial Szemerédi theorem for finite families of commuting transformations, in preparation.

[BMZ] V. Bergelson, R. McCutcheon and Q. Zhang, A Roth theorem for amenable groups (to be submitted).

[Bo1] H. Bohr, Zur theorie der fastperiodischen functionen, I, *Acta Math.* **45** (1924), 29-127.

[Bo2] N. Bogolioùboff, Sur quelques proprietés arithmetiques des presque-périodes, *Ann. Phy. Math. Kiew* **4** (1939).

[Bo3] J. Bourgain, Pointwise ergodic theorems for arithmetic sets, *Publ. Math. IHES* **69** (1989), 5-45.

[Bo4] J. Bourgain, Double recurrence and almost sure convergence, *J. Reine Angew. Math.* **404** (1990), 140-161.

[Br] J.R. Brown, *Ergodic Theory and Topological Dynamics*, Academic Press, 1976.

[C] W. Comfort, Ultrafilters–some old and some new results, *Bull. AMS* **83** (1977), 417-455.

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

[CL2] J. Conze and E. Lesigne, Sur un théorème ergodique pour des mesures diagonales, *C. R. Acad. Sci. Paris, t.* **306**, I (1988), 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.

[D] E. van Douwen, The Čech-Stone compactification of a discrete groupoid, *Topology and its applications* **39** (1991), 43-60.

[E1] R. Ellis, Distal transformation groups, *Pac. J. Math.* **8** (1958), 401-405.

[E2] B. Ellis, Solution of Problem 6612, *Amer. Math. Monthly* **98** (1991), 957-959.

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

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

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

[F3] H. Furstenberg, Poincaré recurrence and number theory, *Bull. AMS (New Series)* **5** (1981), 211-234.

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

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

[FK3] H. Furstenberg and Y. Katznelson, Idempotents in compact semigroups and Ramsey theory, *Israel J. Math.* **68** (1990), 257-270.

[FK4] H. Furstenberg and Y. Katznelson, A density version of the Hales-Jewett theorem, *J. d'Analyse Math.* **57** (1991), 64-119.

[Fo] A.H. Forrest, The construction of a set of recurrence which is not a set of strong recurrence, *Israel J. Math.* **76** (1991), 215-228.

[Fø] E. Følner, Bemaerkning om naestenperiodieitetus definition, *Mat. Tidsskrift B*, 1944.

[FW1] H. Furstenberg and B. Weiss, Topological dynamics and combinatorial number theory, *J. d'Analyse Math.* **34** (1978), 61-85.

[FW2] H. Furstenberg and B. Weiss, A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, to appear in *The Proceedings of the Conference on Ergodic Theory and Probability, June 1993*, Ohio State University Math. Research Institute Publications, de Gruyter, 1995.

[G] J. Gillis, Note on a property of measurable sets, *J. London Math. Soc.* **11** (1936), 139-141.

[GJ] L. Gillman and M. Jerison, *Rings of continuous functions*, Springer-Verlag.

[GRS] R. Graham, B. Rothschild and J. Spencer, *Ramsey Theory*, Wiley, New York, 1980.

[Ha] G.H. Hardy, *A Mathematician's Apology*, Cambridge University Press, 1990.

[H1] D. Hilbert, Über die Irreduzibilität ganzer rationaler Funktionen mit ganzzahligen Koeffizienten, *J. Math.* **110** (1892), 104-129.

[H2] N. Hindman, Finite sums from sequences within cells of a partition of **N**, *J. Combinatorial Theory* (Series A) **17** (1974) 1-11.

[H3] N. Hindman, Ultrafilters and combinatorial number theory, *Number Theory Carbondale*, M. Nathanson, ed., *Lecture Notes in Math.* **751** (1979) 119-184.

[H4] N. Hindman, Partitions and sums and products of integers, *Trans. AMS* **247** (1979), 227-245.

[H5] N. Hindman, The semigroup $\beta\mathbf{N}$ and its applications to number theory, *The Analytical and Topological Theory of Semigroups*, K. H. Hoffman, J. D. Lawson and J. S. Pym, editors, de Gruyter, 1990.

[HJ] A.W. Hales and R.I. Jewett, Regularity and positional games, *Trans. AMS* **106** (1963), 222-229.

[Ho] E. Hopf, *Ergodentheorie,* Chelsea, New York, 1948.

[HR] E. Hewitt and K. Ross, *Abstract Harmonic Analysis I*, Springer-Verlag, 1963.

[K1] A. Y. Khintchine, *Three Pearls of Number Theory*, Graylock Press, Rochester, N.Y., 1952.

[K2] A. Y. Khintchine, Eine Verschärfung des Poincaréschen Wiederkehrsatzes, *Comp. Math.* **1** (1934), 177-179.

[KB] N. Krylov and N. Bogolioùboff, La théorie générale de la measure dans son application à l'étude des systèmes dynamiques de la mécanique non linéare, *Ann. of Math.* **38** (1937), 65-113.

[KN] B. O. Koopman and J. von Neumann, Dynamical systems of continuous spectra, *Proc. Nat. Acad. Sci. U.S.A.* **18** (1932), 255-263.

[Kr] I. Kriz, Large independent sets in shift-invariant graphs. Solution of Bergelson's problem, *Graphs and Combinatorics* **3** (1987), 145-158.

[L1] A. Leibman, Multiple recurrence theorem for nilpotent group actions, *Geom. and Funct. Anal.* **4** (1994), 648-659.

[L2] A. Leibman, Multiple recurrence theorems for nilpotent group actions, Ph.D. dissertation, Technion, Haifa, 1995.

[M] R. McCutcheon, Three results in recurrence, *Ergodic Theory and its Connections with Harmonic Analysis* (edited by K. Petersen and I. Salama), *London Math. Soc. Lecture Notes Series* **205**, Cambridge University Press, 1995, 349-358.

[Mao] Mao Tsetung, *Selected Readings from the Works of Mao Tsetung*, Foreign Language Press, Peking, 1971.

[N] K. Numakura, On bicompact semigroups, *Math. J. of Okayama University* **1** (1952), 99-108.

[Pa] W. Parry, *Topics in Ergodic Theory*, Cambridge Univ. Press, Cambridge, 1981.

[Pe] K. Petersen, *Ergodic Theory*, Cambridge Univ. Press, Cambridge, 1981.

[Ra] R. Rado, Note on combinatorial analysis, *Proc. London Math. Soc.* **48** (1993), 122-160.

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

[S] A. Sárközy, On difference sets of integers III, *Acta. Math. Acad. Sci. Hungar.*, **31** (1978) 125-149.

[Sz1] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, *Acta Math. Acad. Sci. Hungar.* **20** (1969), 89-104.

[Sz2] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199-245.

[W] N. Wiener, *Time Series*, The M.I.T. Press, 1949.