ERGODIC THEORY and DIOPHANTINE PROBLEMS

Vitaly BERGELSON
The Ohio State University, Department of Mathematics
231 West 18th Avenue, Columbus, OH 43210-1174
U.S.A.

Contents

1 Introduction

The topic of these notes is the interplay between ergodic theory, some diophantine problems, and the area of combinatorics called Ramsey Theory.

The first section deals with some classical and well-known diophantine results and their connections with topological and measure-preserving dynamics. Some of the proofs offered in Section 2 are very elementary, while some use ergodic-theoretic machinery which is actually much more sophisticated than the results it gives us as applications. This should not in any way discourage the reader since the author's intention was not to produce proofs that are as elementary as possible (see Appendix), but to show how intertwined the different and seemingly remote areas of mathematics may be.

The combinatorial results discussed in Sections 3, 4, and 5 are more recent, but they are as beautiful and, in our opinion, as important as the diophantine facts dealt with in Section 2.

It was H. Furstenberg, who, with his publication in 1977 of the ergodic-theoretic proof of Szemerédi's theorem (see Section 4), established the link between Ramsey theory and the theory of multiple recurrence. Since then, many open problems of combinatorics and number theory have been solved by the methods of ergodic theory and topological dynamics (see for example, [15], [16], [17], [5], and [6]). As it happens, the developments brought to light many new and intriguing problems which are of interest to both the ergodic theory and combinatorics. (See for example, Section 5 in [3].)

The author hopes that after reading these notes, the reader will develop enough of an interest and curiosity to undertake an in-depth study of Ergodic Ramsey Theory. Such a reader is especially encouraged to look into the book [13] and the recent survey [3].

We conclude this introduction with some general terminological and notational remarks.

An *abstract dynamical system* is a space endowed with some structure and a group or a semigroup of its self-mappings which preserves this structure. *Topological dynamics* concerns itself with compact metric spaces and semigroups of continuous mappings. Measure-preserving dynamics, or *ergodic theory*, works with measure spaces and semigroups of measure-preserving transformations.

Because we are going to be mostly interested in applications to and connections with number theory and combinatorics, the semigroups of structure-preserving mappings we encounter will usually be countable and abelian. However, some material in Section 4 will be developed for the general set-up of countable amenable groups, partly because confining ourselves to abelian groups would not make the presentation easier and partly because countable amenable groups seem to be the rightly general object for developing Ergodic

Ramsey Theory.

We shall denote topological dynamical systems by $(X, \{T_g\}_{g\in G})$ or (X, G) where X will always mean a compact metric space on which a (countable) semigroup G acts by continuous mappings $T_g, g \in G$. In case G is either $\mathbb{N}$ or $\mathbb{Z}$ (i.e. G is generated by a single continuous mapping $T : X \to X$), we shall denote the topological dynamical system by (X, T).

Similarly, a typical notation for a measure-preserving system will be $(X, \mathcal{B}, \mu, \{T_g\}_{g\in G})$, where $(X, \mathcal{B}, \mu)$ is a probability measure space and the transformations $T_g, g \in G$ are measure-preserving (i.e. $\forall A \in \mathcal{B}$ and $\forall g \in G$ one has $\mu(T_g^{-1}A) = \mu(A)$). Again, if G is generated by a single, not necessarily invertible, measure-preserving transformation T, we shall denote the measure-preserving system by $(X, \mathcal{B}, \mu, T)$.

Given a point $x \in X$, its *orbit* under the action $\{T_g\}_{g\in G}$ is defined by $\{T_g x\}_{g\in G}$. In both the topological and measure-preserving situations, it is important to know how the points of an orbit are distributed in X, what can be said about the orbit of a typical point (in this or that sense), how massive is the set of semigroup elements g for which the images $T_g x$ of the point x are close to x, etc.

Finally, we want to emphasize the following important point: even in the purely topological set-up, it is often helpful to introduce an invariant measure. For example, the Bogoliouboff-Kryloff theorem (see Theorem 2.21 below) tells us that for any topological dynamical system (X, G) with G abelian (or, more generally, amenable), an invariant measure exists. When it is unique, the system (X, G) is called *uniquely ergodic*, and one is then able to make strong statements about the uniform distribution of orbits in X. See the discussion at the end of Section 2.

Still another setting that warrants the introduction of an invariant measure is discussed in Section 4 (see Theorems 4.4 and 4.17).

2 Some Diophantine Problems Related to Polynomials and their Connections with Combinatorics and Dynamics

We start this section by giving a simple, "dynamically flavored" proof of the one-dimensional case of Kronecker's theorem on diophantine approximation.

Theorem 2.1 ([24], Theorem 438). *If θ is irrational, α arbitrary, and N and ε are positive, then there are integers n and p such that $n > N$ and $|n\theta - p - \alpha| < \varepsilon$.*

Proof. Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the one-dimensional torus, and let $T_\theta : x \to x + \theta \pmod 1$ be the "rotation" defined by θ. It is easy to see that Kronecker's theorem is equivalent to the following.

Statement. The forward semiorbit of 0, $\{n\theta (\text{mod } 1)\}_{n\in\mathbb{N}}$, is dense in $\mathbb{T}$.

Indeed, let $\alpha - [\alpha] = \alpha' \in (0,1)$, and assume without loss of generality that $\alpha' + \varepsilon < 1$. If for some $n \in \mathbb{N}$ $n\theta \pmod 1 \in (\alpha', \alpha' + \varepsilon)$, then for $p = [n\theta] + [\alpha]$ one has

$$0 < n\theta - p - \alpha < \varepsilon. \tag{1}$$

Clearly, if the semiorbit $\{n\theta \pmod 1\}_{n\in\mathbb{N}}$ is dense, then inequality (1) is satisfied for infinitely many $n \in \mathbb{N}$ (in particular for some $n > N$).

The proof of the Statement is very short. Note that if for some $n_0 \in \mathbb{N}$ one has either $0 < n_0\theta \pmod 1 < \varepsilon$ or $1 - \varepsilon < n_0\theta \pmod 1 < 1$, then the set of multiples $\{nn_0\theta \pmod 1\}_{n\in\mathbb{N}}$ is ε-dense in $[0,1]$. But, any set of $M > [\frac{1}{\varepsilon}] + 1$ points in $[0,1]$ contains a pair of points at a distance $< \varepsilon$. (Just use the pigeon hole principle!). Applying this remark to the set $\{n\theta \pmod 1\}_{n=1}^{M}$, we find $1 \le n_1 < n_2 \le M$ so that either $0 < (n_2 - n_1)\theta \pmod 1 < \varepsilon$ or $1 - \varepsilon < (n_2 - n_1)\theta \pmod 1 < 1$. ■

Question. Where exactly was the irrationality of θ used in the proof?

Definition 2.2 A set $S \subseteq \mathbb{R}$ is called relatively dense, or *syndetic*, if there exists an $L > 0$ such that any interval of length L contains at least one element from S.

Exercise 2.3 Derive from the proof of Kronecker's theorem above that for any irrational θ and any $0 \le a < b \le 1$ the set $\{n \in \mathbb{Z} : a < n\theta \pmod 1 < b\}$ is syndetic.

We now formulate Kronecker's theorem in its general form. Recall that the numbers $x_1, ..., x_m \in \mathbb{R}$ are called *rationally independent* if the relation $\sum_{i=1}^{m} n_i x_i = 0$ with $n_i \in \mathbb{Q}$ is possible only if $n_i = 0$ for all $1 \le i \le m$.

Theorem 2.4 ([24], Theorem 442). *Suppose $\theta_1, \theta_2, ..., \theta_k$ are rationally independent, $\alpha_1, \alpha_2, ..., \alpha_k$ are arbitrary, and N and ε are positive, then there are integers $n > N$ and $p_1, p_2, ..., p_k$ such that $|n\theta_m - p_m - \alpha_m| < \varepsilon$ $(m = 1, 2, ..., k)$.*

Exercise 2.5 Let $f(x) = \sin x + \sin \sqrt{2}x + \sin \sqrt{3}x$, $x \in \mathbb{R}$.
(i) Prove that f is not periodic.
(ii) Prove that for any $\varepsilon > 0$ the set

$$\{\tau \in \mathbb{R} : |f(x+\tau) - f(x)| < \varepsilon \ \forall x \in \mathbb{R}\}$$

is syndetic.
(iii) Functions satisfying (ii) are called *almost periodic.* Prove that if f and g are almost periodic, then $f + g$ is also almost periodic.

Exercise 2.6 Prove that Theorem 2.4 is equivalent to the following:

Statement. Let $\mathbb{T}^k = \mathbb{R}^k/\mathbb{Z}^k$, and for $\theta = (\theta_1, ..., \theta_k) \in \mathbb{R}^k$ define

$$T_\theta : \mathbb{T}^k \to \mathbb{T}^k \text{ by } T_\theta(x_1, ..., x_k) = (x_1 + \theta_1, ..., x_k + \theta_k) \pmod 1.$$

If $\theta_1, ..., \theta_k$, and 1 are rationally independent, then for any $x = (x_1, ..., x_k) \in \mathbb{T}^k$ the forward semiorbit $\{T_\theta^n x\}_{n\in\mathbb{N}}$ is dense in $\mathbb{T}^k$.

One can give a proof of Theorem 2.4 by refining the argument of the proof of Theorem 2.1 above. We prefer to indicate a different and, in a sense, more fruitful approach.

Let X be a compact metric space and μ a probability measure on $\mathcal{B}_X$ (the σ-algebra generated by open sets in X). We say that a sequence $\{x_n\}_{n\in\mathbb{N}}$ of points in X is *uniformly distributed* with respect to μ if for any $f \in C(X)$ one has

$$\frac{1}{N}\sum_{n=1}^{N} f(x_n) \underset{N\to\infty}{\longrightarrow} \int_X f \, d\mu. \tag{2}$$

A useful observation is that if Φ is a countable family of functions in $C(X)$, such that linear combinations of elements of Φ are dense in $C(X)$, then in order to verify the uniform distribution of a sequence $\{x_n\}_{n\in\mathbb{N}}$ it suffices to check that (2) holds for any $f \in \Phi$.

Specifying $X = \mathbb{T}^k$, $\mu = m$ (the Lebesgue measure), and taking into account the fact that finite linear combinations (with complex coefficients) of functions of the form $e^{2\pi i\langle h,t\rangle}$, where $h = (h_1, ..., h_k) \in \mathbb{Z}^k$ and $t = (t_1, ..., t_k) \in \mathbb{T}^k$, are dense in $C(\mathbb{T}^k)$, we obtain the following.

Theorem 2.7 (Weyl criterion). *A sequence $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{T}^k$ is uniformly distributed if and only if for any nonzero $h = (h_1, ..., h_k) \in \mathbb{Z}^k$ one has*

$$\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i\langle h, x_n\rangle} \xrightarrow[N\to\infty]{} 0.$$

Exercise 2.8 Give a detailed proof of Theorem 2.7.

Exercise 2.9 Prove that a sequence $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{T}^k$ is uniformly distributed if and only if for any nonzero $h \in \mathbb{Z}^k$, the sequence $\{\langle h, x_n\rangle \pmod 1\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{T}$.

Exercise 2.10 Prove that a sequence $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{T}$ if and only if for any $0 \le a < b \le 1$ one has

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : a \le x_n \le b\}}{N} = b - a.$$

As Hardy and Wright ([24], 23.10) put it, $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{T}$ "if every subinterval contains its proper quota of points."

Note that even for $k = 1$ Theorem 2.7 gives significantly more than Theorem 2.1. Indeed, since for any irrational θ and $h \in \mathbb{Z}$, $h \neq 0$,

$$\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i h n\theta} = \frac{1}{N} e^{2\pi i h\theta} \frac{e^{2\pi i h N\theta} - 1}{e^{2\pi i h\theta} - 1} \xrightarrow[N\to\infty]{} 0,$$

we have for any $f \in C(\mathbb{T})$

$$\frac{1}{N}\sum_{n=1}^{N} f(n\theta \pmod 1) \xrightarrow[N\to\infty]{} \int_{\mathbb{T}} f dm\ .$$

This implies (cf. Exercise 2.10) that not only does the sequence $\{n\theta \pmod 1\}_{n\in\mathbb{N}}$ visit any subinterval $[a, b] \subseteq [0, 1]$, but it does so with the right frequency. In other words, the sequence $\{n\theta \pmod 1\}$ is *uniformly* dense in $[0, 1]$.

Now let $\theta_1, \theta_2, ..., \theta_k$, and 1 be rationally independent. To see that the sequence $x_n = (n\theta_1, ..., n\theta_k) \pmod 1$, $n = 1, 2, ...$ is uniformly distributed in $\mathbb{T}^k$, it is enough (in accordance with Exercise 2.9) to show that for any nonzero $h = (h_1, h_2, ..., h_k) \in \mathbb{Z}^k$ the sequence $\{\langle h, x_n \rangle \pmod 1\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{T}$. Observe that if $(h_1, h_2, ..., h_k) \neq (0, 0, ..., 0)$, then the number $\gamma = \sum_{i=1}^{k} h_i\theta_i$ is irrational. Hence, the sequence $\{n\gamma \pmod 1\}_{n\in\mathbb{N}}$ is uniformly distributed in $\mathbb{T}$. The identity

$$n\gamma (\mathrm{mod}\ 1) = \Big(n\sum_{i=1}^{k} h_i\theta_i\Big)(\mathrm{mod}\ 1) = \Big(\sum_{i=1}^{k} h_i n\theta_i\Big)(\mathrm{mod}\ 1) = \langle h, x_n\rangle(\mathrm{mod}\ 1)$$

implies that we are done.

Exercise 2.11 Let T_α, where $\alpha = (\alpha_1, ..., \alpha_k) \in \mathbb{R}^k$ be the rotation of $\mathbb{T}^k$ defined by $T_\alpha(x_1, ..., x_k) = (x_1 + \alpha_1, ..., x_k + \alpha_k) \pmod 1$. For a fixed $x \in \mathbb{T}^k$, let $X = \overline{\{T_\alpha^n x, n \in \mathbb{Z}\}}$. Prove that there exists a unique, T_α-invariant Borel measure on X. (T_α-invariance means that for any Borel set $A \subseteq X$, $\mu(T_\alpha^{-1}A) = \mu(A)$.) Show that the sequence $x_n = T_\alpha^n x, n = 1, 2, ...,$ is uniformly distributed with respect to μ.

We will now turn our attention to another refinement of Kronecker's theorem.

Theorem 2.12 ([33]). *If $p(x) = \alpha_s x^s + \alpha_{s-1}x^{s-1} + ... + \alpha_1 x + \alpha_0$ is a polynomial with at least one of its coefficients other than the constant term irrational, then the sequence $\{p(n) \pmod 1\}_{n\in\mathbb{N}}$ is dense in $\mathbb{T}$.*

We shall discuss several different approaches to the proof of Theorem 2.12.

The first possibility is to polynomialize the simple combinatorial proof of Theorem 2.1 above. We illustrate the ideas involved on the special case of $p(x) = \theta x^k$, where θ is irrational, and indicate how to extend the proof to the general case. First notice that to show that the sequence $\{n^k\theta \pmod 1\}_{n\in\mathbb{N}}$ is dense in $\mathbb{T}$, it is enough to prove that for any $\varepsilon > 0$ there exists n_0 satisfying either $0 < n_0^k\theta \pmod 1 < \varepsilon$ or $1 - \varepsilon < n_0^k\theta \pmod 1 < 1$.

Indeed, let $0 < a < \varepsilon$. Assuming that ε is close enough to a, let $N = \max\{n \in \mathbb{N} : n^k a < 1\}$ (so that, under our assumption, $N < \varepsilon^{-\frac{1}{k}}$). The largest possible distance between consecutive numbers of the form $n^k a$, $1 \leq n \leq N$, is not greater than $N^k a - (N-1)^k a < kN^{k-1}\varepsilon < k\varepsilon^{\frac{1}{k}}$. Also, it follows from the definition of N that $1 - N^k a < (N+1)^k a - N^k a < 2^k N^{k-1}\varepsilon < 2^k\varepsilon^{\frac{1}{k}}$.

We see that if for some $n_0 \in \mathbb{N}$, $0 < n_0^k\theta \pmod 1 = a < \varepsilon$, then the multiples $n^k n_0^k\theta \pmod 1 = (nn_0)^k\theta \pmod 1$, $n = 1, 2, ..., N$, with $N = [\varepsilon^{-\frac{1}{k}}]$, are $2^k\varepsilon^{\frac{1}{k}}$-dense in $[0, 1]$. Since k is fixed and ε is arbitrary, it gives the desired density of $\{n^k\theta \pmod 1\}_{n\in\mathbb{N}}$. Now, to show that for any $\varepsilon > 0$ there exists an n_0, satisfying either $0 < n_0^k\theta \pmod 1 < \varepsilon$ or $1-\varepsilon < n_0^k\theta \pmod 1 < 1$, we shall employ the multidimensional version of the celebrated Van der Waerden's theorem on arithmetic progressions, which will be proved in the next section. (An alternative, more elementary approach is offered in the Appendix.)

Theorem 2.13 (Gallai). *If $k \in \mathbb{N}$ and $\mathbb{N}^k = \bigcup_{i=1}^{r} C_i$ is an arbitrary finite partition of $\mathbb{N}^k$, then one of the C_i contains an "affine k-cube" of the form*

$$Q(n_1, n_2, ..., n_k; h) = \{(n_1 + \delta_1 h, n_2 + \delta_2 h, ..., \quad n_k + \delta_k h) : \\ \delta_i \in \{0, 1\},\ i = 1, 2, ..., k\}.$$

To apply Theorem 2.13 to our situation, let us, given $\varepsilon > 0$, induce a partition of $\mathbb{N}^k$ into $r = [\frac{2^k}{\varepsilon}] + 1$ subsets by the following rule:

$$(n_1, n_2, ..., n_k) \in C_i \Leftrightarrow n_1 \cdot n_2 \cdots n_k\theta \pmod 1 \in \Big[\frac{i-1}{r}, \frac{i}{r}\Big),\ i = 1, 2, ..., r.$$

We shall need the following identity:

$$h^k = \sum_{d=0}^{k} \sum_{\substack{\mathcal{D}\subseteq\{1,2,\ldots,k\} \\ |\mathcal{D}|=d}} (-1)^d \prod_{i\in\mathcal{D}} n_i \prod_{i\notin\mathcal{D}} (n_i + h).$$

(The identity looks frightening, but it is just a concise form of writing down a simple counting procedure. The reader is invited to gain some insight by reflecting on the following special case:

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

Let $Q(n_1, n_2, ..., n_k; h)$ be the cube generated by Theorem 2.13. For any vertex of this cube of the form $(n_1 + \delta_1 h, n_2 + \delta_2 h, ..., n_k + \delta_k h)$, where $\delta_i \in \{0, 1\}$, $i = 1, 2, ..., k$, one has $(\prod_{i=1}^{k}(n_i + \delta_i h))\theta \pmod 1 \in [\frac{i-1}{r}, \frac{i}{r})$. Since the sum of the coefficients $\sum_{d=0}^{k} \sum_{\substack{\mathcal{D}\subseteq\{1,2,\ldots,k\} \\ |\mathcal{D}|=d}} (-1)^d$ equals zero (there are precisely 2^k coefficients), one has either $0 < h^k\theta \pmod 1 < \frac{2^k}{r} < \varepsilon$ or $1-\varepsilon < h^k\theta \pmod 1 < 1$. This finishes the proof.

Exercise 2.14 Extend the previous proof to general polynomials by using the following strengthened form of Theorem 2.13: Given $k_1, k_2, ..., k_s \in \mathbb{N}$, suppose $\mathbb{N}^{k_j} = \bigcup_{i=1}^{r_j} C_i^{(j)}$, $1 \leq i \leq s$, are arbitrary finite partitions of the lattices. There exists $h \in \mathbb{N}$ so that the cubes Q_i, $i = 1, 2, ..., s$, whose existence is promised by Theorem 2.13, all have h as the same "edge length".

We now show how one can replace the use of Theorem 2.13 by a simple dynamical argument due to H. Furstenberg (see [13]).

Definition 2.15 Let $(X, \{T_g\}_{g\in G})$ be a topological dynamical system. A point $x \in X$ is called *recurrent* if for any neighborhood V containing x there exists $g \in G$, $g \neq e$ so that $T_g x \in V$.

Exercise 2.16 Show that every point $x \in \mathbb{T}$ is recurrent with respect to the transformation $T_\alpha : x \to x + \alpha \pmod 1$. (This follows from the discussion in the beginning of this section. Note that it is immaterial whether α is rational or irrational.)

Now, fix an irrational number θ and consider the dynamical system on $\mathbb{T}^2$ defined by $T : (x, y) \to (x + \theta, y + 2x + \theta) \pmod 1$. Following Furstenberg, we show that *every* point in $\mathbb{T}^2$ is recurrent under T. It is easy to verify by induction that for any integer n, $T^n(0, 0) = (n\theta, n^2\theta) \pmod 1$, and so the fact that $(0, 0)$ is a recurrent point will imply, in particular, that for any ε, there is an integer $n \neq 0$ such that $|n^2\theta - m| < \varepsilon$, or, equivalently, $|\theta - \frac{m}{n^2}| < \varepsilon$. As it was explained above, this implies, in its turn, that the sequence $\{n^2\theta \pmod 1\}_{n=1}^{\infty}$ is dense in $[0, 1]$.

We remark that the argument below applies actually to a much wider class of dynamical systems; namely, the so-called *group extensions* (see [13]).

To show that a point $(x, y) \in \mathbb{T}^2$ is recurrent, it is enough to show that $(x, 0)$ is recurrent. Indeed, denoting the transformation $(x, y) \to (x, y + t) \pmod 1$ by S_t, we see that if $T^{n_k}(x, 0) \to (x, 0)$ as $k \to \infty$, then

$$T^{n_k}(x, y) = T^{n_k} S_y(x, 0) = S_y T^{n_k}(x, 0) \to S_y(x, 0) = (x, y).$$

Now, to prove that the point $(x,0)$ is recurrent, one argues as follows:

Let $\overline{O((x,0))}$ denote the (forward) orbit closure of $(x,0)$: $\overline{O((x,0))} = \overline{\{T^n(x,0), n \geq 1\}}$. Since every point $x \in \mathbb{T}$ is a recurrent point for the transformation $T_\theta : x \to x + \theta \pmod 1$ (see Exercise 2.16), there exists $y_0 \in \mathbb{T}$ so that $(x, y_0) \in \overline{O((x,0))}$. Using the fact that for any $z \in \mathbb{T}$, $\overline{O(S_y z)} = S_y(\overline{O(z)})$, one gets $(x, 2y_0) \in \overline{O((x,y_0))}$ (we are suppressing the $\pmod 1$ sign), which implies $(x, 2y_0) \in \overline{O((x,0))}$. Repeating this argument, we get that for every $n \geq 1$, $(x, ny_0) \in \overline{O((x,0))}$. Applying again the result of Exercise 2.16 to y_0 and using the fact that $\overline{O((x,0))}$ is closed, we obtain $(x,0) \in \overline{O((x,0))}$. ■

Exercise 2.17 (i) Define $T : \mathbb{T}^3 \to \mathbb{T}^3$ by

$$T(x,y,z) = (x+\theta, y+2x+\theta, z+3y+3x+\theta) \pmod 1.$$

Check that for any $n \in \mathbb{Z}$, $T^n(0,0,0) = (n\theta, n^2\theta, n^3\theta) \pmod 1$. Generalize this to arbitrarily many dimensions.
(ii) Using (i), show that $\{n^k\theta \pmod 1\}_{n\in\mathbb{N}}$ is dense in $[0,1]$ for any $k \in \mathbb{N}$ and irrational θ.

It turns out that the sequence $\{p(n) \pmod 1\}_{n\in\mathbb{N}}$, where the polynomial $p(t) \in \mathbb{R}[t]$ has at least one "non-constant"coefficient irrational, is actually uniformly distributed in $[0,1]$.

We shall now describe two approaches to the proof of this fact. The first one, due to Weyl, is based on the fact that if $p(n)$ is a polynomial of degree $d \geq 1$, then for any $h \neq 0$, the polynomial $p(n+h) - p(n)$ has degree $d-1$. The principle behind the Weyl proof was succinctly formulated by van der Corput in the form of the following proposition (cf. [27]; see also Proposition 4.27 and Exercise 5.5 (i), (ii) below).

Theorem 2.18 (van der Corput's Difference Theorem). *Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of real numbers. If for any $h \in \mathbb{N}$, $h \neq 0$, the sequence $\{x_n - x_{n+h}\}_{n\in\mathbb{N}}$ is uniformly distributed* $\pmod 1$, *then the sequence $\{x_n\}_{n\in\mathbb{N}}$ is uniformly distributed* $\pmod 1$.

Exercise 2.19 Use Theorem 2.18 to prove that if the polynomial $p(n)$ has at least one "non-constant" coefficient irrational, then the sequence $\{p(n)\}_{n\in\mathbb{N}}$ is uniformly distributed $\pmod 1$.

The ergodic approach to equidistribution of polynomials, due to Furstenberg, relies on the notion of unique ergodicity. We give here only some general explanations (see [13], Chapter 3 for the details).

Definition 2.20 Let X be a compact metric space and $T : X \to X$ a continuous mapping. The dynamical system (X, T) is called *uniquely ergodic* if there exists only one T-invariant Borel probability measure on X.

To be able to talk about the uniqueness of invariant measures, one should first make sure that the set of invariant probability measures is non-empty. This is guaranteed by the following:

Theorem 2.21 (Bogoliouboff-Kryloff, [26]). *For any continuous self-map $T : X \to X$ of a compact metric space, there exists a T-invariant Borel probability measure.*

The following exercise indicates a way of proving Theorem 2.21. Recall that given a compact metric space X, the set $M(X)$ of Borel probability measures on X is non-empty, in particular, it contains the point masses (namely, the measures $\mu_x, x \in X$, defined for any Borel set A by $\mu_x(A) = 1$ if $x \in A$ and 0 otherwise), and is compact in the weak*-topology.

Exercise 2.22 Given a measure $\nu \in M(X)$, define $\mu_n = \frac{1}{n} \sum_{k=0}^{n-1} T^k \nu$. Let μ be any weak* limit point of the sequence $\{\mu_n\}_{n=1}^{\infty}$. Show that μ is T-invariant.

Exercise 2.23 Let $T(x) = ax$ be a rotation of a compact metrizable group $(G, \cdot)$. Prove that the Haar measure is the unique invariant measure for (G, T) if and only if the sequence $\{a^n\}_{n \in \mathbb{N}}$ is dense in G.

The uniquely ergodic systems are characterized by the following theorem (for a proof see for example [32], Theorem 6.19).

Theorem 2.24 *Let $T : X \to X$ be a continuous self-mapping of a compact metric space X. The dynamical system (X, T) is uniquely ergodic if and only if $\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$ converges uniformly to a constant.*

Return now to the transformation on $\mathbb{T}^2$ defined by $T(x,y) = (x+\alpha, y+2x+\theta)$ (mod 1) which we dealt with above. One can show that the normalized Lebesgue measure m on $\mathbb{T}^2$ is the unique invariant measure with respect to T. (See, for example, Proposition 3.10 in [13].) It follows from Theorem 2.24 that for any continuous function $f \in C(\mathbb{T}^2)$

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n(x,y)) \underset{N\to\infty}{\longrightarrow} \int_{\mathbb{T}^2} f\,dm \qquad \text{uniformly.}$$

In particular, taking $(x,y) = (0,0)$ and remembering that $T^n(0,0) = (n\alpha, n^2\alpha)$ (mod 1), one obtains the following result:

Proposition 2.25 *The sequence* $\{(n\alpha, n^2\alpha) \text{ (mod 1)}\}_{n=1}^{\infty}$ *is uniformly distributed in* $\mathbb{T}^2$.

Exercise 2.26 Taking for granted the unique ergodicity of the transformations suggested in Exercise 2.17 (i) and (ii), prove, with the help of Exercise 2.9, the Weyl's theorem on equidistribution of polynomials.

3 Ramsey Theory and Topological Dynamics

We start this section by formulating several combinatorial results (cf. [21]). The first of them, van der Waerden's theorem, is actually a 1-dimensional special case of Theorem 2.13.

Theorem 3.1 *If* $\mathbb{Z} = \bigcup_{i=1}^{r} C_i$ *is an arbitrary finite partitioning of* $\mathbb{Z}$*, then one of the* C_i *contains arbitrarily long arithmetic progressions.*

Exercise 3.2 Call a set $S \subset \mathbb{Z}$ *AP-rich* if S contains arbitrarily long arithmetic progressions. Prove that Theorem 3.1 is equivalent to the following statement: If $S \subset \mathbb{Z}$ is AP-rich, then for any finite partition of S, $S = \bigcup_{i=1}^{r} C_i$, one of the C_i is also AP-rich.

Now let F be a finite field and V_F an infinite vector space over F. (Example: $F = \mathbb{Z}_p$ and $V_F = \mathbb{Z}_p^{\infty} = \{(x_1, x_2, ...) : x_i \in \mathbb{Z}_p, i \in \mathbb{N}$, and all but finitely many $x_i = 0\}$.) A set $A \subset V_F$ is a d-dimensional *affine subspace* of V_F if for some $v, x_1, ..., x_d \in V$, where $x_1, ..., x_d$ are linearly independent, $A = v + \text{Span}\{x_1, ..., x_d\}$.

Theorem 3.3 (Geometric Ramsey Theorem, [20]). *If $r \in \mathbb{N}$ and $V_F = \bigcup_{i=1}^{r} C_i$, then one of the C_i contains affine subspaces of arbitrarily large (finite) dimension.*

Exercise 3.4 Call a subset $S \subset V_F$ *AS-rich* (AS stands for affine subspace) if S contains affine subspaces of arbitrarily high dimension. Prove that Theorem 3.3 is equivalent to the following statement: If $S \subset V_F$ is AS-rich and, for some $r \in \mathbb{N}, S = \bigcup_{i=1}^{r} C_i$ is a finite partitioning, then one of the C_i is also AS-rich.

Exercise 3.5 Give an example of a finite partition of $\mathbb{Z}_p^\infty$ (where p is a prime bigger than 2) so that none of the cells of the partition contains an infinite affine subspace. (When $p = 2$ the situation is different. See Exercise 3.7 (ii) below.)

The last combinatorial result we want to discuss is the celebrated Hindman's theorem. To formulate it we first introduce some notation that will be used throughout this section.

Let $\mathcal{F}$ denote the set of all non-empty finite sets in $\mathbb{N}$. If $\alpha, \beta \in \mathcal{F}$, then we write $\alpha < \beta$ if the maximal element in α is smaller than the minimal element of β. We shall say that $\{\alpha_i\}_{i=1}^\infty$ is an increasing sequence in $\mathcal{F}$ if $\alpha_1 < \alpha_2 < \dots$.

Given an increasing sequence $\{\alpha_i\}_{i=1}^\infty$ in $\mathcal{F}$, we shall denote by $FU(\{\alpha_i\}_{i=1}^\infty)$ the set of all finite unions of "atoms" $\alpha_i, i \in \mathbb{N}$. Note that the set $FU(\{\alpha_i\}_{i=1}^\infty)$ has, in a sense, the same structure as $\mathcal{F}$ (the atoms α_i play the same role as the singletons $\{i\}$ in $\mathbb{N}$).

Theorem 3.6 ([22]). *If $\mathcal{F} = \bigcup_{i=1}^{r} C_i$ is an arbitrary finite partition of $\mathcal{F}$, then one of the C_i contains $FU(\{\alpha_i\}_{i=1}^\infty)$ for some increasing sequence $\{\alpha_i\}_{i=1}^\infty$ in $\mathcal{F}$.*

Exercise 3.7 For an infinite subset $\{x_1, x_2, \dots\} \subset \mathbb{N}$, let

$$FS(\{x_i\}_{i=1}^\infty) = \{x_{i_1} + x_{i_2} + \dots + x_{i_k} \ : \ i_1 < i_2 < \dots < i_k, \ k \in \mathbb{N}\}.$$

In other words, $FS(\{x_i\}_{i=1}^\infty)$ is the set of all finite sums of elements of the set $\{x_1, x_2, \dots\}$ having distinct indices.

(i) Prove that Theorem 3.6 is equivalent to the following statement: If $r \in \mathbb{N}$ and $\mathbb{N} = \bigcup_{i=1}^{r} C_i$, then one of the C_i contains $FS(\{x_i\}_{i=1}^{\infty})$ for some infinite subset $\{x_1, x_2, ...\}$.
(ii) Prove that for any finite partition of the vector space $\mathbb{Z}_2^{\infty}$, one of the cells must contain an infinite subspace with the possible exception of the zero vector.

The combinatorial results above have a common feature: they all state that certain structures are undestroyable by finite partitioning. Theorems 3.1, and 3.3 belong to a vast variety of results which form the body of Ramsey Theory and which have the following general form: if a highly organized structure (complete graph, ordered set, vector space, etc.) is finitely partitioned (or as they say, finitely colored), then one of the pieces will still be highly organized.

We are not going to give a proof of Hindman's theorem here. There are many interesting proofs of this theorem in the literature, each of them lending some new insights (see for example [23], [18], [13], or [3]).

We are now going to formulate and prove a dynamical theorem which has Theorems 3.1 and 3.3 as corollaries. Before formulating it, we introduce a few more definitions and some notation.

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

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

The following theorem was first proved in [18]; the proof that we give here is based on a proof of its special case in [10].

Theorem 3.8 *Let X be a compact topological space and G a commutative group of its homeomorphisms such that the dynamical system (X, G) is minimal. For any non-empty open set $V \subseteq X$, $k \in \mathbb{N}$, any IP-systems*

$\{T_\alpha^{(1)}\}_{\alpha\in\mathcal{F}}, ..., \{T_\alpha^{(k)}\}_{\alpha\in\mathcal{F}}$ in G and any $\alpha_0 \in \mathcal{F}$, there exists $\alpha \in \mathcal{F}, \alpha > \alpha_0$ such that $V \cap T_\alpha^{(1)} V \cap ... \cap T_\alpha^{(k)} V \neq \emptyset$.

Proof. We fix a non-empty open $V \subseteq X$ and $S_1, ..., S_r \in G$ with the property that $S_1 V \cup S_2 V \cup \ldots \cup S_r V = X$. (The existence of $S_1, ..., S_k$ is guaranteed by the minimality of (X, G).) The proof proceeds by induction on k. The case $k = 1$ is almost trivial, but we shall do it in detail to set up the notation in a way that indicates the general idea.

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

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

(i) $W_0 = V$;

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

(iii) each $W_n, n \geq 1$, is contained in one of the sets $S_1 V, S_2 V, ..., S_r V$. (We recall that $S_1 V \cup S_2 V \cup ... \cup S_r V = X$.)

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

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

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

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

Assume now that the theorem holds for any k IP-systems in G. Fix a non-empty set V and $k+1$ IP-systems $\{T_\alpha^{(1)}\}_{\alpha\in\mathcal{F}}, ..., \{T_\alpha^{(k+1)}\}_{\alpha\in\mathcal{F}}$. We shall also fix the homeomorphisms $S_1, ..., S_r \in G$ (whose existence is guaranteed by minimality) satisfying $S_1V \cup ... \cup S_rV = G$. We shall inductively construct a sequence $W_0, W_1, ...$ of non-empty open sets in X and an increasing sequence $\alpha_1 < \alpha_2 < ...$ in $\mathcal{F}$ so that

(a) $W_0 = V$,

(b) $(T^{(1)}_{\alpha_n})^{-1}W_n \cup (T^{(2)}_{\alpha_n})^{-1}W_n \cup \ldots \cup (T^{(k+1)}_{\alpha_n})^{-1}W_n \cup \subseteq W_{n-1}$ for all $n \geq 1$, and

(c) each $W_n, n \geq 1$ is contained in one of the sets $S_1V, ..., S_rV$.

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

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

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

$$\begin{aligned} &V \cap (T^{(k+1)}_{\alpha_1})^{-1}T^{(1)}_{\alpha_1}V \cap \ldots \cap (T^{(k+1)}_{\alpha_1})^{-1}T^{(k)}_{\alpha_1}V \\ &\quad = W_0 \cap (T^{(k+1)}_{\alpha_1})^{-1}T^{(1)}_{\alpha_1}W_0 \cap \ldots \cap (T^{(k+1)}_{\alpha_1})^{-1}T^{(k)}_{\alpha_1}W_0 \neq \emptyset. \end{aligned}$$

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

$$T^{(k+1)}_{\alpha_1}W_0 \cap T^{(1)}_{\alpha_1}W_0 \cap \ldots \cap T^{(k)}_{\alpha_1}W_0 \neq \emptyset.$$

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

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

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

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

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

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

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

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

Since, by the construction, each W_n is contained in one of the sets $S_1V, ..., S_rV$, there is $1 \leq t \leq r$ such that infinitely many of the W_n are contained in S_tV. In particular, there exists i as large as we please and $j > i$ so that $W_i \cup W_j \subseteq S_tV$. Let $U = S_t^{-1}W_j$ and $\alpha = \alpha_{i+1} \cup ... \cup \alpha_j$.

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

$$(T_\alpha^{(m)})^{-1}U = (T_{\alpha_{i+1}\cup...\cup\alpha_j}^{(m)})^{-1}S_t^{-1}W_j = S_t^{-1}(T_{\alpha_{i+1}}^{(m)})^{-1}\dots(T_{\alpha_j}^{(m)})^{-1}W_j$$
$$\subseteq S_t^{-1}(T_{\alpha_{i+1}}^{(m)})^{-1}\dots(T_{\alpha_{j-1}}^{(m)})^{-1}W_{j-1} \subseteq \dots \subseteq S_t^{-1}(T_{\alpha_{i+1}}^{(m)})^{-1}W_{i+1} \subseteq S_t^{-1}W_i \subseteq V$$

It follows that $U \cup (T_\alpha^{(1)})^{-1}U \cup \dots \cup (T_\alpha^{(n+1)})^{-1}U \subseteq V$, and this, in turn, implies $V \cap T_\alpha^{(1)}V \cap \dots \cap T_\alpha^{(k+1)}V \neq \emptyset$. ■

Corollary 3.9 *If X is a compact metric space and G a group of its homeomorphisms, then for any k IP-systems $\{T_\alpha^{(1)}\}_{\alpha\in\mathcal{F}}, ..., \{T_\alpha^{(k)}\}_{\alpha\in\mathcal{F}}$ in G, any $\alpha_0 \in \mathcal{F}$, and any $\varepsilon > 0$ there exists $\alpha > \alpha_0$ and $x \in X$ such that the diameter of the set $\{x, T_\alpha^{(1)}x, ..., T_\alpha^{(k)}x\}$ is smaller than ε.*

Proof. If (X, G) is minimal, then the claim follows immediately from Theorem 3.8. If not, then pass to a minimal, non-empty, closed G-invariant subset of X. (It always exists by Zorn's lemma.)

Exercise 3.10 Under the conditions of Corollary 3.9, show that for any $m \in \mathbb{N}$ one can always find $\alpha_1 < \alpha_2 < ... < \alpha_m$ and x such that the set

$$\{T_{\alpha_1}^{(i_1)}T_{\alpha_2}^{(i_2)}\dots T_{\alpha_m}^{(i_m)}x \ : \ i_1, ..., i_m \in \{1, ..., k\}\}$$

has diameter smaller than ε.

Let us now show how to derive Theorems 3.1 and 3.3 from Corollary 3.9. We start with Theorem 3.1. Let $r \in \mathbb{N}, r \geq 2$, and let $\Omega = \{1, 2, ..., r\}^{\mathbb{Z}}$ be the (compact) space of all bilateral sequences with entries from the set $\{1, 2, ..., r\}$ with the product topology. We shall use the standard metric on Ω defined by

$$d(x,y) = \begin{cases} 0, \text{ if } x = y \\ 1, \text{ if } x(0) \neq y(0) \\ \frac{1}{k+1} \text{ otheriwise, where } k \text{ is the maximal natural number with} \\ \quad\quad \text{the property } x(i) = y(i) \text{ for all } |i| < k. \end{cases}$$

The points of Ω are in a natural one-to-one correspondence with the partitions of $\mathbb{Z}$ into r sets: if $x = x(n) \in \Omega$, set $C_i = \{n \in \mathbb{Z} : x(n) = i\}$, $i = 1, 2, ..., r$. Let $T : \Omega \to \Omega$ be the shift homeomorphism defined by $(Tx)(n) = x(n+1)$. Fix a partition $\mathbb{Z} = \bigcup_{i=1}^{r} C_i$, and let $\xi = \xi(n)$ be the corresponding sequence in Ω. Finally, let $X \subseteq \Omega$ be the orbital closure of ξ:

$$X = \overline{\{T^n \xi, n \in \mathbb{Z}\}}$$

Let $\{n_i\}_{i=1}^{\infty}$ be an arbitrary sequence in $\mathbb{Z}$. For $\alpha \in \mathcal{F}$ define $n_\alpha := \sum_{i \in \alpha} n_i$, and consider IP-systems

$$T_\alpha^{(m)} := T^{mn_\alpha}, \ m = 1, 2, ..., k.$$

By Corollary 3.9, for any $\varepsilon > 0$ there exist $\alpha \in \mathcal{F}$ and $x \in X$ such that the diameter of $\{x, T^{n_\alpha} x, ..., T^{kn_\alpha} x\}$ is less than ε. Let $\eta > 0$ be such that $d(x, y) < \eta$ implies that the sequences $x(n)$ and $y(n)$ coincide for $|n| \leq k|n_\alpha|$. Since the orbit $\{T^n \xi, n \in \mathbb{Z}\}$ is dense in X, there exists $m_0 \in \mathbb{Z}$ such that $T^{m_0}\xi$ and x agree on the interval $[-k|n_\alpha|, k|n_\alpha|]$.

It follows that $\xi(m_0) = \xi(m_0 + n_\alpha) = \ldots = \xi(m_0 + kn_\alpha)$. If this common value is j, then clearly C_j contains an arithmetic progression of length k. Note that as a by-product, we showed that the difference of the length k arithmetic progression always to be found in one cell of the partition can be chosen from any prescribed IP-set.

Exercise 3.11 Prove that van der Waerden's theorem is equivalent to the following:

Statement. Let $\mathbb{Z} = \bigcup_{i=1}^{r} C_i$ be an arbitrary finite partition of the integers and $F \in \mathcal{F}$. One of the C_i necessarily contains an affine image of F. In other words, one of the C_i contains a set of the form $a + nF = \{a + nx : x \in F\}$ for some $a \in \mathbb{Z}$ and $n \in \mathbb{N}$.

The above statement is a special case of the following multidimensional version of van der Waerden's theorem which also follows from Theorem 3.8.

Theorem 3.12 *For any $d \in \mathbb{N}$ and finite subset $F \subset \mathbb{Z}^d$, if $\mathbb{Z}^d = \bigcup_{i=1}^{r} C_i$ is a finite partition of $\mathbb{Z}^d$, then one of the C_i contains a set of the form $a + nF = \{a + nx : x \in F\}$ for some $a \in \mathbb{Z}^d, n \in \mathbb{N}$.*

Exercise 3.13 Prove that Theorems 3.12 and 2.13 are equivalent.

We move now to the derivation of Theorem 3.3. Let V_F be a vector space over a finite field F of characteristic p. Without loss of generality, we shall assume that V_F is countable. As an abelian group, V_F has a natural representation as the direct sum of countably many copies of F:

$$F^\infty = \{g = (a_1, a_2, ...,) : a_i \in F \text{ and all but finitely many } a_i = 0\}.$$

Fix an IP-system $\{g_\alpha\}_{\alpha\in\mathcal{F}}$ such that $\mathrm{Span}\{g_\alpha, \alpha \in \mathcal{F}\}$ is an infinite subset in V_F. We will show a stronger fact that if V_F is partitioned into r subsets $C_1, C_2, ..., C_r$, then one of the C_i contains an affine subspace of arbitrarily large dimension which is generated by elements of $\{g_\alpha\}_{\alpha\in\mathcal{F}}$. Let $\Omega = \{1, 2, ...r\}^{V_F}$. In other words, Ω is the set of all functions defined on V_F which take values in $\{1, 2, ..., r\}$. With its product topology, Ω is a compact topological space. Introduce a metric on Ω analogous to the one used for $\{1, ..., r\}^{\mathbb{Z}}$ above. For $g = (a_1, a_2, ...) \in V_F$, let $|g|$ be the minimal natural number such that $a_i = 0$ for all $i > |g|$. For $x = x(g)$ and $y = y(g)$ in Ω, define

$$d(x,y) = \begin{cases} 0, \text{ if } x = y \\ 1, \text{ if } x(\mathbf{0}) \neq y(\mathbf{0}) \\ \frac{1}{k+1} \text{ otheriwise, where } k \text{ is the maximal natural number with} \\ \quad\text{the property } x(g) = y(g) \text{ for all } |g| < k \end{cases}$$

(where $\mathbf{0}$ denotes the element $(0, 0, ...) \in V_F$). Let $V_F = \bigcup_{i=1}^{r} C_i$ be a partition of V_F, and define $\xi \in \Omega$ by $\xi(g) = i \Leftrightarrow g \in C_i$.

We show first that one of the C_i contains an affine line (i.e. a one-dimensional affine subspace). For $h \in V_F$, define $T_h : \Omega \to \Omega$ by $(T_h x)(g) = x(gh)$. Clearly T_h is a homeomorphism of Ω for every $h \in V_F$. Let $X \subseteq \Omega$ be the orbital closure of $\xi(g)$: $X = \overline{\{T_h\xi, h \in V_F\}}$.

Use now the IP-system $\{g_\alpha\}_{\alpha\in\mathcal{F}}$ to define an IP-system of homeomorphisms of X. Put $T_\alpha := T_{g_\alpha}, \alpha \in \mathcal{F}$, and for each $c \in F, c \neq 0$, define an IP-system by $T_\alpha^{(c)} = T_{cg_\alpha}, \alpha \in \mathcal{F}$. This way we get $q-1$ (where $q = |F|$) IP-systems of commuting homeomorphisms of Ω (and of X). Applying Corollary 3.9 to the space X and these IP-systems and taking $\varepsilon < 1$, we get a point $x \in X$ and $\alpha_1 \in \mathcal{F}$ such that the diameter of $\{T_{c\alpha_1}x, c \in \mathcal{F}\}$ is less than 1. This implies that $x(0) = x(cg_{\alpha_1})$ for every $c \in F$. Since the orbit

$\{T_h\xi, h \in V_F\}$ is dense in X, there exists $h_0 \in V_F$ such that $(T_{h_0}\xi)(g)$ and $x(g)$ agree on all g satisfying $|g| \le |g_{\alpha_1}|$. If $\xi(h_0) = i$ then C_i contains the affine line $\{h_0 + cg_{\alpha_1}, c \in F\}$ (in view of our assumptions on $\{g_\alpha\}_{\alpha\in\mathcal{F}}$, we, of course, took care to choose α_1 so that $g_{\alpha_1} \ne \mathbf{0}$.) The statement about affine spaces of arbitrary dimension follows by iteration (cf. Exercise 3.10) and is left to the reader.

4 Density Ramsey Theory and Ergodic Theory of Multiple Recurrence

In this section, we concern ourselves with *density* Ramsey Theory and its links to ergodic theory. While the main theme of *partition* Ramsey theory is to look for nontrivial patterns in one cell of an arbitrary finite partition, the typical density Ramsey theory statement concerns an appropriately defined notion of largeness: any large subset of a highly organized structure contains large, highly organized substructures. Two basic properties are usually required from the notion of largeness:

(i) if A is large and $A = \bigcup_{i=1}^{r} C_i$, then at least one of the C_i is large;

(ii) the family of large subsets of a given set with a particular structure is invariant under some natural semigroup of structure preserving transformations.

We shall discuss now the density version of Theorems 3.1 and 3.3. (As for the density version of Theorem 3.6, see the discussion in [3], Section 4.) As Graham, Rothschild and Spencer put it in [21], "for all Ramsey theorems, one can express (but not always prove) the corresponding density statements." For a set $E \subseteq \mathbf{Z}$ define its *upper density* by

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

Clearly the property of a set of integers to have positive upper density satisfies property (i) above. It is also invariant with respect to the shift: for any $k \in \mathbf{Z}, \bar{d}(E+k) = \bar{d}(E)$, where $E + k = \{x + k, x \in E\}$.

We formulate now a density version of van der Waerden's theorem. We say *a* density rather than *the* density version since $\bar{d}$ is not the only notion of largeness which leads to a density generalization of van der Waerden's theorem. See Exercise 4.2 below.

Theorem 4.1 (Szemerédi, [31]). *If a set $E \subseteq \mathbf{Z}$ has positive upper density, then it contains arbitrarily long arithmetic progressions.*

Exercise 4.2 (i) Derive from Theorem 4.1 the following finitistic version of it: For any $\varepsilon > 0$ and $k \in \mathbf{N}$ there exists $N = N(\varepsilon, k) \in \mathbf{N}$ such that if $I = [a, b] \subset \mathbf{Z}$ is an interval with $|b - a| \geq N$ and $E \subseteq I$ satisfies $\frac{|E|}{|I|} > \varepsilon$, then E contains a k-term arithmetic progression.
(ii) For any $E \subseteq \mathbf{Z}$, let the quantity $d^*(E) = \limsup_{N-M\to\infty} \frac{|E\cap\{M,M+1,\ldots,N\}|}{N-M+1}$ denote its *upper Banach density.* Call a set $E \subseteq \mathbf{Z}$ *d^*-large* if $d^*(E) > 0$. Clearly, E is d^*-large if and only if for some sequence of intervals

$$I_N = [a_N, b_N] \subseteq \mathbf{Z} \text{ with } |b_N - a_N| \to \infty, \ \limsup_{N\to\infty} \frac{|E \cap I_N|}{|I_N|} > 0.$$

(In other words, each such sequence $I_N, N = 1, 2, \ldots$, defines a notion of largeness (check!), and to be d^*-large means to be large with respect to some sequence of intervals of increasing length). Prove that Theorem 4.1 implies that a d^*-large set contains arbitrarily long arithmetic progressions.

The original proof of Theorem 4.1 in [31] is a brilliant and highly nontrivial piece of combinatorial reasoning. A different, ergodic theoretical proof was given by Furstenberg in [12], thereby starting a new branch of mathematics, Ergodic Ramsey Theory. Soon the methods of ergodic theory turned out to be very useful in proving some natural density conjectures and have led to some strong results, which so far have no conventional combinatorial proofs (cf. [15], [16], [17], [4], [5], [6], [29], [7], [8]).

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 4.3 (Furstenberg, [12]). *Let $(X, \mathcal{B}, \mu, T)$ be an invertible probability measure preserving system. For any $k \in \mathbf{N}$ and $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $n \in \mathbf{N}$ such that*

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

To see that Theorem 4.1 follows from Theorem 4.3, one needs the following:

Theorem 4.4 *(Furstenberg's Correspondence Principle.) Let $E \subseteq \mathbf{Z}$ with $d^*(E) > 0$. Then there exist an invertible probability measure preserving system $(X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{B}$ with $\mu(A) = d^*(E)$ such that for any $k \in \mathbf{N}$ and $n_1, n_2, ..., n_k \in \mathbf{Z}$ one has:*

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

Exercise 4.5 Prove that Theorem 4.1, in its turn, implies Theorem 4.3.

We shall formulate now a density version of Theorem 3.3. Let V_F be a countable vector space over a finite field F. As before, it will be convenient to work with the realization of V_F as a direct sum of countably many copies of F:

$$F^\infty = \{g = (a_1, a_2, \ldots) : a_i \in F, i \in \mathbf{N} \text{ and } a_i = 0 \text{ for all but finitely many } i\}.$$

Let $F_n = \{g = (a_1, a_2, \ldots) : a_i = 0\ \forall i > n\}$. For a set $S \subseteq V_F$ define its *upper density*, $\bar{d}(s)$ by

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

Theorem 4.6 *Let $S \subseteq V_F$ with $\bar{d}(S) > 0$. Then S contains affine subspaces of arbitrarily high dimension.*

Exercise 4.7 Define in $V_F = F^\infty$ the notion of upper Banach density similar to that defined in Exercise 4.2 (ii) for subsets of $\mathbf{Z}$. Derive from 4.6 a finitistic statement similar to that of Exercise 4.2 (i) and a statement obtained by replacing in the formulation of Theorem 4.6 the upper density by the upper Banach density.

Similarly to the situation with Szemerédi's theorem, Theorem 4.6 follows from a measure theoretical theorem dealing with multiple recurrence:

Theorem 4.8 *Let $\{T_g\}_{g\in V_F}$ be a measure preserving action of $V_F = F^\infty$ on a probability measure space $(X, \mathcal{B}, \mu)$. Let $A \in \mathcal{B}, \mu(A) > 0$. Then for some $g \in V_F, g \neq e$ one has $\mu(\bigcap_{c\in F} T_{cg}A) > 0$.*

Corollary 4.9 *Under the assumptions of Theorem 4.8, for any k there exist $g_1, \ldots, g_k$ in V_F such that* $\dim\Big(\mathrm{Span}\{g_1, \ldots, g_k\}\Big) = k$ *and* $\mu(\bigcap_{i=1}^{k} \bigcap_{c \in F} T_{cg_i} A) > 0$.

Exercise 4.10 Derive Corollary 4.9 from Theorem 4.8.
Hint: if $g_1, \ldots, g_k$ have already been found, apply Theorem 4.8 to the set $A_k = \bigcap_{i=1}^{k} \bigcap_{c \in F} T_{cg_i} A$ and to the (sub) action $\{T_g\}_{g \in V_F^{(k)}}$, where, for some fixed m satisfying $m \geq \max_{1 \leq i \leq k} |g_i|$,

$$V_F^{(k)} = \{g = (a_1, a_2, \ldots) \in V_F \ : \ a_1 = a_2 = \ldots = a_m = 0\}$$

To derive Theorem 4.6 from Theorem 4.8 (or, rather, from Corollary 4.9) one utilizes the Furstenberg correspondence principle for V_F:

Theorem 4.11 *Let $S \subseteq V_F$ with $\bar{d}(S) > 0$. Then there exist a measure preserving system with "time" V_F, $(X, \mathcal{B}, \mu, \{T_g\}_{g \in V_F})$ and a set $A \in \mathcal{B}$ with $\mu(A) = \bar{d}(S)$ such that for any $k \in \mathbf{N}$ and any $g_1, g_2, ..., g_k \in V_F$ one has:*

$$\bar{d}(S \cap (S - g_1) \cap \ldots \cap (S - g_k)) \geq \mu(A \cap T_{g_1}^{-1} A \cap \ldots \cap T_{g_k}^{-1} A).$$

The apparent similarity of Theorems 4.4 and 4.11 hints that there is a general Furstenberg correspondence principle which encompasses both theorems. This is indeed so and the "right" class of groups to which it applies are countable *amenable* groups. Among the many equivalent definitions of amenability the one of major importance to ergodic theory is the following one.

Definition 4.12 *A countable group G is called amenable, if it has a (left) Følner sequence, namely, a sequence of finite sets $\Phi_n \subset G, n \in \mathbf{N}$ with $|\Phi_n| \to \infty$ and such that $\frac{|\Phi_n \cap g\Phi_n|}{|\Phi_n|} \to 1$ for all $g \in G$.*

Remark 4.13 Strictly speaking, Definition 4.12 is a definition of *left* amenability. *Right* amenability is defined as the property of group G of possessing a *right Følner sequence*, i.e. a sequence of finite sets $\Phi_n \subset G, n \in \mathbf{N}$ with $|\Phi_n| \to \infty$ such that $\frac{|\Phi_n \cap \Phi_n g|}{|\Phi_n|} \to 1$ for all $g \in G$. It is known, however, that the notions of right and left amenability coincide. At the same time one should be aware of the fact that in non-abelian groups not every right Følner sequence is necessarily a left Følner sequence and vice versa. Since in these notes we are mostly concerned with abelian groups, these subtleties will not bother us too much.

Exercise 4.14 (i) Let $M_n = [a_n^{(1)}, b_n^{(1)}] \times ... \times [a_n^{(d)}, b_n^{(d)}] \subset \mathbf{Z}^d, n \in \mathbf{N}$, and assume that $|b_n^{(i)} - a_n^{(i)}| \to \infty, i = 1, 2, ..., d$. Show that $M_n, n \in \mathbf{N}$ is a Følner sequence in $\mathbf{Z}^d$.
(ii) Check that the sets $F_n, n \in \mathbf{N}$ defined before the formulation of Theorem 4.6 form a Følner sequence in V_F.
(iii) Give an example of a left Følner sequence in a non-abelian group which is not a right Følner sequence.

The set of countable amenable groups is quite rich and includes all countable abelian groups and many classes of non-abelian ones, for example, solvable, locally finite, etc. On the other hand, the free group on two or more generators and the groups $SL(n, \mathbf{Z}), n \geq 2$ are not amenable.

The following version of the classical von Neumann ergodic theorem for amenable groups is merely an illustration of the principle that many results of conventional ergodic theory of one-parameter actions extend naturally to amenable groups.

Theorem 4.15 *Let H be a Hilbert space and let $\{U_g\}_{g\in G}$ be an antirepresentation of a countable amenable group G as a group of unitary operators on H (i.e. $U_{g_1}U_{g_2} = U_{g_2g_1}$ for all $g_1, g_2 \in G$). Let $H_{\mathrm{c}} = \{f \in H : U_g f = f \; \forall g \in G\}$ and let P be the orthogonal projection on H_{inv}. For any left Følner sequence $\Phi_n, n \in \mathbf{N}$ in G one has:*

$$\lim_{n\to\infty} \left\| \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} U_g f - Pf \right\| = 0.$$

Proof. In complete analogy with the case of a single operator (or rather the $\mathbf{Z}$-action generated by a single unitary operator) one checks that the orthogonal complement of $H_{\mathrm{inv}} \in H$, call it H_{erg}, coincides with the space $\overline{\mathrm{Span}\{f - U_g f : f \in H, g \in G\}}$. It remains to show that on H_{erg} the limit in question is zero. It is enough to prove it for the elements of the form $f - U_{g_0} f$. We have:

$$\left\| \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} U_g(f - U_{g_0} f) \right\| = \left\| \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} U_g f - \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} U_{g_0 g} f \right\|$$
$$= \left\| \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} U_g f - \frac{1}{|\Phi_n|} \sum_{g\in g_0\Phi_n} U_g f \right\| \leq \frac{|\Phi_n \triangle g_0\Phi_n|}{|\Phi_n|} \|f\|.$$

Since the definition of left Følner sequence clearly implies $\frac{|\Phi_n \triangle g_0 \Phi_n|}{|\Phi_n|} \to 0$, we are done.

Exercise 4.16 Recall that a measure preserving action $\{T_g\}_{g \in G}$ on a probability space $(X, \mathcal{B}, \mu)$ is *ergodic* if any set $A \in \mathcal{B}$ which satisfies $\mu(T_g A \triangle A) = 0$ for all $g \in G$ has measure zero or measure one. Derive from Theorem 4.15 the following statement:

If $\{T_g\}_{g \in G}$ is an ergodic measure preserving action of an amenable group G, then for any (left or left) Følner sequence $\Phi_n, n \in \mathbf{N}$ and any $A_1, A_2 \in B$ one has:

$$\frac{1}{|\Phi_n|} \sum_{g \in \Phi_n} \mu(A_1 \cap T_g A_2) \to \mu(A_1)\mu(A_2).$$

We shall formulate and prove now the Furstenberg correspondence principle for countable amenable groups.

Theorem 4.17 *Let G be a countable amenable group and $\Phi_n, n \in \mathbf{N}$ a left Følner sequence in G. Let $E \subseteq G$ have positive upper density with respect to $\{\Phi_n\}_{n=1}^{\infty}$: $\bar{d}(E) = \limsup_{n \to \infty} \frac{|E \cap \Phi_n|}{|\Phi_n|} > 0$ (for convenience we are suppressing in the notation $\bar{d}(E)$ the dependence on $\{\Phi_n\}$). Then there exists a probability measure preserving system with "time" G, $(X, \mathcal{B}, \mu, \{T_g\}_{g \in G})$ and a set $A \in \mathcal{B}$ with $\mu(A) = \bar{d}(E)$ such that for any $k \in \mathbf{N}$ and $g_1, g_2, ..., g_k \in G$ one has*

$$\bar{d}(E \cap g_1^{-1}E \cap ... \cap g_k^{-1}E) \geq \mu(A \cap T_{g_1}^{-1}A \cap ... \cap T_{g_k}^{-1}A).$$

Before giving a proof of Theorem 4.17 we shall need a crucial fact linking sets of positive upper density in G with invariant linear functionals on the space $B(G)$ of all complex-valued bounded functions on G. We recall first some definitions. Let G be a group and Y a closed subset of the space $B(G)$ of all bounded complex valued functions equipped with the uniform norm $\| \cdot \|_\infty$. Assume additionally that Y is closed under complex conjugation and contains all the constants. A linear functional $L : B(G) \to \mathbf{C}$ is a *left-invariant mean*, if it has the following additional properties:

(i) $L(\bar{f}) = \overline{L(f)} \quad \forall f \in Y$.

(ii) if $f \geq 0$ then $L(f) \geq 0$ and $L(1) = 1$.

(iii) For all $g \in G$ and $f \in Y$ $\ L({}_g f) = L(f)$ where ${}_g f(t) = f(gt)$.

Exercise 4.18 For $f \in Y$ define $\check{f}(t) = f(t^{-1})$. Show that, if L is a left-invariant mean on Y, the functional $\check{L}$ defined on $\check{Y} = \{\check{f} : f \in Y\}$ by $\check{L}(\check{f}) = L(f)$ is a *right-invariant* mean, i.e. for all $g \in G$ and $f \in \check{Y}$, $\check{L}(f^g) = \check{L}(f)$, where $f^g(t) = f(tg)$.

One can show that a countable group G is amenable if and only if the space $B(G)$ admits a left-invariant mean. We shall need this fact only in one direction; it follows from the following result.

Proposition 4.19 *Let G be a countable group and $\Phi_n, n \in \mathbf{N}$ a left Følner sequence in G. Assume that a set $E \subseteq G$ has positive upper density with respect to $\{\Phi_n\}$: $\bar{d}(E) = \limsup_{n\to\infty} \frac{|E\cap\Phi_n|}{|\Phi_n|} > 0$. There exists a left-invariant mean L on $B(G)$ satisfying the following conditions:*

(i) $L(1_E) = \bar{d}(E)$

(ii) *for any $k \in \mathbf{N}$ and any $g_1, g_2, .., g_k \in G$*

$$\bar{d}(E \cap g_1^{-1}E \cap ... \cap g_k^{-1}E) \geq L(1_E \cdot 1_{g_1^{-1}E} \cdot ... \cdot 1_{g_k^{-1}E}).$$

Proof. Let P be the (countable!) family of subsets of G having the form $\bigcap_{i=1}^{k} g_i^{-1}E$, where $g_i \in G$, $i = 1, \ldots, k$. By using the diagonal procedure we arrive at a subsequence $\{\Phi_{n_i}\}_{i=1}^{\infty}$ of our Følner sequence such that for our fixed set E we have $\bar{d}(E) = \lim_{i\to\infty} \frac{|E\cap\Phi_{n_i}|}{|\Phi_{n_i}|}$ and for any $S \in P$ the limit $L(S) = \lim_{i\to\infty} \frac{|S\cap\Phi_{n_i}|}{|\Phi_{n_i}|} = \lim_{i\to\infty} \frac{1}{|\Phi_{n_i}|} \sum_{g\in\Phi_{n_i}} 1_S(g)$ exists. Notice that for any $g_1, ..., g_k \in G$ this gives

$$\bar{d}(\bigcap_{j=1}^{k} g_j^{-1}E) = \limsup_{n\to\infty} \frac{\left|(\bigcap_{j=1}^{k} g_j^{-1}E) \cap \Phi_n\right|}{|\Phi_n|} \geq \lim_{i\to\infty} \frac{\left|(\bigcap_{j=1}^{k} g_j^{-1}E) \cap \Phi_{n_i}\right|}{|\Phi_{n_i}|} = L(\bigcap_{j=1}^{k} 1_{g_j^{-1}E}).$$

Extending out by linearity we get a linear functional L on the subspace $Y_0 \subset B_{\mathbf{R}}(G)$ of finite linear combinations of characteristic functions of sets in P. To extend L from Y_0 to $B_{\mathbf{R}}(G)$ define Minkowski functional $p(\varphi)$ by $p(\varphi) = \limsup_{i\to\infty} \frac{1}{|\Phi_{n_i}|} \sum_{g\in\Phi_{n_i}} \varphi(g)$. Clearly, for all $\varphi_1, \varphi_2 \in B_{\mathbf{R}}(G)$, $p(\varphi_1+\varphi_2) \leq p(\varphi_1) + p(\varphi_2)$ and for any non-negative t, $p(t\varphi) = tp(\varphi)$. Also, on Y_0, $L(\varphi) = p(\varphi)$. By Hahn-Banach theorem there is an extension of L (which we denote by L as well) to $B_{\mathbf{R}}(G)$ satisfying $L(\varphi) \leq p(\varphi)\ \forall \varphi \in B_{\mathbf{R}}(G)$. This L naturally extends to a functional on the space $B(G)$ of complex-valued bounded functions and we are done.

Remark 4.20 For the proof of the Furstenberg correspondence principle we shall need only the existence of linear functional satisfying the conditions (i) and (ii) and defined on the uniformly closed and closed under conjugation algebra of functions on G which is generated by the characteristic function 1_E and its shifts. As the first half of the proof above shows this could be achieved without involving the Hahn-Banach theorem.

Proof of Theorem 4.17. Let $f(h) = 1_E(h)$ be the characteristic function of E. Let $\mathcal{A}$ be the uniformly closed and closed under the conjugation algebra generated by the function f and its shifts of the form ${}_gf(h) = f(gh)$. $\mathcal{A}$ is a separable (linear combinations with rational coefficients are dense in $\mathcal{A}$), commutative C^*-algebra with respect to sup norm. By Gelfand representation theorem, $\mathcal{A}$ is isomorphic to the space $C(X)$ of continuous functions on a compact metric space X. Let L be a right invariant mean on $B(G)$ satisfying the condition (i) and (ii) of the Proposition 4.19 above. The linear functional L induces a positive linear functional $\widetilde{L}$ on $C(X)$. By the Riesz representation theorem there exists a regular Borel measure μ on the Borel σ-algebra $\mathcal{B}$ of X such that for any $\varphi \in \mathcal{A}$

$$L(\varphi) = \widetilde{L}(\widetilde{\varphi}) = \int_X \widetilde{\varphi} d\mu,$$

where $\widetilde{\varphi}$ denotes the image of φ in $C(X)$. Now, since the Gelfand transform establishing the isomorphism between $\mathcal{A}$ and $C(X)$ preserves the algebraic operations and since the characteristic functions of sets are the only idempotents in $C(X)$, it follows that the function $\widetilde{f}$ in $C(X)$ which corresponds to $f(h) = 1_E(h)$ is the characteristic function of a set $A \subset X : \widetilde{f}(x) = 1_A(x)$.

This gives

$$\bar{d}(E) = L(1_E) = \widetilde{L}(1_A) = \int\limits_X 1_A d\mu = \mu(A)$$

Finally, notice that the shift operators $\varphi(h) \to \varphi(gh)$, $\varphi \in \mathcal{A}$, $g \in G$, form an anti-action of G on $\mathcal{A}$, which induces an anti-action $\{T_g\}_{g\in G}$ on $C(X)$ defined for $\varphi \in \mathcal{A}$ by $(T_g)\widetilde{\varphi} = \widetilde{{}_g\varphi}$, where ${}_g\varphi(h) = \varphi(gh)$.

Now, the transformations $T_g, g \in G$ are C^*-isomorphisms of $C(X)$ (since they are induced by C^*-isomorphisms $\varphi \to {}_g\varphi$ of $\mathcal{A}$). It is known that algebra isomorphisms of $C(X)$ are induced by homeomorphisms of X which we will, by slight abuse of notation, also denote by $T_g, g \in G$. The homeomorphisms $T_g\colon X \to X$ form an action of G and preserve the measure μ. Indeed, let $C \in \mathcal{B}$ and let $\varphi \in \mathcal{A}$ be the preimage of 1_C (so that $\widetilde{\varphi} = 1_C$). Then we have:

$$\begin{aligned}\mu(C) &= \int\limits_X 1_C(x)d\mu(x) = \widetilde{L}(\widetilde{\varphi}) = L(\varphi) = L({}_g\varphi) = \widetilde{L}(\widetilde{{}_g\varphi}) \\ &= \widetilde{L}(\widetilde{\varphi}(T_g x)) = \int\limits_X 1_C(T_g x)d\mu(x) = \int\limits_X 1_{T_g^{-1}C}(x)d\mu(x) = \mu(T_g^{-1}C)\end{aligned}$$

Notice also that since $L(1) = 1, \mu(X) = L(1_X) = 1$. It follows that $\{T_g\}_{g\in G}$ is a measure preserving action on the probability space $(X, \mathcal{B}, \mu)$. Taking into account that the functional L satisfies the conditions of the Proposition 4.19, we have for $f = 1_E, g_0 = e$ and any $g_1, ..., g_k \in G$:

$$\begin{aligned}\bar{d}(\bigcap_{i=0}^{k} g_i^{-1}E) \geq L(\prod_{i=0}^{k} {}_{g_i}f) = \widetilde{L}(\prod_{i=0}^{k} \widetilde{{}_{g_i}f}) = \widetilde{L}(\prod_{i=0}^{k}((T_{g_i})\tilde{f})) \\ = \int\limits_X \prod_{i=0}^{k} 1_{T_{g_i}^{-1}A} d\mu = \mu(\bigcap_{i=0}^{k} T_{g_i}^{-1}A).\end{aligned}$$

We are done. ■

Remark 4.21 Let us describe an alternative way of proving Theorem 4.17 which avoids the use of Gelfand transform (cf. [14]). Given a Følner sequence Φ_n, $n \in \mathbf{N}$ and $E \subseteq G$ having positive upper density $\bar{d}(E)$ with respect to $\{\Phi_n\}$, identity E with $1_E = \xi \in \Omega = \{0,1\}^G$ and take the orbital closure $X = \overline{\{T_g\xi, g \in G\}}$, where the shift transformations $T_g : \Omega \to \Omega$, $g \in G$, are defined by $T_g f(h) = f(gh)$. By utilizing a procedure similar to that employed in the proof of Proposition 4.19, one constructs a functional L on $C(X)$ which, in addition to $L(1_A) = \bar{d}(E)$ (where $A = \{\varphi \in X : \varphi(e) = 1\}$; notice that $1_A \in C(X)$), satisfies the conditions (i) $L(F) \geq 0$ for $F \geq 0$, (ii) $L(1) = 1$, (iii) $L(F \circ T_g) = L(F)$ $\forall g \in G$. It is admittedly somewhat confusing to deal with functions $F \in C(X)$ which are defined on the space X which itself consists of functions, mapping G into $\{0,1\}$, but one has to get used to that! Now, by Riesz representation theorem such a (positive, normalized) linear functional is given by a probability measure μ on Borel sets of X. The condition $L(1_A) = d(E)$ implies $\mu(A) = \bar{d}(E)$ and condition (iii) implies that μ is T_g invariant for every $g \in G$.

Unfortunately, the scope of these notes does not allow us to present full proofs of Theorems 4.3 and 4.8. The reader is referred to Furstenberg's original paper [12] which contains the proof of Theorem 4.3 and to [16] where both theorems are derived from a very general ergodic *IP-Szemerédi theorem.* We shall, however, be able to give a proof of the following result which contains among its corollaries some nontrivial cases of Theorems 4.3 and 4.8.

Theorem 4.22 *Let* $(G,+)$ *be a countable abelian group and* $(X,\mathcal{B},\mu,\{T_g\}_{g\in G})$ *a probability measure preserving system. For any Følner sequence* $\{\Phi_n\}_{n=1}^{\infty}$ *in* G *and any* $A \in \mathcal{B}$ *with* $\mu(A) > 0$ *one has*

$$\lim_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} \mu(A \cap T_{-g}A \cap T_g A) > 0.$$

Corollary 4.23 *Under the conditions and notation of Theorem 4.22, the set*

$$R_A = \{g \in G : \mu(A \cap T_{-g}A \cap T_g A) > 0\}$$

is syndetic. In other words, for some finite set $F \subset G$ *one has:*

$$F + R_A = \{x + y : x \in F, y \in R_A\} = G.$$

Corollary 4.24 *Let* $E \subseteq \mathbf{Z}$ *be* d^**-large, and let* $A \subseteq \mathbf{Z}^2$ *be defined by*

$$A = \{(a,d) : \{a, a+d, a+2d\} \subset E\}.$$

A is a large subset of $\mathbf{Z}^2$ *in the sense that for some sequence of rectangles* $M_n = [a_n^{(1)}, b_n^{(1)}] \times [a_n^{(2)}, b_n^{(2)}]$, $n \in \mathbf{N}$ *with* $|b_n^{(i)} - a_n^{(i)}| \to \infty$, $i = 1,2$ *one has* $\lim_{n\to\infty} \frac{|A\cap M_n|}{|M_n|} > 0$.

Corollary 4.25 (cf. [1]). *Let $E \subseteq \mathbf{Z}_3^\infty$ with $\bar{d}(E) > 0$. Then E is AS-rich, i.e. contains affine subspaces of arbitrarily high dimension.*

Exercise 4.26 Derive Corollaries 4.23, 4.24 and 4.25 from Theorem 4.22.

We shall preface the proof of Theorem 4.22 with some remarks regarding the machinery that one needs for the proof.

First of all, we are going to take for granted the theorem about *ergodic decomposition.* Recall that a probability measure preserving system $(X, \mathcal{B}, \mu, \{T_g\}_{g\in G})$ is called *ergodic*, if there are no nontrivial invariant sets: $\mu(T_g A \triangle A) = 0$ for all $g \in G$ implies $\mu(A) = 0$ or $\mu(A) = 1$.

Under some mild regularity assumptions (which are satisfied for the spaces we are working with like those featured in the proof(s) of Theorem 4.17) one can assume that if our measure preserving system $(X, \mathcal{B}, \mu, \{T_g\}_{g\in G})$ is not ergodic, then there exists a family of invariant probability measures $\{\mu_\omega\}_{\omega\in\Omega}$ indexed by another probability space $(\Omega, \mathcal{D}, \nu)$ such that with respect to each μ_ω the measure preserving system $(X, \mathcal{B}, \mu_\omega, \{T_g\}_{g\in G})$ is ergodic and such that for any $f \in L^1(X, \mathcal{B}, \mu)$ one has:

$$\int_X f d\mu = \int_\Omega (\int_X f d\mu_\omega) d\nu(\omega)$$

To illustrate the usefulness of the theorem about the ergodic decomposition, let us show that Theorem 4.22 follows from its special, ergodic case.

Fix $A \in \mathcal{B}$ with $\mu(A) > 0$ and a Følner sequence $\Phi_n, n \in \mathbf{N}$. Let $\{\mu_\omega\}_{\omega\in\Omega}$ be the ergodic decomposition of the measure μ and assume that Theorem 4.22 is valid for each $\mu_\omega, \omega \in \Omega$. Since

$$\mu(A) = \int_X 1_A d\mu = \int_\Omega (\int_X 1_A d\mu_\omega) d\nu(\omega) = \int_\Omega \mu_\omega(A) d\nu(\omega)$$

there exist $\delta > 0$ and a measurable set $C \subset \Omega,\ \nu(C) > 0$, such that for any $\omega \in \mathbf{C}$ one has $\mu_\omega(A) > \delta$.

We have:

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}\mu(A\cap T_{-g}A\cap T_gA)\\
&\quad=\lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}\int_\Omega \mu_\omega(A\cap T_{-g}A\cap T_gA)d\nu(\omega)\\
&\quad=\int_\Omega \lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}\mu_\omega(A\cap T_{-g}A\cap T_gA)d\nu(\omega)\\
&\quad\geq\int_C \lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}\mu_\omega(A\cap T_{-g}A\cap T_gA)d\nu(\omega)>0.
\end{aligned}$$

Another useful tool that we shall employ in the proof is a version of *van der Corput trick*:

Proposition 4.27 ([9]). *Let H be a Hilbert space and let $\{v_g\}$ be a bounded family of elements of H indexed by a countable abelian group G. Let $\{\Phi_n\}_{n\in\mathbf{N}}$ be a Følner sequence in G. If*

$$\lim_{m\to\infty}\frac{1}{|\Phi_m|^2}\sum_{h,k\in\Phi_m}\left|\lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}\langle v_{g+h},v_{g+k}\rangle\right|=0,$$

then

$$\lim_{n\to\infty}\left\|\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}v_g\right\|=0.$$

Finally, we shall need the following splitting theorem (see, for example, [25]).

Proposition 4.28 *Let H be a Hilbert space and $U_g:H\to H$, $g\in G$, a unitary action of a countable abelian group G. Then $H=H_{\mathrm{c}}\bigoplus H_{\mathrm{wm}}$, where the orthogonal, $\{U_g\}$-invariant spaces H_{c} and H_{wm} are characterized in the following way:*

$$\begin{aligned}
H_{\mathrm{c}}&=\{v\in H:\text{the orbit }\{U_gv,g\in G\}\text{ is precompact in the norm topology}\}\\
&=\overline{\mathrm{Span}\{v\in H:\text{there exist }\lambda_g\in\mathbf{C}\text{ with }U_gv=\lambda_gv,g\in G\}},
\end{aligned}$$

$$\begin{aligned}
H_{\mathrm{wm}}=\{v\in H:&\text{ for any Følner sequence }\{\Phi_n\}\text{ and any }v'\in H,\\
&\lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}|\langle U_gv,v'\rangle|=0\}.
\end{aligned}$$

Exercise 4.29 Prove that the following statements are equivalent:
(i) $v \in H_{\text{wm}}$,
(ii) for any $v' \in H$ one has $\lim_{n\to\infty} \frac{1}{|\Phi_n|^2} \sum_{h,k\in\Phi_n} |\langle U_h v, U_k v'\rangle|^2 = 0$.

Exercise 4.30 Prove that a non-zero $v \in H_{\text{c}}$ if and only if for any $\varepsilon > 0$ the set $\{g \in G : \|U_g v - v\| < \varepsilon\|\}$ is syndetic in G.

Proof of Theorem 4.22. We shall give the detailed proof for the special case when $2G = \{2g : g \in G\} = G$. The same proof works with only minor changes for the case when $2G$ is a subgroup of finite index in G (this condition is clearly satisfied by the groups $\mathbf{Z}$ and $\mathbf{Z}_3^\infty$, dealt with in Corollaries 4.24 and 4.25). We leave the treatment of the case when $2G$ has infinite index in G to the reader.

In light of the remarks above, we may and will assume that the action $\{T_{2g}\}_{g\in G}$ is ergodic (remembering that we work under the assumption $2G = G$). Let $H = L^2(X, \mathcal{B}, \mu)$ and $f = 1_A$. We shall show first that the limit

$$\lim_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} f(T_{-g}x) f(T_g x)$$

exists in the norm of $H = L^2(X, \mathcal{B}, \mu)$. It will follow then that the limit

$$\lim_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} \mu(A \cap T_{-g}A \cap T_g A) = \lim_{n\to\infty} \frac{1}{|\Phi_n|} \int_X f(x) f(T_{-g}x) f(T_g x) d\mu$$

also exists and we will be left only with showing that it is positive.

Denote by U_g, $g \in G$ the unitary action on H induced by T_g, $g \in G$: $(U_g\varphi)(x) := \varphi(T_g x)$, $\varphi \in H$, and utilize the splitting $H = H_{\text{c}} \bigoplus H_{\text{wm}}$ described in the Proposition 4.28 above.

Let $f = f_{\text{c}} + f_{\text{wm}}$ be the decomposition of $f = 1_A$ corresponding to this splitting. We shall show first that f_{c}, and hence f_{wm}, are bounded functions. First of all, notice that since f is a real-valued function, and since the operators U_g, $g \in G$ send real-valued functions into real-valued functions, the components f_{c} and f_{wm} are also real-valued. We claim that $f_{\text{c}} \geq 0$. Indeed, f_{c} minimizes the distance from H_{c} to f and the function $\bar{f} = \max\{f_{\text{c}}, 0\}$ (which also has precompact orbit and hence belongs to H_{c})

would do as well in minimizing this distance. Similarly one shows that $f_{\rm c} \le 1$. Consider now the average

$$\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}f\ U_gf = \frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}f_{\rm c}U_gf_{\rm c} + \frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}f_{\rm c}U_gf_{\rm wm} + \frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}f_{\rm wm}U_gf_{\rm c} + \frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}f_{\rm wm}U_gf_{\rm wm}$$

Note that the limits of the first three expressions (namely, those involving $f_{\rm c}$) exist in view of Theorem 4.15. Consider, for example, the expression

$$\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}f_{\rm c}U_gf_{\rm wm}.$$

Since $f_{\rm c}$ is a linear combination (potentially infinite) of eigenfunctions, namely, functions φ, satisfying, for some $\lambda_g \in \mathbf{C}$, $g \in G$, the equation $U_g\varphi = \lambda_g\varphi$, $g \in G$, it is enough to prove that for each such φ the limit

$$\lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}\varphi U_gf_{\rm wm}$$

exists in norm. But

$$\lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}\varphi U_gf_{\rm wm} = \varphi\cdot\lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} \lambda_{-g}U_gf_{\rm wm} = \varphi\cdot\lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} \widetilde{U}_gf_{\rm wm} = \varphi\cdot Pf_{\rm wm}$$

where P is the orthogonal projection on the space of invariant functions of the action $\widetilde{U}_g = \lambda_{-g}U_g$, $g \in G$.

It is easy to see that since $\widetilde{U}_g\psi = \psi$ implies $U_g\psi = \lambda_g\psi$, the projection $Pf_{\rm wm}$ belongs to $H_{\rm c}$. But $f_{\rm wm} \perp H_{\rm c}$ and we get $Pf_{\rm wm} = 0$. In complete analogy one shows that

$$\lim_{n\to\infty}\left\|\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}f_{\rm wm}U_gf_{\rm c}\right\| = 0.$$

We shall show now that

$$\lim_{n\to\infty}\left\|\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n} U_{-g}f_{\rm wm}U_gf_{\rm wm}\right\| = 0.$$

This will follow from the van der Corput trick (Proposition 4.27) and from Exercise 4.29.

Let $v_g = U_{-g} f_{\mathrm{wm}} U_g f_{\mathrm{wm}}$. We have, using the ergodicity of $\{U_{2g}\}$:

$$\begin{aligned}
\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}\langle v_{g+h}, v_{g+k}\rangle &= \frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}\int U_{-g-h}f_{\mathrm{wm}}U_{g+h}f_{\mathrm{wm}}U_{-g-k}f_{\mathrm{wm}}U_{g+k}f_{\mathrm{wm}}d\mu \\
&= \frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}\int U_{-g}(U_{-h}f_{\mathrm{wm}}U_{-k}f_{\mathrm{wm}})U_g(U_h f_{\mathrm{wm}}U_k f_{\mathrm{wm}})d\mu \\
&= \frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}\int (U_{-h}f_{\mathrm{wm}}U_{-k}f_{\mathrm{wm}})U_{2g}(U_h f_{\mathrm{wm}}U_k f_{\mathrm{wm}})d\mu \\
&\underset{n\to\infty}{\longrightarrow} \int U_{-h}f_{\mathrm{wm}}U_{-k}f_{\mathrm{wm}}d\mu \int U_h f_{\mathrm{wm}}U_k f_{\mathrm{wm}}d\mu \\
&= \Big(\int U_h f_{\mathrm{wm}}U_k f_{\mathrm{wm}}d\mu\Big)^2 = |\langle U_h f_{\mathrm{wm}}, U_k f_{\mathrm{wm}}\rangle|^2.
\end{aligned}$$

Now use the characterization of H_{wm} (and Exercise 4.29):

$$\lim_{n\to\infty}\frac{1}{|\Phi_n|^2}\sum_{h,k\in\Phi_n}|\langle U_h f_{\mathrm{wm}}, U_k f_{\mathrm{wm}}\rangle|^2 = 0.$$

So, it follows that

$$\lim_{n\to\infty}\Big\|\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}U_{-g}f_{\mathrm{wm}}U_g f_{\mathrm{wm}}\Big\| = 0,$$

and that

$$\lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}U_{-g}fU_g f = \lim_{n\to\infty}\frac{1}{|\Phi_n|}\sum_{g\in\Phi_n}U_{-g}f_{\mathrm{c}}U_g f_{\mathrm{c}}.$$

Noticing that the products of bounded functions from H_{c} belong to H_{c} and that $f_{\mathrm{wm}} \perp H_{\mathrm{c}}$, we have:

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} \mu(A \cap T_{-g}A \cap T_g A) \\
&\quad = \lim_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} \int_X f(x) f(T_g x) f(T_{-g} x) d\mu \\
&\quad = \int_X f \cdot \Big(\lim_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} U_g f U_{-g} f \Big) d\mu \\
&\quad = \int_X (f_{\mathrm{c}} + f_{\mathrm{wm}}) \Big(\lim_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} U_g f_{\mathrm{c}} U_{-g} f_{\mathrm{c}} \Big) d\mu \\
&\qquad\qquad = \lim_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{g\in\Phi_n} \int_X f_{\mathrm{c}} U_g f_{\mathrm{c}} U_{-g} f_{\mathrm{c}} d\mu.
\end{aligned}$$

Now, the positivity of the last expression follows from Exercise 4.30. ■

5 Polynomial Ergodic Theorems and Ramsey Theory

This section's intention is to give a glimpse of a relatively new subject: ergodic theorems along polynomials. The natural limitations allow us only to discuss a few results and applications, but the interested reader is referred, for the details, to the recent survey [3], as well as to the papers [2], [4], [5], [6], [7], [8], [11], [19], [28], [29].

The following theorem, due to Furstenberg (see [12], [13], [14]), shows that measure preserving systems exhibit regular behavior along polynomials.

Theorem 5.1 *If $p(t) \in \mathbf{Q}[t]$, $p(\mathbf{Z}) \subseteq \mathbf{Z}$, and $p(0) = 0$, then for any invertible measure preserving system $(X, \mathcal{B}, \mu, T)$ and $A \in \mathcal{B}$, with $\mu(A) > 0$, one has:*

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

Applying Furstenberg's correspondence principle, one gets the following corollary, independently obtained by Sárközy using methods of analytic number theory.

Theorem 5.2 *([12], [13], [14], [30]). If $p(t) \in \mathbf{Q}[t]$, $p(\mathbf{Z}) \in \mathbf{Z}$, $p(0) = 0$, and $E \subseteq \mathbf{Z}$ has positive upper Banach density, then for some $x, y \in E$ and $n \in \mathbf{N}$ one has $x - y = p(n)$.*

Both Theorems 5.1 and 5.2 are quite striking if one takes into account that the set of values of $p(n)$ is a "small" subset of $\mathbf{Z}$ for any polynomial $p(n)$ whose degree is larger than 1. One can view Theorem 5.1 as a polynomial refinement of the classical Poincaré recurrence theorem.

The next natural question is whether Theorem 4.3 has a polynomial generalization. The answer is *yes* (see [5]). We shall formulate here a special case of the main result from [5].

Theorem 5.3 *Assume that $(X, \mathcal{B}, \mu, T)$ is an invertible probability measure-preserving system, $k \in \mathbf{N}$, $A \in \mathcal{B}$ with $\mu(A) > 0$, and $p_i(t) \in \mathbf{Q}[t]$ are polynomials satisfying $p_i(\mathbf{Z}) \subseteq \mathbf{Z}$ and $p_i(0) = 0$, $1 \le i \le k$. Then*

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

Remark 5.4 Actually, one can show ([7]) that:

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

This result is itself a special case of two different and far-reaching extensions. See [8] and [29].

We now return to Theorem 5.1. The original proof of Furstenberg was based on the spectral theorem and Weyl's theorem on the equidistribution of polynomials (see Section 2). Later, a few more proofs appeared based on different ideas. See [2], [4] and [3].

The following sequence of exercises, supplied with hints, leads to a proof of Theorem 5.1 in stages.

Exercise 5.5 (i) Let H be a Hilbert space and $U: H \to H$ a unitary operator. Let

$$H_{\text{rat}} = \overline{\text{Span}\{f \in H : \text{there exists } i \in \mathbf{N} \text{ with } U^i f = f\}},$$
$$H_{\text{tot.erg}} = \left\{ f \in H : \text{for all } i \in \mathbf{N}, \left\| \tfrac{1}{N} \textstyle\sum_{n=0}^{N-1} U^{in} f \right\| \to 0 \right\}.$$

Show that $H = H_{\text{rat}} \oplus H_{\text{tot.erg}}$.
(ii) Prove the following version of the van der Corput trick: If $\{v_n\}_{n\in\mathbf{N}}$ is a bounded sequence in H such that for any $h \in \mathbf{N}$ $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\langle v_{n+h}, v_n\rangle = 0$, then

$$\|\frac{1}{N}\sum_{n=1}^{N} v_n\| \to 0.$$

Hint: for any $\varepsilon > 0$ and $M \in \mathbf{N}$, if N is large enough, then one has

$$\left\|\frac{1}{N}\sum_{n=1}^{N} v_n - \frac{1}{NM}\sum_{n=1}^{N}\sum_{h=1}^{M} v_{n+h}\right\| < \varepsilon.$$

(iii) (Aside.) Derive from (ii) the van der Corput difference theorem (Theorem 2.18).

Hint: take $H = \mathbf{C}$ and use the Weyl criterion (2.7).

(iv) Prove that for any polynomial $p(t) \in \mathbf{Q}[t]$ with $p(\mathbf{Z}) \subseteq \mathbf{Z}$ and any $f \in H$, the limit $\lim_{N\to\infty} \|\frac{1}{N}\sum_{n=1}^{N} U^{p(n)} f\|$ exists.
Hint: use (ii) to show that if $\deg p(n) > 0$, and $f \in H_{\text{tot.erg}}$, then the limit in question is 0. If $f \in H_{\text{rat}}$, verify that it is enough to check the existence of the limit for f satisfying $U^i f = f$ for some i.
(v) Let $H = L^2(X, \mathcal{B}, \mu)$, $(U\varphi)(x) := \varphi(Tx)$, $\varphi \in H$. Let $f = 1_A$ (where $A \in \mathcal{B}$ with $\mu(A) > 0$). Let f_a, $a \in \mathbf{N}$, be the orthogonal projection of f onto the subspace

$$H_a = \{g \in H : U^a g = g\} \subseteq H_{\text{rat}}$$

Notice that each H_a contains the constants and show that $f_a \geq 0$, $f_a \not\equiv 0$.
(vi) Assume that $p(0) = 0$ in addition to the assumptions of (iv). Conclude the proof by showing that for any $a \in \mathbf{N}$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\langle U^{p(n)} f_a, f_a\rangle > 0,$$

and that it implies

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\mu(A \cap T^{p(n)}A) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\langle U^{p(n)} f, f\rangle > 0$$

(vii) Prove that Theorem 5.1 remains true if one replaces the condition $p(0) = 0$ by the following, weaker one: $\{p(n) : n \in \mathbf{Z}\} \cap a\mathbf{Z} \neq \emptyset$ for all $a \in \mathbf{N}$.

We conclude this section with an application of the polynomial theorem to partition Ramsey theory (cf. [3], p. 53 and [7], Theorem 0.4).

Theorem 5.6 *Let $p(t) \in \mathbf{Q}[t]$, $p(\mathbf{Z}) \subseteq \mathbf{Z}$, $p(0) = 0$. For any finite partition of $\mathbf{N}$, $\mathbf{N} = \bigcup_{i=1}^{r} C_i$, one can find i, $1 \leq i \leq r$, and $x, y, z \in C_i$ so that $x - y = p(z)$.*

Remark 5.7 By Theorem 5.2 any set of positive upper Banach density in $\mathbf{N}$ contains x, y with $x - y = p(n)$ for *some* n. The crux of Theorem 5.6 is that for any finite partition of $\mathbf{N}$ one cell has the additional property that in the equation $x - y = p(z)$, all three parameters are from the same set.

Proof of Theorem 5.6. Let $\mathbf{N} = \bigcup_{i=1}^{r} C_i$ be a given partition. Reindexing if necessary, we may assume that the first k sets $C_1, C_2, \ldots, C_k$ are such that $\bar{d}(C_i) > 0$, $i = 1, 2, \ldots, k$, and $\bar{d}(\bigcup_{i=1}^{k} C_i) = 1$. Let $(X_i, \mathcal{B}_i, \mu_i, T_i)$ and $A_i \in \mathcal{B}_i$ with $\mu(A_i) = \bar{d}(C_i)$, $i = 1, 2, \ldots, k$, be the measure preserving systems and sets guaranteed by the Furstenberg correspondence principle (Theorem 4.4). Form the product system $(X, \mathcal{B}, \mu, T)$ where $X = X_1 \times \ldots \times X_k$, $\mathcal{B}$ is the σ-algebra generated by $\mathcal{B}_1 \times \ldots \times \mathcal{B}_k$, $T = T_1 \times \ldots \times T_k$, and $\mu = \mu_1 \otimes \ldots \otimes \mu_k$. Finally, let $A = A_1 \times \ldots \times A_k \in \mathcal{B}$. Applying Theorem 5.1 to the system $(X, \mathcal{B}, \mu, T)$ and the set A, we obtain that the set $R = \{n \in \mathbf{N} : \mu(A \cap T^{p(n)} A) > 0\}$ has positive lower density (i.e. $\liminf_{n \to \infty} \frac{|R \cap [-n,n]|}{2n+1} > 0$). It follows that $R \cap (\bigcup_{i=1}^{k} C_i) \neq \emptyset$. Let i_0, $1 \leq i_0 \leq k$, be such that for some $n \in C_{i_0}$

$$\begin{aligned}\mu(A \cap T^{p(n)} A) &= \mu\Big((A_1 \times \ldots \times A_k) \cap (T_1^{p(n)} \times \ldots \times T_k^{p(n)})(A_1 \times \ldots \times A_k)\Big) \\ &= \mu\Big((A_1 \cap T_1^{p(n)} A_1) \times \ldots \times (A_k \cap T_k^{p(n)} A_k)\Big) > 0.\end{aligned}$$

Applying again Furstenberg's correspondence principle, we get

$$\bar{d}(C_{i_0} \cap (C_{i_0} - p(n))) > 0.$$

If $y \in C_{i_0} \cap (C_{i_0} - p(n))$, then $x = y + p(n) \in C_{i_0}$, and this establishes the partition regularity of the equation $x - y = p(z)$. ∎

6 Appendix

In this appendix, we shall give an elementary proof of Theorem 2.12 which makes the claim to be, in Erdös' terminology, from THE BOOK, or at least from THE BOOK OF ELEMENTARY PROOFS. Actually, since the readers of these notes surely have no problem with the method of mathematical induction, we shall confine ourself to the following special case, whose proof has all the ingredients of the proof of the general case.

Theorem 6.1 *For any irrational* α*, the sequence* $\{n^2\alpha \pmod 1\}_{n\in\mathbf{N}}$ *is dense in* $[0,1]$.

In light of the (elementary!) discussion in Section 2, Theorem 6.1 clearly follows from the following.

Theorem 6.2 *For any* $\alpha \in \mathbf{R}$ *there is a sequence* $\{n_k\} \subseteq \mathbf{N}$ *such that* $n_k^2\alpha \pmod 1 \underset{k\to\infty}{\longrightarrow} 0$.

Before giving the promised elementary proof of Theorem 6.2, we shall reveal the source of our inspiration. It is $\beta\mathbf{N}$, the Stone-Čech compactification of $\mathbf{N}$, and, specifically, those elements of $\beta\mathbf{N}$ which are called *idempotents*. Section 4 of [3] contains all information about $\beta\mathbf{N}$ and idempotent ultrafilters that we are going to need. (Readers who for this or that reason do not like the ultrafilters can skip the following ultrafilter proof of Theorem 6.2 and go straight to the elementary proof below.) Let $p \in \beta\mathbf{N}$ be any idempotent ultrafilter. The only property of p that we are interested in is given by the following easy proposition.

Theorem 6.3 (Theorem 3.8 in [3]). *Let* X *be a compact Hausdorff space and let* $\{x_n\}_{n\in\mathbf{N}}$ *be a sequence in* X*. Let* $p \in \beta\mathbf{N}$ *be an idempotent ultrafilter. Then*

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

Ultrafilter Proof of Theorem 6.2. Let $X = [0,1]$ with the conventional metric. (Another possibility would be to work with $X = \mathbf{T}$.) Fix an idempotent $p \in \beta\mathbf{N}$. All we have to do is to show that either $p\text{-}\lim_{n\in\mathbf{N}}(n^2\alpha \pmod 1) = 0$ or $p\text{-}\lim_{n\in\mathbf{N}}(n^2\alpha \pmod 1) = 1$. (The reader is invited to check that the latter case is possible and is as good for our purposes.) Let us show first that for any $\gamma \in \mathbf{R}$ one has either $\underset{n\in\mathbf{N}}{p\text{-lim}}(n\gamma \pmod 1) = 0$ or $\underset{n\in\mathbf{N}}{p\text{-lim}}(n\gamma \pmod 1) = 1$. Indeed, if $\underset{n\in\mathbf{N}}{p\text{-lim}}\, n\gamma \pmod 1 = c$, then we have

$$\begin{aligned} c = p\text{-lim}_{n\in\mathbf{N}}(n\gamma \pmod 1) &= p\text{-lim}_{n\in\mathbf{N}}\, p\text{-lim}_{k\in\mathbf{N}}((n+k)\gamma \pmod 1) \\ &= p\text{-lim}_{n\in\mathbf{N}}((n\gamma + c) \pmod 1) = 2c \pmod 1. \end{aligned}$$

So, $c = 2c \pmod 1$ and hence $c \in \{0,1\}$. But the same proof works for any polynomial! Indeed, let $\underset{n\in\mathbf{N}}{p\text{-lim}}\, n^2\alpha \pmod 1 = c$. We have:

$$\begin{aligned} c = p\text{-lim}_{n\in\mathbf{N}}(n^2\alpha \pmod 1) &= p\text{-lim}_{n\in\mathbf{N}}\, p\text{-lim}_{k\in\mathbf{N}}((n+k)^2\alpha \pmod 1) \\ &= p\text{-lim}_{n\in\mathbf{N}}\, p\text{-lim}_{k\in\mathbf{N}}((n^2\alpha + k(2n\alpha) + k^2\alpha) \pmod 1) \\ &= p\text{-lim}_{n\in\mathbf{N}}((n^2\alpha + c) \pmod 1) = 2c \pmod 1. \end{aligned}$$

So, again, $c = 2c \pmod 1$ which implies $c \in \{0,1\}$ and we are done.

Now we shall show how this proof may be elementarized. The most important hint which the perspicacious reader may extract from this proof is that if for some sequence $\{n_k\} \subseteq \mathbf{N}$ one has $n_k^2\alpha \pmod 1 \underset{k\to\infty}{\longrightarrow} c$, then, along the finite sums of elements from $\{n_k\}$, one can approach $lc \pmod 1$ for any $l \in \mathbf{N}$. This is all that one needs since the sequence $\{lc \pmod 1\}_{l\in\mathbf{N}}$ has 0 as a limit point.

Elementary Proof of Theorem 6.2. Let $\{n_k\}$ be an increasing sequence of positive integers satisfying $n_k\alpha \pmod 1 \underset{k\to\infty}{\longrightarrow} 0$. Passing, if needed, to a subsequence, assume that simultaneously $n_k^2\alpha \pmod 1 \underset{k\to\infty}{\longrightarrow} c \in [0,1]$. If $c = 0$, we are done. If $c = 1$, we are also done, since it is easy to see that for appropriately chosen m_k one will have $(m_k n_k)^2\alpha \pmod 1 \underset{k\to\infty}{\longrightarrow} 0$. So assume that $c \in (0,1)$. Again, if $c = r/s$ is a rational number, then replacing n_k by sn_k we are done, so assume without loss of generality that c is irrational, and let us show how, for any $l \in \mathbf{N}$ and any $\varepsilon > 0$, to find $m \in \mathbf{N}$ with $lc \pmod 1 - \varepsilon < m^2\alpha \pmod 1 < lc \pmod 1 + \varepsilon$. Assuming, as we may, that ε is so small that $\varepsilon < lc \pmod 1 < 1 - \varepsilon$, let us show that such an m can be found among the numbers of the form $n_{k_1} + n_{k_2} + \ldots + n_{k_l}$ with $k_1 \le k_2 \le \ldots \le k_l$. Let us do it, for simplicity, for $l = 3$. Let $k_1 \in \mathbf{N}$ be such that for any $k \ge k_1$ one has $c - \frac{\varepsilon}{6} < n_k^2\alpha \pmod 1 < c + \frac{\varepsilon}{6}$. Choose now $k_3 \geq k_2 \geq k_1$ so that, in addition, $0 < n_{k_2}\alpha \pmod 1 < \frac{\varepsilon}{12n_{k_1}}$ and $0 < n_{k_3}\alpha \pmod 1 < \frac{\varepsilon}{12n_{k_2}}$. One trivially checks that such a choice of k_1, k_2, and k_3 guarantees that

$$3c \pmod 1 - \frac{\varepsilon}{2} < (n_{k_1} + n_{k_2} + n_{k_3})^2\alpha \pmod 1 < 3c \pmod 1 + \varepsilon. \quad \blacksquare$$

Acknowledgment. I wish to express my sincere appreciation to Mrs. Gladys Cavallone, Steve Cook, Paul Larick and Sasha Leibman for their help in preparing these notes. I also would like to thank Francois Blanchard, Alejandro Maass and Arnaldo Nogueira for organizing, under the aegis of CIMPA and UNESCO, the wonderful workshop in Temuco.

This work was partially supported by NSF under grant DMS-970605.

References

[1] V. Bergelson, Ergodic Ramsey theory, logic and combinatorics, edited by S. Simpson, *Contemporary Mathematics* **65**, pp. 63-87 (1987).

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

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

[4] V. Bergelson, H. Furstenberg and R. McCutcheon, IP-sets and polynomial recurrence, *Ergodic Theory and Dynamical Systems* **16**, pp. 963-974 (1996).

[5] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems, *Journal of AMS* **9**, pp. 725-753 (1996).

[6] V. Bergelson and A. Leibman, Set-polynomials and a polynomial extension of Hales-Jewett theorem, *Annals of Math.*, to appear.

[7] V. Bergelson and R. McCutcheon, Uniformity in polynomial Szemerédi theorem, *Ergodic Theory of* $\mathbf{Z}^d$*-actions*, edited by M. Pollicott and K. Schmidt, London Math. Soc. Lecture Note Series **228**, pp. 273-296 (1996).

[8] V. Bergelson and R. McCutcheon, An IP polynomial Szemerédi theorem for finite families of commuting transformations, *Memoirs of AMS*, to appear.

[9] V. Bergelson, R. McCutcheon and Q. Zhang, A Roth theorem for amenable groups, *American Journal of Mathematics* **119**, pp. 1173-1211 (1997).

[10] A. Blaszczyk, S. Plewik and S. Turek, Topological multidimensional van der Waerden theorem, *Comment. Math. Univ. Carolinae* **30**, pp. 783-787 (1989).

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

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

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

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

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

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

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

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

[19] H. Furstenberg and B. Weiss, A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, *Convergence in Ergodic Theory and Probability*, edited by V. Bergelson, P. March and J. Rosenblatt, Ohio State University Math. Research Institute Publications, de Gruyter (1995).

[20] R. Graham, K.Leeb and B. Rothschild, Ramsey's theorem for a class of categories, *Advances in Math.* **8**, pp. 417-433 (1972). Errata **10**, pp. 326-327 (1973).

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

[22] D. Hilbert, Über die Irreduzibilität Ganzer Rationaler Funktionen mit Ganzzahligen Koeffizienten, *J. Math.* **110**, pp. 104-129 (1892).

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

[24] G. Hardy and E. Wright, *An Introduction to the Theory of Numbers*, fifth edition, Clarendon Press, Oxford (1979).

[25] V. Krengel, *Ergodic Theorems*, de Gruyter (1985).

[26] N. Kryloff and N. Bogoliouboff, La théorie générale de la measure dans son application à l'étude des systémes dynamiques de la mécanique non linéaire, *Ann. of Math.* **38**, pp. 65-113 (1937).

[27] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, John Wiley (1974).

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

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

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

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

[32] P. Walters, *An Introduction to Ergodic Theory*, Springer (1982).

[33] H. Weyl, Über die Gleichverteilung von Zahlen mod1, *Math. Ann.* **77**, pp. 313-352 (1916).