Ergodic Methods in Combinatorics

Stockholm Master Class in Dynamical Systems 2025

Florian K. Richter

May 19–23, 2025

Contents

1. Introduction

Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ denote the set of positive integers.

The notion of density in $\mathbb{N}$ measures the relative proportion of a set with respect to all of $\mathbb{N}$; it formalizes the intuitive concept of comparative size in the positive integers. With the help of this notion, we can turn vague statements such as "almost no integer is the sum of two squares" or "the probability for two integers to be coprime is $\frac{6}{\pi^2}$" into meaningful mathematical theorems.

Definition 1. The *lower density* and *upper density* of a set $E \subseteq \mathbb{N}$ are defined respectively as

$$\underline{d}(E) = \liminf_{N\to\infty} \frac{|E \cap \{1,\ldots,N\}|}{N} \qquad \text{and} \qquad \overline{d}(E) = \limsup_{N\to\infty} \frac{|E \cap \{1,\ldots,N\}|}{N}.$$

Observe that $\underline{d}(E) \leqslant \overline{d}(E)$ always. If $E \subseteq \mathbb{N}$ is such that $\underline{d}(E) = \overline{d}(E)$ then the limit

$$d(E) = \lim_{N\to\infty} \frac{|E \cap \{1,\ldots,N\}|}{N}$$

exists and we call this number the *density* of E.

Example 2. Here are some typical examples of subsets of $\mathbb{N}$ and their associated densities:

- $d(\mathbb{N}) = 1$;
- $d(a\mathbb{N} + b) = a^{-1}$ for all $a \in \mathbb{N}$ and $b \in \mathbb{N} \cup \{0\}$;
- $d(\{\lfloor \alpha n + \beta \rfloor : n \in \mathbb{N}\}) = \min\{\alpha^{-1}, 1\}$ for all $\alpha > 0$ and $\beta \geqslant 0$;
- $d(\{n \in \mathbb{N} : n \text{ is squarefree}\}) = \frac{6}{\pi^2}$;
- $d(\mathbb{P}) = 0$, where $\mathbb{P} = \{2, 3, 5, 7, 11, \ldots\}$ denotes the set of prime numbers.

Remark 3. Some of the basic properties of the lower and upper density on $\mathbb{N}$ include:

- (unit range). $0 \leqslant \underline{d}(E) \leqslant \overline{d}(E) \leqslant 1$.
- (monotonicity). If $D \subseteq E$ then $\underline{d}(D) \leqslant \underline{d}(E)$ and $\overline{d}(D) \leqslant \overline{d}(E)$.
- (sub-additivity of upper density). $\overline{d}(D \cup E) \leqslant \overline{d}(D) + \overline{d}(E)$.
- (super-additivity of lower density). If $D \cap E = \emptyset$ then $\underline{d}(D) + \underline{d}(E) \leqslant \underline{d}(D \cup E)$.
- (complement property). $\overline{d}(E^c) = 1 - \underline{d}(E)$ and $\underline{d}(E^c) = 1 - \overline{d}(E)$.
- (shift-invariance). $\underline{d}(E - m) = \underline{d}(E)$ and $\overline{d}(E - m) = \overline{d}(E)$, where $E - m = \{n \in \mathbb{N} : n + m \in E\}$.

In arithmetic combinatorics, it is common to think of subsets of the integers with positive density as "large" sets, since such sets are not merely infinite but occupy a non-negligible portion of the integers. A central question in the field can be phrased as follows:

> *What kinds of arithmetic structures must appear in subsets of the integers that have positive density?*

This question seeks to understand the extent to which largeness alone forces the presence of specific arithmetic patterns. For example, do such sets necessarily contain arithmetic progressions $\{a, a+d, \ldots, a+(k-1)d\}$, or sumsets $B + C = \{b + c : b \in B,\ c \in C\}$, or solutions to linear equations such as $x + y = w + z$? In exploring these questions, we will uncover a rich body of research on the interplay between density and arithmetic structure, involving numerous deep results and techniques coming from combinatorics, number theory, and ergodic theory.

1.1. Difference patterns in sets of positive density

Given a set $E \subseteq \mathbb{N}$, we call $E - E = \{b - a : a, b \in E,\ a < b\}$ the *difference set* of E, and we refer to any element in $E - E$ as a *difference* appearing in E. An important topic in additive combinatorics and number theory is understanding which type of differences must occur in sets of positive density. It turns out that sets of positive density cannot avoid certain difference patterns. A key result in this direction is the Furstenberg–Sárközy theorem, which highlights the inevitability of polynomial differences.

Theorem 4 (Furstenberg-Sárközy theorem, [Sár78, Fur77]). *Let p be a polynomial satisfying $p(\mathbb{N}) \subseteq \mathbb{N}$ and $p(0) = 0$. If $\overline{d}(E) > 0$, then $E - E \cap \{p(n) : n \in \mathbb{N}\} \neq \emptyset$, i.e., there exist $a, b \in E$ and $n \in \mathbb{N}$ such that $b - a = p(n)$.*

An important special case of Theorem 4 is when $p(n) = n^2$. In this case, the theorem asserts that any subset of the natural numbers with positive upper density must contain two distinct elements whose difference is a perfect square.

1.2. Arithmetic progressions in sets of positive density

Another foundational result in this area is Roth's theorem on arithmetic progressions.

Theorem 5 (Roth's theorem, [Rot53]). *If $\overline{d}(E) > 0$, then E contains a 3-term arithmetic progression $\{a, a + d, a + 2d\}$.*

Roth's theorem resolved the first nontrivial case of a conjecture posed by Erdős and Turán in the 1930s, which asserted that any subset of the integers with positive upper density must contain arbitrarily long arithmetic progressions. While Roth established the case of three-term progressions, the full conjecture was ultimately proved by Endre Szemerédi in 1975, in what is now known as Szemerédi's theorem.

Theorem 6 (Szemerédi's theorem, [Sze75]). *If $\overline{d}(E) > 0$, then E contains a k-term arithmetic progression for all $k \geqslant 2$.*

Theorem 6 has been generalized in many different ways. One of the most striking results in this direction is the following landmark theorem.

Theorem 7 (Green-Tao theorem, [GT08]). *The prime numbers $\mathbb{P}$ contain a k-term arithmetic progression for all $k \geqslant 2$.*

The following is a well-known conjecture in combinatorial number theory and offers a simultaneous generalization of both Szemerédi's theorem and the Green-Tao theorem.

Conjecture 8 (Erdős's conjecture on arithmetic progressions). *If $\sum_{n \in E} \frac{1}{n} = \infty$, then E contains a k-term arithmetic progression for all $k \geqslant 2$.*

The $k = 3$ case of Erdős's conjecture on arithmetic progressions was resolved by Bloom and Sisask in [BS21] (see also [BS23, KM23]); the claim remains open for $k \geqslant 4$.

In spirit, the Furstenberg-Sárközy theorem is closely related to Szemerédi's theorem in that both guarantee the presence of a specific arithmetic structure in sets of positive upper density. A unifying generalization that encompasses both results is the so-called polynomial Szemerédi theorem, which can be stated as follows:

Theorem 9 (Polynomial Szemerédi theorem, [BL96]). *Let $p_1, p_2, \dots, p_k$ be polynomials satisfying $p_i(\mathbb{N}) \subseteq \mathbb{N}$ and $p_i(0) = 0$. If $\overline{d}(E) > 0$, then there exist $a, n \in \mathbb{N}$ such that $\{a + p_1(n), \dots, a + p_k(n)\} \subseteq E$.*

Configurations of the from $\{a + p_1(n), \dots, a + p_k(n)\}$ are often referred to as *polynomial progressions*, since they represent a polynomial extension of an arithmetic progression. Observe that the case $k = 1$ of Theorem 9 corresponds to the Furstenberg-Sárközy theorem, and the case $p_i(n) = (i-1)n$, $i = 1, \dots, k$, corresponds to Szemerédi's theorem.

1.3. Sumsets in sets of positive density

In the late 1970s and early 1980s, inspired by Szemerédi's seminal work on arithmetic progressions in sets of positive density, Paul Erdős asked if sets of positive density contain sumsets.

The notion of a *sumset* refers to two different, but closely related, concepts. First, given two sets of non-negative integers B and C, the sumset $B + C$ is the set of all possible pairwise sums of elements from B and C, that is,

$$B + C = \{b + c : b \in B,\ c \in C\}.$$

Second, given a sequence of integers $b_1 < b_2 < b_3 < \dots$, its (restricted) sumset consists of all possible sums of distinct elements from the sequence, i.e., $\{b_i + b_j : i, j \in \mathbb{N},\ i \neq j\}$.

Erdős sought to characterize the types of sets that necessarily contain infinite sumsets, a line of inquiry that culminated in two influential conjectures. Both conjectures have since been resolved and we state them as theorems below. We begin with the second of the two conjectures, chronologically speaking, since it is the more accessible one.

Theorem 10 (Erdős's 2$^{\text{nd}}$ sumset conjecture, [MRR19]). *If $\overline{d}(E) > 0$, then there exist two infinite sets $B, C \subseteq \mathbb{N}$ such that $B + C \subseteq E$.*

Theorem 11 (Erdős's 1$^{\text{st}}$ sumset conjecture, [KMRR24]). *If $\overline{d}(E) > 0$, then there exist $b_1 < b_2 < b_3 < \dots \in \mathbb{N}$ and $t \in \mathbb{N}$ such that $\{b_i + b_j : i, j \in \mathbb{N},\ i \neq j\} \subseteq E - t$.*

Note that Theorem 11 contains Theorem 10 as a special case.

Exercises

(1) Suppose $E = \{f(n) : n \in \mathbb{N}\}$ where $f : \mathbb{N} \to \mathbb{N}$ is an increasing function. Prove that

$$\underline{d}(E) = \liminf_{n\to\infty} \frac{n}{f(n)} \qquad \text{and} \qquad \overline{d}(E) = \limsup_{n\to\infty} \frac{n}{f(n)}.$$

(2) Show that $\overline{d}$ is not additive: There exist disjoint sets $E, D \subseteq \mathbb{N}$ such that $\overline{d}(D \cup E) < \overline{d}(D) + \overline{d}(E)$.

(3) Show there is no probability measure on the σ-algebra of all subsets of $\mathbb{N}$ such that $\mu(a\mathbb{N}) = a^{-1}$ holds for every $a \in \mathbb{N}$. (*Hint:* Show that

$$\mu(\{k\}) \leqslant \mu\left(\bigcap_{p \in \mathbb{P}_{>k}} (\mathbb{N} \setminus p\mathbb{N})\right) = \prod_{p \in \mathbb{P}_{>k}} \left(1 - \frac{1}{p}\right)$$

and use the fact that $\sum_{p\in\mathbb{P}} \frac{1}{p} = \infty$.)

(4) Show that Erdős's conjecture on arithmetic progressions implies Szemerédi's theorem.

(5) Prove that if $\overline{d}(E) > \delta$ then for any finite set $F \subseteq \mathbb{N}$ there exists $n \in \mathbb{N}$ such that $|(n+F) \cap E| > \delta |F|$.
(6) Let $r \geqslant 3$ and $a_1, \ldots, a_r \in \mathbb{Z} \setminus \{0\}$. Show that the equation $a_1 x_1 + \cdots + a_r x_r = 0$ is density regular[1] if and only if $a_1 + \cdots + a_r = 0$.

[1]An equation $a_1 x_1 + \cdots + a_r x_r = 0$ is said to be *density regular* if for every set $E \subseteq \mathbb{N}$ with positive upper density there exist distinct $x_1, \ldots, x_r \in E$ satisfying the equation.

2. Furstenberg's correspondence principle

2.1. Measure preserving systems

Definition 12 (Measure preserving transformation)**.** Given a probability space $(X,\mathscr{A},\mu)$, we say that a measurable map $T: X \to X$ *preserves the measure* or is a *measure preserving transformation* if for every $A \in \mathscr{A}$ we have $\mu(T^{-1}A) = \mu(A)$.

Recall, the *push-forward* of a measure μ under a transformation T is defined via

$$T\mu(A) = \mu(T^{-1}A), \qquad \forall A \in \mathscr{A}.$$

We say that the measure μ is *invariant* under the map T if $T\mu = \mu$. Observe that T preserves the measure μ (as specified in Definition 12) if and only if μ is invariant under T; the two statements express the same property and we will use them interchangeably throughout.

The basic object of study in ergodic theory is a *measure preserving system*, which we now define.

Definition 13 (Measure preserving system)**.** A *measure preserving system* is a quadruple $(X,\mathscr{A},\mu,T)$ where $(X,\mathscr{A},\mu)$ is a probability space and $T: X \to X$ is a measure preserving transformation.

Here are two important examples of measure preserving systems.

Example 14 (Circle rotations)**.** Let $X = [0,1)$, endowed with the Borel σ-algebra $\mathscr{A} = \mathscr{B}_{[0,1)}$ and the Lebesgue measure $\mu = m_{[0,1)}$. Given $\alpha \in \mathbb{R}$ we consider the map $T = T_\alpha: [0,1) \to [0,1)$ given by $Tx = x + \alpha \bmod 1$. The fact that T preserves the Lebesgue measure follows from the basic properties of Lebesgue measure.

Example 15 (The doubling map)**.** Take $(X,\mathscr{A},\mu) = ([0,1),\mathscr{B}_{[0,1)},m_{[0,1)})$, where $\mathscr{B}_{[0,1)}$ is the Borel σ-algebra on $[0,1)$, and $m_{[0,1)}$ the Lebesgue measure. Let $T\colon [0,1) \to [0,1)$ be the *doubling map* $T(x) = 2x \bmod 1$. Let us show that this transformation preserves the measure: Given an interval $[a,b) \subseteq [0,1)$, the pre-image $T^{-1}([a,b))$ is the union of two intervals, each half the length of the original interval:

$$T^{-1}\big([a,b)\big) = \left[\frac{a}{2},\frac{b}{2}\right) \cup \left[\frac{a+1}{2},\frac{b+1}{2}\right).$$

This shows that the Lebesgue measure of $[a,b)$ and $T^{-1}([a,b))$ are identical. Since T^{-1} preserves the measure of all intervals and since intervals generate the Borel σ-algebra on $[0,1)$, it follows that T is a measure-preserving transformation.

More generally, for any positive integer p the map $T(x) = px \bmod 1$ preserves the Lebesgue measure, giving rise to a class of measure-preserving systems whose dynamical behavior is closely related to the base-p digit expansions of the real numbers.

2.2. Poincaré's recurrence theorem

The theory of recurrence in ergodic theory studies the return of points or sets to their original position (or neighborhood) within measure-preserving dynamical systems. Here is the first theorem of recurrence in ergodic theory.

Theorem 16 (Poincaré's recurrence theorem)**.** *Let $(X,\mathscr{A},\mu,T)$ be a measure preserving system and let $A \in \mathscr{A}$ with $\mu(A) > 0$. Then for some $n \in \mathbb{N}$ we have*

$$\mu(A \cap T^{-n}A) > 0. \tag{2.1}$$

Proof. Since T is measure preserving, for any $n \in \mathbb{N}$ the set $T^{-n}A$ has the same measure as the set A. Since the ambient space X has measure 1 and $A, T^{-1}A, T^{-2}A, \ldots$ is an infinite sequence of sets with the same (positive) measure, by the pigeonhole principle there must exist $i > j$ with $\mu(T^{-i}A \cap T^{-j}A) > 0$. Letting $n = i - j$, we obtain

$$\mu(A \cap T^{-n}A) = \mu\big(T^{-j}(A \cap T^{-n}A)\big) = \mu(T^{-i}A \cap T^{-j}A) > 0,$$

completing the proof. □

Definition 17. Given a measure preserving system $(X, \mathscr{A}, \mu, T)$, the *orbit* of a point $x \in X$ under T is the set $\mathscr{O}_T(x) = \{x, Tx, T^2x, T^3x, \ldots\}$.

Proposition 18. *Let X be a compact metric space (with metric $d_X : X \times X \to [0,\infty)$), $\mathscr{B}_X$ the σ-algebra of Borel sets on X, μ a Borel probability measure on X, and $T : X \to X$ a measure preserving transformation. Then the orbit of μ-a.e. point returns arbitrarily close to its initial position, i.e., μ-a.e. $x \in X$ satisfies $\inf_{n\in\mathbb{N}} d_X(x, T^n x) = 0$.*

Proof. Consider the set $C_\varepsilon = \{x \in X : \inf_{n\in\mathbb{N}} d_X(x, T^n x) \geqslant \varepsilon\}$; we need to show that $\mu(\bigcup_{\varepsilon>0} C_\varepsilon) = 0$. By way of contradiction, assume $\mu(\bigcup_{\varepsilon>0} C_\varepsilon) > 0$. Then, by the monotone convergence theorem, there exists $\varepsilon > 0$ such that $\mu(C_\varepsilon) > 0$. Since X is a compact metric space, we can cover X using finitely many balls of radius $\varepsilon/2$. Hence, there exists a ball of radius $\varepsilon/2$, say B, such that $\mu(B \cap C_\varepsilon) > 0$. Define $A = B \cap C_\varepsilon$. In light of Theorem 16, there exists $m \in \mathbb{N}$ such that $A \cap T^{-m}A$ is non-empty. But if $y \in A \cap T^{-m}A$ then y and $T^m y$ are less than ε apart, because both belong to B, contradicting the assumption that $y \in C_\varepsilon$. □

2.3. Sets of recurrence

Poincaré's recurrence theorem (Theorem 16) revealed that in a measure-preserving dynamical system almost every point returns arbitrarily close to its initial position at some time n. This naturally leads to a deeper question: what can we say about the set of n for which this happens? How large or structured is this set of return times? The following definition sets the stage for investigating this question.

Definition 19. A set $R \subseteq \mathbb{N}$ is a *set of recurrence* if for all measure preserving systems $(X, \mathscr{A}, \mu, T)$ and all sets $A \in \mathscr{A}$ with $\mu(A) > 0$ there exists $n \in R$ with $\mu(A \cap T^{-n}A) > 0$.

Poincaré's recurrence theorem (Theorem 16) asserts that $\mathbb{N}$ is a set of recurrence. Below we collect some other examples and non-examples of sets of recurrence (without proof).

sets of recurrence	**sets of non-recurrence**
$\mathbb{N}$	Any finite set
$a\mathbb{N}$	$a\mathbb{N} + b$ for any $b \not\equiv 0 \bmod a$
$\{n^2 : n \in \mathbb{N}\}$	$\{n^2 + 1 : n \in \mathbb{N}\}$
$\mathbb{P} + t$ for $t \in \{-1, 1\}$	$\mathbb{P} + t$ for any $t \in \mathbb{Z} \setminus \{\pm 1\}$
$D - D$ for any infinite $D \subseteq \mathbb{N}$	$\{2^n : n \in \mathbb{N}\}$

Proposition 20. *Suppose X is a compact metric space (with metric $d_X : X \times X \to [0,\infty)$), $\mathscr{B}_X$ is the σ-algebra of Borel subsets of X, μ is a Borel probability measure on X, and $T : X \to X$ is a measure preserving transformation. If $R \subseteq \mathbb{N}$ is a set of recurrence then μ-a.e. $x \in X$ satisfies $\inf_{n\in R} d_X(x, T^n x) = 0$.*

Proof. The proof is very similar to the proof of Proposition 18. □

2.4. Intersective sets

One of the goals of additive combinatorics is to study the arithmetic and combinatorial properties of difference sets $E-E=\{b-a : a,b\in E,\ b>a\}$. In this section, we investigate this question under the assumption that the set E has positive density. The following definition provides a framework for this exploration.

Definition 21. A set $R\subseteq\mathbb{N}$ is *intersective* if for all sets $E\subseteq\mathbb{N}$ with $\overline{d}(E)>0$, there exists $n\in R$ with $n\in E-E$.

Note that if R is an intersective set then one can remove a finite amount of elements from R and it remains an intersective set. It follows that a set R is intersective if and only if for all sets E of positive upper density the intersection $R\cap(E-E)$ has infinite cardinality.

For any $a\in\mathbb{N}$ the set $a\mathbb{N}=\{an : n\in\mathbb{N}\}$ is intersective because any set E of positive upper density contains two numbers of the same residue class mod a. This and other examples and non-examples of intersective sets are summarized in the following table:

intersective sets	**non-intersective sets**
$\mathbb{N}$	Any finite set
$a\mathbb{N}$	$a\mathbb{N}+b$ for any $b\not\equiv 0 \bmod a$
$\{n^2 : n\in\mathbb{N}\}$	$\{n^2+1 : n\in\mathbb{N}\}$
$\mathbb{P}+t$ for $t\in\{-1,1\}$	$\mathbb{P}+t$ for any $t\in\mathbb{Z}\backslash\{\pm1\}$
$D-D$ for any infinite $D\subseteq\mathbb{N}$	$\{2^n : n\in\mathbb{N}\}$

2.5. The correspondence principle

Furstenberg's correspondence principle tells us that we can model any subset of the integers as an orbit in a dynamical system.

Theorem 22 (Furstenberg's correspondence principle)**.** *For any set $E\subseteq\mathbb{N}$ there exist a compact metric space X, a Borel probability measure μ on X, a continuous, measure preserving map $T\colon X\to X$, a clopen (i.e., closed and open) set $A\subseteq X$, and a point $x\in X$ with dense orbit (i.e., $\overline{\mathcal{O}_T(x)}=X$) such that:*

(i) $\mu(A)=\overline{d}(E)$;

(ii) $E=\{n\in\mathbb{N} : T^n x\in A\}$.

Proof. Consider $\{0,1\}^{\mathbb{Z}}$, the space of all $\{0,1\}$-valued sequences $(z_n)_{n\in\mathbb{Z}}$, which is a compact metric space. On this space, let T denote the left-shift operator $(z_n)_{n\in\mathbb{N}}\mapsto(z_{n+1})_{n\in\mathbb{N}}$, that is, we shift every element in the sequence one position to the left. Define $x=1_E$, where $1_E\in\{0,1\}^{\mathbb{Z}}$ is the indicator function of the set E, and take $X=\overline{\mathcal{O}_T(x)}$. Observe that $T(X)\subseteq X$, which means we can view T as a transformation on X. Our set A will be the set of all sequences $(z_n)_{n\in\mathbb{Z}}\in X$ with $z_0=1$; this is a clopen subset of X. Moreover, with this definition of x, T, and A, we have $E=\{n\in\mathbb{N} : T^n x\in A\}$ as desired.

Finally, we need to construct a measure. We first define the *finite* measures

$$\mu_N=\frac{1}{N}\sum_{n=1}^{N}\delta_{T^n 1_E}.$$

We are placing a point mass on each of the points in the orbit of $x=1_E$ — the points 1_E, 1_{E-1}, 1_{E-2}, and so on — and averaging them. Choose a sequence (N_k) such that

$$\overline{d}(E) = \lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} 1_E(n) = \lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N} 1_{E-n}(0).$$

Such a sequence exists because $\overline{d}(E)$ is defined as the lim sup of this quantity over all N, and there must be a subsequence approaching the lim sup.

When endowed with the weak-$*$ topology, the space of Borel probability measures on a compact metric space is itself a compact metric space. In particular, every sequence of measures has a convergent subsequence. Let μ be any weak-$*$ accumulation point of the sequence of measures μ_{N_k}. (Recall, a sequence of measures ν_n converges to ν in the weak-$*$ topology if and only if for all continuous functions f we have $\int f\,d\nu_n \to \int f\,d\nu$.) We now have our measure μ. The fact that $\mu(A) = \overline{d}(E)$ follows from the above relation and the definition of A (since $\delta_{1_{E-n}}(A)$ is 1 if 1_{E-n} begins with a 1, or equivalently if $E-n$ contains 0, and is 0 otherwise). This completes the proof. □

Theorem 23. *Let $R \subseteq \mathbb{N}$. The following are equivalent:*

(i) R is a set of recurrence.

(ii) R is an intersective set.

(iii) For all sets $E \subseteq \mathbb{N}$ with $\overline{d}(E) > 0$, there exists $m \in R$ such that $\{n \in E : n+m \in E\}$ is infinite.

Note that (iii) simply asserts that m appears infinitely often as a difference in E.

Proof of Theorem 23. Let us first prove that (ii) $\Longrightarrow$ (i). Assume, for contradiction, that R is not a set of recurrence, which means there exists a measure preserving system $(X, \mathscr{A}, \mu, T)$ and a measurable set $A \in \mathscr{A}$ with positive measure such that

$$\mu(A \cap T^{-m}A) = 0, \qquad \forall m \in R. \tag{2.2}$$

Define $A' = A \setminus (\bigcup_{m\in R} T^{-m}A)$ and note that $\mu(A) = \mu(A')$ due to (2.2), and

$$A' \cap T^{-m}A' = \emptyset, \qquad \forall m \in R. \tag{2.3}$$

By Fatou's lemma, we have

$$\int \limsup_{N\in\mathbb{N}} \frac{1}{N}\sum_{n=1}^{N} 1_{A'}(T^n x)\,\mathrm{d}\mu \geqslant \limsup_{N\in\mathbb{N}} \frac{1}{N}\sum_{n=1}^{N} \int 1_{A'}(T^n x)\,\mathrm{d}\mu = \mu(A') = \mu(A) > 0.$$

This means there exists a point $x \in X$ such that

$$\limsup_{N\in\mathbb{N}} \frac{1}{N}\sum_{n=1}^{N} 1_{A'}(T^n x) > 0. \tag{2.4}$$

If we now consider the set $E = \{n \in \mathbb{N} : T^n x \in A'\}$, then (2.4) implies $\overline{d}(E) > 0$. Since R is an intersective set, there exists $m \in R$ with $m \in E - E$. Using the definition of E, we can thus find some $n \in \mathbb{N}$ with $T^n x, T^{n+m}x \in A'$. This implies that $T^n x \in A' \cap T^{-m}A'$, contradicting (2.3).

Since the implication (iii) $\Longrightarrow$ (ii) is immediate, it remains to prove (i) $\Longrightarrow$ (iii). Let $E \subseteq \mathbb{N}$ with $\overline{d}(E) > 0$ be given. By Theorem 22, there exist a compact metric space X, a Borel probability measure μ, a measure preserving map $T\colon X \to X$, a clopen set $A \subseteq X$, and a point $x \in X$ with $\overline{\mathcal{O}_T(x)} = X$ such that $\mu(A) > 0$ and $E = \{n \in \mathbb{N} : T^n x \in A\}$. Since R is a set of recurrence, there exists some $m \in R$ with

$$\mu(A \cap T^{-m}A) > 0.$$

In particular, $A \cap T^{-m}A$ is not the empty set. Since A is clopen and the orbit of x is dense, there exist infinitely many $n \in \mathbb{N}$ such that $T^n x \in A \cap T^{-m}A$. Using $E = \{n \in \mathbb{N} : T^n x \in A\}$, we conclude that there are infinitely many n for which $n, n+m \in E$, as desired. □

Exercises

(2.1) Show that if $R \subseteq \mathbb{N}$ is intersective then for every $\alpha \in \mathbb{R}$ and every $\varepsilon > 0$, there exist $m \in R$ and $n \in \mathbb{Z}$ such that

$$\left|\alpha - \frac{n}{m}\right| \leqslant \frac{\varepsilon}{m}.$$

(*Hint:* Consider the measure preserving system described in Example 14.)

(2.2) Prove that for any infinite set $D \subseteq \mathbb{N}$, the set of differences $D - D$ is an intersective set.

(2.3) Suppose $R \subseteq \mathbb{N}$ is intersective. Show that the following sets are also intersective:

(a) $aR = \{an : n \in R\}$ for any $a \in \mathbb{N}$.

(b) $R/a = \{n \in \mathbb{N} : an \in R\}$ for any $a \in \mathbb{N}$.

(c) $R \backslash F$ for any finite subset $F \subseteq R$.

(2.4) Show that the measure μ constructed in the proof of the Furstenberg's correspondence principle is T-invariant. (This detail was glossed over in the proof, despite being essential to its correctness.)

(2.5) Show that if $R \subseteq \mathbb{N}$ is intersective and is decomposed as $R = R_1 \cup R_2$, then either R_1 or R_2 is intersective.

(2.6) Prove the following strengthening of Poincaré's recurrence theorem: For any measure-preserving system $(X, \mathscr{A}, \mu, T)$, any $A \in \mathscr{A}$, and any $\varepsilon > 0$, there exists some $n \in \mathbb{N}$ such that

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

3. The Furstenberg-Sárközy theorem

The purpose of this section is to prove the Furstenberg-Sárközy theorem (Theorem 4). Instead of proving the theorem directly, we will derive it from the following polynomial recurrence result.

Theorem 24 (Furstenberg's polynomial recurrence theorem, [Fur77]). *Let p be a polynomial satisfying $p(\mathbb{N}) \subseteq \mathbb{N}$ and $p(0) = 0$. For any measure preserving system $(X, \mathscr{A}, \mu, T)$ and any $A \in \mathscr{A}$ with $\mu(A) > 0$ there exist $n \in \mathbb{N}$ such that $\mu(A \cap T^{-p(n)}A) > 0$.*

3.1. The Furstenberg-Sárközy theorem – equivalent forms

Proposition 25. *Let p be a polynomial satisfying $p(\mathbb{N}) \subseteq \mathbb{N}$ and $p(0) = 0$. The following are equivalent:*

(i) (Furstenberg-Sárközy theorem – infinitary version). Any $E \subseteq \mathbb{N}$ with positive upper density contains $\{a, a + p(n)\}$ for some $a, n \in \mathbb{N}$.

(ii) (Furstenberg-Sárközy theorem – finitary version). For every $\delta > 0$ there exists $N(\delta, p) \in \mathbb{N}$ such that if $N \geqslant N(\delta)$ then any set $E \subseteq \{1, \ldots, N\}$ with $|E| \geqslant \delta N$ contains $\{a, a + p(n)\}$ for some $a, n \in \mathbb{N}$.

(iii) (Furstenberg's polynomial recurrence theorem). For any measure preserving system $(X, \mathscr{A}, \mu, T)$ and any $A \in \mathscr{A}$ with $\mu(A) > 0$ there exist $n \in \mathbb{N}$ such that $\mu(A \cap T^{-p(n)}A) > 0$.

Proof. Note that (i) says $\{p(n) : n \in \mathbb{N}\}$ is an intersective set and (iii) says $\{p(n) : n \in \mathbb{N}\}$ is a set of recurrence. By Theorem 23, these two assertions are equivalent.

It remains to prove the equivalence (i) $\Longleftrightarrow$ (ii). Note that the direction (ii) $\Longrightarrow$ (i) is immediate.

For the reverse direction, we use a proof by contradiction. Assume that (ii) is false, which means there exist $\delta > 0$ and an increasing sequence $N_1 < N_2 < \ldots \in \mathbb{N}$ such that for every $i \in \mathbb{N}$ there is a set $E_i \subseteq \{1,\ldots,N_i\}$ with $|E_i| \geqslant \delta N_i$ admitting no arrangement of the from $\{a, a+p(n)\}$ for $a, n \in \mathbb{N}$. By replacing $(N_i)_{i\in\mathbb{N}}$ with a subsequence of itself, we can assume without loss of generality that $N_{i+1} > 4N_i$. Define a new set E via

$$E = \bigcup_{i\in\mathbb{N}} (N_i + E_i).$$

A straightforward calculation reveals that

$$\begin{aligned}\overline{d}(E) &= \limsup_{N\to\infty} \frac{|E \cap \{1,\ldots,N\}|}{N} \\ &\geqslant \limsup_{i\to\infty} \frac{|E \cap \{1,\ldots,2N_i\}|}{2N_i} \\ &\geqslant \limsup_{i\to\infty} \frac{|(N_i + E_i)|}{2N_i} \\ &= \limsup_{i\to\infty} \frac{|E_i|}{2N_i} \geqslant \frac{\delta}{2}.\end{aligned}$$

By (i), the set $\{p(n) : n \in \mathbb{N}\}$ is intersective, which in view of Theorem 23 means that there exists some $m \in \mathbb{N}$ such that $\{n \in E : n + p(m) \in E\}$ is infinite. In particular, we can find $n, m \in \mathbb{N}$ with $n > p(m)$ and such that $n, n+p(m) \in E$. Take $i \in \mathbb{N}$ with $n \in N_i + E_i$. Since $N_{i+1} > 4N_i$, the smallest element of $N_{i+1} + E_{i+1}$ is at least twice as large as the largest element in $N_i + E_i$. So from $n \in N_i + E_i$ and $n > p(m)$ it follows that $n + p(m) \in N_i + E_i$. This means that $p(m)$ appears as a difference in E_i, yielding a contradiction to the assumption that E_i does not contain differences of this shape. □

3.2. van der Corput's difference theorem

Theorem 26 (van der Corput's difference theorem). *Let $\mathscr{H}$ be a Hilbert space with inner product $\langle .,. \rangle$ and norm $\|.\|$, and let $(x_n)_{n=1}^{\infty}$ be a bounded sequence taking values in $\mathscr{H}$. If*

$$\limsup_{H\to\infty} \limsup_{N\to\infty} \left| \frac{1}{H} \sum_{h=1}^{H} \frac{1}{N} \sum_{n=1}^{N} \langle x_{n+h}, x_n \rangle \right| = 0 \tag{3.1}$$

then

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} x_n \right\| = 0.$$

Proof. First, observe that we have

$$\limsup_{D\to\infty} \limsup_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} x_n - \frac{1}{D} \sum_{d=1}^{D} \frac{1}{N} \sum_{n=1}^{N} x_{n+d} \right\| = 0.$$

Hence it suffices to show that

$$\limsup_{D\to\infty} \limsup_{N\to\infty} \left\| \frac{1}{D} \sum_{d=1}^{D} \frac{1}{N} \sum_{n=1}^{N} x_{n+d} \right\| = 0.$$

Using the Cauchy-Schwarz inequality we have

$$\begin{aligned}
\limsup_{D\to\infty}\limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}\frac{1}{D}\sum_{d=1}^{D}x_{n+d}\right\|^2 &\leqslant \limsup_{D\to\infty}\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left\|\frac{1}{D}\sum_{d=1}^{D}x_{n+d}\right\|^2\\
&= \limsup_{D\to\infty}\limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\frac{1}{D^2}\sum_{d_1,d_2=1}^{D}\langle x_{n+d_1},x_{n+d_2}\rangle\\
&\leqslant \limsup_{D\to\infty}\limsup_{N\to\infty}\left|\frac{1}{D^2}\sum_{d=1}^{D}2(D-d)\left(\frac{1}{N}\sum_{n=1}^{N}\langle x_n,x_{n+d}\rangle\right)\right|.
\end{aligned}$$

Note that for any sequence $y_1,\dots,y_D$,

$$\frac{1}{D^2}\sum_{d=1}^{D}2(D-d)y_d=\frac{1}{D}\sum_{H=1}^{D}\left(\frac{1}{H}\sum_{h=1}^{H-1}y_h\right).$$

All the above combined gives

$$\begin{aligned}
\limsup_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}x_n\right\| &\leqslant \limsup_{D\to\infty}\lim_{N\to\infty}\left|\frac{1}{D}\sum_{H=1}^{D}\left(\frac{1}{H}\sum_{h=1}^{H-1}\frac{1}{N}\sum_{n=1}^{N}\langle x_n,x_{n+h}\rangle\right)\right|\\
&\leqslant \limsup_{D\to\infty}\frac{1}{D}\sum_{H=1}^{D}\limsup_{N\to\infty}\left|\frac{1}{H}\sum_{h=1}^{H-1}\frac{1}{N}\sum_{n=1}^{N}\langle x_n,x_{n+h}\rangle\right|\\
&\leqslant \limsup_{H\to\infty}\left|\frac{1}{H}\sum_{h=1}^{H-1}\frac{1}{N}\sum_{n=1}^{N}\langle x_n,x_{n+h}\rangle\right|.
\end{aligned}$$

The claim follows. □

3.3. The ergodic theorem

Given a measure-preserving system $(X,\mathscr{A},\mu,T)$ and a function $f\in L^2(X,\mathscr{A},\mu)$, the notation Tf denotes the composition $f\circ T$, which is also in $L^2(X,\mathscr{A},\mu)$ due to the measure-preserving property of T. This defines a linear operator $T:L^2(X,\mathscr{A},\mu)\to L^2(X,\mathscr{A},\mu)$, called the *Koopman operator*, acting as an isometry.

Definition 27. The system $(X,\mathscr{A},\mu,T)$ is *ergodic* if the only functions $f\in L^2(X,\mathscr{A},\mu)$ satisfying $Tf=f$ are those that are constant almost everywhere.

We denote by $\mathscr{H}_{\mathrm{inv}}$ the space of almost everywhere invariant functions in $L^2(X,\mathscr{A},\mu)$,

$$\mathscr{H}_{\mathrm{inv}}=\{f\in L^2(X,\mathscr{A},\mu):Tf=f\}.$$

In view of Definition 27, the system $(X,\mathscr{A},\mu,T)$ is ergodic if and only if $\mathscr{H}_{\mathrm{inv}}$ consists only of almost everywhere constant functions.

Theorem 28 (Mean ergodic theorem). *Let $(X,\mathscr{A},\mu,T)$ be a measure preserving system. For every $f\in L^2(X,\mathscr{A},\mu)$ we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}U_T^n f=f_{\mathrm{inv}}\qquad \text{in } L^2\text{-norm,}\tag{3.2}$$

where f_{inv} is the orthogonal projection of f onto $\mathscr{H}_{\mathrm{inv}}$.

Remark 29. Note that if $(X,\mathscr{A},\mu,T)$ is an ergodic system then $f_{\mathrm{inv}} = \int_X f \,\mathrm{d}\mu$ and hence (3.2) becomes

$$\underbrace{\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} T^n f}_{\text{time average}} = \underbrace{\int f \,\mathrm{d}\mu}_{\text{space average}} \qquad \text{in } L^2(X,\mathscr{A},\mu).$$

We say that $L^2(X,\mathscr{A},\mu)$ is the *orthogonal direct sum* of $\mathscr{H}_1$ and $\mathscr{H}_2$, and write $L^2(X,\mathscr{A},\mu) = \mathscr{H}_1 \oplus \mathscr{H}_2$, if $\mathscr{H}_1$ and $\mathscr{H}_2$ are closed subspaces of $\mathscr{H}$ satisfying:

- $\langle f_1, f_2\rangle = 0$ whenever $f_1 \in \mathscr{H}_1$ and $f_2 \in \mathscr{H}_2$, and
- for every $f \in \mathscr{H}$, there exist $f_1 \in \mathscr{H}_1$ and $f_2 \in \mathscr{H}_2$ such that $f = f_1 + f_2$.

Note that in this case, f_1 and f_2 are uniquely determined by f. In fact, f_1 equals the orthogonal projection of f onto the subspace $\mathscr{H}_1$, whereas f_2 is the orthogonal projection of f onto $\mathscr{H}_2$.

The following is an immediate corollary of the mean ergodic theorem.

Corollary 30. *Let $(X,\mathscr{A},\mu,T)$ be a measure preserving system. Then $L^2(X,\mathscr{A},\mu) = \mathscr{H}_{\mathrm{inv}} \oplus \mathscr{H}_{\mathrm{erg}}$, where $\mathscr{H}_{\mathrm{inv}}$ is as defined above and*

$$\mathscr{H}_{\mathrm{erg}} = \left\{ f \in L^2(X,\mathscr{A},\mu) : \lim_{N\to\infty} \left\| \frac{1}{N}\sum_{n=1}^{N} T^n f \right\|_{L^2} = 0 \right\}.$$

3.4. Totally ergodic systems

Definition 31. A measure preserving system $(X,\mathscr{A},\mu,T)$ is *totally ergodic* if for every $m \in \mathbb{N}$, the measure preserving system $(X,\mathscr{B},\mu,T^m)$ is ergodic.

Theorem 32. *Let $(X,\mathscr{A},\mu,T)$ be a measure preserving system. Then $(X,\mathscr{A},\mu,T)$ is totally ergodic if and only if for all non-constant polynomials p with $p(\mathbb{N}) \subseteq \mathbb{N}$ and any $f \in L^2(X,\mathscr{A},\mu)$ we have*

$$\underbrace{\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} T^{p(n)} f}_{\text{polynomial time average}} = \underbrace{\int f \,\mathrm{d}\mu.}_{\text{space average}} \qquad \text{in } L^2(X,\mathscr{A},\mu). \tag{3.3}$$

Proof. If the system is not totally ergodic, then there exists $q \in \mathbb{N}$ and a non-constant $f \in L^2(X,\mathscr{A},\mu)$ such that $T^q f = f$. Thus

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} T^{p(n)} f = f \neq \int_X f \,\mathrm{d}\mu$$

contradicting (3.3).

To prove the converse direction, let $(X,\mathscr{B},\mu,T)$ be totally ergodic and let $f \in L^2(X,\mathscr{A},\mu)$ be arbitrary. We proceed by induction on the degree of p. If $p(n) = qn + r$ is a linear polynomial, then the result follows from the ergodic theorem applied to the ergodic transformation T^q. So assume that p has degree at least 2. Eq. (3.3) holds for f if and only if it holds for $f - c$ where c is a constant; therefore, after subtracting $\int_X f \,\mathrm{d}\mu$ from f, we can assume that $\int_X f \,\mathrm{d}\mu = 0$. Letting $x_n = T^{p(n)} f$, we need to show that $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} x_n = 0$, and to this end we will invoke the van der Corput difference theorem (Theorem 26) applied to the Hilbert space $L^2(X,\mathscr{A},\mu)$. Fixing $d \in \mathbb{N}$ we can compute

$$\langle x_{n+d}, x_n\rangle = \int_X T^{p(n+d)} f \cdot T^{p(n)}\overline{f} \,\mathrm{d}\mu = \int_X T^{p(n+d)-p(n)} f \cdot \overline{f} \,\mathrm{d}\mu = \left\langle T^{p(n+d)-p(n)} f, f \right\rangle.$$

Since $n \mapsto p(n+d) - p(n)$ is a polynomial of degree smaller than the degree of p, we can use the induction hypothesis (together with the fact that convergence in $L^2(X,\mathscr{A},\mu)$ implies convergence in the weak topology) to conclude

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle x_{n+d},x_n\rangle = \left\langle \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{p(n+d)-p(n)}f,f\right\rangle = 0.$$

This establishes the hypothesis (3.1) of the van der Corput's difference theorem, so we conclude that $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}x_n = 0$, as desired. □

3.5. Proof of the Furstenberg's polynomial recurrence theorem

Theorem 33 (cf. [Ber96, p. 14]). *Let $(X,\mathscr{A},\mu,T)$ be a measure preserving system. Then $L^2(X,\mathscr{A},\mu) = \mathscr{H}_{\text{rat}}\oplus\mathscr{H}_{\text{tot erg}}$, where*

$$\mathscr{H}_{\text{rat}} = \overline{\{f\in L^2(X,\mathscr{A},\mu):\ \exists q\in\mathbb{N},\ T^q f = f\}},$$
$$\mathscr{H}_{\text{tot erg}} = \left\{f\in L^2(X,\mathscr{A},\mu): \forall q\in\mathbb{N},\ \lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}T^{qn}f\right\|_{L^2} = 0\right\}.$$

Proof. For every $q\in\mathbb{N}$ define

$$\mathscr{H}_{\text{inv},q} = \{f\in L^2(X,\mathscr{A},\mu): T^q f = f\}\quad\text{and}\quad \mathscr{H}_{\text{erg},q} = \left\{f\in L^2(X,\mathscr{A},\mu): \lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N}T^{qn}f\right\|_{L^2} = 0\right\}.$$

By Corollary 30, we have $L^2(X,\mathscr{A},\mu) = \mathscr{H}_{\text{inv},q}\oplus\mathscr{H}_{\text{erg},q}$ for all $q\in\mathbb{N}$. Therefore,

$$L^2(X,\mathscr{A},\mu) = \overline{\bigcup_{q\in\mathbb{N}}\mathscr{H}_{\text{inv},q}}\oplus\bigcap_{q\in\mathbb{N}}\mathscr{H}_{\text{erg},q}.$$

It is now straightforward to verify that $\mathscr{H}_{\text{rat}} = \overline{\bigcup_{q\in\mathbb{N}}\mathscr{H}_{\text{inv},q}}$ and $\mathscr{H}_{\text{tot erg}} = \bigcap_{q\in\mathbb{N}}\mathscr{H}_{\text{erg},q}$. □

Proof of Theorem 24. Using Theorem 33, we can decompose $1_A = f+g$ with $f\in\mathscr{H}_{\text{rat}}$ and $g\in\mathscr{H}_{\text{tot erg}}$. Since $\mathscr{H}_{\text{rat}}$ contains the constant functions, using the Cauchy-Schwarz inequality we have $\langle 1_A,f\rangle = \|f\|_{L^2}^2 \geqslant \langle f,1\rangle^2 = \mu(A)^2$. Find $h\in\mathscr{H}_{\text{rat}}$ such that $T^q h = h$ for some $q\in\mathbb{N}$, and such that $\|f-h\|_{L^2} < \varepsilon/2$. In particular it follows that $\langle 1_A,h\rangle > \mu(A)^2-\varepsilon/2$.

Note that $T^{p(qn)}h = h$ for all $n\in\mathbb{N}$. As in the proof of Theorem 32, an application of the van der Corput difference theorem implies that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{p(qn)}g = 0.$$

Finally, we have

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mu(A\cap T^{-p(qn)}A) &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left\langle 1_A, h+T^{p(qn)}(f-h)+T^{p(qn)}g\right\rangle\\ &= \left\langle 1_A, h+\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{p(qn)}(f-h)+T^{p(qn)}g\right\rangle\\ &\geqslant \langle 1_A,h\rangle-\varepsilon/2\\ &\geqslant \mu(A)^2-\varepsilon.\end{aligned}$$

□

4. Roth's theorem

In this section, we give an ergodic-theoretic proof of Roth's theorem (Theorem 5). The next result is an ergodic theorem from which Roth's theorem can be derived. It is, in fact, equivalent to Roth's theorem, and can be viewed as a "degree-two" generalization of Poincaré's recurrence theorem.

Theorem 34 (Furstenberg's double recurrence theorem, [Fur77]). *For any measure preserving system $(X,\mathscr{A},\mu,T)$ and any $A\in\mathscr{A}$ with $\mu(A)>0$ there exist $n\in\mathbb{N}$ such that $\mu(A\cap T^{-n}A\cap T^{-2n}A)>0$.*

4.1. Roth's theorem – equivalent forms

Proposition 35. *The following are equivalent:*

(i) (Roth's theorem – infinitary version). Any $E\subseteq\mathbb{N}$ with positive upper density contains a 3-term arithmetic progressions.

(ii) (Roth's theorem – finitary version). For every $\delta>0$ there exists $N(\delta)\in\mathbb{N}$ such that if $N\geqslant N(\delta)$ then any set $E\subseteq\{1,\ldots,N\}$ with $|E|\geqslant\delta N$ contains a 3-term arithmetic progression.

(iii) (Furstenberg's double recurrence theorem). For any measure preserving system $(X,\mathscr{A},\mu,T)$ and any $A\in\mathscr{A}$ with $\mu(A)>0$ there exist $n\in\mathbb{N}$ such that $\mu(A\cap T^{-n}A\cap T^{-2n}A)>0$.

The proof of Proposition 35 is very similar to the proof of Proposition 25 and omitted.

4.2. Weak mixing systems

Definition 36. A measure preserving system $(X,\mathscr{A},\mu,T)$ is called *weak mixing* if for all $f,g\in L^2(X,\mathscr{A},\mu)$ we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left|\langle T^n f,g\rangle-\left(\int f\,\mathrm{d}\mu\right)\cdot\left(\int g\,\mathrm{d}\mu\right)\right|=0.$$

Theorem 37. *Let $(X,\mathscr{A},\mu,T)$ be a weakly mixing measure preserving system. Then for any $f,g\in L^2(X,\mathscr{A},\mu)$ we have*

$$\underbrace{\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^n f T^{2n}g}_{\text{degree−two time average}}=\underbrace{\left(\int f\,\mathrm{d}\mu\right)\cdot\left(\int g\,\mathrm{d}\mu\right)}_{\text{space average}},\qquad \text{in } L^2(X,\mathscr{A},\mu).$$

Proof. By replacing g with $g-\int_X g\,\mathrm{d}\mu$, we can assume without loss of generality that $\int_X g\,\mathrm{d}\mu=0$. With the goal of using Theorem 26, let $x_n=T^n f\cdot T^{2n}g$. We have

$$\langle x_{n+h},x_n\rangle=\int_X T^{n+h}f\cdot T^{2n+2h}g\cdot T^n\overline{f}\cdot T^{2n}\overline{g}\,\mathrm{d}\mu=\int_X\left(\overline{f}\cdot T^h f\right)\cdot T^n\left(\overline{g}\cdot T^{2h}g\right)\mathrm{d}\mu.$$

Using ergodicity and taking an average in n we get

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle x_{n+h},x_n\rangle=\left(\int_X\overline{f}\cdot T^h f\,\mathrm{d}\mu\right)\cdot\left(\int_X\overline{g}\cdot T^{2h}g\,\mathrm{d}\mu\right).$$

Thus, we have

$$\lim_{H\to\infty}\lim_{N\to\infty}\left|\frac{1}{H}\sum_{h=1}^{H}\frac{1}{N}\sum_{n=1}^{N}\langle x_{n+h},x_n\rangle\right|=\lim_{H\to\infty}\left|\frac{1}{H}\sum_{h=1}^{H}\left(\int_X\overline{f}\cdot T^h f\,\mathrm{d}\mu\right)\cdot\left(\int_X\overline{g}\cdot T^{2h}g\,\mathrm{d}\mu\right)\right|$$

$$\leqslant \|f\|_{L^2}^2 \cdot \left(\lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \left| \int_X \overline{g} \cdot T^{2h} g \, \mathrm{d}\mu \right| \right).$$

The last expression is 0 because of Definition 36. □

4.3. The Jacobs–de Leeuw–Glicksberg decomposition

Definition 38. Let $(X, \mathscr{A}, \mu, T)$ be a measure preserving system and let $f \in L^2(X, \mathscr{A}, \mu)$. We say that f is a *weak mixing* function if for all $g \in L^2(X, \mathscr{A}, \mu)$ one has

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left| \langle T^n f, g \rangle \right| = 0.$$

Notice that a weak-mixing function f always satisfies $\int f \, \mathrm{d}\mu = 0$. Moreover, a system is weak mixing if and only if every function f with $\int f \, \mathrm{d}\mu = 0$ is a weak-mixing function.

Definition 39. Given a measure preserving system $(X, \mathscr{A}, \mu, T)$, a non-zero function $f \in L^2(X, \mathscr{A}, \mu)$ is an *eigenfunction* if there exists a constant λ, called the *eigenvalue*, such that $Tf = \lambda f$. The set of all eigenvalues of $(X, \mathscr{A}, \mu, T)$ is called the *point-spectrum of T* and denoted by $\sigma(T)$.

Since T is an isometry, all its eigenvalues must have absolute value 1. In other words, for any measure preserving system $(X, \mathscr{A}, \mu, T)$ we have $\sigma(T) \subseteq \mathbb{S}^1$.

Theorem 40 (Jacobs–de Leeuw–Glicksberg decomposition). *Let $(X, \mathscr{A}, \mu, T)$ be a measure preserving system. Then $L^2(X, \mathscr{A}, \mu) = \mathscr{H}_{\mathrm{c}} \oplus \mathscr{H}_{\mathrm{wm}}$, where*

$$\mathscr{H}_{\mathrm{c}} = \overline{\operatorname{span}\left\{ f \in L^2(X, \mathscr{A}, \mu) : f \text{ is an eigenfunction} \right\}},$$
$$\mathscr{H}_{\mathrm{wm}} = \left\{ f \in L^2(X, \mathscr{A}, \mu) : f \text{ is a weak mixing function} \right\}.$$

Sketch of Proof. We provide here only a sketch of the proof; for a complete argument see [EFHN15, Section 16.3].

Fix $f \in L^2(A, \mathscr{A}, \mu)$. Let $H_f = \overline{\operatorname{span}\{T^n f : n \in \mathbb{N} \cup \{0\}\}}$, which is the smallest closed and T-invariant subspace of $L^2(A, \mathscr{A}, \mu)$ containing f. By the spectral theorem, there exists a finite Borel measure ν on $\mathbb{S}^1$ (called the spectral measure of f) and an isometric isomorphism $\Phi \colon H_f \to L^2(\mathbb{S}^1, \mathscr{B}_{\mathbb{S}^1}, \nu)$ such that for all $g \in H_f$ we have

$$\Phi(T^n g)(z) = z \cdot \Phi(g)(z), \qquad \text{for } \nu\text{-a.e. } z \in \mathbb{S}^1.$$

We can split ν into its discrete and continuous components, i.e., $\nu = \nu_{\mathrm{discrete}} + \nu_{\mathrm{continuous}}$. This induces a splitting

$$L^2(\mathbb{S}^1, \mathscr{B}_{\mathbb{S}^1}, \nu) = L^2(\mathbb{S}^1, \mathscr{B}_{\mathbb{S}^1}, \nu_{\mathrm{discrete}}) \oplus L^2(\mathbb{S}^1, \mathscr{B}_{\mathbb{S}^1}, \nu_{\mathrm{continuous}}).$$

Through the isomorphism Φ, this induces a splitting

$$H_f = H_{\mathrm{c},f} \oplus H_{\mathrm{wm},f},$$

where $\Phi(H_{\mathrm{c},f}) = L^2(\mathbb{S}^1, \mathscr{B}_{\mathbb{S}^1}, \nu_{\mathrm{discrete}})$ and $\Phi(H_{\mathrm{wm},f}) = L^2(\mathbb{S}^1, \mathscr{B}_{\mathbb{S}^1}, \nu_{\mathrm{continuous}})$. In particular, we can write f uniquely as

$$f = f_{\mathrm{c}} + f_{\mathrm{wm}},$$

where $f_{\mathrm{c}} \in H_{\mathrm{c},f}$ and $f_{\mathrm{wm}} \in H_{\mathrm{wm},f}$. Note that the spectral measure associated to f_{c} is discrete and the spectral measure associated to f_{wm} is continuous. One can then show that any element in $L^2(A,\mathscr{A},\mu)$ whose spectral measure is discrete belongs to $\mathscr{H}_{\mathrm{c}}$ (using elementary methods) and any element in $L^2(A,\mathscr{A},\mu)$ whose spectral measure is continuous belongs to $\mathscr{H}_{\mathrm{wm}}$ (using Wiener's lemma). This shows that $H_{\mathrm{c},f} \subseteq \mathscr{H}_{\mathrm{c}}$ as well as $H_{\mathrm{wm},f} \subseteq \mathscr{H}_{\mathrm{wm}}$. □

Remark 41. Functions in $\mathscr{H}_{\mathrm{c}}$ are often refereed to as *compact functions*, because (through some work) one can show that $\mathscr{H}_{\mathrm{c}}$ are exactly those functions f whose orbit closure $\overline{\{T^n f : n \in \mathbb{N} \cup \{0\}\}}$ is a compact subset of $L^2(X,\mathscr{A},\mu)$ (with respect to the norm-topology). From this characterization, it also follows that if $f \in L^2(X,\mathscr{A},\mu)$ satisfies $a \leqslant f(x) \leqslant b$ for μ-a.e. $x \in X$, where $a \leqslant b\mathbb{R}$, then f_{c}, the orthogonal projection of f onto $\mathscr{H}_{\mathrm{c}}$, also has $a \leqslant f(x) \leqslant b$ for μ-a.e. $x \in X$.

4.4. Proof of Furstenberg's double recurrence theorem

Lemma 42. *Let $(X,\mathscr{A},\mu,T)$ be a measure preserving system and let $f,g \in L^2(X,\mathscr{A},\mu)$. If at least one of the functions f or g is weakly mixing, then we have*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} T^n f \cdot T^{2n} g = 0, \qquad \text{in } L^2(X,\mathscr{A},\mu).$$

The proof of Lemma 42 follows the same arguments as the proof of Theorem 37.

Lemma 43. *Let $(X,\mathscr{A},\mu,T)$ be a measure preserving system. Then for any $f_{\mathrm{c}} \in \mathscr{H}_{\mathrm{c}}$ with $f_{\mathrm{c}}(x) \in [0,1]$ for μ-a.e. $x \in X$ and $\int f_{\mathrm{c}}\,\mathrm{d}\mu > 0$ we have*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int f_{\mathrm{c}} \cdot T^n f_{\mathrm{c}} \cdot T^{2n} f_{\mathrm{c}}\,\mathrm{d}\mu > 0.$$

Proof. Given $\varepsilon > 0$, we can find $r \in \mathbb{N}$ and eigenfunctions $g_1,\ldots,g_r$ such that if $g = g_1 + \ldots + g_r$ then

$$\|f_{\mathrm{c}} - g\|_{L^2} \leqslant \varepsilon.$$

Let $\lambda_1,\ldots,\lambda_r \in \mathbb{S}^1$ denote the eigenvalues corresponding to the eigenfunctions $g_1,\ldots,g_r$. Let $B \subseteq \mathbb{N}$ be the Bohr set

$$B = \big\{n \in \mathbb{N} : |\lambda_i^n - 1| \leqslant \varepsilon \|g_i\|_{L^2} r^{-1},\ i = 1,\ldots,r\big\}.$$

Then for any $n \in B$ we have

$$\begin{aligned}
\|T^n f_{\mathrm{c}} - f_{\mathrm{c}}\|_{L^2} &\leqslant \|T^n g - g\|_{L^2} + 2\varepsilon \\
&\leqslant \|(T^n g_1 - g_1)\|_{L^2} + \ldots + \|(T^n g_r - g_r)\|_{L^2} + 2\varepsilon \\
&= |\lambda_1^n - 1| \cdot \|g_1\|_{L^2} + \ldots + |\lambda_r^n - 1| \cdot \|g_r\|_{L^2} + 2\varepsilon \\
&\leqslant 3\varepsilon.
\end{aligned}$$

This also implies that $\|T^{2n} f_{\mathrm{c}} - f_{\mathrm{c}}\|_{L^2} \leqslant 6\varepsilon$. Thus, we have

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int f_{\mathrm{c}} \cdot T^n f_{\mathrm{c}} \cdot T^{2n} f_{\mathrm{c}}\,\mathrm{d}\mu &\geqslant \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_B(n) \int f_{\mathrm{c}} \cdot T^n f_{\mathrm{c}} \cdot T^{2n} f_{\mathrm{c}}\,\mathrm{d}\mu \\
&\geqslant \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_B(n) \left(\int f_{\mathrm{c}}^3\,\mathrm{d}\mu - 9\varepsilon\right) \\
&\geqslant d(B) \cdot \left(\int f_{\mathrm{c}}^3\,\mathrm{d}\mu - 9\varepsilon\right).
\end{aligned}$$

If ε is sufficiently small, then this last quantity is positive and we are done. □

Proof of Theorem 34. Using Theorem 40, we can decompose $1_A = f_c + f_{wm}$ with $f_c \in \mathscr{H}_c$ and $f_{wm} \in \mathscr{H}_{wm}$. Note that since $1_A(x) \in [0,1]$, we have $f_c(x) \in [0,1]$ for μ-a.e. $x \in X$ (cf. Remark 41). Moreover, as $\langle 1_A, 1\rangle = \langle f_c, 1\rangle$, we have $\int_X f_c \, d\mu > 0$. Using Lemma 42, we have

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

$$= \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int 1_A \cdot T^n f_c \cdot T^{2n} f_c \, d\mu.$$

Note that $T^n f_c \cdot T^{2n} f_c \in \mathscr{H}_c$ and hence it is orthogonal to f_{wm}. This implies that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int 1_A \cdot T^n f_c \cdot T^{2n} f_c \, d\mu = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int f_c \cdot T^n f_c \cdot T^{2n} f_c \, d\mu.$$

The claim now follows from Lemma 43. □

5. Sumsets in sets of positive density

In this section, we present a proof of Theorem 10.

5.1. Improved correspondence principle

We say a measure preserving system $(X, \mathscr{A}, \mu, T)$, where X is a compact metric space and T is a continuous map on X, has *continuous eigenfunctions* if every eigenfunction has a continuous representative in its L^2 equivalency class.

Lemma 44. *One can assume without loss of generality that the system obtained in Furstenberg's correspondence principle (Theorem 22) is ergodic and has continuous eigenfunctions.*

The proof of Lemma 44 is omitted. A proof can be obtained by combining [KMRR24, Proposition 2.3] and [KMRR24, Proposition 3.1].

5.2. Reduction to a dynamical statement

In what follows, we say a family of sets has the *large intersection property* if any finite sub-family intersects in a set that has infinite cardinality.

Lemma 45. *Let $E \subseteq \mathbb{N}$. Then E contains $B + C$ for some infinite $B, C \subseteq \mathbb{N}$ if and only if there exists an increasing sequence of positive integers $(s_i)_{i\in\mathbb{N}}$ such that*

$$1_L(n) = \lim_{i\to\infty} 1_E(n + s_i)$$

exists for all $n \in \mathbb{N}$, and the family of sets $(L \cap (E - s_i))_{i\in\mathbb{N}}$ has the large intersection property.

Proof. First, suppose there is a sequence $(s_i)_{i\in\mathbb{N}}$ such that $1_L(n) = \lim_{i\to\infty} 1_E(n + s_i)$ exists and $(L \cap (E - s_i))_{i\in\mathbb{N}}$ has the large intersection property. We can use this to inductively construct sequences $(b_i)_{i\in\mathbb{N}}$ and $(c_j)_{j\in\mathbb{N}}$ with

$$b_i + c_j \in E, \qquad \forall i, j \in \mathbb{N}.$$

Take $b_1 = s_1$ and let c_1 be any element in $L \cap (E - b_1) = L \cap (E - s_1)$. If $c_1, \dots, c_n \in L$ and $b_1, \dots, b_n \in \{s_i : i \in \mathbb{N}\}$ have already been found, then let c_{n+1} be any element in

$$\bigcap_{i=1}^{n} \big(L \cap (E - b_i)\big) \tag{5.1}$$

that satisfies $c_{n+1} > c_n$; such an element exists because $b_1, \dots, b_n \in \{s_i : i \in \mathbb{N}\}$ and hence the intersection in (5.1) has infinite cardinality by the large intersection property. Since $c_1, \dots, c_{n+1} \in L$ we have

$$1 = 1_L(c_j) = \lim_{i \to \infty} 1_E(c_j + s_i), \qquad j = 1, \dots, n+1.$$

Therefore $c_j + s_i \in E$ for all but finitely many $i \in \mathbb{N}$. This allows us to find some $b_{n+1} \in \{s_i : i \in \mathbb{N}\}$ for which $b_{n+1} > b_n$ and $c_j + b_{n+1} \in E$ for all $j = 1, \dots, n+1$. It is now straightforward to verify that $b_i + c_j \in E$ for all $i, j \in \mathbb{N}$.

For the prove of the reverse direction, assume we have two infinite sequences $(b_i)_{i \in \mathbb{N}}$ and $(c_j)_{j \in \mathbb{N}}$ such that

$$b_i + c_j \in E, \qquad \forall i, j \in \mathbb{N}.$$

By refining $(b_i)_{i \in \mathbb{N}}$ if necessary, we can assume that

$$1_L(n) = \lim_{i \to \infty} 1_E(n + b_i)$$

exists for all $n \in \mathbb{N}$. Then, since $C = \{c_j : j \in \mathbb{N}\}$ is a subset of $L \cap (E - b_i)$ for all $i \in \mathbb{N}$, the family of sets $(L \cap (E - b_i))_{i \in \mathbb{N}}$ has the large intersection property. □

Here is a dynamical result from which Theorem 10 follows.

Theorem 46. *Let X be a compact metric space and $T \colon X \to X$ a continuous map. Let μ be a Borel probability measure on X and suppose $(X, \mathscr{A}, \mu, T)$ is ergodic and has continuous eigenfunctions. Let $x \in X$ be a point with a dense orbit. Then for any clopen set $A \subseteq X$ with $\mu(A) > 0$ there exist a point $y \in X$, a sequence $s_1 < s_2 < \ldots \in \mathbb{N}$, and a Borel probability measure λ on $X \times X$ such that*

- *(i) $T^{s_i} x \to y$ as $i \to \infty$;*
- *(ii) λ is supported on the orbit closure of (x, y) under $T \times T$;*
- *(iii) for any $k \in \mathbb{N}$ we have $\lambda\big(\big(T^{-s_1} A \cap \ldots \cap T^{-s_k} A\big) \times A\big) > 0$.*

Proof that Theorem 46 implies Theorem 10. Suppose $E \subseteq N$ has positive upper density. By Theorem 22, there exist a compact metric space X, a Borel probability measure μ on X, a continuous measure preserving transformation $T \colon X \to X$, a point $x \in X$ with dense orbit, and a clopen set $A \subseteq X$ such that $\mu(A) \geqslant \overline{d}(E)$ and $E = \{n \in \mathbb{N} : T^n x \in A\}$. Let $\mathscr{A}$ denote the Borel σ-algebra on X. In view of Lemma 44, we can assume without loss of generality that $(X, \mathscr{A}, \mu, T)$ is ergodic and has continuous eigenfunctions. Since all the prerequisites of Theorem 46 are met, there exist a point $y \in X$, a sequence $s_1 < s_2 < \ldots \in \mathbb{N}$, and a Borel probability measure λ on $X \times X$ so that (i) – (iii) are satisfied. Now define $L = \{n \in \mathbb{N} : T^n y \in A\}$. Since $T^{s_i} x \to y$ as $i \to \infty$ we have

$$\begin{aligned} 1_L(n) &= 1_A(T^n y) \\ &= \lim_{i \to \infty} 1_A(T^{n+s_i} x) \\ &= \lim_{i \to \infty} 1_E(n + s_i). \end{aligned}$$

Moreover, for any $k \in \mathbb{N}$,

$$\begin{aligned}(L \cap (E - s_1)) \cap \ldots \cap (L \cap (E - s_k)) \\ &= \big((E - s_1) \cap \ldots \cap (E - s_k)\big) \cap L \\ &= \big\{n \in \mathbb{N} : (T \times T)^n(x, y) \in \big(T^{-s_1}A \cap \ldots \cap T^{-s_k}A\big) \times A\big\}.\end{aligned}$$

Since A is a clopen set, λ is supported on the orbit closure of (x, y), and $\lambda\Big(\big(T^{-s_1}A \cap \ldots \cap T^{-s_k}A\big) \times A\Big) > 0$, we conclude that the set of n for which $(T \times T)^n(x, y)$ belongs to $\big(T^{-s_1}A \cap \ldots \cap T^{-s_k}A\big) \times A$ is infinite. It follows that the family of sets $(L \cap (E - s_i))_{i \in \mathbb{N}}$ has the large intersection property. By Lemma 45, E contains $B + C$ for some infinite $B, C \subseteq \mathbb{N}$ as desired. □

5.3. Proof of Theorem 46

It remains to prove Theorem 46.

Proof of Theorem 46. Let $(I_k)_{k \in \mathbb{N}}$ be a sequence of intervals in $\mathbb{N}$ such that x is generic for μ along $(I_k)_{k \in \mathbb{N}}$; such a sequence exists because the orbit of x is dense in X. By replacing $(I_k)_{k \in \mathbb{N}}$ with a subsequence of itself if necessary, we can assume without loss of generality that μ-almost every point in X is generic for μ along $(I_k)_{k \in \mathbb{N}}$. Since $\mathrm{supp}(\mu)$, the topological support of μ, has full measure, there exists at least one point $y \in \mathrm{supp}(\mu)$ that is generic for μ along $(I_k)_{k \in \mathbb{N}}$. Further refining $(I_k)_{k \in \mathbb{N}}$ if necessary, we can assume that (x, y) is generic along $(I_k)_{k \in \mathbb{N}}$ with respect to the transformation $T \times T$ for a measure λ on $X \times X$.

Using Theorem 40, we can write

$$1_A = f_{\mathrm{c}} + f_{\mathrm{wm}}$$

for $f_{\mathrm{c}} \in \mathscr{H}_{\mathrm{c}}$ and $f_{\mathrm{wm}} \in \mathscr{H}_{\mathrm{wm}}$. For convenience, write $f_{\mathrm{c},1} = f_{\mathrm{c}} \otimes 1$ and $f_{\mathrm{c},2} = 1 \otimes f_{\mathrm{c}}$. We make the following claim.

Claim 1. *For all $\delta > 0$ there exists $\eta(\delta) > 0$ such that $d(T^s x, y) < \eta(\delta)$ implies $\|T^s f_{\mathrm{c},1} - f_{\mathrm{c},2}\|_{L^2(\lambda)} \leqslant \delta$.*

Proof of Claim 1. Suppose $\delta > 0$ is given. Since $f_{\mathrm{c}} \in \mathscr{H}_{\mathrm{c}}$, we know that f_{c} can be approximated by eigenfuncitons. In particular, there exist eigenfunctions $g_1, \ldots, g_r$ such that if $g = g_1 + \ldots + g_r$ then $\|g - f_{\mathrm{c}}\|_{L^2} \leqslant \delta/3$. Next, we choose $\eta(\delta) > 0$ sufficiently small such that $|g_i(y) - g_i(z)| \leqslant \delta/3r$ whenever $d(z, y) < \eta(\delta)$; this is possible because g_i is a continuous function on a compact space. Since g_i is an eigenfunction, we additionally obtain $\sup_{n \in \mathbb{N}} |g_i(T^n y) - g_i(T^n z)| \leqslant \delta/3r$ whenever $d(z, y) < \eta(\delta)$, which implies

$$\sup_{n \in \mathbb{N}} |g(T^n y) - g(T^n z)| \leqslant \delta/3 \quad \text{whenever} \quad d(z, y) < \eta(\delta).$$

It follows that for all $s \in \mathbb{N}$ with $d(T^s x, y) < \eta(\delta)$ we have $\sup_{n \in \mathbb{N}} |g(T^{n+s}x) - g(T^n y)| \leqslant \delta/3$. By denseness of $(T^n x, T^n y)$, it follows that

$$\|T^s f_{\mathrm{c},1} - f_{\mathrm{c},2}\|_{L^2(\lambda)} \leqslant \|(T^s g \otimes 1) - (1 \otimes g)\|_{L^2(\lambda)} + \frac{2\delta}{3} \leqslant \sup_{n \in \mathbb{N}} |g(T^{n+s}x) - g(T^n y)| + \frac{2\delta}{3} \leqslant \delta,$$

as desired. △

We are left with finding a sequence $s_1 < s_2 < \ldots \in \mathbb{N}$ such that $T^{s_i} x \to y$ as $i \to \infty$ and

$$\lambda\Big(\big(T^{-s_1}A \cap \ldots \cap T^{-s_k}A\big) \times A\Big) > 0 \tag{5.2}$$

for all $k \in \mathbb{N}$. We will build this sequence by induction on k. For convenience write $A_1 = A \times X$, $A_2 = X \times A$, $f_{wm,1} = f_{\mathrm{wm}} \otimes 1$, and $f_{\mathrm{wm},2} = 1 \otimes f_{\mathrm{wm}}$. Then (5.2) can be rewritten as

$$\lambda\Big(A_2 \cap T^{-s_1} A_1 \cap \ldots \cap T^{-s_k} A_1\Big) > 0. \tag{5.3}$$

Suppose $s_1 < \ldots < s_k$ for which (5.3) is satisfied have already been found, i.e., the set $Y := A_2 \cap (T \times T)^{-s_1} A_1 \cap \ldots \cap (T \times T)^{-s_k} A_1$ satisfies $\lambda(Y) > 0$. (If we haven't found any s_i yet, let $k = 0$ and $Y = A_2$.) Our goal is to find s_{k+1}.

Claim 2. *We have $\langle 1_Y, f_{\mathrm{c},2}\rangle > 0$.*

Proof of Claim 2. Let $P \colon L^2(X \times X, \mathscr{A} \otimes \mathscr{A}, \lambda) \to L^2(X \times X, \mathscr{A} \otimes \mathscr{A}, \lambda)$ denote the orthogonal projection onto the subspace $1 \otimes \mathscr{H}_{\mathrm{c}} = \{1 \otimes f : f \in \mathscr{H}_{\mathrm{c}}\}$. Note that

$$1_{A_2} = 1 \otimes 1_A = 1 \otimes f_{\mathrm{c}} + 1 \otimes f_{\mathrm{wm}} = f_{\mathrm{c},2} + f_{\mathrm{wm},2},$$

which shows that $f_{\mathrm{c},2} = P 1_{A_2}$. Since $Y \subseteq A_2$, we have $0 \leqslant 1_Y \leqslant 1_{A_2}$. It follows that $P 1_Y \leqslant P 1_{A_2} = f_{\mathrm{c},2}$ and hence

$$\langle 1_Y, f_{\mathrm{c},2}\rangle = \langle P 1_Y, f_{\mathrm{c},2}\rangle \geqslant \langle P 1_Y, P 1_Y\rangle = \|P 1_Y\|^2_{L^2(\lambda)}.$$

But $\|P 1_Y\|^2_{L^2(\lambda)} > 0$ because $\lambda(Y) > 0$, finishing the proof of Claim 2. △

Now take $\delta := \frac{1}{3}\langle 1_Y, f_{\mathrm{c},2}\rangle$ and let $V := \{s \in \mathbb{N} : d(T^s x, y) < \min\{\eta(\delta), 2^{-k}\}\}$, where $\eta(\delta)$ is as in *Claim 1*. Also, take $D := \{s \in \mathbb{N} : |\langle T^s f_{\mathrm{wm},1}, 1_Y\rangle| < \delta\}$. Since $y \in \mathrm{supp}(\mu)$ and x is generic for μ along $(I_k)_{k\in\mathbb{N}}$, the set V has positive density with respect to $(I_k)_{k\in\mathbb{N}}$. Moreover, since $f_{\mathrm{wm},1}$ is weak mixing (in $L^2(X \times X, \mathscr{A} \otimes \mathscr{A}, \lambda)$) we conclude that the set D has full density with respect to $(I_k)_{k\in\mathbb{N}}$. This implies that the sets D and V intersect in an infinite set; let s_{k+1} be any element in this intersection with $s_{k+1} > s_k$. It follows that

$$\begin{aligned}\langle 1_Y, T^{s_{k+1}} 1_{A_1}\rangle &= \langle 1_Y, T^{s_{k+1}} f_{\mathrm{c},1}\rangle + \langle 1_Y, T^{s_{k+1}} f_{\mathrm{wm},1}\rangle \\ &\geqslant \langle 1_Y, f_{\mathrm{c},2}\rangle - 2\delta \\ &> 0.\end{aligned}$$

Also, since we ensured that $d(T^{s_{k+1}} x, y) < 2^{-k}$, we end up with a sequence $s_1 < s_2 < \ldots$ that satisfies $T^{s_i} x \to y$ as $i \to \infty$. This completes the proof. □

References

[Ber96] V. Bergelson, Ergodic Ramsey theory – an update, in *Ergodic theory of $\mathbb{Z}^d$ actions (Warwick, 1993–1994)*, *London Math. Soc. Lecture Note Ser.* **228**, Cambridge Univ. Press, Cambridge, 1996, pp. 1–61. https://doi.org/10.1017/CBO9780511662812.002.

[BL96] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems, *J. Amer. Math. Soc.* **9** no. 3 (1996), 725–753. https://doi.org/10.1090/S0894-0347-96-00194-4.

[BS21] T. F. Bloom and O. Sisask, Breaking the logarithmic barrier in roth's theorem on arithmetic progressions, 2021. Available at https://arxiv.org/abs/2007.03528.

[BS23] T. F. BLOOM and O. SISASK, The kelley–meka bounds for sets free of three-term arithmetic progressions, *Essential Number Theory* **2** no. 1 (2023), 15–44. https://doi.org/10.2140/ent.2023.2.15.

[EFHN15] T. EISNER, B. FARKAS, M. HAASE, and R. NAGEL, *Operator theoretic aspects of ergodic theory*, *Graduate Texts in Mathematics* **272**, Springer, Cham, 2015. https://doi.org/10.1007/978-3-319-16898-2.

[Fur77] H. FURSTENBERG, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204–256. https://doi.org/10.1007/BF02813304.

[GT08] B. GREEN and T. TAO, The primes contain arbitrarily long arithmetic progressions, *Ann. of Math. (2)* **167** no. 2 (2008), 481–547. MR 2415379. https://doi.org/10.4007/annals.2008.167.481.

[KM23] Z. KELLEY and R. MEKA, Strong bounds for 3-progressions, in *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, 2023, pp. 933–973. https://doi.org/10.1109/FOCS57990.2023.00059.

[KMRR24] B. KRA, J. MOREIRA, F. K. RICHTER, and D. ROBERTSON, A proof of Erdős's $B+B+t$ conjecture, *Commun. Am. Math. Soc.* **4** (2024), 480–494. MR 4780876. https://doi.org/10.1090/cams/34.

[MRR19] J. MOREIRA, F. K. RICHTER, and D. ROBERTSON, A proof of a sumset conjecture of Erdős, *Ann. of Math. (2)* **189** no. 2 (2019), 605–652. https://doi.org/10.4007/annals.2019.189.2.4.

[Rot53] K. ROTH, On certain sets of integers, *Journal of the London Mathematical Society* **28** no. 1 (1953), 104–109. https://doi.org/10.1112/jlms/s1-28.1.104.

[Sár78] A. SÁRKÖZY, On difference sets of sequences of integers. I, *Acta Math. Acad. Sci. Hungar.* **31** no. 1–2 (1978), 125–149. MR 0466059. https://doi.org/10.1007/BF01896079.

[Sze75] E. SZEMERÉDI, On sets of integers containing k elements in arithmetic progression, *Acta Arithmetica* **27** no. 1 (1975), 199–245 (eng). Available at http://eudml.org/doc/205339.