ERGODIC RAMSEY THEORY – NOTES

JOEL MOREIRA

These notes are being written for the TCC module on Ergodic Ramsey Theory, which runs in the Fall of 2021. Ergodic Ramsey Theory is a relatively young subject of mathematics whose purpose is to apply techniques, methods and ideas from ergodic theory, and more the general theory of dynamical systems, to problems that arise in Ramsey theory, combinatorics, and number theory. The main interface between the dynamical and the combinatorial realms is provided by the Correspondence Principle of Furstenberg, first introduced in [5] to give a new ergodic theoretic proof of Szemerédi's theorem on arithmetic progressions.

The module will start by introducing some of the problems from Ramsey theory that we will consider, as well as the preliminary results from Ergodic theory, and then, after introducing the Furstenberg Correspondence Principle we will go over the ergodic theoretic proof of Szemerédi's theorem. The last half (or third) of the course focuses on more recent developments in ergodic Ramsey theory (still to be decided).

We will not follow any single textbook from beginning to end, but both Furstenberg's book [6] and Einsiedler-Ward's book [4] share the same spirit of introducing ergodic theory both as a theory on its own and as a tool to approach problems in combinatorics and number theory. Bergelson's survey [1] with the same title as these notes obviously shares a great deal of content. A more advanced text on this subject is the recent book of Host and Kra [10], which goes into much more depth. For an introductory text to general ergodic theory, Walters [18] is an excellent source, which can be complemented with Glasner's [7] or Cornfeld-Fomin-Sinai's [3]. For an introductory text to general Ramsey Theory, the book [9] by Graham, Rothschild and Spencer of that title is still one of the best sources.

1. Ramsey Theory

Ramsey theory is a branch of combinatorics which, roughly speaking, explores structures that persist when partitioned. Instead of trying to give a more precise description, we illustrate this principle with a few examples of results from Ramsey theory.

Theorem 1.1 (Schur, [15]). *Given a finite coloring of $\mathbb{N}$, one can always find $x, y \in \mathbb{N}$ with $x, y, x+y$ all having the same colour.*

To be clear, a finite coloring of $\mathbb{N}$ is a function $f : \mathbb{N} \to F$, where F is a finite set (whose elements are the "colours"). Two elements x, y of $\mathbb{N}$ have the same color if $f(x) = f(y)$.

In fact, it is not necessary to color all of $\mathbb{N}$ before one finds a ***monochromatic*** (i.e. with a single color) triple of the form $\{x, y, x+y\}$. Here's an alternative formulation of Schur's theorem.

Theorem 1.2 (Schur, again). *For every $r \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that whenever the set $\{1, \dots, N\}$ is colored with r colors there is a monochromatic triple of the form $\{x, y, x+y\} \subset \{1, \dots, N\}$.*

The difference between Theorems 1.1 and 1.2 is that in the latter, N is chosen depending only on the number of colors r. To estimate the smallest N in terms of r is a difficult and interesting problem, but the purely qualitative Theorem 1.2 as formulated turns out to be equivalent to the apparently weaker Theorem 1.1 (and not in the uninteresting sense that any two true statements are tautological equivalent).

Exercise 1.3. Prove that Theorems 1.1 and 1.2 are equivalent. *[Hint: One implication is easy. For the other, suppose you have counterexamples to Theorem 1.2 for every N, then you can use them to find a counterexample to Theorem 1.1.]*

The most common way to solve the previous exercise is to use, explicitly or implicitly, the so-called ***compactness principle***, which in this case is simply the statement that the set of all colorings of $\mathbb{N}$

Date: October 20, 2021.

into r colors is a compact set. The compactness principle allows one to formulate many Ramsey theoretic statements in an ***infinitary*** form, such as Theorem 1.1. This is the form that ergodic theory can handle, but it is useful to keep in mind that the statements are equivalent to their ***finitistic*** forms.

The next theorem was considered by Khinchine as one of "Three Pearls in Number Theory" [12].

Theorem 1.4 (Van der Waerden, [17]). *In any finite coloring of $\mathbb{N}$ there exist arbitrarily long monochromatic arithmetic progressions.*

In other words, for any $k \in \mathbb{N}$ there are $x, y \in \mathbb{N}$ such that the arithmetic progression $\{x, x+y, x+2y, \ldots, x+ky\}$ is monochromatic.

There is a natural finitistic form of van der Waerden's theorem.

Exercise 1.5. *Show that Theorem 1.4 is equivalent to the following statement:*

"For any $r, k \in \mathbb{N}$ there exists N such that for any coloring of the set $\{1, \ldots, N\}$ with r colors there exists a monochromatic arithmetic progression of the form $\{x, x+y, x+2y, \ldots, x+ky\} \subset \{1, \ldots, N\}$."

There is yet another equivalent formulation of van der Warden's theorem with a more geometric flavour.

Exercise 1.6. *Show that Theorem 1.4 is equivalent to the following statement:*

"For any finite coloring of $\mathbb{N}$ and for any finite set $F \subset \mathbb{N}$ there exists a monochromatic affine image of F, i.e. there exist $a, b \in \mathbb{N}$ such that the set $aF + b := \{ax + b : x \in F\}$ is monochromatic."

The next result was conjectured by Erdős and Turán as an attempt to better understand the true nature of van der Waerden's theorem. After some initial progress it was finally settled by Szemerédi in a remarkably involved combinatorial proof. In order to state it we need the notion of (upper) density.

Definition 1.7 (Upper density). *Given a set $A \subset \mathbb{N}$ its* **upper density**, *denoted $\bar{d}(A)$ is the quantity*

$$\bar{d}(A) = \limsup_{N\to\infty} \frac{1}{N}\big|A \cap \{1, \ldots, N\}\big|.$$

Replacing $\limsup$ *with* $\liminf$ *we obtain the analogous notion of* **lower density**.

Here and elsewhere in these notes, when X is a finite set we denote by $|X|$ its cardinality.

Exercise 1.8. *Show that upper density is subadditive and shift invariant, i.e. if $A, B \subset \mathbb{N}$ and $n \in \mathbb{N}$ then $\bar{d}(A \cup B) \leq \bar{d}(A) + \bar{d}(B)$, and $\bar{d}(A - n) = \bar{d}(A)$, where $A - n := \{x \in \mathbb{N} : x + n \in A\}$.*

Theorem 1.9 (Szemerédi, [16]). *If $A \subset \mathbb{N}$ has positive upper density, then it contains arbitrarily long arithmetic progressions.*

Note that Szemerédi's theorem implies van der Waerden's theorem, since for any finite coloring of $\mathbb{N}$ one can use Exercise 1.8 to deduce that at least one of the colors has positive density.

Here is the finitistic form of Szemerédi's theorem.

Exercise 1.10. *Show that Theorem 1.9 is equivalent to the following statement:*

"For any $\delta > 0$ and $k \in \mathbb{N}$ there exists N such that any set $A \subset \{1, \ldots, N\}$ with $|A| > \delta N$ contains an arithmetic progression of the form $\{x, x+y, x+2y, \ldots, x+ky\}$."

Exercise 1.11. *Let $k \in \mathbb{N}$. Show that there exists $\delta < 1$ such that any set $A \subset \mathbb{N}$ with $\bar{d}(A) > \delta$ contains an arithmetic progression of the form $\{x, x+y, x+2y, \ldots, x+ky\}$.*

More than twenty years before Szemerédi's theorem was first proved, Roth obtained the special case corresponding to arithmetic progressions of length 3.

Theorem 1.12 (Roth, [13]). *Any set $A \subset \mathbb{N}$ with $\bar{d}(A) > 0$ contains a 3-term arithmetic progression.*

Roth's proof of Theorem 1.12 made use of Fourier Analysis, and would later inspire Gowers to obtain a full proof of Szemerédi's theorem [8] by developing what is now called "Higher order Fourier Analysis". Another Ramsey theoretic result that can be obtained using Fourier Analysis is the following.

Theorem 1.13 (Sárközy, [14]). *If $A \subset \mathbb{N}$ has $\bar{d}(A) > 0$, then there exist $x, y \in A$ whose difference is a perfect square.*

Theorem 1.13 is connected with the study of *sets of differences* of large sets. In this context, we think of a set $A \subset \mathbb{N}$ with positive upper density as a *large set*, and are interested in understand the structure of the set of differences $A - A := \{x - y : x, y \in A\}$. A related concept is that of intersective sets:

Definition 1.14. *A set $R \subset \mathbb{Z}$ is a called* **intersective** *if for every $A \subset \mathbb{N}$ with $\bar{d}(A) > 0$, the intersection $(A - A) \cap R$ is non-empty.*

Using this terminology, Theorem 1.13 states that the set of perfect squares is an intersective set.

Exercise 1.15. *Show that the following are intersective sets.*

- *Any set with lower density* 1*.*
- *The set $k\mathbb{N}$ of all multiples of k, for an arbitrary $k \in \mathbb{N}$.*
- $(*)$ *Any set of differences $I - I$ for any infinite set I (not necessarily with positive upper density).*

Exercise 1.16. *Show that the following are not intersective sets.*

- *The odd numbers.*
- *The set $\mathbb{N} \setminus (k\mathbb{N})$ of numbers not divisible by k, for an arbitrary $k \in \mathbb{N}$.*

Sárközy's theorem can be extended to more general polynomials than $p(x) = x^2$. The exact extent of this generalization was only fully understood after work of Furstenberg [5, 6] and of Kamae and Mendes-France [11].

Theorem 1.17. *Let $p \in \mathbb{Z}[x]$ be a polynomial with integer coefficients and no constant term. Then the set $R := \{p(n) : n \in \mathbb{N}\}$ is intersective if and only if it contains a multiple of any $k \in \mathbb{N}$ (in other words, if p has a root modulo k for every k).*

Notice that an easy sufficient condition on a polynomial to have a root modulo k for every k, is to satisfy $p(0) = 0$.

One can interpret Sárközy's theorem as stating that any set $A \subset \mathbb{N}$ with positive upper density contains a 2-term arithmetic progression whose common difference is a perfect square. From this angle it makes sense to ask about longer arithmetic progressions. The following powerful theorem of Bergelson and Leibman gives an affirmative answer.

Theorem 1.18 (Polynomial Szemerédi theorem, [2])**.** *Let $p_1, \ldots, p_k \in \mathbb{Z}[x]$ satisfy $p_i(0) = 0$. Then any set $A \subset \mathbb{N}$ with $\bar{d}(A) > 0$ contains a "polynomial progression" of the form*

$$\big\{x, x + p_1(y), x + p_2(y), \ldots, x + p_k(y)\big\}.$$

Observe that by taking $p_i(y) = iy$ one recovers Szemerédi's theorem from Theorem 1.18.

2. Ergodic theory background

In this section we collect some of the basic definitions and facts about ergodic theory that we will need later on.

Definition 2.1 (Measure preserving transformation)**.** *Given two probability spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$, we say that a map*[1] *$T : X \to Y$* **preserves the measure** *or is a* **measure preserving transformation** *if for every $B \in \mathcal{B}$, the set $T^{-1}B := \{x \in X : Tx \in B\}$ is in $\mathcal{A}$ and satisfies $\mu(T^{-1}B) = \nu(B)$.*

A map between probability spaces induces a linear operator between the corresponding L^p spaces.

Exercise 2.2. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be probability spaces and let $T : X \to Y$ be a measurable map.*

- *Show that T preserves the measure if and only if for every $f \in L^2(Y)$, the function $f \circ T$ belongs to $L^2(X)$ and satisfies*
$$\int_X f \circ T \, \mathsf{d}\mu = \int_Y f \, \mathsf{d}\nu. \tag{2.1}$$
- *If both μ and ν are Radon measures, show that T preserves the measure if and only if (2.1) holds for every $f \in C(Y)$. [Hint: $C(Y)$ is dense in $L^2(Y)$.]*

[1] To be completely precise, T may be defined only on a full measure subset of X.

The basic object in ergodic theory is a ***measure preserving system*** (m.p.s. for short), which we now define.

Definition 2.3 (Measure preserving system)**.** *A measure preserving system is a quadruple $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is a probability space and $T : X \to X$ is a measure preserving transformation.*

Example 2.4 (Circle rotation)**.** *Let $X = [0, 1)$, endowed with the Borel σ-algebra $\mathcal{B}$ and the Lebesgue measure μ. Given $\alpha \in \mathbb{R}$ we consider the map $T = T_\alpha : X \to X$ given by $Tx = x + \alpha \bmod 1$. The fact that T preserves the measure μ follows from the basic properties of Lebesgue measure.*

Alternatively, we can identify the space X with the compact group $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ in the obvious way. The Lebesgue measure on $[0, 1)$ gets identified with the Haar measure on $\mathbb{T}$, and T becomes the map $Tx = x + \tilde{\alpha}$ (where $\tilde{\alpha} = \alpha + \mathbb{Z} \in \mathbb{T}$). This map clearly preserves the Haar measure.

The reason to call this system a circle rotation is that the group $\mathbb{T}$ is isometrically isomorphic to the circle $S^1 \subset \mathbb{C}$, viewed as a group under multiplication. The map T under this identification becomes the rotation $T : z \mapsto \theta z$, where $\theta = e^{2\pi i \alpha} \in S^1$.

The above example can be extended to "rotations" on any compact group X, endowed with the Borel σ-algebra $\mathcal{B}$ and Haar measure μ. Taking any $\alpha \in X$, the map $T : x \mapsto x + \alpha$ preserves μ and hence $(X, \mathcal{B}, \mu, T)$ is a measure preserving system, called a ***group rotation*** or a ***Kronecker system***.

Example 2.5 (Doubling map)**.** *Again take $(X, \mathcal{B}, \mu)$ to be the unit interval $X = [0, 1]$ equipped with its Borel σ-algebra and Lebesgue measure. Let $T : X \to X$ be the doubling map $Tx = 2x \bmod 1$.*

At first sight it may seem that the doubling map doubles the measure, but in fact it preserves the measure! For instance, given an interval $[a, b] \subset [0, 1]$, the pre-image $T^{-1}[a, b]$ is the union of two intervals, each half the length of the original interval:

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

Exercise 2.6. *Show that the doubling map does indeed preserve the Lebesgue measure. [Hint: use Exercise 2.2]*

Here is the first theorem of ergodic theory.

Theorem 2.7 (Poicaré recurrence theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then for some $n \in \mathbb{N}$ we have*

$$\mu(A \cap T^{-n}A) > 0. \tag{2.2}$$

Proof. The sets $A, T^{-1}A, T^{-2}A, \ldots$ all have the same (positive) measure, and all live in X which has measure 1. Therefore we must have $\mu(T^{-i}A \cap T^{-j}A) > 0$ for some $i > j$. Finally, letting $n = i - j$, observe that

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

□

While Poicaré's recurrence theorem is a simple result, it has a lot of potential for extensions, which in turn reveal a lot about the structure of measure preserving systems. For instance, one may ask how small can we choose n? How large is the set of n for which (2.2) holds? How large can we make the measure of the intersection be?

In order to address some of these questions, we make the following definition.

Definition 2.8. *A set R of natural numbers is called a* **set of recurrence** *if for every measure preserving system $(X, \mathcal{B}, \mu, T)$ and every $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists $n \in R$ such that $\mu(A \cap T^{-n}A) > 0$.*

With this notion we can reformulate Poicaré's recurrence theorem as stating that $\mathbb{N}$ is a set of recurrence.

Exercise 2.9. *Show that the set $2\mathbb{N}$ of even numbers is a set of recurrence but the set $2\mathbb{N} - 1$ of odd numbers is not.*

Here is a more sophisticated result, due to Furstenberg, which will be proved later in the course.

Theorem 2.10. *The set $Q := \{m^2 : m \in \mathbb{N}\}$ of perfect squares is a set of recurrence. In fact, for every m.p.s. $(X, \mathcal{B}, \mu, T)$, every $A \in \mathcal{B}$ and for every $\epsilon > 0$ there exists a perfect square $n = m^2 \in \mathbb{N}$ such that*

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

It turns out that the notion of sets of recurrence coincides with the notion of intersective sets.

Proposition 2.11. *A set $R \subset \mathbb{N}$ is a set of recurrence if and only if it is intersective (see Definition 1.14).*

Proposition 2.11 provides the first connection we've encountered between combinatorics and Ramsey theory; to prove it we will need the Furstenberg Correspondence Principle.

Exercise 2.12. $(*)$ *Show that if $R \subset \mathbb{N}$ is a set of recurrence and is decomposed as $R = A \cup B$ then either A or B is a set of recurrence. [Hint: Proceed by contradiction and take the product system of the two presumed counter-examples.]*

2.1. **Ergodicity.** The word ergodic arises from Boltzman's "ergodic hypothesis" in termodynamics, which describes a system where, *over long periods of time, the time spent by a system in some region of the phase space of microstates with the same energy is proportional to the volume of this region*[2]. In the language of measure preserving systems, the ergodic hypothesis would imply that the proportion of time that the orbit of a point (i.e. the sequence $x, Tx, T^2x, \dots$) is in a set A, tends to $\mu(A)$. This is in fact the conclusion of the ergodic theorem, which will be discussed below.

However, there is an obvious obstruction to the ergodic hypothesis: suppose $(X_i, \mathcal{A}_i, \mu_i, T_i)$ is a measure preserving system for each $i = 1, 2$ with X_1 and X_2 disjoint. Now let $Y = X_1 \cup X_2$, let $\mathcal{B}$ be the σ-algebra generated by $\mathcal{A}_1 \cup \mathcal{A}_2$, let $\nu = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2$ and let $S : Y \to Y$ be the map that maps $x \in X_i$ to $T_i x$, for $i = 1, 2$. Then $(Y, \mathcal{B}, \nu, S)$ is a measure preserving system, but a point $x \in X_1$ (or, more precisely, its orbit) will never visit X_2, even though $\mu(X_2) = 1/2 > 0$. A system is ergodic when it avoids this behavior.

Definition 2.13. *A measure preserving system $(X, \mathcal{B}, \mu, T)$ is* **ergodic** *if every set $A \in \mathcal{B}$ satisfying $T^{-1}A = A$ is trivial in the sense that either $\mu(A) = 0$ or $\mu(A) = 1$.*

Proposition 2.14. *A measure preserving system $(X, \mathcal{B}, \mu, T)$ is ergodic if and only if every $f \in L^2$ which is invariant in the sense that $f \circ T = f$ a.e. is constant a.e.*

Proof. For every $A \in \mathcal{B}$ the indicator function 1_A is in L^2, and hence we obtain the "only if" implication.

For the converse implication, suppose the system is ergodic and $f \in L^2$ is invariant. Then for every $t \in \mathbb{R}$, the set $A_t := \{x \in X : f(x) > t\}$ is invariant and hence has either measure 0 or 1. Let $r = \inf\{t : \mu(A_t) = 0\}$. Then $\mu(A_r) = 0$ because $A_r = \bigcup_{n \geq 1} A_{r+1/n}$. On the other hand $\mu(A_t) = 1$ for every $t < r$ and hence $\mu(\{x : f(x) \geq r\}) = 1$. We conclude that $f = r$ a.e. ☐

The ergodic theorems assert, roughly speaking, that ergodic systems satisfy the ergodic hypothesis. Given a measure preserving system $(X, \mathcal{B}, \mu, T)$, the set $I \subset L^2(X)$ consisting of (almost everywhere) T-invariant functions, i.e. $I := \{f \in L^2(X) : f \circ T = f\}$ is a closed subspace. Therefore we can consider the orthogonal projection $P_I : L^2(X) \to I$ defined so that $P_I f$ is the element of I which is closest to f. It is not hard to show that P_I is a linear operator, and that it satisfies $\langle f - P_I f, g \rangle = 0$ for every $g \in I$. Here and in these notes, the inner product in L^2 is defined by

$$\langle f, g \rangle = \int_X f(x)\overline{g(x)}\, \mathsf{d}\mu(x).$$

Theorem 2.15 (Birkhoff's pointwise ergodic theorem, L^2 version)**.** *Let $P_I : L^2(X) \to I$ denote the orthogonal projection onto the subspace of T-invariant functions. Then for every $f \in L^2$*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f \circ T^n = P_I f \qquad \text{a.e.}. \tag{2.3}$$

If the system is ergodic, then I consists only of the constant functions and $P_I f = \int_X f \, \mathsf{d}\mu$ a.e. Therefore for ergodic systems we have the following corollary.

[2] https://en.wikipedia.org/wiki/Ergodic_hypothesis

Corollary 2.16. *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system. Then for every $A \in \mathcal{B}$ and almost every $x \in X$,*

$$\lim_{N\to\infty} \frac{1}{N} \Big| \big\{ n \in \{1, \dots, N\} : T^n x \in A \big\} \Big| = \mu(A).$$

Proof. Apply Theorem 2.15 to the indicator function 1_A of A and observe that, for each $x \in X$,

$$\sum_{n=1}^{N} (1_A \circ T^n)(x) = \Big| \big\{ n \in \{1, \dots, N\} : T^n x \in A \big\} \Big|.$$

□

A different version of the ergodic theorem was obtained by von Neumann, usually called the mean ergodic theorem because it deals with convergence in L^2 (or more generally in L^p) instead of almost everywhere convergence. This version has the advantage that it holds even if one changes the averaging scheme from $\{1, \dots, N\}$ to any sequence of intervals $\{a_N, a_N + 1, \dots, a_N + N\}$. Moreover, the simpler proof of von Neumann's theorem can be easily modified to apply to measure preserving actions of any amenable group.

Theorem 2.17 (von Neumann's mean ergodic theorem, L^2 version)**.** *Let $P_I : L^2(X) \to I$ denote the orthogonal projection onto the subspace of T-invariant functions. Then for every $f \in L^2$*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} f \circ T^n = P_I f \qquad \textit{in } L^2(X). \tag{2.4}$$

Remark 2.18.

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} h_n = c$$

means that for every $\epsilon > 0$ there exists some K such that if $M, N \in \mathbb{N}$ satisfy $N - M > K$, then $\left| \frac{1}{N-M} \sum_{n=M}^{N} h_n - c \right| < \epsilon$. This mode of convergence is often used in ergodic theory and is called a **uniform Cesàro limit** *or a* **uniform Cesàro average***, as opposed to the kind of averages used in the pointwise ergodic theorem, called simply* **Cesàro averages***.*

Exercise 2.19. *Show that*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} h_n = c$$

is equivalent to

$$\forall\ (I_N)_{N\in\mathbb{N}} \qquad \lim_{N\to\infty} \frac{1}{|I_N|} \sum_{n\in I_N} h_n = c$$

where $(I_N)_{N\in\mathbb{N}}$ is a sequence of intervals $I_N = \{a_N + 1, a_N + 2, \dots, a_N + b_N\}$ whose lengths b_N tend to infinity.

Given a measure preserving system $(X, \mathcal{B}, \mu, T)$, the ***Koopman operator*** $\Phi_T : L^2(X) \to L^2(X)$ is the linear operator defined by the equation $\Phi_T f := f \circ T$. Since T is measure preserving, it follows that Φ_T is an isometry, i.e., $\langle \Phi_T f, \Phi_T g \rangle = \langle f, g \rangle$. Therefore Theorem 2.17 is a corollary of the following.

Theorem 2.20 (von Neumann's mean ergodic theorem, Hilbert space version)**.** *Let H be a Hilbert space, let $\Phi : H \to H$ be an isometry and let $I \subset H$ be the subspace of invariant vectors, i.e. $I = \{f \in H : \Phi f = f\}$. Let $P : H \to I$ be the orthogonal projection onto I. Then for every $f \in H$,*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} \Phi^n f = P f \qquad \textit{in norm} \tag{2.5}$$

Proof. If $f \in I$ then (2.5) holds trivially (with both sides equal to f).

On the other hand, if $f = g - \Phi g$ for some $g \in H$, then for any $h \in I$ we have

$$\langle f, h\rangle = \langle g, h\rangle - \langle \Phi g, h\rangle = \langle g, h\rangle - \langle g, \Phi h\rangle = 0$$

hence f is orthogonal to I and so $Pf = 0$. Moreover we have that $\sum_{n=M}^{N} \Phi^n f = \Phi^M g - \Phi^{N+1} g$, which has norm at most $2\|g\|$, and so the limit in the left hand side of (2.5) is also 0.

Call J the subspace of the vectors of the form $g - \Phi g$. We claim that $H = I \oplus J$ and this concludes the proof. To prove the claim, letting $f \perp J$, we have:

$$\begin{aligned} \|f - \Phi f\| &= \|f\|^2 + \|\Phi f\|^2 - 2Re\langle f, \Phi f\rangle \\ &= 2\|f\|^2 - 2Re\langle f, \Phi f\rangle - 2Re\langle f, f - \Phi f\rangle = 2\|f\|^2 - 2Re\langle f, f\rangle = 0 \end{aligned}$$

so $f \in I$ and hence $I = J^{\perp}$ and this finishes the proof. ☐

Corollary 2.21. *A measure preserving system* $(X, \mathcal{B}, \mu, T)$ *is ergodic if and only if for every* $A, B \in \mathcal{B}$*,*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} \mu(T^{-n}A \cap B) = \mu(A)\mu(B). \tag{2.6}$$

Proof. If the system is not ergodic, then there exists $A \in \mathcal{B}$ with $\mu(A) \in (0, 1)$ which is invariant. Therefore, taking $B = X \setminus A$, we see that $T^{-n}A \cap B = \emptyset$ for every n, contradicting (2.6).

Let $f = 1_A$ and $g = 1_B$. Observe that $1_{T^{-n}A} = f \circ T^n = \Phi_T^n f$. Therefore $\mu(T^{-n}A \cap B) = \int_X \Phi_T^n 1_A \cdot 1_B \, \mathsf{d}\mu = \langle \Phi_T^n 1_A, 1_B\rangle$. Since strong (or norm) convergence in L^2 implies weak convergence, it follows from(2.4) that

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} \mu(T^{-n}A \cap B) = \langle P_I f, g\rangle.$$

Finally, in view of ergodicity, we have that $P_I f$ is the constant $\int_X f \, \mathsf{d}\mu = \mu(A)$, and (2.6) follows from the fact that $\int_X \mu(A) g \, \mathsf{d}\mu = \mu(A)\mu(B)$. ☐

Setting $A = B$ in Corollary 2.21 we see that, in ergodic system, one can improve Poincaré's recurrence theorem by finding $n \in \mathbb{N}$ such that $\mu(T^{-n}A \cap A)$ is arbitrarily close to $\mu^2(A)$. One can in fact obtain a stronger version of this fact, which also applies to non-ergodic systems.

Definition 2.22. *A set* $S \subset \mathbb{N}$ *is called* **syndetic** *if it has bounded gaps. More precisely,* S *is syndetic if there exists* $L \in \mathbb{N}$ *such that every interval* $\{n, n+1, \ldots, n+L-1\}$ *of length* L *contains some element of* S*.*

Exercise 2.23. *Let* (a_n) *be a sequence of non-negative real numbers and let* $a \in \mathbb{R}$*. Show that if*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} a_n = a,$$

then for every $\epsilon > 0$ *the set*

$$\{n \in \mathbb{N} : a_n \geq a - \epsilon\}$$

is syndetic.

Theorem 2.24 (Khintchine's recurrence theorem)**.** *Let* $(X, \mathcal{B}, \mu, T)$ *be a measure preserving system, let* $A \in \mathcal{B}$ *and let* $\epsilon > 0$*. Then there exists* $n \in \mathbb{N}$ *such that* $\mu(A \cap T^{-n}A) > \mu^2(A) - \epsilon$*, and moreover the set*

$$\left\{n \in \mathbb{N} : \mu(A \cap T^{-n}A) > \mu^2(A) - \epsilon\right\}$$

is syndetic.

Proof. Applying Theorem 2.17 to the indicator function 1_A of A we have

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} \mu(T^{-n}A \cap A) = \int_X P_I 1_A \cdot 1_A \, \mathsf{d}\mu.$$

Since P_I is an orthogonal projection it follows that $\int_X P_I 1_A \cdot 1_A \, \mathsf{d}\mu = \|P_I 1_A\|^2$. We now use the Cauchy-Schwarz inequality to get

$$\|P_I 1_A\|^2 \geq \left(\int_X P_I 1_A \, \mathsf{d}\mu \right)^2 = \mu(A)^2.$$

$\square$

References

[1] V. Bergelson. Ergodic Ramsey theory–an update. In *Ergodic theory of $\mathbb{Z}^d$ actions*, volume 228 of *London Math. Soc. Lecture Note Ser.*, pages 1–61. Cambridge Univ. Press, Cambridge, 1996.

[2] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9(3):725–753, 1996.

[3] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinaĭ. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskiĭ.

[4] Manfred Einsiedler and Thomas Ward. *Ergodic theory with a view towards number theory*, volume 259 of *Graduate Texts in Mathematics*. Springer-Verlag London, Ltd., London, 2011.

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

[6] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, Princeton, N.J., 1981.

[7] E. Glasner. *Ergodic theory via joinings*, volume 101 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2003.

[8] T. Gowers. A new proof of Szemerédi's theorem. *GAFA*, 11:465–588, 2001.

[9] R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey theory*. John Wiley & Sons, Inc., New York, second edition, 1990.

[10] B. Host and B. Kra. *Nilpotent structures in ergodic theory*, volume 236 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2018.

[11] T. Kamae and M. Mendès France. Van der Corput's difference theorem. *Israel J. Math.*, 31(3-4):335–342, 1978.

[12] A. Y. Khinchin. *Three pearls of number theory*. Graylock Press, Rochester, N. Y., 1952.

[13] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:245–252, 1953.

[14] A. Sárközy. On difference sets of sequences of integers. I. *Acta Math. Acad. Sci. Hungar.*, 31(1–2):125–149, 1978.

[15] I. Schur. Über die kongruenz $x^m + y^m \equiv z^m (\text{mod } p)$. *Jahresbericht der Deutschen Math. Verein.*, 25:114–117, 1916.

[16] E. Szemerédi. On the sets of integers containing no k elements in arithmetic progressions. *Acta Arith.*, 27:299–345, 1975.

[17] B. L. van der Waerden. Beweis einer baudetschen vermutung. *Nieuw. Arch. Wisk.*, 15:212–216, 1927.

[18] P. Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1982.