ERGODIC THEORY – NOTES

JOEL MOREIRA

Contents

1. Measure preserving systems 2
1.1. Probability spaces 2
1.2. Measure preserving systems 2
1.3. Examples 4
2. Recurrence and ergodicity 5
2.1. Recurrence 5
2.2. Ergodicity 6
2.3. Mean ergodic theorem 7
2.4. Consequences of the ergodic theorem 8
3. Furstenberg's correspondence principle 9
3.1. Intersective sets are sets of recurrence 9
3.2. Furstenberg's correspondence principle 10
3.3. Szemerédi's theorem in arithmetic progressions 12
4. Polynomial recurrence 13
4.1. Uniform distribution 13
4.2. The van der Corput trick 14
4.3. Total ergodicity 15
4.4. Polynomial recurrence via the spectral theorem 17
4.5. Polynomial recurrence via L^2 decomposition 20
5. Mixing and eigenfunctions 21
5.1. Mixing and weak-mixing 21
5.2. Eigenfunctions and discrete spectrum 24
5.3. Compact functions 25
5.4. The Kronecker factor 26
5.5. The Jacobs-de Leeuw-Glicksberg decomposition 27
6. Special cases of Multiple recurrence 29
6.1. Multiple recurrence in weak mixing systems 29
6.2. Multiple recurrence in compact systems 31
6.3. An example of a skew-product 32
6.4. Roth's theorem 32
7. Disintegration of measures and the ergodic decomposition 34
7.1. Conditional expectation 34
7.2. Disintegration of measures 36
7.3. The ergodic decomposition 36
7.4. Proof of Theorem 7.3.1 38
7.5. Reducing multiple recurrence to ergodic systems 39
8. Proof of Furstenberg's Multiple Recurrence theorem 40
8.1. Outline of the proof 40
8.2. Weak-mixing and compact extensions 42
8.3. Sz lifts through weak mixing extensions 45
8.4. Sz lifts through compact extensions I 47

Date: July 21, 2020.

8.5. Sz lifts through compact extensions 2 49
8.6. The dichotomy between compactness and weak mixing 51
References 53

These are notes for a course in ergodic theory focused on applications to combinatorics and number theory. We will not follow any single textbook from beginning to end, but both Furstenberg's book [7] 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. A more advanced text on this subject is the recent book of Host and Kra [11]. For an introductory text to general ergodic theory, Walters [15] is an excellent source, which can be complemented with Glasner's [9] or Cornfeld-Fomin-Sinai's [2].

1. Measure preserving systems

The material in this section is contained, for instance, in Chapters 0 and 1 of [15] and in Chapter 2 of [4].

1.1. Probability spaces.

Definition 1.1.1.

- *A σ-algebra on a set X is a collection $\mathcal{B}$ of subsets of X closed under complements (i.e. $B \in \mathcal{B} \iff (X \setminus B) \in \mathcal{B}$), countable unions (i.e. if $B_1, B_2, \dots \in \mathcal{B}$ then $\bigcup_{n=1}^\infty B_n \in \mathcal{B}$) and such that $X \in \mathcal{B}$.*
- *A probability measure on a σ-algebra $\mathcal{B}$ (over some set X) is a function $\mu : \mathcal{B} \to [0,1]$ such that $\mu(X) = 1$ and whenever $B_1, B_2, \dots$ is a sequence of pairwise disjoint sets in $\mathcal{B}$,*
$$\mu\left(\bigcup_{n=1}^\infty B_n\right) = \sum_{n=1}^\infty \mu(B_n).$$

A **probability space** *is a triple $(X, \mathcal{B}, \mu)$ where X is a set, $\mathcal{B}$ is a σ-algebra on X and $\mu : \mathcal{B} \to [0,1]$ is a probability measure.*

Given two probability spaces $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$, a function $f : X \to Y$ is ***measurable*** if for every $C \in \mathcal{C}$, the pre-image $f^{-1}(C) := \{x \in X : f(x) \in C\}$ is in $\mathcal{B}$.

When the underlying set X is equipped with a topology it is natural to consider the Borel σ-algebra, i.e. the smallest σ-algebra containing all the open sets. In this case we often restrict attention to Radon probability measures, which we now define. Throughout these notes we will use the notation $C(X) := \{f : X \to \mathbb{C} : f \text{ is continuous}\}$. When X is compact, the space $C(X)$ is equipped with the topology arising from the supremum norm $\|f\| := \sup_{x\in X} |f(x)|$.

Definition 1.1.2. *Let X be a compact Hausdorff topological space and let $\mathcal{B}$ be the Borel σ-algebra. A probability measure $\mu : \mathcal{B} \to [0,1]$ is* **Radon** *if the linear functional $\lambda : C(X) \to \mathbb{C}$ defined as $\lambda(f) = \int_X f \, \mathsf{d}\mu$ is continuous.*

The Riesz-Markov-Kakutani representation theorem states that any positive linear functional $\lambda : C(X) \to \mathbb{C}$ (positive means that $\lambda(f) \geq 0$ whenever $f \geq 0$) gives rise to a unique Radon probability measure.

1.2. Measure preserving systems.

Definition 1.2.1 (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 1.2.2. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be probability spaces and let $T : X \to Y$ be a measurable map.*

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

- *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*

$$\int_X f \circ T \, \mathsf{d}\mu = \int_Y f \, \mathsf{d}\nu. \tag{1.1}$$

- *If both μ and ν are Radon measures, show that T preserves the measure if and only if (1.1) holds for every $f \in C(Y)$. [Hint: $C(Y)$ is dense in $L^2(Y)$.]*

Definition 1.2.3 (Koopman operator)**.** *Given a measure preserving transformation $T : X \to Y$ between the probability spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$, we call the linear operator $\Phi_T : L^2(Y) \to L^2(X)$ given by*

$$\Phi_T f = f \circ T$$

the associated **Koopman operator***.*

The fact that $f \circ T \in L^2(X)$ follows from the fact that T preserves the measure (cf. Exercise 1.2.2).

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

Definition 1.2.4 (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.*

The morphisms in the category of measure preserving systems are factor maps.

Definition 1.2.5 (Factor map)**.** *Let $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ be m.p.s. and let $\phi : X \to Y$. Then ϕ is a* **factor map** *if it is surjective, preserves the measure and intertwines T and S, in the sense that $S \circ \phi = \phi \circ T$.*

More generally, one can allow ϕ to be a surjective map between full measure sets $X_0 \in \mathcal{A}$ and $Y_0 \in \mathcal{B}$ such that $T^{-1}X_0 = X_0$ and $S^{-1}Y_0 = Y_0$, and the relation $S \circ \phi = \phi \circ T$ only needs to hold in X_0.

We say that the system $(Y, \mathcal{B}, \nu, S)$ is a ***factor*** of $(X, \mathcal{A}, \mu, T)$ if there is a factor map $\phi : X \to Y$. We will also say that, in this case, $(X, \mathcal{A}, \mu, T)$ is an ***extension*** of $(Y, \mathcal{B}, \nu, S)$.

Since in measure theory, sets with 0 measure are considered negligible, the definition of factor maps, and of isomorphism, needs to flexible enough so that in particular two systems which differ only on a 0 measure set should be isomorphic. There are several ways to formalize this, unfortunately not all equivalent.

Definition 1.2.6 (Isomorphism)**.** *Two measure preserving systems $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ are* **isomorphic** *if there exists a factor maps $\phi : X \to Y$ and $\psi : Y \to X$ such that $\phi \circ \psi$ and $\psi \circ \phi$ are almost everywhere the identity.*

For some purposes, the following weaker notion of when two measure preserving systems are "the same" is more convenient.

Definition 1.2.7 (Conjugacy)**.** *Two measure preserving systems $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ are* **conjugate** *if there exists a factor map $\phi : X \to Y$ whose Koopman operator $\Phi : L^2(Y) \to L^2(X)$ given by $\Phi f = f \circ \phi$ is an isomorphism of Hilbert spaces.*

Exercise 1.2.8. *Show that if two systems are isomorphic, then they are conjugate.*

The distinction between isomorphism and conjugacy is due only to the fact that the underlying probability spaces may not be isomorphic, even though the *dynamics* are. In fact in many cases there is no distinction:

Theorem 1.2.9 ([15, Theorems 2.5 and 2.6])**.** *Let $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ be measure preserving systems where both X and Y are separable complete metric spaces and both $\mathcal{A}$ and $\mathcal{B}$ are the Borel σ-algebras. Then the systems are conjugate if and only if they are isomorphic.*

Given a m.p.s. $(X, \mathcal{A}, \mu, T)$, any σ-subalgebra $\mathcal{B} \subset \mathcal{A}$ which is invariant in the sense that $T^{-1}B \in \mathcal{B}$ for every $B \in \mathcal{B}$ induces a factor of $(X, \mathcal{A}, \mu, T)$, namely the system $(X, \mathcal{B}, \mu, T)$ (by an abuse of notation we denote the restriction of μ to $\mathcal{B}$ also by μ). Conversely, ***any*** factor of $(X, \mathcal{A}, \mu, T)$ can be described in this form, up to conjugacy. Indeed, given a factor map $\phi : X \to Y$ between the m.p.s. $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{C}, S, \nu)$, one can consider the σ-algebra $\mathcal{B} := \{\phi^{-1}C : C \in \mathcal{C}\} \subset \mathcal{A}$. Then the quadruple $(X, \mathcal{B}, \mu, T)$ is a measure preserving system and it is conjugate (but not necessarily isomorphic) to $(Y, \mathcal{C}, S, \nu)$.

1.3. **Examples.** The following examples will illustrate the above definitions and will guide us for much of the course.

Example 1.3.1 (Trivial system)**.** *Let $X = \{0\}$ be a singleton, and let $\mathcal{A}$ and μ be the only σ-algebra and probability measure on X and let $T : X \to X$ be the identity map. Then $(X, \mathcal{A}, \mu, T)$ is a (rather trivial) measure preserving system, called the* **one point system***.*

Example 1.3.2 (Identity systems)**.** *Let $X = Y = [0, 1]$, let $\mathcal{A}$ be the Borel σ-algebra and let $\mathcal{B}$ be the trivial σ-algebra $\mathcal{B} = \{\emptyset, [0, 1]\}$. Let μ be the Lebesgue measure, let ν be the restriction of μ to $\mathcal{B}$ and let $T = S = Id$. Then both tuples $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ are measure preserving systems. Since the measure preserving transformation is the identity, both of these systems are called* **identity systems***.*

It is easy to see that $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ are **not** *isomorphic, and in fact not even conjugate. However, $(Y, \mathcal{B}, \nu, S)$ is conjugate (but not isomorphic) to the one point system.*

Example 1.3.3 (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 (almost everywhere) 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 1.3.4 (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 1.3.5. *Show that the doubling map does indeed preserve the Lebesgue measure. [Hint: use Exercise 1.2.2]*

More generally, given a matrix $A \in GL(n, \mathbb{Z})$ (or in fact any matrix with integer coefficients and non zero determinant) one can construct the measure preserving system $(X, \mathcal{B}, \mu, T)$ where $X = [0, 1]^n$ is the unit cube in n dimensions, $\mathcal{B}$ is the Borel σ-algebra, μ is the Lebesgue measure and $Tx = Ax \bmod \mathbb{Z}^n$. One way to show that T indeed preserves the measure is by identifying X with the n-dimensional torus $\mathbb{T}^n := \mathbb{R}^n/\mathbb{Z}^n$ (as probability spaces) in the obvious way, and then using Fourier analysis to reducing the problem to establishing (1.1) for characters.

Even more generally, if X is a compact abelian group, $\mathcal{B}$ is the Borel σ-algebra, μ is the Haar measure and $T : X \to X$ is a group automorphism, then $(X, \mathcal{B}, \mu, T)$ is a measure preserving system.

Example 1.3.6 (Bernoulli system)**.** *Let $X = \{0, 1\}^{\mathbb{N}}$ be the space of all (one-sided) infinite strings of 0's and 1's. Give $\{0, 1\}$ the discrete topology and let X be have the product topology[2]. In view of Tychonoff's theorem, X is compact.*

Let $\mathcal{B}$ be the Borel σ-algebra on X. Given $p \in (0, 1)$, let μ_0 be the measure on $\{0, 1\}$ given by $\mu_0(\{1\}) = p$ (and hence $\mu_0(\{0\}) = 1 - p$) and then let $\mu = \mu^{\mathbb{N}}$ be the product probability measure on X. Equivalently, μ

[2]This means that a set $U \subset X$ is open iff for every $x \in U$ there exists $n \in \mathbb{N}$ such that if $y \in X$ satisfies $y_i = x_i$ for all $i \leq n$, then $y \in U$.

can be described as the unique measure satisfying

$$\mu\Big(\big\{(x_n)_{n=1}^\infty \in X : x_1 = a_1, \ldots, x_m = a_m\big\}\Big) = \prod_{i=1}^m \mu_0\big(\{a_i\}\big)$$

for every $m \in \mathbb{N}$ and $a_1, \ldots, a_m \in \{0,1\}$.

Exercise 1.3.7. *Show that for $p = 1/2$ the Bernoulli system is isomorphic to the doubling map. [Hint: Write a number $x \in (0,1)$ in its binary expansion as an infinite string of 0's and 1's.]*

More generally, Bernoulli systems can be defined over any finite set $\mathscr{A}$, usually called the alphabet. A measure preserving system is called a ***Bernoulli system*** if it is isomorphic to a system of the form $(X, \mathcal{B}, \mu, T)$ where $X = \mathscr{A}^{\mathbb{N}}$, $\mathcal{B}$ is the σ-algebra of Borel sets, T is the left shift and $\mu = \mu_0^{\mathbb{N}}$ is the product measure of some arbitrary probability measure μ_0 on $\mathscr{A}$.

If T and S are two measure preserving transformations on the same probability space, then so is their composition $T \circ S$. For instance the map $x \mapsto 2x + \alpha \bmod 1$ is a measure preserving transformation on the interval $[0,1]$ (equipped with Lebesgue measure). Another way to construct new measure preserving systems out of given ones is by taking products.

Given functions $f : X \to \mathbb{C}$ and $g : Y \to \mathbb{C}$ we can form the function $f \otimes g : X \times Y \to \mathbb{C}$ defined by $(f \otimes g)(x,y) = f(x)g(y)$.

Exercise 1.3.8. *Let X and Y be compact Hausdorff spaces. Then the subspace of $C(X \times Y)$ consisting of functions of the form $f \otimes g$ with $f \in C(X)$ and $g \in C(Y)$ is dense. [Hint: Use the Stone-Weierstrass theorem]*

Definition 1.3.9 (Product system)**.** *Given two measure preserving systems $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$, we define their* **product** *to be the m.p.s. $(Z, \mathcal{C}, \lambda, R)$, where $Z = X \times Y$; $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$ is the defined as the smallest σ-algebra on Z containing all the rectangles $A \times B$ for $A \in \mathcal{A}$ and $B \in \mathcal{B}$; $\lambda = \mu \otimes \nu$ is the unique measure on $(Z, \mathcal{C})$ which satisfies $\lambda(A \times B) = \mu(A)\nu(B)$ for every $A \in \mathcal{A}$ and $B \in \mathcal{B}$; and $R : Z \to Z$ is the map $R(x,y) = (Tx, Sy)$.*

Exercise 1.3.10.

- *Show that the product of two group rotations is also a group rotation.*
- *Show that the product of two Bernoulli systems is also a Bernoulli system.*

Example 1.3.11 (Skew-product)**.** *Let $X = [0,1]^2$, let $\mathcal{A}$ be the Borel σ-algebra and let μ be the Lebesgue measure. Fix $\alpha \in \mathbb{R}$ and let $T : X \to X$ be the map $T(x,y) = (x + \alpha \bmod 1, y + x \bmod 1)$. Then $(X, \mathcal{A}, \mu, T)$ is a measure preserving system called a* **skew-product***.*

To see why T preserves the measure, observe that in view of Exercises 1.2.2 and 1.3.8 it suffices to check that for any $f, g \in C([0,1])$

$$\int_0^1 \int_0^1 f(x + \alpha \bmod 1) g(x + y \bmod 1) \,\mathsf{d}y \,\mathsf{d}x = \int_0^1 \int_0^1 f(x) g(y) \,\mathsf{d}y \,\mathsf{d}x,$$

which can be directly established.

Exercise 1.3.12. *Let $\pi : [0,1]^2 \to [0,1]$ be the projection onto the first coordinate. Show that π is a factor map between the skew product $(X, \mathcal{A}, \mu, T)$ described in Example 1.3.11 and the circle rotation by α, described in Example 1.3.3.*

Identify the σ-subalgebra $\mathcal{B}$ of $\mathcal{A}$ which turns $(X, \mathcal{B}, \mu, T)$ into a system conjugate to the circle rotation. Are they isomorphic?

2. Recurrence and ergodicity

2.1. Recurrence.

Here is the first theorem of ergodic theory.

Theorem 2.1.1 (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*

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

Proof. The sets $A, T^{-1}A, T^{-2}A, \dots$ 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

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

□

Corollary 2.1.2. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and let $A \in \mathcal{B}$. Then for μ-a.e. $x \in A$ there exists $n \in \mathbb{N}$ such that $T^n x \in A$, i.e. x returns to A at time n.*

Proof. Let $B := \{x \in A : T^n x \notin A \text{ for all } n \in \mathbb{N}\}$; we need to show that $\mu(B) = 0$. If $\mu(B) > 0$, then by Theorem 2.1.1 one can find $m \in \mathbb{N}$ such that $B \cap T^{-m}B$ has positive measure and, in particular, is non-empty. But if $y \in B \cap T^{-m}B$, then $T^m y \in B \subset A$, contradicting the fact that $y \in B$. This contradiction implies that $\mu(B) = 0$. □

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, how small can we choose n? How large is the set of n for which (2.1) 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 2.1.3. *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 2.1.4. *Show that the set $2\mathbb{N}$ of even numbers is a set of recurrence. [Hint: Consider the m.p.s. $(X, \mathcal{B}, \mu, T^2)$.]*

Exercise 2.1.5. *Show that the set $2\mathbb{N} - 1$ of odd numbers is not a set of recurrence. [Hint: Use a m.p.s. with 2 points.]*

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

Theorem 2.1.6. *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, in Theorem 2.1.6, one can replace the condition that n is a perfect square with the condition that $n + 1$ is a prime number!

Exercise 2.1.7. *Show that Theorem 2.1.6 does not hold if the condition that n is a perfect square gets replaced with the condition that n is a prime number. [Hint: Consider a measure preserving system where X has only 4 points.]*

Exercise 2.1.8. *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.*

In order to prove Theorem 2.1.6 and to answer the other questions raised above, we need to build up some more ergodic theory.

2.2. **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*[3]. 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

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

$i = 1, 2$. 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 2.2.1. *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 2.2.2. *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 asserts, 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 2.2.3 (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 $f \in L^2$*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f \circ T^n = P_I f \qquad a.e.. \tag{2.2}$$

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.

Corollary 2.2.4. *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system. Then for every $A \in \mathcal{B}$ and almost every $x \in X$,*

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

Proof. Apply Theorem 2.2.3 to the indicator function 1_A of A and observe that, for each $x \in X$,

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

□

2.3. **Mean ergodic theorem.** 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 $\{1, \ldots, N\}$ to any sequence of intervals $\{a_N, a_N + 1, \ldots, a_N + N\}$. Moreover, the simpler proof of von Neumann's theorem can be easily modified to apply to measure preserving actions of any amenable group.

Theorem 2.3.1 (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 $f \in L^2$*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} f \circ T^n = P_I f \qquad in\ L^2(X). \tag{2.3}$$

Remark 2.3.2.
$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N}h_n=c$$
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 2.3.3. *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* $I_N=\{a_N+1,a_N+2,\ldots,a_N+b_N\}$ *whose lengths* b_N *tend to infinity.*

Recall that 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 2.3.1 is a corollary of the following.

Theorem 2.3.4 (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* $f\in H$*,*
$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N}\Phi^n f=Pf\qquad \textit{in norm}\tag{2.4}$$

Proof. If $f\in I$ then (2.4) 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 (2.4) 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, let $f\perp J$, 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. □

2.4. **Consequences of the ergodic theorem.**

Corollary 2.4.1. *A measure preserving system* $(X,\mathcal{B},\mu,T)$ *is ergodic if and only if for every* $A,B\in\mathcal{B}$*,*
$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N}\mu(T^{-n}A\cap B)=\mu(A)\mu(B).\tag{2.5}$$

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 (2.5).

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(2.3) 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 (2.5) follows from the fact that $\int_X \mu(A) g\,\mathsf{d}\mu = \mu(A)\mu(B)$. □

Setting $A = B$ in Corollary 2.4.1 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 2.4.2. *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 2.4.3. *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 2.4.4 (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*

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

is syndetic.

Proof. Applying Theorem 2.3.1 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

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

□

3. Furstenberg's correspondence principle

3.1. **Intersective sets are sets of recurrence.** There is no good way to put a probability measure on $\mathbb{N}$, since it is a countable set and measures are countably additive. The upper density is one alternative way to describe how large a subset of $\mathbb{N}$ is

Definition 3.1.1 (Upper density)**.** *Given a set $A \subset \mathbb{N}$, its* **upper density** *is the quantity*

$$\bar{d}(A) := \limsup_{N\to\infty} \frac{\big|A \cap [1, N]\big|}{N} \tag{3.1}$$

When the limit in (3.1) exists we denote it by $d(A)$ and call it simply the **density** *of A.*

It is clear that both the set of even numbers and the set of odd numbers have density $1/2$. More generally, for every $a, b \in \mathbb{N}$, we have $\bar{d}(a\mathbb{N}+b) = 1/a$. But one can also compute the upper density of some less regular sets; for instance, it can be shown that the set of squarefree numbers has density $6/\pi^2$.

Given sets $A, B \subset \mathbb{N}$ and $n \in \mathbb{N}$ we define $A - n := \{m : m + n \in A\}$ and $A - B := \{a - b : a \in A, b \in B, a > b\}$.

Exercise 3.1.2. *Show that for every $A, B \in \mathbb{N}$ and $n \in \mathbb{N}$*

- *$\bar{d}(A \cup B) \leq \bar{d}(A) + \bar{d}(B)$.*
- *$\bar{d}(A - n) = \bar{d}(A)$.*

Definition 3.1.3. *A set $I \subset \mathbb{N}$ is called an* **intersective set** *if it has nonempty intersection with every set of the form $A - A$ where $A \subset \mathbb{N}$ satisfies $\bar{d}(A) > 0$.*

Our first connection between combinatorics and ergodic theory is the fact that a set $I \subset \mathbb{N}$ is intersective if and only if it is a set of recurrence. In the rest of this subsection we prove one of the implications, the proof of the other requires the Furstenberg correspondence principle and is left to the next subsection.

We start with a lemma by Bergelson.

Lemma 3.1.4 (Intersectivity lemma). *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\epsilon > 0$ and let $A_1, A_2, \ldots$ be a sequence in $\mathcal{B}$ with $\mu(A_n) \geq \epsilon$ for every $n \in \mathbb{N}$. Then there exists a set $L \subset \mathbb{N}$ with $\bar{d}(L) \geq \epsilon$ and such that for every finite set $F \subset L$,*

$$\mu\left(\bigcap_{n\in F} A_n\right) > 0.$$

Proof. For each $x \in X$ consider the set $L_x := \{n \in \mathbb{N} : x \in A_n\}$. We have

$$\bar{d}(L_x) = \limsup_{N\to\infty} \frac{\big|L_x \cap [1,N]\big|}{N} = \limsup_{N\to\infty} \frac{\big|\{n \in [1,N] : x \in A_n\}\big|}{N} = \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_{A_n}(x).$$

Since the function 1_{A_n} is measurable, it follows that the function $x \mapsto \bar{d}(L_x)$ is also measurable. From Fatou's lemma we obtain

$$\int_X \bar{d}(L_x)\,\mathsf{d}\mu(x) \geq \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \int_X 1_{A_n}\,\mathsf{d}\mu = \limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu(A_n) \geq \epsilon.$$

Therefore the set $B := \{x \in X : \bar{d}(L_x) \geq \epsilon\}$ has positive measure.

Notice that for every $x \in B$ and every finite subset $F \subset L_x$ we have $x \in \bigcap_{n\in F} A_n$. But we want to ensure that this intersection has positive measure. To this end, consider the collection

$$\mathcal{F} := \left\{F \subset \mathbb{N} : F \text{ is finite and } \mu\left(\bigcap_{n\in F} A_n\right) = 0\right\}.$$

This is a countable collection and therefore the set

$$C := \bigcup_{F\in\mathcal{F}} \bigcap_{n\in F} A_n$$

satisfies $\mu(C) = 0$. Finally, taking $x \in B \setminus C$ we conclude that any intersection of the form $\bigcap_{n\in F} A_n$ containing x must have positive measure, and hence L_x satisfies the desired property. $\square$

Corollary 3.1.5. *Let $I \subset \mathbb{N}$. If I is intersective, then it is a set of recurrence.*

Proof. Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and let $A \in \mathcal{B}$ with $\mu(A) > 0$. To show that I is a set of recurrence, we need to find $n \in I$ such that $\mu(A \cap T^{-n}A) > 0$. Apply Lemma 3.1.4 to the sequence $A, T^{-1}A, T^{-2}A, \ldots$ and let $L \subset \mathbb{N}$ with $\bar{d}(L) \geq \mu(A)$ be the set given by that lemma.

Since I is intersective, there exists $n \in I$ and $a, b \in L$ such that $a - b = n$. It follows from the conclusion of Lemma 3.1.4 that $\mu(T^{-a}A \cap T^{-b}A) > 0$, and since $T^{-a}A \cap T^{-b}A = T^{-b}(T^{-n}A \cap A)$ we conclude that indeed $\mu(T^{-n}A \cap A) > 0$ as desired. $\square$

3.2. **Furstenberg's correspondence principle.** Let $S : \mathbb{N} \to \mathbb{N}$ be the map $S : n \mapsto n + 1$. It follows from Exercise 3.1.2 that $\bar{d}(S^{-1}A) = \bar{d}(A)$. In this sense the quadruple $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \bar{d}, S)$ behaves like a measure preserving system, but of course this is not the case since $\bar{d}$ is not a measure. The idea behind the Furstenberg correspondence principle is to make this idea precise in a certain sense.

Furstenberg introduced the correspondence principle in 1977 paper [6] giving an ergodic theoretic proof of Szemerédi's theorem in arithmetic progressions (which is discussed in the next subsection). Since then it has become the fundamental tool in a new area of mathematics, called ***ergodic Ramsey theory***, which analyses problems in combinatorics by transferring them to ergodic theoretic world, and then proving the resulting ergodic theoretic statement.

Lemma 3.2.1 (Furstenberg correspondence principle)**.** *Given $E \subset \mathbb{N}$ there exists a measure preserving system $(X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{B}$ with $\mu(A) = \bar{d}(E)$ such that for every finite $F \subset \mathbb{N}$,*

$$\mu\left(\bigcap_{n\in F} T^{-n}A\right) \leq \bar{d}\left(\bigcap_{n\in F} S^{-n}E\right) \tag{3.2}$$

Proof. Denote by $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ the set on non-negative integers. Let $X = \{0, 1\}^{\mathbb{N}_0}$ and let $\mathcal{B}$ be the Borel σ-algebra. Let $T : X \to X$ be the shift map, so that $T : (x_n)_{n=0}^{\infty} \mapsto (x_{n+1})_{n=0}^{\infty}$. Let A denote the set of all sequences $(x_n)_{n\in\mathbb{N}_0}$ with $x_0 = 1$.

All that remains to do construct is the measure. First find a sequence (N_j) such that $\bar{d}(E) = \lim_{j\to\infty} \frac{|E\cap[1,N_j]|}{N_j}$. Let $\omega = (\omega_n)_{n=0}^{\infty} \in X$ be the indicator function of E, so that $\omega_n = 1$ iff $n \in E$. For each $j \in \mathbb{N}$ let μ_j be the probability measure on $(X, \mathcal{B})$ defined by

$$\mu_j = \frac{1}{N_j}\sum_{n=1}^{N_j} \delta_{T^n\omega},$$

where δ_y is the Dirac point mass at the point y, i.e. δ_y is a probability measure on $(X, \mathcal{B})$ satisfying $\delta_y(B) = 1$ if $y \in B$ and $\delta_y(B) = 0$ otherwise. Observe that

$$\bar{d}(E) = \lim_{j\to\infty} \frac{|E \cap [1, N_j]|}{N_j} = \lim_{j\to\infty} \frac{1}{N_j}\sum_{n=1}^{N_j} \omega_n = \lim_{j\to\infty} \mu_j(A)$$

Finally, take a weak* limit point μ of the sequence $(\mu_j)_{j=1}^{\infty}$, observing that it takes values in the weak* compact set of probability measures on $(X, \mathcal{B})$. In particular $\mu(A) = \bar{d}(E)$.

To show that T preserves the measure μ, recall that a ***cylinder set*** in X is a set of the form $\{(x_n)_{n=0}^{\infty} \in X : x_0 = a_0, \dots, x_n = a_n\}$ for some $n \in \mathbb{N}$ and $a_0, \dots, a_n \in \{0, 1\}$. In view of the Stone-Weierstrass theorem, finite linear combinations of indicator functions of cylinder sets form a dense subset of $C(X)$, so it suffices to show that $\mu(T^{-1}B) = \mu(B)$ for every cylinder set B. But one can directly compute that $\left|\mu_j(B) - \mu_j(T^{-1}B)\right| < 2/N_j$ when B is a cylinder and this proves that $(X, \mathcal{B}, \mu, T)$ is a measure preserving system.

It is left to establish (3.2). For $n \in \mathbb{N}$ we have that $T^{-n}A = \{x \in X : x_n = 1\}$. Therefore for a finite set $F \subset \mathbb{N}$ and $j \in \mathbb{N}$ we have

$$\begin{aligned}\mu_j\left(\bigcap_{n\in F} T^{-n}A\right) &= \frac{1}{N_j}\Big|\big\{m \in [1, N_j] : (T^m\omega)_n = 1\ \forall n \in F\big\}\Big| = \frac{1}{N_j}\Big|\big\{m \in [1, N_j] : n + m \in E\ \forall n \in F\big\}\Big| \\ &= \frac{1}{N_j}\left|\left\{m \in [1, N_j] : m \in \bigcap_{n\in F}(E - n)\right\}\right| = \frac{1}{N_j}\left|\bigcap_{n\in F} S^{-n}E \cap [1, N_j]\right|\end{aligned}$$

Therefore, given a finite set $F \subset \mathbb{N}$,

$$\bar{d}\left(\bigcap_{n\in F} S^{-n}E\right) \geq \lim_{j\to\infty} \frac{1}{N_j}\left|\bigcap_{n\in F} S^{-n}E \cap [1, N_j]\right| = \lim_{j\to\infty} \mu_j\left(\bigcap_{n\in F} T^{-n}A\right) = \mu\left(\bigcap_{n\in F} T^{-n}A\right).$$

$\square$

The correspondence principle allows one to convert many ergodic theoretic results into combinatorial statements. For instance, by combining the correspondence principle with Khintchine's recurrence theorem (Theorem 2.4.4 above) we deduce the following.

Corollary 3.2.2. *If $E \subset \mathbb{N}$ has positive upper density, then $E - E$ is syndetic.*

Proof. Apply the correspondence principle to find a measure preserving system $(X, \mathcal{B}, \mu, T)$ and a set $A \subset X$ with $\mu(A) = \bar{d}(E)$ and satisfying (3.2). Then apply Theorem 2.4.4 with any $\epsilon < \mu^2(A)$ to conclude that the set $S := \{n \in \mathbb{N} : \mu(A \cap T^{-n}A) > 0\}$ is syndetic. In view of (3.2) it follows that for every $n \in S$, $\bar{d}(E \cap S^{-n}E) > 0$. Letting x be any number in the intersection $E \cap S^{-n}E$ it follows that $x \in E$ and that $x + n \in E$; therefore $n \in E - E$. We conclude that $S \subset E - E$ and hence $E - E$ is syndetic. $\square$

We now conclude the proof of the fact that a set of natural numbers is intersective if and only if it is a set of recurrence.

Theorem 3.2.3. *Let $R \subset \mathbb{N}$. Then R is a set of recurrence if and only if it is an intersective set.*

Proof. One direction was established in Corollary 3.1.5. To prove the other direction, suppose that R is a set of recurrence and let $E \subset \mathbb{N}$ satisfy $\bar{d}(E) > 0$. We need to show that $(E - E) \cap R \neq \emptyset$. Applying the correspondence principle we find a measure preserving system $(X, \mathcal{B}, \mu, T)$ and a set $A \subset X$ with $\mu(A) = \bar{d}(E)$ and satisfying (3.2). Since R is a set of recurrence there exists $n \in R$ such that $\mu(A \cap T^{-n}A) > 0$. In view of (3.2) it follows that $\bar{d}(E \cap S^{-n}E) > 0$. Letting x be any number in the intersection $E \cap S^{-n}E$ it follows that $x \in E$ and that $x + n \in E$; therefore $n \in E - E$. $\square$

3.3. **Szemerédi's theorem in arithmetic progressions.** A (finite) arithmetic progression is a set of the form $\{a, a+b, a+2b, \ldots, a+kb\}$ for some $a, b, k \in \mathbb{N}$. The ***length*** of an arithmetic progression is its cardinality. The study of arithmetic progressions as a combinatorial entity began with van der Waerden's theorem.

Theorem 3.3.1 (van der Waerden)**.** *If $\mathbb{N}$ is partitioned into finitely many sets $\mathbb{N} = C_1 \cup \cdots \cup C_r$, then one of the C_i contains arbitrarily long arithmetic progressions.*

Van der Waerden's theorem is one of the oldest results in the area now called Ramsey theory, preceding Ramsey's theorem itself by a decade. In 1936 Erdős and Turán conjectured that in fact every set with positive upper density contains arbitrarily long arithmetic progressions, which is a strengthening of van der Waerden's theorem. The conjecture was solved 40 years later by Szemerédi.

Theorem 3.3.2 (Szemerédi)**.** *Every $E \subset \mathbb{N}$ with $\bar{d}(E) > 0$ contains arbitrarily long arithmetic progressions.*

Szemerédi's theorem and its several proofs have a truly remarkable history, which this paragraph is too small to contain. In these notes we are mainly concerned with the ergodic theoretic proof of Furstenberg. To see how Szemerédi's theorem can be interpreted in ergodic theoretic language, note that

$$\{a, a+b, a+2b, \ldots, a+kb\} \subset E \iff a \in E \cap S^{-b}E \cap S^{-2b}E \cap \cdots \cap S^{-kb}E$$

Thus Szemerédi's theorem can be rephrased as: if $E \subset \mathbb{N}$ has $\bar{d}(E) > 0$ then for every $k \in \mathbb{N}$ there exists $b \in \mathbb{N}$ such that $E \cap S^{-b}E \cap S^{-2b}E \cap \cdots \cap S^{-kb}E \neq \emptyset$. Furstenberg proved a strengthening of this statement.

Theorem 3.3.3. *Let $E \subset \mathbb{N}$ with $\bar{d}(E) > 0$ and let $k \in \mathbb{N}$. Then*

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \bar{d}\big(E \cap S^{-n}E \cap S^{-2n}E \cap \cdots \cap S^{-kn}E\big) > 0.$$

Furstenberg's result in fact uses uniform Cesáro averages. In view of the correspondence principle, Theorem 3.3.3 (and hence Szemerédi's theorem) follows from the following multiple recurrence theorem.

Theorem 3.3.4 (Furstenberg's multiple recurrence theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then for every $k \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that*

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

In fact

$$\lim_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^{N} \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > 0. \tag{3.3}$$

Remark 3.3.5. *In [6], Furstenberg didn't show that the limit in (3.3) exists, he showed only that the* lim inf *is positive. This is enough information to recover Szemerédi's theorem. The existence of the limit remained an open problem until 2005, when it was established by Host and Kra in [10].*

As a corollary of Theorem 3.3.4 we obtain the following strengthening of Szemerédi's theorem.

Corollary 3.3.6. *Let $E \subset \mathbb{N}$ have $\bar{d}(E) > 0$ and let $k \in \mathbb{N}$. Then the set of common differences of arithmetic progressions of length $k+1$ contained in E, i.e., the set*

$$\big\{n \in \mathbb{N} : \exists a \in \mathbb{N} \text{ s.t. } \{a, a+n, a+2n, \ldots, a+kn\} \subset E\big\}$$

is syndetic.

Proof. Using the correspondence principle (Lemma 3.2.1) we find a m.p.s. $(X,\mathcal{B},\mu,T)$ and a set $A\in\mathcal{B}$ with $\mu(A)=\bar{d}(E)$ and satisfying (3.2). Apply the multiple recurrence theorem (Theorem 3.3.4) to A. Combining (3.3) and (3.2) it follows that if there exists $L\in\mathbb{N}$ such that whenever $N-M>L$ one has

$$\frac{1}{N-M}\sum_{n=M}^{N}\bar{d}\big(E\cap S^{-n}E\cap\cdots\cap S^{-kn}E\big)>0.$$

Therefore the set $S:=\{n:\bar{d}\big(E\cap S^{-n}E\cap\cdots\cap S^{-kn}E\big)>0\}$ is syndetic, and for every $n\in S$ and any $a\in E\cap S^{-n}E\cap\cdots\cap S^{-kn}E$ we have $\{a,a+n,a+2n,\dots,a+kn\}\subset E$. □

The proof of Theorem 3.3.4 will require a deeper understanding of the structure of measure preserving systems, which we start analyzing in the next section.

4. Polynomial recurrence

In this section we prove Theorem 2.1.6. While this result is more sophisticated than the recurrence results already established, it is still fairly simple to prove. The main purpose of this section is then to introduce some of the tools and ideas that will appear again later, with a simple application as motivation.

One of these tools, and a fundamental lemma in ergodic Ramsey theory, is the van der Corput lemma, sometimes called the van der Corput trick. My survey with Bergelson [1] explores the many directions in which one can make use of this trick. In order to properly talk about the van der Corput lemma we need some basic notions in the theory of uniform distribution.

4.1. **Uniform distribution.**

The original lemma due to van der Corput [3] is concerned with uniform distribution in the unit interval.

Definition 4.1.1. *A sequence $(x_n)_{n=1}^\infty$ taking values in $[0,1]$ is said to be* **uniformly distributed** *or* **equidistributed** *if for every interval $(a,b)\subset[0,1]$,*

$$\lim_{N\to\infty}\frac{1}{N}\Big|\big\{n\in[1,N]:x_n\in(a,b)\big\}\Big|=b-a. \tag{4.1}$$

Due to the fact that there are uncountably many intervals (a,b) inside $[0,1]$, it is not clear that uniformly distributed sequences even exist. However, we have the following criterion by Weil [16] (for a proof, see [12, Theorems 1.1.1 and 1.2.1]).

Lemma 4.1.2 (Weyl criterion)**.** *Let $(x_n)_{n=1}^\infty$ be a sequence taking values in $[0,1]$. The following are equivalent.*

(1) *$(x_n)_{n=1}^\infty$ is uniformly distributed.*
(2) *For every continuous function $f\in C[0,1]$,*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}f(x_n)=\int_0^1 f(t)\,\mathsf{d}t.$$

(3)

$$\forall\ h\in\mathbb{N}\qquad\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i h x_n}=0. \tag{4.2}$$

Example 4.1.3. *Let $\alpha\in\mathbb{R}\setminus\mathbb{Q}$. Then the sequence $x_n=n\alpha \bmod 1$ is uniformly distributed. Indeed, for every $h,N\in\mathbb{N}$ we have*

$$\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i h x_n}=\frac{1}{N}\sum_{n=1}^{N}\left(e^{2\pi i h\alpha}\right)^n=\frac{1}{N}\cdot\frac{e^{2\pi i h\alpha(N+1)}-e^{2\pi i h\alpha}}{e^{2\pi i h\alpha}-1}$$

and the last expression converges to 0 as $N\to\infty$.

Exercise 4.1.4. *Show that the sequence $x_n=\sqrt{n}\bmod 1$ is uniformly distributed.*

Exercise 4.1.5. *Show that the sequence $x_n=\log n\bmod 1$ is* **not** *uniformly distributed.*

Here is the original version of the van der Corput trick.

Lemma 4.1.6. *Let $(x_n)_{n=1}^\infty$ be a sequence taking values in $\mathbb{R}$. If for every $m \in \mathbb{N}$ the sequence $n \mapsto x_{n+m} - x_n \bmod 1$ is uniformly distributed, then also the sequence $n \mapsto x_n \bmod 1$ is uniformly distributed.*

We will prove a more general result below. As a corollary of Lemma 4.1.6 we obtain Weyl's equidistribution theorem.

Corollary 4.1.7. *Let $f \in \mathbb{R}[t]$ be a polynomial with real coefficients. If at least one of the coefficients of f, other than the constant term, is irrational, then $f(n) \bmod 1$ is uniformly distributed.*

Proof. We proceed by induction on the degree $d = d(f)$ of the largest degree term of f with an irrational coefficient. If $d = 1$, then the sequence $f(n) \bmod 1$ is the sum of a periodic sequence (say of period p) and the sequence $n \mapsto n\alpha \bmod 1$ where α is the irrational coefficient of degree 1. Since $p\alpha$ is still irrational, one can adapt the argument in Example 4.1.3 to show that $f(n) \bmod 1$ is indeed uniformly distributed when $d = 1$.

Next suppose that $d > 1$. For each $m \in \mathbb{N}$, the sequence $g_m : n \mapsto f(n+m) - f(n)$ is itself a polynomial with $d(g_m) = d(f) - 1$ by induction, $g_m \bmod 1$ is uniformly distributed, and in view of Lemma 4.1.6, so is $f(n) \bmod 1$. □

4.2. **The van der Corput trick.** For the purposes of ergodic Ramsey theory, the most useful version of the van der Corput trick deals with sequences of vectors in a Hilbert space.

Lemma 4.2.1. *Let H be a Hilbert space and let $(x_n)_{n=1}^\infty$ be a bounded sequence taking values in H. If for every $d \in \mathbb{N}$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle x_{n+d}, x_n \rangle = 0 \tag{4.3}$$

then

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

There is also a version for uniform Cesàro averages, which can be proved in the same way. Before proving Lemma 4.2.1, let's see how it implies van der Corput's original lemma

Proof of Lemma 4.1.6. Let $(x_n)_{n=1}^\infty$ be a sequence taking values in $\mathbb{R}$ such that for every $m \in \mathbb{N}$ the sequence $n \mapsto x_{n+m} - x_n \bmod 1$ is uniformly distributed. Using the Weyl's criterion in the form of (4.2), and using the notation $e(x)$ to denote $e^{2\pi i x}$, it follows that for every $m, h \in \mathbb{N}$, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e\big(h(x_{n+m} - x_n)\big) = 0$.

Let $H = \mathbb{C}$ be the one dimensional Hilbert space, with the usual inner product $\langle z, w\rangle := z\bar{w}$. Fix $h \in \mathbb{N}$. Letting $u_n = e(hx_n)$ we deduce that $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle u_{n+m}, u_n\rangle = 0$. Using Lemma 4.2.1 it follows that $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} u_n = 0$. In other words, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e(hx_n) = 0$, and since $h \in \mathbb{N}$ was arbitrary it follows from (4.2) that $x_n \bmod 1$ is indeed uniformly distributed. □

We will sometimes need a version of the van der Corput trick slightly stronger than Lemma 4.2.1. For completeness, we state simultaneously a version for regular Cesàro averages, as well as a version for uniform Cesàro averages.

Lemma 4.2.2. *Let H be a Hilbert space and let $(x_n)_{n=1}^\infty$ be a bounded sequence taking values in H. If*

$$\lim_{D\to\infty} \frac{1}{D} \sum_{d=1}^{D} \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=1}^{N} \langle x_{n+d}, x_n \rangle \right| = 0 \qquad \left(\textit{or } \lim_{D\to\infty} \frac{1}{D} \sum_{d=1}^{D} \limsup_{N-M\to\infty} \left| \frac{1}{N-M} \sum_{n=M}^{N} \langle x_{n+d}, x_n \rangle \right| = 0 \right)$$

then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n = 0 \qquad \left(\textit{resp. } \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N} x_n = 0 \right)$$

We now prove Lemma 4.2.1. The proof can be adapted to yield the stronger form of Lemma 4.2.2, but since the adaptation is mostly of a technical nature, we will omit it from this notes. The interested reader can find a very complete proof of Lemma 4.2.2 in [4, Theorem 7.11].

Proof of Lemma 4.2.1. For any $\epsilon > 0$ and any $D \in \mathbb{N}$, if $N \in \mathbb{N}$ is large enough we have

$$\left\| \frac{1}{N} \sum_{n=1}^{N} x_n - \frac{1}{D} \sum_{d=1}^{D} \frac{1}{N} \sum_{n=1}^{N} x_{n+d} \right\| < \frac{\epsilon}{2}$$

Hence it suffices to show that, if D is large enough,

$$\limsup_{N\to\infty} \left\| \frac{1}{D} \sum_{d=1}^{D} \frac{1}{N} \sum_{n=1}^{N} x_{n+d} \right\| < \frac{\epsilon}{2}$$

Using the Cauchy-Schwarz inequality we have

$$\begin{aligned} \limsup_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} \frac{1}{D} \sum_{d=1}^{D} x_{n+d} \right\|^2 &\leq \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left\| \frac{1}{D} \sum_{d=1}^{D} x_{n+d} \right\|^2 \\ &= \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \frac{1}{D^2} \sum_{d_1,d_2=1}^{D} \langle x_{n+d_1}, x_{n+d_2} \rangle \\ &\leq \frac{1}{D^2} \sum_{d_1,d_2=1}^{D} \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \langle x_{n+d_1}, x_{n+d_2} \rangle \qquad (4.4) \end{aligned}$$

Note that, for $d_1 \neq d_2$, it follows from (4.3) that $\frac{1}{N}\sum_{n=1}^{N} \langle x_{n+d_1}, x_{n+d_2}\rangle \to 0$ as $N \to \infty$. We conclude that the quantity in (4.4) is bounded by $\frac{D}{D^2} = \frac{1}{D}$ which is arbitrarily small for large enough D. $\square$

4.3. **Total ergodicity.**

Definition 4.3.1. *A measure preserving system $(X, \mathcal{B}, \mu, T)$ is* **totally ergodic** *if for every $n \in \mathbb{N}$, the measure preserving system $(X, \mathcal{B}, \mu, T^n)$ is ergodic.*

A convenient notation we will often use from now on is the following: given a m.p.s. $(X, \mathcal{B}, \mu, T)$ and a function $f \in L^2(X)$, we denote by Tf the composition $f \circ T$ (another way to think about this is, as an abuse of language, to denote by T the associated Koopman operator).

Example 4.3.2. *Recall the circle rotation $(X, \mathcal{B}, \mu, T)$ described in Example 1.3.3, where $X = [0, 1]$, $\mathcal{B}$ is the Borel σ-algebra, μ is the Lebesgue measure and $T : x \mapsto x + \alpha \bmod 1$. This system is totally ergodic if and only if α is irrational. Indeed, if α is rational, say $\alpha = p/q$, then $q\alpha$ is an integer and hence T^q is the identity map on $[0, 1]$, which is trivially not ergodic.*

On the other hand, if α is irrational, then the system is ergodic. To see this we use the ergodic theorem. Then we need to show that for every $f \in L^2$ the average

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f$$

is a constant function. But this is easy to check for functions $t \mapsto e(nt)$ with $n \in \mathbb{Z}$, and finite linear combinations of functions of this kind form a dense subset of L^2.

Finally, for every $n \in \mathbb{N}$, the measure preserving system $(X, \mathcal{B}, \mu, T^n)$ is the circle rotation by $n\alpha$; since $n\alpha$ is also irrational when α is, the system $(X, \mathcal{B}, \mu, T)$ is totally ergodic in this case.

Example 4.3.3. *Let $X = \{0, 1\}$, $\mathcal{B}$ the discrete σ-algebra, μ the normalized counting measure and $T : x \mapsto x + 1 \bmod 2$. In other words $(X, \mathcal{B}, \mu, T)$ is a transposition on 2 points. Then this system is ergodic, since the only sets with measure in $(0, 1)$ are the singletons $\{0\}$ and $\{1\}$, and neither of them is invariant. However, the system is not totally ergodic, since T^2 is the identity map and leaves both singletons (which have positive measure) invariant.*

While Example 4.3.3 seems rather trivial, it turns out that finite systems are in some sense the only obstruction to total ergodicity:

Theorem 4.3.4. *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic system. Then it is totally ergodic if and only if it does not contain any non-trivial finite factor.*

Proof. Let $(Y, \mathcal{B}, \nu, S)$ be a non-trivial finite system and suppose that there is a factor map $\pi : X \to Y$. Let $y \in Y$ be such that $\nu(\{y\}) \in (0,1)$ and let $A = \pi^{-1}(\{y\})$. Then $\mu(A) = \nu(\{y\}) \in (0,1)$. Let $n = |Y|!$. Then S^n acts trivially on Y, and in particular $S^{-n}\{y\} = \{y\}$. Therefore $T^{-n}A = A$ and we conclude that $(X, \mathcal{A}, \mu, T^n)$ is not ergodic.

To prove the converse direction, suppose that $(X, \mathcal{A}, \mu, T)$ is not totally ergodic. Let $n \in \mathbb{N}$ be such that T^n is not ergodic and let $A \in \mathcal{A}$ be such that $\mu(A) \in (0,1)$ and $T^{-n}A = A$. It follows that the σ-algebra generated by the sets $A, T^{-1}A, \ldots, T^{-(n-1)}A$ is invariant under T, finite and non-trivial. In view of the discussion following Theorem 1.2.9, we conclude that $(X, \mathcal{A}, \mu, T^n)$ has a non-trivial finite factor. Here is a more explicit construction of this factor.

Assume that n is the smallest for which T^n is not ergodic. The function $f = \sum_{i=0}^{n-1} T^i 1_A$ satisfies $Tf = \sum_{i=1}^{n-1} T^i 1_A + T^n 1_A = f$, so by ergodicity it must be a constant, namely $\int_X f \, \mathsf{d}\mu$ which equals $n\mu(A)$. Since f only takes integer values, $n\mu(A)$ must be an integer. If $\mu(A) > 1/n$ then $\mu(A \cap T^{-i}A) > 0$ for some $i \in \{1, \ldots, n-1\}$. If $\mu(A \cap T^{-i}A) = \mu(A)$ then $T^{-i}A = A$ a.e., contradicting the minimality of n. Therefore $0 < \mu(A \cap T^{-i}A) < \mu(A)$. But since A is invariant under T^{-n}, then so is $\tilde{A} = A \cap T^{-i}A$. If $\mu(\tilde{A}) > 1/n$ we can repeat the argument until we find a set $C \in \mathcal{A}$, invariant under T^n with measure $\mu(C) = 1/n$ (using the fact that any set invariant under T^{-n} must have measure k/n for some $k \in [0, n]$).

Now let $Y = \{0, \ldots, n-1\}$, let $\mathcal{B}$ be the discrete σ-algebra on Y, let ν be the normalized counting measure and let $S : y \mapsto y + 1 \bmod n$. Then $(Y, \mathcal{B}, \nu, S)$ is a measure preserving system and $\pi : X \to Y$ given by $\pi(x) = \{i \in Y : x \in T^{-i}C\}$ is a factor map.

□

Exercise 4.3.5. *Show that Theorem 4.3.4 holds even without the ergodicity assumption.*

Exercise 4.3.6. *Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving system and let $(Y, \mathcal{B}, \nu, S)$ be a factor. Prove that:*

- *If $(X, \mathcal{A}, \mu, T)$ is ergodic, then so is $(Y, \mathcal{B}, \nu, S)$.*
- *If $(X, \mathcal{A}, \mu, T)$ is totally ergodic, then so is $(Y, \mathcal{B}, \nu, S)$.*

In Theorem 4.3.4 we characterized the obstructions to total ergodicity. On the other hand, when a system $(X, \mathcal{B}, \mu, T)$ is totally ergodic, we obtain from the ergodic theorem the following corollary.

Corollary 4.3.7. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Then it is totally ergodic if and only if for every $f \in L^2(X)$ and every $q, r \in \mathbb{N}$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{qn+r} f = \int_X f \, \mathsf{d}\mu. \qquad \text{in } L^2(X) \tag{4.5}$$

Proof. If the system is not totally ergodic, then there exists $q \in \mathbb{N}$ and a non-constant $f \in L^2(X)$ such that $T^q f = f$. Thus (4.5) implies that the system is totally ergodic.

To prove the converse direction, let $(X, \mathcal{B}, \mu, T)$ be totally ergodic and let $f \in L^2(X)$ and $q, r \in \mathbb{N}$ be arbitrary. Applying the ergodic theorem (Theorem 2.3.1) to the (ergodic) system $(X, \mathcal{B}, \mu, T^q)$ we conclude that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{qn+r} f = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} (T^q)^n (T^r f) = \int_X T^r f \, \mathsf{d}\mu = \int_X f \, \mathsf{d}\mu.$$

□

Remark 4.3.8. *A measure preserving system $(X, \mathcal{B}, \mu, T)$ is called* **invertible** *if T is invertible a.e. and the inverse is measurable and measure preserving. In this situation we can allow q and r in Corollary 4.3.7 to be negative, but if the system is not invertible, then the expression $T^n f$ does not make sense for a negative value of n.*

Nevertheless, Corollary 4.3.7 still makes sense when $r < 0$, even if the system is not invertible. Indeed, in this case the expression $qn + r$ is positive for all but finitely many values of n, and since we take an average over $\mathbb{N}$ we can just ignore those finitely many values.

One could interpret the expression T^{qn+r} appearing in (4.5) as $T^{p(n)}$ where p is a linear polynomial. The following theorem reveals the power of the van der Corput trick, which allows one to upgrade Corollary 4.3.7 to general polynomials.

Theorem 4.3.9. *Let $(X,\mathcal{B},\mu,T)$ be a totally ergodic system and let $p\in\mathbb{Z}[x]$ be such that either the system is invertible or the polynomial has a positive leading coefficient. Then for every $f\in L^2(X)$,*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}T^{p(n)}f=\int_X f\,\mathsf{d}\mu.\qquad \text{in } L^2(X) \tag{4.6}$$

Proof. We proceed by induction on the degree of p. If p is linear, then the result follows from Corollary 4.3.7, so assume that p has degree at least 2. Eq. (4.6) holds for f if and only if it holds for $f-c$ where c is a constant; therefore, after subtracting $\int_X f\,\mathsf{d}\mu$ from f we can assume that $\int_X f\,\mathsf{d}\mu=0$. 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 lemma (Lemma 4.2.1). Fixing $d\in\mathbb{N}$ we can compute

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

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 $L^2(X)$ 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 (4.3) of the van der Corput lemma, so we conclude that $\displaystyle\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}x_n=0$, as desired. □

Remark 4.3.10. *Both Corollary 4.3.7 and Theorem 4.3.9 have versions for uniform Cesàro averages, which can be proved in the exact same way. The choice to present the regular Cesàro versions was made with the hope that the main ideas became more transparent.*

4.4. **Polynomial recurrence via the spectral theorem.** An immediate corollary of Theorem 4.3.9 is the following recurrence theorem for totally ergodic systems.

Corollary 4.4.1. *For any totally ergodic system $(X,\mathcal{B},\mu,T)$, any set $A\in\mathcal{B}$, any polynomial $p\in\mathbb{Z}[x]$ (with a positive leading coefficient) and any $\epsilon>0$, there exists $n\in\mathbb{N}$ such that $\mu(A\cap T^{-p(n)}A)>\mu^2(A)-\epsilon$.*

In this subsection we see how this result can be extended to general measure preserving systems.

First, by looking at the rotation on two point from Example 4.3.3, and considering the polynomial $p(n)=2n-1$, we see that Corollary 4.4.1 does not hold without the total ergodicity assumption for every polynomial.

Definition 4.4.2. *A polynomial $p\in\mathbb{Z}[x]$ is called* **divisible** *or* **intersective** *if for every $k\in\mathbb{N}$ there exists $n\in\mathbb{N}$ such that $p(n)$ is a multiple of k.*

If $p(0)=0$ or, more generally, p has an integer root, then it is divisible. However there are polynomials, such as $p(x)=(x^2-3)(x^2-5)(x^2-15)$ which have no integer root but are divisible. It is easy to see that if p is not divisible, then there exists a finite system where recurrence does not occur at times of the form $p(n)$. In other words, if p is not divisible, then the set $\{p(n):n\in\mathbb{N}\}$ is not a set of recurrence. The converse of this observation is the content of the following theorem, which significantly extends Theorem 2.1.6.

Theorem 4.4.3. *Let $(X,\mathcal{B},\mu,T)$ be a measure preserving system, let $A\in\mathcal{B}$, let $\epsilon>0$ and let $p\in\mathbb{Z}[x]$ be a divisible polynomial with a positive leading coefficient. Then there exists $n\in\mathbb{N}$ such that $\mu(A\cap T^{-p(n)}A)>\mu^2(A)-\epsilon$.*

Observe the following immediate corollary.

Corollary 4.4.4. *Let $p\in\mathbb{Z}[x]$ have positive leading coefficient. Then the set $\{p(n):n\in\mathbb{N}\}$ is a set of recurrence if and only if p is divisible.*

In view of Theorem 3.2.3, for every divisible polynomial p and every set $A \subset \mathbb{N}$ with positive upper density contains two numbers a, b whose difference $a-b$ is of the form $a-b=p(n)$ for some n.

We will present two proofs of Theorem 4.4.3 which illustrate different useful techniques in obtaining recurrence results. For the first proof we will need the spectral theorem for unitary operators. More precisely, we will use the following corollary, which can be proved independently.

Theorem 4.4.5 (Herglotz theorem)**.** *Let H be a Hilbert space and let $U : H \to H$ be an isometry. For every $f \in H$ there exists a finite measure ν on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ whose Fourier transform satisfies*

$$\hat{\nu}(n) := \int_0^1 e(-nx)\,\mathsf{d}\nu(x) = \langle U^n f, f\rangle \qquad \forall\, n \in \mathbb{N}$$

Sketch of the proof. For each $N \in \mathbb{N}$ let $\phi_N : \mathbb{T} \to [0, \infty)$ be the function

$$\phi_N(t) := \frac{1}{N}\left\|\sum_{n=1}^{N} e(nt)U^n f\right\|^2 = \|f\|^2 + \frac{1}{N}\sum_{n=1}^{N-1}(N-n)\Big(e(nt)\big\langle U^n f, f\big\rangle + \overline{e(nt)\big\langle U^n f, f\big\rangle}\Big)$$

and let ν_N be the measure on $\mathbb{T}$ with density ϕ_N. Let $(N_k)_{k=1}^{\infty}$ be an increasing sequence such that $\big(\nu_{N_k}\big)_{k=1}^{\infty}$ converges as $k \to \infty$ to a measure ν in the weak* topology. Finally we can compute, for $m \in \mathbb{N}$,

$$\begin{aligned}
\hat{\nu}(m) &= \int_0^1 e(-mt)\,\mathsf{d}\nu(t) = \lim_{k\to\infty}\int_0^1 e(-mt)\,\mathsf{d}\nu_k(t) = \lim_{k\to\infty}\int_0^1 e(-mt)\phi_{N_k}(t)\,\mathsf{d}t \\
&= \lim_{k\to\infty}\frac{1}{N_k}\sum_{n=1}^{N_k-1}(N_k-n)\Big(\big\langle U^n f, f\big\rangle\int_0^1 e\big((n-m)t\big)\,\mathsf{d}t + \overline{\big\langle U^n f, f\big\rangle}\int_0^1 e\big((-n-m)t\big)\,\mathsf{d}t\Big) \\
&= \lim_{k\to\infty}\frac{1}{N_k}(N_k-m) = \big\langle U^m f, f\big\rangle
\end{aligned}$$

□

We are now ready to give a first proof of Theorem 4.4.3.

Proof of Theorem 4.4.3. Apply Herglotz theorem (Theorem 4.4.5) to the indicator function 1_A of A and let ν be the corresponding measure on $\mathbb{T}$. Then for every $n \in \mathbb{N}$, $\mu(A \cap T^{-n}A) = \int_X 1_A \cdot T^n 1_A\,\mathsf{d}\mu = \hat{\nu}(n)$. Using the bounded convergence theorem and the sum of a geometric series we observe that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\hat{\nu}(n) = \int_0^1 \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} e(-nt)\,\mathsf{d}\nu(t) = \int_0^1 1_{\{0\}}(t)\,\mathsf{d}\nu(t) = \nu(\{0\}).$$

On the other hand, using the mean ergodic theorem (Theorem 2.3.1) together with the Cauchy-Schwarz inequality as in the proof of Khintchine's theorem (Theorem 2.4.4) we have that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\hat{\nu}(n) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mu(A\cap T^{-n}A) = \|P_I 1_A\|^2 \geq \mu(A)^2.$$

Putting the two together it follows that $\nu(\{0\}) \geq \mu^2(A)$. Unfortunately, it is not true in general that the average of $\mu(A \cap T^{-p(n)}A)$ over all $n \in \mathbb{N}$ is (at least) $\mu^2(A)$. Instead we need to locate a subset of $\mathbb{N}$ on which the average of $\mu(A \cap T^{-n}A)$ is at least $\mu^2(A) - \epsilon$. It turns out that this subset of $\mathbb{N}$ is determined by the rational points in $(0, 1)$ to which ν attributes positive measure.

The measure ν is finite, thus for every $\epsilon > 0$ there exists a finite set $F \subset \mathbb{Q} \cap (0, 1)$ such that

$$\sum_{t\in\mathbb{Q}\setminus(F\cup\{0\})}\nu(\{t\}) < \epsilon.$$

Let $k \in \mathbb{N}$ be such that $kt \in \mathbb{N}$ for every $t \in F$. Since p is divisible, there exists $n_0 \in \mathbb{N}$ such that $p(n_0) \equiv 0 \bmod k$. Let $q : n \mapsto p(n_0 + kn)$ and observe that $q \in \mathbb{Z}[x]$ satisfies $q(n) \equiv p(n_0) \equiv 0 \bmod k$ for every $n \in \mathbb{N}$.

Next observe that, using again the bounded convergence theorem,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mu(A\cap T^{-q(n)}A)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\hat{\nu}\big(q(n)\big)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int_0^1 e\big(-tq(n)\big)\,\mathsf{d}\nu(t)=\int_0^1\psi(t)\,\mathsf{d}\nu(t),\tag{4.7}$$

where $\psi(t)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}e\big(-tq(n)\big)$. We are going to split the last integral into three pieces, according to the partition of the interval

$$[0,1)=\{0\}\cup\big(\mathbb{Q}\cap(0,1)\big)\cup\big((\mathbb{R}\setminus\mathbb{Q})\cap(0,1)\big).$$

The first piece can be directly evaluated as $\nu(\{0\})$. The third piece turns out to be 0. Indeed, for each irrational $t\in(0,1)$, the polynomial $n\mapsto tq(n)$ has an irrational coefficient. In view of Weyl's equidistribution theorem (Corollary 4.1.7) the sequence $n\mapsto tq(n)\bmod 1$ is uniformly distributed, and by Weyl's criterion (Lemma 4.1.2) this implies that $\psi(t)=0$, and in particular $\int_{(\mathbb{R}\setminus\mathbb{Q})\cap(0,1)}\psi(t)\,\mathsf{d}\nu(t)=0$. Thus (4.7) becomes

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mu(A\cap T^{-q(n)}A)=\nu(\{0\})+\int_{\mathbb{Q}\cap(0,1)}\psi(t)\,\mathsf{d}\nu(t).$$

Since $\nu(\{0\})\geq\mu^2(A)$, it suffices to show that the integral over the rationals is $>-\epsilon$. To do this, we further split the integral into two

$$\int_{\mathbb{Q}\cap(0,1)}\psi(t)\,\mathsf{d}\nu(t)=\int_F\psi(t)\,\mathsf{d}\nu(t)+\int_{\mathbb{Q}\cap(0,1)\setminus F}\psi(t)\,\mathsf{d}\nu(t).$$

But from the construction of F it follows that $\nu(\mathbb{Q}\cap(0,1)\setminus F)<\epsilon$. Since ψ has supremum at most 1, it follows that the second integral is $>-\epsilon$. Finally, since for every $t\in F$ and $n\in\mathbb{N}$, $tq(n)$ is an integer, we conclude that $\psi(t)=1$ for every $t\in F$ and hence $\int_F\psi\,\mathsf{d}\nu=\nu(F)\geq 0$, and therefore $\int_{\mathbb{Q}\cap(0,1)}\psi(t)\,\mathsf{d}\nu(t)>-\epsilon$ as needed. $\square$

The above proof of Theorem 4.4.3 (and hence of Corollary 4.4.4) can be modified to yield other sets of recurrence.

Theorem 4.4.6. *Let $\mathbb{P}$ denote the set of prime numbers and let $a\in\mathbb{Z}$. Then the set $\mathbb{P}+a$ is a set of recurrence if and only if $|a|=1$.*

In order to prove Theorem 4.4.6 we need some information about the primes, which comes in the form of a theorem of Vinogradov [14].

Theorem 4.4.7. *Let $(p_n)_{n=1}^{\infty}$ be the increasing enumeration of the primes and let $\alpha\in\mathbb{R}\setminus\mathbb{Q}$. Then the sequence $n\mapsto p_n\alpha\bmod 1$ is uniformly distributed.*

Moreover, if $a,b\in\mathbb{N}$ are coprime and $(q_n)_{n=1}^{\infty}$ is the increasing enumeration of the primes which are congruent to $a\bmod b$, then for every $\alpha\in\mathbb{R}\setminus\mathbb{Q}$ the sequence $n\mapsto q_n\alpha\bmod 1$ is uniformly distributed.

Proof of Theorem 4.4.6. If $|a|\neq 1$, then there exists k such that no element of $\mathbb{P}+a$ is a multiple of k. For $a=0$ we can take $k=4$, and for $|a|>1$ we can take $k=a^3$. In any case, taking a cycle on k points we see that $\mathbb{P}+a$ is not a set of recurrence.

If $|a|=1$, we proceed as in the proof of Theorem 4.4.3. Let ν be the measure on $\mathbb{T}$ given by applying Herglotz theorem to 1_A. As above, $\nu(\{0\})\geq\mu^2(A)$, and for every $\epsilon>0$ we can find a finite set $F\subset\mathbb{Q}\cap(0,1)$ such that $\nu(\mathbb{Q}\cap(0,1)\setminus F)<\epsilon$. Let $k\in\mathbb{N}$ be such that $kt\in\mathbb{N}$ for every $t\in F$ and let (q_n) be the increasing enumeration of the primes which are congruent to $-a\bmod k$ (this is an infinite set by Dirichlet's theorem). As in (4.7),

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mu(A\cap T^{-(q_n+a)}A)=\int_0^1\psi(t)\,\mathsf{d}\nu(t),$$

where $\psi(t)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}e\big(t(q_n+a)\big)$. Theorem 4.4.7 implies that $n\mapsto q_nt\bmod 1$ is uniformly distributed for every $t\in\mathbb{R}\setminus\mathbb{Q}$, which in turn implies that $\psi(t)=0$ for any such t. Since $(q_n+a)t\in\mathbb{N}$ for every $t\in F$, it follows that $\psi(t)=1$ for every such t. Finally, since $\nu(\mathbb{Q}\cap(0,1)\setminus F)<\epsilon$ we conclude that

$$\int_0^1\psi(t)\,\mathsf{d}\nu(t)=\int_{\{0\}}\psi(t)\,\mathsf{d}t+\int_{(0,1)\setminus\mathbb{Q}}\psi(t)\,\mathsf{d}t+\int_{\mathbb{Q}\cap(0,1)\setminus F}\psi(t)\,\mathsf{d}t+\int_F\psi(t)\,\mathsf{d}t\geq\mu^2(A)-\epsilon.$$

$\square$

4.5. **Polynomial recurrence via L^2 decomposition.** We saw in the proof of von Neumann's mean ergodic theorem that for any measure preserving system $(X, \mathcal{B}, \mu, T)$, the Hilbert space $L^2(X)$ can be decomposed as the direct sum of the orthogonal subspaces $I := \{f \in L^2(X) : Tf = f\}$ and $J := \overline{\{f - Tf : f \in L^2(X)\}}$. The idea of decomposing $L^2(X)$ into orthogonal pieces with opposite dynamical properties, and more generally, of decomposing a measure preserving system in some sense into a "structured" and a "mixing" components turns out to be very fruitful.

Consider the following subspaces of $L^2(X)$:

$$H_{rat} := \overline{\{f \in L^2(X) : T^k f = f \text{ for some } k \in \mathbb{N}\}};\ H_{te} := \left\{ f \in L^2(X) : \forall\, k \in \mathbb{N},\ \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{kn} f = 0 \right\}$$

Observe that in a totally ergodic system the space H_{rat} consists only of constant functions, while the space H_{te} contains every function with 0 integral. The following proposition generalizes this observation.

Proposition 4.5.1. *For any measure preserving system $(X, \mathcal{B}, \mu, T)$, the spaces H_{rat} and H_{te} are orthogonal and $L^2(X) = H_{rat} \oplus H_{te}$.*

Proof. Let $f \in L^2(X)$ be such that $T^k f = f$ for some $k \in \mathbb{N}$ and let $g \in H_{te}$. Then $\langle f, g\rangle = \langle T^k f, T^k g\rangle = \langle f, T^k g\rangle$. Iterating this observation we deduce that $\langle f, g\rangle = \langle f, T^{kn} g\rangle$ for every $n \in \mathbb{N}$. Averaging over n we then deduce

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

showing that H_{rat} and H_{te} are orthogonal.

Now suppose that $f \in L^2(X)$ is orthogonal to H_{rat}, we need to show that $f \in H_{te}$. But for every $k \in \mathbb{N}$, the space H_{rat} contains the invariant subspace I_k for the system $(X, \mathcal{B}, \mu, T^k)$. It follows that f is orthogonal to I_k for every k, and in view of the mean ergodic theorem, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{kn} f = 0$, so that indeed $f \in H_{te}$. $\square$

Here is how to use this decomposition to prove Theorem 4.4.3.

Proof of Theorem 4.4.3. Decompose $1_A = f + g$ with $f \in H_{rat}$ and $g \in H_{te}$. Since H_{rat} contains the constants, using the Cauchy-Schwarz inequality we have $\langle 1_A, f\rangle = \|f\|^2 \geq \langle f, 1\rangle^2 = \mu(A)^2$. Find $h \in H_{rat}$ such that $T^k h = h$ for some $k \in \mathbb{N}$, and such that $\|f - h\| < \epsilon/2$. In particular it follows that $\langle 1_A, h\rangle > \mu(A)^2 - \epsilon/2$.

Using divisibility of p, find $a \in \mathbb{N}$ such that $p(a) \equiv 0 \bmod k$ and consider the polynomial $q(n) = p(a + kn)$. Then $T^{q(n)} h = h$ for all $n \in \mathbb{N}$. As in the proof of Theorem 4.3.9, an application of the van der Corput trick implies that

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

Finally, we have

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

$\square$

5.1. **Mixing and weak-mixing.** As we saw in Corollary 2.4.1, a measure preserving system is ergodic if and only if any two sets became asymptotically independent on average. For certain systems, this asymptotic independence occurs even without averaging, and we call this property ***mixing***.

Definition 5.1.1. *A measure preserving system $(X\,B,\mu,T)$ is* **mixing** *or* **strong-mixing** *if for every $A,B\in\mathcal{B}$,*

$$\lim_{n\to\infty}\mu(T^{-n}A\cap B)=\mu(A)\mu(B).$$

Proposition 5.1.2. *Let $(X,\mathcal{B},\mu,T)$ be a measure preserving system. Then the following are equivalent.*

- *The system is mixing.*
- *For every $f,g\in L^2(X)$, $\lim_{N\to\infty}\int_X T^nf\cdot g\,\mathsf{d}\mu=\int_X f\,\mathsf{d}\mu\int_X g\,\mathsf{d}\mu$.*
- *For every $f\in L^2(X)$ with $\int_X f\,\mathsf{d}\mu=0$, the orbit T^nf converges to 0 in the weak topology.*

Proof. The equivalence between the first two follows from the fact that the set of finite linear combinations of indicator functions is dense in L^2. The equivalence between the last two is immediate, after replacing f with $\tilde f:=f-\int_X f\,\mathsf{d}\mu$ and noticing that $\int_X\tilde f\,\mathsf{d}\mu=0$. □

It should be clear that every mixing system is ergodic, but the opposite is not true. There is also a notion of higher order mixing.

Definition 5.1.3. *A measure preserving system $(X\,B,\mu,T)$ is* **mixing of order** k *if for every $A_1,\dots,A_k\in\mathcal{B}$ and every sequences $\left(n_i^{(1)}\right)_{i=1}^{\infty},\dots,\left(n_i^{(k)}\right)_{i=1}^{\infty}$ with $\lim_{i\to\infty}n_i^{(r)}-n_i^{(s)}=\infty$ for every $1\le r,s\le k$*

$$\lim_{i\to\infty}\mu\Big(T^{-n_i^{(1)}}A_1\cap T^{-n_i^{(2)}}A_2\cap\cdots\cap T^{-n_i^{(k)}}A_k\Big)=\mu(A_1)\mu(A_2)\cdots\mu(A_k).$$

Notice that mixing of order 2 is the same a strong-mixing. It is clear that k-mixing implies $k-1$-mixing; it is in fact a major open problem in ergodic theory whether the converse holds, even for $k=3$.

A weaker notion of mixing is ***weak-mixing***. While at first it seems a less natural notion, it turns out to be fundamental in the structure of arbitrary measure preserving systems. One reason for this is that there are many equivalent definitions of weak-mixing. Here is the simplest to state.

Definition 5.1.4. *Let $\mathbf{X}=(X,\mathcal{B},\mu,T)$ be a measure preserving system and let $\mathbf{X}\times\mathbf{X}$ be the self product system, defined in Definition 1.3.9. The system $\mathbf{X}$ is* **weak-mixing** *or* **weakly mixing** *if and only if $\mathbf{X}\times\mathbf{X}$ is ergodic.*

The following theorem states several equivalent properties to weak-mixing which explain the name.

Theorem 5.1.5. *Let $\mathbf{X}=(X,\mathcal{B},\mu,T)$ be a measure preserving system. Then the following are equivalent*

1. *$\mathbf{X}$ is weak mixing.*
2. *For every ergodic m.p.s. $\boldsymbol{Y}$, the product $\boldsymbol{X}\times\boldsymbol{Y}$ is ergodic.*
3. *For any two sets $A,B\in\mathcal{B}$ we have $\displaystyle\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\big|\mu(A\cap T^{-n}B)-\mu(A)\mu(B)\big|=0$*
4. *For any $f,g\in L^2$ we have $\displaystyle\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left|\int_X T^nf\cdot g\,\mathsf{d}\mu-\int_X f\,\mathsf{d}\mu\int_X g\,\mathsf{d}\mu\right|=0$*
5. *For any $A,B\in\mathcal{B}$ there exists a subset $E\subset\mathbb{N}$ with upper density $\bar d(E)=0$ such that*
$$\lim_{\substack{n\to\infty\\ n\notin E}}\mu(A\cap T^{-n}B)=\mu(A)\mu(B)$$
Moreover if $\mathcal{B}$ is separable we can choose E independent of A,B.

Condition (3) explains why it is called weak mixing, and makes it clear that every mixing system is weak mixing, and that every weak mixing system is ergodic. Not every weak-mixing system is strong-mixing, but examples are not easy to come by. On the other hand, it is easy to show, using the definition that irrational circle rotations are ergodic but not weakly mixing.

Condition (2) implies that if $\mathbf{X}$ is weak mixing, then $\mathbf{X}\times\mathbf{X}\times\mathbf{X}\times\mathbf{X}$ is ergodic, and hence $\mathbf{X}\times\mathbf{X}$ is weak mixing. Therefore any self product $\mathbf{X}\times\mathbf{X}$ is weak mixing if and only if it is ergodic.

Proof of Theorem 5.1.5.

(1)⇒(4) Replacing f with $f - \int_X f \, \mathsf{d}\mu$ we can assume that $\int_X f \, \mathsf{d}\mu = 0$. Using the Cauchy-Schwartz inequality we have

$$\limsup_{N\to\infty} \left(\frac{1}{N} \sum_{n=1}^{N} \left| \int_X T^n f \cdot g \, \mathsf{d}\mu \right| \right)^2 \leq \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left| \int_X T^n f \cdot g \, \mathsf{d}\mu \right|^2.$$

Using the hypothesis that $\mathbf{X} \times \mathbf{X}$ is ergodic, and applying the von Neumann's Ergodic Theorem (Theorem 2.3.1) to the functions $f \otimes \overline{f} \in L^2(X \times X)$ and $g \otimes \overline{g} \in L^2(X \times X)$ we obtain

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{X\times X} (f \otimes \overline{f}) \circ (T \times T)^n \cdot g \otimes \overline{g} \, \mathsf{d}(\mu \otimes \mu) = \int_{X\times X} f \otimes \overline{f} \, \mathsf{d}(\mu \otimes \mu) \int_{X\times X} g \otimes \overline{g} \, \mathsf{d}(\mu \otimes \mu).$$

Observe that $\int_{X\times X} f \otimes \overline{f} \, \mathsf{d}(\mu \otimes \mu) = \left| \int_X f \, \mathsf{d}\mu \right|^2 = 0$, so the previous equation can be rewritten as

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left| \int_X T^n f \cdot g \, \mathsf{d}\mu \right|^2 = 0,$$

finishing the proof.

(4)⇒(3) This is immediate by letting $f = 1_A$ and $g = 1_B$.

(3)⇒(5) Fix $m \in \mathbb{N}$ and set $A_m := \{n \in \mathbb{N} : |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| > 1/m\}$. Observe that

$$\frac{1}{N} \sum_{n=1}^{N} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| \geq \frac{1}{m} \frac{|A_m \cap [1, N]|}{N}$$

Taking the limit as $N \to \infty$ we conclude that $\bar{d}(A_m) = 0$ for all $m \in \mathbb{N}$. For each $m \in \mathbb{N}$ let $N_m \in \mathbb{N}$ be such that for all $N > N_m$ we have $|A_m \cap [1, N]| \leq N/m$ and make

$$E = \bigcup_{m=1}^{\infty} (A_m \cap [N_m + 1, N_{m+1}])$$

Now observe that $A_k \subset A_{k+1}$ for all $k \in \mathbb{N}$, hence for each $N \in \mathbb{N}$, choosing m such that $N \in [N_m + 1, N_{m+1}]$ we have $E \cap [1, N] \subset A_m \cap [1, N]$ and hence $|E \cap [1, N]| \leq N/m$. Taking $N \to \infty$ (note that also $m \to \infty$ because all A_m have 0 density) we conclude that $\bar{d}(E) = 0$.

Finally, for each $m \in \mathbb{N}$, let $N > N_m$, then if $N \notin E$ we also have $N \notin A_m$ and so $|\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| < 1/m$ concluding the proof.

In the case when $\mathcal{B}$ is separable, let $\{B_n\}_{n=1}^{\infty}$ be a countable dense family. For each $m = (m_1, m_2) \in \mathbb{N}^2$ let $E_m \subset \mathbb{N}$ be such that $\bar{d}(E_m) = 0$ and $\lim_{n\to\infty} \mu(T^{-n}B_{m_1} \cap B_{m_2}) \to \mu(B_{m_1})\mu(B_{m_2})$ for $n \notin E_m$. As above we construct a set E of 0 density such that for all $m \in \mathbb{N}^2$ there exists $N = N(m) \in \mathbb{N}$ such that $E_m \setminus [1, N] \subset E$.

It is not hard to check that this set E satisfies the conditions, we omit the details.

(5)⇒(3) Assuming (5), for every ϵ the set $\left\{ n \in \mathbb{N} : \left| \mu(A \cap T^{-n}B) - \mu(A)\mu(B) \right| > \epsilon \right\}$ has density 0. On the other hand $\left| \mu(A \cap T^{-n}B) - \mu(A)\mu(B) \right| \leq 1$ for every $n \in \mathbb{N}$, and hence

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left| \mu(A \cap T^{-n}B) - \mu(A)\mu(B) \right| \leq \epsilon.$$

Since ϵ is arbitrary we conclude that (3) holds.

(3)⇒(4) Condition (3) is the special case of (4) when f and g are indicator functions. It is not hard to see that if (4) holds for pairs (f_1, g) and (f_2, g), then it holds for the pair $(af_1 + bf_2, g)$. Since every L^2 function is approximated by finite linear combinations of indicator functions, we deduce that (4) holds whenever g is an indicator function. But similarly, if (4) holds for (f, g_1) and (f, g_2), it holds for $(f, ag_1 + bg_2)$, and hence the same argument shows that it must hold for any $f, g \in L^2$.

(4)⇒(2) Let $\mathbf{Y} = (Y, \mathcal{A}, S, \nu)$. In order to show that $\mathbf{X} \times \mathbf{Y}$ is ergodic, we will show that for any $f, g \in L^2(X \times Y)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{X\times Y} (T \times S)^n f \cdot g \,\mathsf{d}(\mu \otimes \nu) = \int_{X\times Y} f \,\mathsf{d}(\mu \otimes \nu) \int_{X\times Y} g \,\mathsf{d}(\mu \otimes \nu). \tag{5.1}$$

Since finite linear combinations of tensor functions of the form $(f_1 \otimes f_2)(x, y) = f_1(x)f_2(y)$ form a dense subset of $L^2(X \times Y)$, it suffices to establish (5.1) when both f and g are tensor functions. Let $f(x, y) = f_1(x)f_2(y) \in L^2(X \times Y)$ and $g(x, y) = g_1(x)g_2(y) \in L^2(X \times Y)$ be arbitrary tensor functions. Then (5.1) can be written as

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X T^n f_1 \cdot g_1 \,\mathsf{d}\mu \int_Y S^n f_2 \cdot g_2 \,\mathsf{d}\nu = \int_X f_1 \,\mathsf{d}\mu \int_Y f_2 \,\mathsf{d}\nu \int_X g_1 \,\mathsf{d}\mu \int_Y g_2 \,\mathsf{d}\nu. \tag{5.2}$$

Since (5.2) is linear in f_2 we can, splitting $f_2 = \int_Y f_2 \,\mathsf{d}\nu + \big(f_2 - \int_Y f_2 \,\mathsf{d}\nu\big)$, separate the proof of (5.2) in two cases: when f_2 is a constant and when $\int_Y f_2 \,\mathsf{d}\nu = 0$. For the first case, since $\mathbf{Y}$ is ergodic, it follows that f_2 is a constant, and hence the left hand side of (5.2) is

$$\int_Y f_2 \,\mathsf{d}\nu \int_Y g_2 \,\mathsf{d}\nu \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X T^n f_1 \cdot g_1 \,\mathsf{d}\mu.$$

But now, using (4), it is clear that (5.2) holds in this case.

Next we establish (5.2) in the case that $\int_Y f_2 \,\mathsf{d}\nu = 0$. Applying Cauchy-Schwarz with f_2, g_2 and using (4) we get

$$\begin{aligned}\left|\frac{1}{N} \sum_{n=1}^{N} \int_X T^n f_1 \cdot g_1 \,\mathsf{d}\mu \int_Y S^n f_2 \cdot g_2 \,\mathsf{d}\nu\right| &\leq \frac{1}{N} \sum_{n=1}^{N} \left|\int_X T^n f_1 \cdot g_1 \,\mathsf{d}\mu \int_Y S^n f_2 \cdot g_2 \,\mathsf{d}\nu\right| \\ &\leq \|f_2\| . \|g_2\| \frac{1}{N} \sum_{n=1}^{N} \left|\int_X T^n f_1 \cdot g_1 \,\mathsf{d}\mu\right|\end{aligned}$$

Using (4) we conclude that this quantity converges to 0 as $N \to \infty$, establishing (5.2).

(2)⇒(1) It suffices to show that if (2) holds, then $\mathbf{X}$ is ergodic. To see this assume that $\mathbf{X}$ is not ergodic and let $A \in \mathcal{B}$ be an invariant set such that $0 < \mu(A) < 1$. Let $\mathbf{Y} = (Y, S)$ be the (ergodic) one point system. Then $A \times Y$ is invariant for $T \times S$ and so $\mathbf{X} \times \mathbf{Y}$ wouldn't also be ergodic.

□

Remark 5.1.6. *Conditions (3) and (4) can be formulated using strong Cesàro averages, and the proof presented holds in that case as well. Therefore we obtain two other equivalent properties to weak mixing.*

Throughout this section we will add more properties to the list of equivalent characterizations of weak mixing. We already saw that every weak mixing system is ergodic. It turns out that it must in fact be totally ergodic.

Theorem 5.1.7. *Let $k \in \mathbb{N}$. A system $(X, \mathcal{B}, \mu, T)$ is weak mixing if and only if the system $(X, \mathcal{B}, \mu, T^k)$ is weak mixing.*

Proof. First suppose that $(X, \mathcal{B}, \mu, T)$ is weak mixing. To show that $(X, \mathcal{B}, \mu, T^k)$ is weak mixing we will use Condition (5) from Theorem 5.1.5. Let $A, B \in \mathcal{B}$ and let $E \subset \mathbb{N}$ be the set with 0 density satisfying

$$\lim_{\substack{n\to\infty \\ n\notin E}} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

Let $\tilde{E} := \{m \in \mathbb{N} : mk \in E\}$. It is clear that $\bar{d}(\tilde{E}) = 0$ and that

$$\lim_{\substack{m\to\infty \\ m\notin \tilde{E}}} \mu\Big(A \cap \big(T^k\big)^{-m} B\Big) = \lim_{\substack{m\to\infty \\ m\notin \tilde{E}}} \mu\big(A \cap T^{-mk} B\big) = \mu(A)\mu(B).$$

To prove the converse, suppose that $(X, \mathcal{B}, \mu, T^k)$ is weak mixing. To show that $(X, \mathcal{B}, \mu, T)$ is weak mixing we will use Condition (1) from Theorem 5.1.5. Indeed, if $(X, \mathcal{B}, \mu, T) \times (X, \mathcal{B}, \mu, T)$ were not ergodic, there would exist a $T \times T$ invariant set $A \subset X \times X$ with $(\mu \otimes \mu)(A) \in (0,1)$. But A would also be invariant under $T^k \times T^k = (T \times T)^k$, and hence $(X, \mathcal{B}, \mu, T^k) \times (X, \mathcal{B}, \mu, T^k)$ would not be ergodic, contradicting the assumption. □

5.2. **Eigenfunctions and discrete spectrum.** Given a measure preserving system $(X, \mathcal{B}, \mu, T)$, a non-zero function $f \in L^2(X)$ is an ***eigenfunction*** if there exists a constant λ, called the ***eigenvalue*** such that $Tf = \lambda f$. Equivalently, f is an eigenvector for the Koopman operator $\Phi_T : L^2(X) \to L^2(X)$ and λ is the associated eigenvalue. Since the Koopman operator of a measure preserving transformation is an isometry, all eigenvalues have absolute value 1. If $\mathbf{X}$ is an ergodic system and f is an eigenfunction, then the function $|f|$ is invariant and therefore constant. It is usual to normalize the eigenfunctions so that they take values in the unit circle S^1.

If f and g are two eigenfunctions associated with different eigenvalues λ_f and λ_g respectively, then they are orthogonal, since

$$\langle f, g\rangle = \langle Tf, Tg\rangle = \lambda_f \overline{\lambda_g} \langle f, g\rangle,$$

and hence $\langle f, g\rangle = 0$. On the other hand, since f and g are bounded, the produt fg is also in L^2 and satisfies $T(fg) = \lambda_f \lambda_g fg$. This shows that $\lambda_f \lambda_g$ is also an eigenvalue, and in fact the collection of all eigenvalues of a m.p.s. is a group.

Example 5.2.1. *Let $\mathbf{X}$ be a circle rotation, with $T : x \mapsto x+\alpha \bmod 1$. Then the function $f(x) = e(x) = e^{2\pi i x}$ is an eigenfunction with eigenvalue $e(\alpha)$. The eigenfunctions of $\mathbf{X}$ are in fact the functions $x \mapsto ce(nx)$ with $n \in \mathbb{Z}$ and $c \in \mathbb{C}$, and the eigenvalues are the numbers $e(n\alpha)$ with $n \in \mathbb{Z}$.*

The above example generalizes in a strong form to arbitrary group rotations (cf. Example 1.3.3 and the paragraph following it).

Proposition 5.2.2. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be a group rotation. This means that X is a compact abelian group, $\mathcal{B}$ is the Borel σ-algebra, μ is the normalized Haar measure and $T : x \mapsto x + \alpha$ for some element $\alpha \in X$. Then there exists an orthonormal basis for $L^2(X)$ consisting of eigenfunctions.*

Proof. Recall that a character of a compact abelian group is a continuous homomorphism χ from X into the multiplicative group $S^1 \subset \mathbb{C}$. Finite linear combinations of characters are dense in L^2, and any two distinct characters are orthogonal. Finally, if χ is a character, $\chi(x + \alpha) = \chi(\alpha)\chi(x)$, so χ is an eigenfunction with eigenvalue $\chi(\alpha)$. □

Systems with the property described in Proposition 5.2.2 are said to have discrete spectrum.

Definition 5.2.3. *A measure preserving system $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ has* **discrete spectrum** *if there exists an orthonormal basis of $L^2(X)$ consisting of eigenfunctions.*

It should be clear that if two m.p.s. are conjugate, then they have the same group of eigenvalues. The converse is not true in general, but it is true for systems with discrete spectrum.

Theorem 5.2.4. *Two ergodic systems with discrete spectrum have the same group of eigenvalues if and only if they are conjugate.*

The proof of Theorem 5.2.4 will be omitted, it can be consulted in [15, Theorem 3.4] or in [4, Theorem 6.13].

Remark 5.2.5. *The problem of deciding when two measure preserving systems are conjugate was one of the earliest and most influential problems in ergodic theory. Theorem 5.2.4 is one of the very few results which establish an if and only if condition for conjugacy (for the class of systems with discrete spectrum). The other class of systems for which necessary and sufficient conditions for conjugacy have been found is the class of Bernoulli systems (the condition being famously whether the systems have the same entropy).*

It is now known that for general measure preserving systems it is hopeless to try to obtain useful necessary and sufficient conditions for when two systems are conjugate, as it has been showed that the conjugacy relation is not a Borel set [5].

It turns out that the converse of Proposition 5.2.2 is also true.

Theorem 5.2.6. *A measure preserving system $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ has discrete spectrum if and only if it is conjugate to a group rotation.*

Proof. That every group rotation has discrete spectrum is the content of Proposition 5.2.2. To prove the converse, let $H \subset S^1$ be the group of eigenvalues of $\mathbf{X}$, endowed with the discrete topology. Its Pontryagin dual K is therefore a compact abelian group. Let $\alpha : H \to \mathbb{C}$ be the identity function $\alpha(h) = h$. It is clear that α is a continuous character on H and so $\alpha \in K$. Let $S : K \to K$ be the rotation by α, i.e. $S : k \mapsto k\alpha$. Then the system $\mathbf{K} = (K, \mathrm{Borel}, \mathrm{Haar}, S)$ is a group rotation, and hence, using Proposition 5.2.2, it follows that it has discrete spectrum.

Using Theorem 5.2.4 we need only show that the group of eigenvalues of $\mathbf{K}$ is precisely H in order to finish the proof. The proof of Proposition 5.2.2 tells us that the eigenvalues of $\mathbf{K}$ are precisely the values of $\chi(\alpha)$ as χ runs over the characters of K. But by Pontryagin duality, the characters of $K = \hat{H}$ are the point evaluations at H, and α is the identity map, so the eigenvalues of $\mathbf{K}$ are indeed precisely the elements of H. □

As mentioned after Example 1.3.3, group rotations are sometimes called Kronecker systems. Therefore the terms "system with discrete spectrum", "group rotation" and "Kronecker system" are all synonymous. In the next section we introduce yet another equivalent characterization of such systems.

5.3. **Compact functions.** Given a measure preserving system $\mathbf{X} = (X, \mathcal{B}, \mu, T)$, let $H_{eig} \subset L^2(X)$ denote the closed subspace generated by all the eigenfunctions. Notice that at the very least, H_{eig} contains the one dimensional subspace of constant functions. The system $\mathbf{X}$ has discrete spectrum precisely when $H_{eig} = L^2(X)$. In general, there is a way to tell whether an arbitrary function $f \in L^2(X)$ belongs to H_{eig}, just by looking at its orbit.

Definition 5.3.1. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and let $f \in L^2(X)$. We say that f is a* **compact** *or* **almost periodic** *function if the orbit closure $\overline{\{T^n f : n \in \mathbb{N}\}} \subset L^2$ is compact as a subset of L^2 with the strong topology. The set of all compact functions is denoted by H_c.*

Theorem 5.3.2. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. Then $H_{eig} = H_c$.*

We need the following lemma.

Lemma 5.3.3. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and let $\phi : L^2(X) \times L^2(X) \to L^2(X)$ be a function that commutes with T (so that $\phi(Tf, Tg) = T\phi(f, g)$) and is uniformly continuous.*
Then that for $f, g \in H_c$ also $\phi(f, g) \in H_c$.

We remark that, for instance, the function $\phi(f, g) = f + g$ satisfies the hypothesis.

Proof. Let $f, g \in H_c$ and fix $\epsilon > 0$. Let $\delta > 0$ be such that if $f_1, f_2, g_1, g_2 \in H$ are such that $\|f_1 - f_2\| < \delta$ and $\|g_1 - g_2\| < \delta$ then $\|\phi(f_1, g_1) - \phi(f_2, g_2)\| < \epsilon$.

Next let $\{B_i\}$ and $\{C_i\}$ be finite covers of the orbit closure of f and g (respectively) by balls with diameter less than δ. Then for any $n \in \mathbb{N}$ we have $T^n f \in B_i$ and $T^n g \in C_j$ for some i, j depending on n. Therefore $T^n\phi(f, g) = \phi(T^n f, T^n g) \in \phi(B_i, C_j)$ and by construction $\phi(B_i, C_j)$ has diameter at most ϵ. This implies that the orbit of $\phi(f, g)$ is contained in the union of finitely many sets with diameter ϵ (namely $\{\phi(B_i, C_j)\}_{i,j}$) and therefore is totally bounded. Hence $\phi(f, g)$ is a compact function. □

Proof of Theorem 5.3.2. In view of 5.3.3, the set H_c is a subspace. To show that it is closed, let $f \in \overline{H_c}$, we will show that f is compact. Let $\epsilon > 0$ and let $g \in H_c$ be such that $\|f - g\| < \epsilon/2$. Then $\|T^n f - T^n g\| < \epsilon/2$ for every $n \in \mathbb{N}$. Let $x_1, \ldots, x_n$ be such that the orbit of g is contained in the union of the balls of radius $\epsilon/2$ centered at the points $x_1, \ldots, x_n$. Then the orbit of f is contained in the union of the balls of radius ϵ centered at the points $x_1, \ldots, x_n$, and since ϵ was arbitrary, it follows that the orbit closure of f is compact, and hence H_c is closed.

Since every eigenfunction is trivially compact, it follows from the previous paragraph that $H_{eig} \subset H_c$. To show the reverse inclusion, let $f \in H_c$, and denote by Φ the Koopman operator (i.e. $\Phi g = g \circ T$). Then the orbit closure $Y := \overline{\{\Phi^n f : n \in \mathbb{N}\}} \subset L^2(X)$ is a compact set, invariant under Φ. Since Φ is an isometry, the pair (Y, Φ) is a transitive isometric topological system, and hence Y can be endowed with the structure of an abelian group[4]. Moreover we can take $f \in Y$ to be the identity and $\phi : Y \to Y$ to be a rotation (by the element $\Phi f \in Y$).

Let ν be the Haar measure on Y. For each character $\chi : Y \to S^1$ of Y, the function $f_\chi \in L^2(X)$ defined by $f_\chi = \int_Y \chi(y) y \, \mathsf{d}\nu(y)$ is an eigenfunction for Φ. More generally, for each function $\psi \in L^\infty(Y)$ we can define the function $f_\psi \in L^2(X)$ by $f_\psi = \int_Y \psi(y) y \, \mathsf{d}\nu(y)$. Since finite linear combinations of characters approximate any L^∞ function, we can ψ to be the normalized (i.e. so that $\int_Y \psi \, \mathsf{d}\nu = 1$) indicator function of a small ball around $f \in Y$. As the radius of the ball converges to 0, it is clear that f_ψ is converging to f in L^2. In this way we find a sequence of finite linear combinations of eigenvalues which converges to f in L^2, and hence $f \in H_{eig}$ as desired. □

We now collect all the equivalent characterizations of Kronecker systems we have seen so far.

Corollary 5.3.4. *An ergodic m.p.s.* $\mathbf{X}$ *is a Kronecker system if and only if any of the following equivalent statements hold:*

- $\mathbf{X}$ *has discrete spectrum.*
- $\mathbf{X}$ *is conjugate to a compact group rotation.*
- $\mathbf{X}$ *is a* **compact system**, *i.e. every* $f \in L^2$ *is a compact function.*

5.4. **The Kronecker factor.** Even when a system is not itself a Kronecker system, eigenfunctions still give rise to a factor map into a group rotation.

Proposition 5.4.1. *Let* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *be ergodic and suppose that there exists an eigenfunction* $f \in L^2(X)$ *with* $Tf = e(\alpha) f$ *for some* $\alpha \in (0,1)$. *Then there exists a non-trivial group rotation which is a factor of* $\mathbf{X}$.

Proof. If α is rational, then $\mathbf{X}$ is not totally ergodic and hence by Theorem 4.3.4 it has a finite factor, which can be interpreted as a rotation on a finite group. Next suppose that α is irrational, and let $S : t \mapsto t + \alpha \bmod 1$. We claim that the system $([0,1), \mathcal{A}, \nu, S)$ is factor of $\mathbf{X}$, where $\mathcal{A}$ is the Borel σ-algebra and ν is the Lebesgue measure. Let $\pi : X \to [0,1)$ be such that $f(x) = e(\pi(x))$ for every $x \in X$. Since $f(Tx) = e\big(\pi(x) + \alpha\big) = e\big(S\pi(x)\big)$ for almost every $x \in X$, it follows that $\pi(Tx) = S\pi(x)$. To see that π preserves the measure, we create a new measure $\tilde\nu$ on $[0,1)$ given by $\tilde\nu(A) = \mu(\pi^{-1}A)$. This measure $\tilde\nu$ is invariant under S (because $S \circ \pi = \pi \circ T$ and μ is invariant under T), and therefore it is not hard to check that it must be the Haar measure ν on $[0,1)$. □

Proposition 5.4.1 implies that whenever an ergodic m.p.s. has an eigenfunction, it must have a non-trivial Kronecker factor. It is in fact true that all eigenfunctions of a system can be realized in the same Kronecker system.

Definition 5.4.2. *Let* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *be an ergodic measure preserving system. Its* **maximal Kronecker factor** *(sometimes called only the* **Kronecker factor***) is the factor* $(X, \mathcal{K}, \mu, T)$, *where* $\mathcal{K}$ *is the smallest* σ*-algebra with respect to which every eigenfunction is measurable.*

In the spirit of identifying a factor of $\mathbf{X}$ *with the corresponding* σ*-subalgebra of* $\mathcal{B}$, *we sometimes call* $\mathcal{K}$ *itself the Kronecker factor.*

The claim that $(X, \mathcal{K}, \mu, T)$ is a factor requires that $\mathcal{K}$ be invariant under T, but this is easy to check since for every eigenfunction f the function Tf is also an eigenfunction (with the same eigenvalue). Less obvious is the fact that the maximal Kronecker factor is indeed a Kronecker system, and the maximal factor with this property.

Theorem 5.4.3. *Let* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *be an ergodic system and let* $\mathcal{K} \subset \mathcal{B}$ *be its Kronecker factor. Then* $\mathcal{K} = \{A \in \mathcal{B} : 1_A \in H_{eig}\}$.

Proof. Let $\mathcal{A} := \{A \in \mathcal{B} : 1_A \in H_{eig}\}$. If $A \in \mathcal{B}$ is such that $1_A \in H_{eig}$, then 1_A is approximated by finite linear combinations of eigenfunctions and hence it must be a measurable function with respect to $\mathcal{K}$, which is equivalent to $A \in \mathcal{K}$. This shows that $\mathcal{A} \subset \mathcal{K}$.

To show the reverse inclusion we need to show that $\mathcal{A}$ is a σ-algebra and that every eigenfunction is measurable with respect to $\mathcal{A}$. Since $1 \in \mathcal{A}$, and H_{eig} is a subspace, if follows that $\mathcal{A}$ is closed under

[4]The group structure can be described as follows: for $g, h \in Y$, suppose $\Phi^{n_i} f \to g$ and $\Phi^{m_i} f \to h$. Then $g \bullet h := \lim \Phi^{n_i + m_i} f$.

complements. In view of Theorem 5.3.2, and applying Lemma 5.3.3 to the function $\phi(f,g)=\min(f,g)$ we conclude that $\mathcal{A}$ is closed under finite intersections. But since H_{eig} is closed we conclude that $\mathcal{A}$ is closed under countable intersections and hence $\mathcal{A}$ is a σ-algebra. Since every constant function is in H_{eig}, the fact that every eigenfunction is measurable with respect to $\mathcal{A}$ follows from again applying Lemma 5.3.3 to the function $\phi(f,g)=\min(f,g)$. $\square$

Corollary 5.4.4. *Let $\mathbf{X}=(X,\mathcal{B},\mu,T)$ be an ergodic measure preserving system, let $\mathbf{Y}$ be its maximal Kronecker factor and let $\pi:\mathbf{X}\to\mathbf{Y}$ be the factor map. Then $\mathbf{Y}$ is a Kronecker system and any Kronecker system $\mathbf{Z}$ which is a factor of $\mathbf{X}$ via a factor map $\tilde{\pi}:\mathbf{X}\to\mathbf{Z}$ is also a factor of $\mathbf{Y}$, via a factor map $\pi':\mathbf{Y}\to\mathbf{Z}$ which satisfies $\pi'\circ\pi=\tilde{\pi}$.*

Proof of Corollary 5.4.4. In view of Theorem 5.4.3, the system $\mathbf{Y}$ has discrete spectrum and therefore by Theorem 5.2.4 it is a Kronecker system. By identifying factors with σ-subalgebras of $\mathcal{B}$, the second statement follows immediately from the fact that in a factor which is a Kronecker system, every function is compact (by Theorem 5.3.2), and being a compact function lifts through pullbacks under factor maps. $\square$

5.5. **The Jacobs-de Leeuw-Glicksberg decomposition.** We are now ready to describe the relation between eigenfunctions and weak mixing. First observe that if $(X,\mathcal{B},\mu,T)$ is a measure preserving system and $f\in L^2(X)$ is a (non-constant) eigenfunction, then the function $f\otimes\overline{f}\in L^2(X\times X)$ is invariant and non-constant. It follows that if a system has non-constant eigenfunctions, then it can not be weak mixing. It turns out that the converse is also true, and the goal of this subsection is to prove a strengthening of this statement.

Theorem 5.5.1. *A measure preserving system $\mathbf{X}=(X,\mathcal{B},\mu,T)$ is weak mixing if and only if its Kronecker factor is the trivial system (i.e. the system with only one point).*

Theorem 5.5.1 is an immediate a corollary of a decomposition of $L^2(X)$ (for a general m.p.s.) into two orthogonal components, called the Jacobs-de Leeuw-Glicksberg decomposition, given below in Theorem 5.5.3. Here is a more direct proof.

Proof of Theorem 5.5.1. As observed in the first paragraph of this section, if the system is weak mixing it can not have any eigenfunction, and hence its Kronecker factor must be the trivial system. Conversely, suppose that the system is not weak mixing and let $H\in L^2(X\times X)$ be a bounded non-constant invariant function (for instance, the indicator function of an invariant set). Consider the operator $\phi:L^2(X)\to L^2(X)$ given by

$$(\phi f)(x)=\int_X H(x,y)f(y)\,\mathsf{d}\mu(y).$$

The Hilbert-Schmidt norm[5] is finite, and therefore ϕ is a ***compact operator***, i.e. the image of a bounded set under ϕ is pre-compact (i.e. its closure is compact) in the norm topology. Since H is invariant, it follows that $\phi Tf=T\phi f$ for every $f\in L^2(X)$, and hence the image under ϕ of the orbit $\{T^nf:n\in\mathbb{N}\}$ of f is the orbit of its image. From compactness of ϕ is follows that ϕf is a compact function for every $f\in L^2(X)$. Finally, since H is non-constant, there exists $f\in L^2$ with ϕf not constant. We conclude that H_c is non-trivial and hence $\mathbf{X}$ has a non-trivial Kronecker factor. $\square$

Definition 5.5.2. *Let $(X,\mathcal{B},\mu,T)$ be a m.p.s. and let $f\in L^2(X)$. We say that f is a* **weak-mixing** *function if*

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

The set of all weak-mixing functions is denoted by H_{wm}.

Notice that, in view of Theorem 5.1.5, a system is weak-mixing if and only if every function f with 0 integral is a weak-mixing function.

Here is the Jacobs-de Leeuw-Glicksberg decomposition:

[5] Given a Hilbert space H with an orthonormal basis $\{e_n\}$, and a bounded operator $\phi:H\to H$, the ***Hilbert-Schmidt norm*** of ϕ is $\|\phi\|_{HS}:=\sqrt{\sum_n\|\phi e_n\|^2}$.

Theorem 5.5.3 (Jacobs-de Leeuw-Glicksberg)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s.. Then H_c and H_{wm} are closed invariant subspaces of $L^2(X)$, are orthogonal and $L^2(X) = H_c \oplus H_{wm}$.*

Proposition 5.5.4. *Let $f \in H_{wm}$ and $g \in H$. Then*

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

Proof. We will use the van der Corput trick with $u_n := \overline{\langle T^n f, g\rangle} T^n f$. We have

$$\begin{aligned}
& \lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=0}^{N-1} \langle u_{n+h}, u_n\rangle \right| \\
= \quad & \lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=0}^{N-1} \overline{\langle T^{n+h} f, g\rangle} \langle T^n f, g\rangle \langle T^{n+h} f, T^n f\rangle \right| \\
\leq \quad & \lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H} \|f\|^2 \cdot \|g\|^2 \cdot \left|\langle U^h f, f\rangle\right| \\
= \quad & 0
\end{aligned}$$

By Lemma 4.2.2 we conclude that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \overline{\langle T^n f, g\rangle} T^n f = 0$$

and thus

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\langle T^n f, g\rangle|^2 = \lim_{N\to\infty} \left\langle \frac{1}{N} \sum_{n=1}^{N} \overline{\langle T^n f, g\rangle} T^n f, g \right\rangle = 0$$

□

We can now prove that H_c and H_{wm} are orthogonal sets:

Lemma 5.5.5. *Let $f \in H_{wm}$ and $g \in H_c$. Then $\langle f, g\rangle = 0$.*

Proof. Without loss of generality, assume that $\|f\| = 1$. Fix $\epsilon > 0$. Let $g_1, ..., g_n$ be such that the balls $B(g_i, \epsilon)$ cover the orbit of g. For each m let $i(m)$ be such that $T^m g \in B(g_{i(m)}, \epsilon)$. We have

$$|\langle f, g\rangle| = |\langle U^m f, U^m g\rangle| \leq \epsilon + |\langle U^m f, g_{i(m)}\rangle| \leq \epsilon + \sum_{i=1}^{n} |\langle U^m f, g_i\rangle|$$

Thus we have

$$|\langle f, g\rangle| = \frac{1}{M} \sum_{m=1}^{M} |\langle U^m f, U^m g\rangle| \leq \epsilon + \sum_{i=1}^{n} \frac{1}{M} \sum_{m=1}^{M} |\langle U^m f, g_i\rangle|$$

Since $f \in H_{wm}$, using Proposition 5.5.4 with for M large enough we have that the second term can be made smaller than ϵ. Thus we have $|\langle f, g\rangle| \leq 2\epsilon$ and since $\epsilon > 0$ was arbitrary we conclude that indeed $\langle f, g\rangle = 0$. □

We now need a converse of the previous lemma:

Lemma 5.5.6. *Let $g \in H$ be not weak mixing. Then there exist some $f \in H_c$ such that $\langle f, g\rangle \neq 0$.*

Proof. Let $U : L^2(X) \to L^2(X)$ be the Koopman operator associated with T. Consider the operator $\phi_g : f \mapsto \langle f, g\rangle g$. This is rank one and thus has bounded Hilbert-Schmidt norm. In the Hilbert space HS of all operators on L^2 with bounded Hilbert-Schmidt norm define the unitary operator $V : HS \to HS$ by $V(\psi) = U\psi U^{-1}$. Thus $V(\phi_g) = \phi_{Ug}$.

By the Mean Ergodic Theorem (Theorem 2.3.1), we have that

$$\psi_g := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} V^n \phi_g$$

exists, is in HS and is invariant under V. In other words ψ_g commutes with U.

We will prove that $f = \psi_g g$ satisfies the claims. Note that $\langle (V^n\phi_g)g, g\rangle = |\langle g, U^n g\rangle|^2$ so

$$\langle f, g\rangle = \left\langle \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}(V^n\phi_g)g, g\right\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}|\langle g, U^n g\rangle|^2 > 0$$

The last inequality is from the definition of weak mixing function (which g is not) and we can pass the limit outside the inner product because convergence in the Hilbert-Schmidt norm implies convergence in the weak operator topology.

Now all that remains to prove is that $f \in H_c$. But since ψ_g is an Hilbert-Schmidt operator (and hence compact) and commutes with U, the orbit of $f = \psi_g g$ (under U) is the image under ψ_g of the orbit of g. But since U is unitary, the orbit of g is bounded and therefore the orbit of f is pre-compact. In other words, f is a compact vector.

□

We are now ready to prove the Jacobs-de Leeuw-Glicksberg Decomposition

Proof of Theorem 5.5.3. It follows from Theorem 5.3.2 that H_c is a closed invariant subspace of $L^2(X)$, the fact that H_{wm} is a closed invariant subspace follows easily from the definition and the triangle inequality. By Lemma's 5.5.5 and 5.5.6 we conclude that $f \in H_{wm}$ if and only if it is orthogonal to H_c, hence $H_{wm} = H_c^{\perp}$ and this concludes the proof. □

6. Special cases of Multiple recurrence

As mentioned in Section 3, Szemerédi's theorem (Theorem 3.3.2) has an equivalent formulation in terms of multiple recurrence in measure preserving systems (Theorem 3.3.4); the equivalence is established using Furstenberg's correspondence principle (Lemma 3.2.1) and Bergelson's intersectivity lemma (Lemma 3.1.4). In this section we will go over special cases of Theorem 3.3.4 which illustrate the main ideas that go into the proof. For convenience, here is the statement again.

Theorem 3.3.4. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then for every $k \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that*

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

In fact

$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N}\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > 0. \tag{3.3}$$

6.1. Multiple recurrence in weak mixing systems.

When working in multiple recurrence, several averages will appear. To reduce the notational footprint, the following convention will be convenient.

Definition 6.1.1. *Given a sequence (u_n) of complex numbers, or vectors in a Banach space, we denote by*

$$C\text{-}\lim_n u_n := \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}u_n,$$

assuming the limit exists. We also use

$$UC\text{-}\lim_n u_n := \lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N}u_n,$$

assuming the limit exists. We will also sometimes use some variations such as

$$UC\text{-}\limsup_n u_n := \limsup_{N-M\to\infty}\left|\frac{1}{N-M}\sum_{n=M}^{N}u_n\right|.$$

In this subsection we prove the following theorem.

Theorem 6.1.2. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be a weak-mixing system, let $k \in \mathbb{N}$ and let $f_1, \ldots, f_k \in L^\infty(X)$. Then*

$$UC\text{-}\lim_n \prod_{i=1}^k T^{ni} f_i = \prod_{i=1}^k \int_X f_i \, \mathsf{d}\mu \qquad \text{in } L^2(X). \tag{6.1}$$

Observe that the case $k = 1$ of Theorem 6.1.2 is the mean ergodic theorem. Since strong convergence implies weak convergence, it follows from (6.1) that for any $f_0 \in L^\infty$ we have

$$UC\text{-}\lim_n \int_X f_0 \cdot \prod_{i=1}^k T^{ni} f_i = \prod_{i=0}^k \int_X f_i \, \mathsf{d}\mu.$$

In fact we get this weak convergence to occur in the strong Cesàro sense.

Corollary 6.1.3. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be a weak-mixing system, let $k \in \mathbb{N}$ and let $f_0, \ldots, f_k \in L^\infty(X)$. Then*

$$UC\text{-}\lim_n \left| \int_X f_0 \cdot \prod_{i=1}^k T^{ni} f_i \, \mathsf{d}\mu - \prod_{i=0}^k \int_X f_i \, \mathsf{d}\mu \right| = 0. \tag{6.2}$$

Proof. Splitting f_k as the sum of a constant and function with 0 integral, we reduce the proof to those two cases. If f_k is a constant, then (6.2) follows immediately by induction. Assume now that $\int_X f_k \, \mathsf{d}\mu = 0$. Therefore we need to show that

$$UC\text{-}\lim_n \left| \int_X f_0 \cdot \prod_{i=1}^k T^{ni} f_i \, \mathsf{d}\mu \right|^2 = 0.$$

Expanding the square

$$\begin{aligned} UC\text{-}\lim_n \left| \int_X f_0 \cdot \prod_{i=1}^k T^{ni} f_i \, \mathsf{d}\mu \right|^2 &= UC\text{-}\lim_n \int_{X \times X} \left(f_0 \cdot \prod_{i=1}^k T^{ni} f_i \right) \otimes \overline{\left(f_0 \cdot \prod_{i=1}^k T^{ni} f_i \right)} \, \mathsf{d}(\mu \otimes \mu) \\ &= UC\text{-}\lim_n \int_{X \times X} f_0 \otimes \overline{f_0} \cdot \prod_{i=1}^k (T \times T)^{ni} (f_i \otimes \overline{f_i}) \, \mathsf{d}(\mu \otimes \mu) \end{aligned}$$

Using Theorem 5.1.5 we deduce that the product $\mathbf{X} \times \mathbf{X}$ is weak-mixing. Therefore it follows from Theorem 6.1.2 that

$$UC\text{-}\lim_n \left| \int_X f_0 \cdot \prod_{i=1}^k T^{ni} f_i \, \mathsf{d}\mu \right|^2 = \prod_{i=0}^k \int_{X \times X} f_i \otimes \overline{f_i} \, \mathsf{d}(\mu \otimes \mu) = \prod_{i=0}^k \left| \int_X f_i \, \mathsf{d}\mu \right|^2 = 0.$$

□

Observe that the case $k = 1$ of Corollary 6.1.3 is Condition (4) in Theorem 5.1.5. In this sense, Corollary 6.1.3 can be interpreted as stating that weak-mixing implies weak-mixing of all orders, contrasting with the (still open) problem of whether (strong) mixing implies mixing of all orders.

The next corollary shows that a strong form of Theorem 3.3.4 holds for weak mixing systems.

Corollary 6.1.4. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be a weak-mixing system, let $k \in \mathbb{N}$ and let $A \in \mathcal{B}$. Then*

$$UC\text{-}\lim_n \mu\big(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A\big) = \mu(A)^{k+1}.$$

Proof. Apply Theorem 6.1.2 with all $f_i = 1_A$. □

The main idea in the proof of Theorem 6.1.2 is to use the van der Corput trick.

Proof of Theorem 6.1.2. We proceed by induction on k. The case $k = 1$ follows from the mean ergodic theorem (Theorem 2.3.1). Assume now that $k > 1$ and the result has been established for $k - 1$. Splitting f_k as the sum of a constant and function with 0 integral, we reduce the proof to those two cases. If f_k is a constant, then (6.1) follows immediately by induction.

Assume next that $\int_X f_k \, \mathsf{d}\mu = 0$. Since the right hand side of (6.1) is 0, we will use the van der Corput trick. Let $u_n := \prod_{i=1}^k T^{ni} f_i$. We have

$$\langle u_{n+h}, u_n \rangle = \int_X \prod_{i=1}^k T^{(n+h)i} f_i \cdot \overline{T^{ni} f_i} \, \mathsf{d}\mu = \int_X \prod_{i=1}^k T^{ni}\big(T^{hi} f_i \cdot \overline{f_i}\big) \, \mathsf{d}\mu = \int_X T^h f_1 \cdot \overline{f_1} \prod_{i=2}^k T^{n(i-1)}\big(T^{hi} f_i \cdot \overline{f_i}\big) \, \mathsf{d}\mu,$$

where the last equality follows from the fact that T^n preserves the measure. After using induction hypothesis on the $k-1$ functions $\big\{T^h f_2 \cdot \overline{f_2}, \dots, T^{h(k-1)} f_k \cdot \overline{f_k}\big\}$ and taking averages we get

$$\begin{aligned} UC\text{-}\lim_n \langle u_{n+h}, u_n \rangle &= UC\text{-}\lim_n \int_X T^h f_1 \cdot \overline{f_1} \prod_{i=2}^k T^{n(i-1)}\big(T^{hi} f_i \cdot \overline{f_i}\big) \, \mathsf{d}\mu \\ &= \prod_{i=1}^k \int_X T^{hi} f_i \cdot \overline{f_i} \, \mathsf{d}\mu. \end{aligned}$$

Finally, taking an average on h and using Theorem 5.1.7 and condition (4) from Theorem 5.1.5 we obtain

$$\left| C\text{-}\lim_h UC\text{-}\lim_n \langle u_{n+h}, u_n \rangle \right| = \left| C\text{-}\lim_h \prod_{i=1}^k \int_X T^{hi} f_i \cdot \overline{f_i} \, \mathsf{d}\mu \right| \leq \prod_{i=1}^{k-1} \|f_i\|^2 \cdot C\text{-}\lim_h \left| \langle T^{hk} f_k, f_k \rangle \right| = 0$$

□

6.2. **Multiple recurrence in compact systems.** In this section we prove the following special case of Theorem 3.3.4.

Theorem 6.2.1. *Let* $\mathbf{X}$ *be a compact system (i.e. with discrete spectrum), let* $k \in \mathbb{N}$ *and let* $f \in L^\infty(X)$*. Then for every* $\epsilon > 0$*, the set*

$$\left\{ n \in \mathbb{N} : \int_X \prod_{i=0}^k T^{ni} f \, \mathsf{d}\mu > \int_X f^{k+1} \, \mathsf{d}\mu - \epsilon \right\}$$

is syndetic.

To see why Theorem 6.2.1 implies that Theorem 3.3.4 holds for compact systems, apply Theorem 6.2.1 to the indicator function 1_A of a set $A \in \mathcal{B}$ with $\mu(A) > 0$ and let S be the syndetic set of n for which $\mu(A \cap T^{-n} A \cap \cdots \cap T^{-kn} A) > \mu(A)/2$. Let $L \in \mathbb{N}$ be bound on the gaps of S (so that every interval of length L contains an element of S). Then for $N - M$ large enough we have $|[M, N] \cap S| \geq (N - M - L)/L > (N - M)/(2L)$ and hence

$$UC\text{-}\lim_n \mu(A \cap \cdots \cap T^{-kn} A) \geq \lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n \in S \cap [M,N]} \mu(A \cap \cdots \cap T^{-kn} A) \geq \frac{\mu(A)}{4L} > 0.$$

The following standard lemma from topological dynamics will be used in the proof of Theorem 6.2.1.

Lemma 6.2.2. *Let* X *be a compact metric space, let* $T : X \to X$ *be an isometry and let* $U \subset X$ *be open. Then for every* $x \in X$*, the set* $\{n \in \mathbb{N} : T^n x \in U\}$ *is syndetic.*

Proof of Theorem 6.2.1. Let $\mathbf{X}$, k, f and $\epsilon > 0$ be as in the statement of the theorem. By rescaling we may assume that $\|f\|_\infty \leq 1$. Since f is a compact function, it follows from Lemma 6.2.2 that the set $S := \big\{n \in \mathbb{N} : \|T^n f - f\| < \epsilon/k^2\big\}$ is syndetic. Observe that for each $n \in S$ and $i \in \{0, 1, \dots, k\}$ we have

$$\|T^{in} f - f\| \leq \|T^{in} f - T^{(i-1)n} f\| + \|T^{(i-1)n} f - T^{(i-2)n} f\| + \cdots + \|T^n f - f\| = i\|T^n f - f\| \leq \frac{\epsilon}{k}.$$

Using the Cauchy-Schwarz inequality repeatedly we conclude that for every $n \in S$

$$\int_X \prod_{i=0}^k T^{ni} f \, \mathsf{d}\mu = \int_X f \cdot \prod_{i=1}^k T^{ni} f \, \mathsf{d}\mu \geq \int_X f^2 \cdot \prod_{i=2}^k T^{ni} f \, \mathsf{d}\mu - \frac{\epsilon}{k} \geq \int_X f^3 \cdot \prod_{i=3}^k T^{ni} f \, \mathsf{d}\mu - \frac{2\epsilon}{k} \geq \cdots \geq \int_X f^{k+1} \, \mathsf{d}\mu - \epsilon.$$

□

An alternative way to think about (and prove, at least in the ergodic case) Theorem 6.2.1 is to invoke Corollary 5.3.4, which tells us that $(X, \mathcal{B}, \mu, T)$ is conjugate to a compact abelian group rotation. Therefore, by modifying X in a zero measure set we can assume that X itself is a compact abelian group, and the group of isometries of X is itself compact. Therefore, Lemma 6.2.2 implies that T^n is very close to the identity map on X (uniformly, and hence in the strong operator topology) for a syndetic set of n, and hence all T^{in} for $i = 0, \dots, k$ are close to the identity for a syndetic set of n.

6.3. **An example of a skew-product.** In this subsection we will consider the system $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ where $X = \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ is the two dimensional torus, $\mathcal{B}$ is the Borel σ-algebra, μ is the Haar/Lebesgue measure and $T : X \to X$ is the map $T(x, y) = (x + \alpha, y + x)$, where $\alpha \in \mathbb{T}$ is irrational.

We start by checking that $\mathbf{X}$ is indeed a measure preserving system, in other words, that T does preserve μ. It suffices to show that for every $f \in C(X)$, $\int_X f \, \mathsf{d}\mu = \int_X Tf \, \mathsf{d}\mu$. One can use the Stone-Weierstrass theorem (or Fourier analysis) to reduce this to the case when f is a character of $\mathbb{T}^2$, i.e. a function of the form $f(x, y) = e_{k,m}(x, y) = e(kx + my)$ with $k, m \in \mathbb{Z}$. Now we can directly compute $Te_{k,m}(x, y) = e_{k,m}(T(x, y)) = e_{k,m}(x + \alpha, y + x) = e(kx + n\alpha + my + mx) = e(k\alpha)e_{k+m,m}(x, y)$. In particular, if $(k, m) = (0, 0)$ (i.e. if f is constant) then clearly $\int_X f \, \mathsf{d}\mu = \int_X Tf \, \mathsf{d}\mu$ and otherwise then $\int_X f \, \mathsf{d}\mu = 0 = \int_X Tf \, \mathsf{d}\mu$.

Next we check that $\mathbf{X}$ is ergodic. In order to establish ergodicity, it suffices to show that for every $f \in L^2(X)$, the ergodic averages $\frac{1}{N}\sum_{n=1}^N T^n f$ converge to $\int_X f \, \mathsf{d}\mu$. By approximating an arbitrary L^2 function by finite linear combinations of characters, it suffices to verify this identity when f is a character. Indeed, when $(k, m) \neq (0, 0)$, an easy induction shows that $T^n e_{k,m}(x, y) = e\big(p_{x,y,k,m}(n)\big)$, where

$$p_{x,y,k,m}(n) = \frac{n(n-1)}{2} m\alpha + nmx + nk\alpha + kx + ym$$

is a polynomial with leading coefficient either $\frac{m\alpha}{2}$ (in case $m \neq 0$) or $k\alpha$. In either case, the leading coefficient is irrational, so Corollary 4.1.7 implies that the sequence $n \mapsto p_{x,y,k,m}(n)$ is uniformly distributed, and then Lemma 4.1.2 implies that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N T^n e_{k,m}(x, y) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N e\big(p_{x,y,k,m}(n)\big) = 0.$$

Next we claim that the Kronecker factor of $\mathbf{X}$ is (conjugate to) the rotation by α on $\mathbb{T}$. Indeed, for every $k \in \mathbb{N}$, the function $e_{k,0}$ satisfies $Te_{k,0}(x, y) = e(kx + k\alpha) = e(k\alpha)e_{k,0}(x, y)$, so it is an eigenfunction. The functions $e_{k,0}$ generate the ***vertical*** σ-algebra, $\mathcal{K} := \mathcal{B}_{\mathbb{T}} \otimes \{\emptyset, X\} = \{A \times \mathbb{T} : A \subset \mathbb{T} \text{ is Borel}\} \subset \mathcal{B}$. Observe that $\mathcal{K}$ is T-invariant and hence a factor of $\mathbf{X}$, which is conjugate to the rotation by α on $\mathbb{T}$.

To show that $\mathcal{K}$ is indeed the Kronecker factor, we show that every function orthogonal to $L^2(\mathcal{K})$ is weak mixing. Since every such function is approximated by finite linear combinations of characters of the form $e_{k,m}$ with $k \neq 0$, it suffices to show that every function of the form $e_{k,m}$ is weak mixing. As we saw above, $Te_{k,m} = e(k\alpha)e_{k+m,m}$, so that $T^n e_{k,m} = ce_{k+nm,m}$, where $c = e(k\alpha + n(n-1)m\alpha/2)$ is a constant which depends on k, m, n but not on x, y. Since any two distinct characters are orthogonal, it follows that whenever $n \in \mathbb{N}$ and $m \neq 0$, $e_{m,k}$ and $T^n e_{m,k}$ are orthogonal. This immediately implies that $e_{m,k}$ is weak mixing whenever $m \neq 0$ and hence $\mathcal{K}$ is indeed the Kronocker factor.

6.4. **Roth's theorem.** In 1953, Roth established the following result.

Theorem 6.4.1. *Let $A \subset \mathbb{N}$ have positive upper density. Then it contains a 3 term arithmetic progression.*

As seen in Section 3, the Furstenberg correspondence principle can be used to reduce Theorem 6.4.1 to a statement about multiple recurrence in measure preserving systems:

Theorem 6.4.2. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be a measure preserving system and let $A \in \mathcal{B}$ have $\mu(A) > 0$. Then there exists $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}A \cap T^{-2n}A) > 0$. In fact*

$$\underset{n}{UC\text{-}\lim}\ \mu(A \cap T^{-n}A \cap T^{-2n}A) > 0. \tag{6.3}$$

In this subsection we will prove Theorem 6.4.2 in the case $\mathbf{X}$ is ergodic. The assumption of ergodicity is not a serious one and will be removed in the next section, but it requires the use of the Ergodic Decomposition theorem.

The idea to prove Theorem 6.4.2 is to decompose the indicator function 1_A of A as the sum of a compact function and a weak mixing function $1_A = f_c + f_w$, using Theorem 5.5.3. Then (6.3) splits into 4 terms:

$$\mu(A\cap T^{-n}A\cap T^{-2n}A) = \int_X 1_A\cdot T^n 1_A\cdot T^{2n}1_A\,\mathrm{d}\mu = \int_X 1_A\cdot T^n(f_c+f_w)\cdot T^{2n}(f_c+f_w)\,\mathrm{d}\mu.$$

$$= \int_X 1_A\cdot T^n f_c\cdot T^{2n}f_c\,\mathrm{d}\mu + \int_X 1_A\cdot T^n f_c\cdot T^{2n}f_w\,\mathrm{d}\mu + \int_X 1_A\cdot T^n f_w\cdot T^{2n}f_c\,\mathrm{d}\mu + \int_X 1_A\cdot T^n f_w\cdot T^{2n}f_w\,\mathrm{d}\mu$$

Then using Theorem 6.2.1 we will show that the first term has a positive average, while the other three terms can be shown to have 0 average.

To put everything together into a proof we will need two lemmas.

Lemma 6.4.3. *Let $\mathbf{X} = (X,\mathcal{B},\mu,T)$ be a measure preserving system, let $f\in L^2(X)$ and let $f = f_c + f_w$ be the decomposition of f into a compact and a weak mixing components given by Theorem 5.5.3. If f takes values in $[0,1]$, then so does f_c.*

Proof. It is easy to see that the real part $g_0 := \Re f_c$ is a compact function (as its orbit closure is the real part of the orbit closure of f_c, and the map $\Re : L^2\to L^2$ is continuous). Since H_c contains the constant functions, it follows from Lemma 5.3.3 that the function $g_1 = \min(g_0,1)$ is also in H_c, and so is $g = \max(g_1,0)$. Clearly g takes values in $[0,1]$, and it is also clear that $\|g_1 - f\| \le \|f_c - f\|$, and hence that $\|g - f\| \le \|g_1 - f\| \le \|g_0 - f\| \le \|f_c - f\|$.

We will show that $g = f_c$. Since $f - f_c \in H_{wm}$, we have $f - f_c \perp H_c$. In particular $\langle f - f_c, f_c - g\rangle = 0$ and hence

$$\|f - f_c\|^2 \ge \|f-g\|^2 = \|(f-f_c) + (f_c - g)\|^2 = \|f - f_c\|^2 + \|f_c - g\|^2,$$

which implies that $\|f_c - g\|^2 = 0$ as desired. □

Exercise 6.4.4. *Under the conditions of Lemma 6.4.3, show that if f takes values in a convex set $C\subset\mathbb{C}$, then so does f_c.*

Lemma 6.4.5. *Let $\mathbf{X} = (X,\mathcal{B},\mu,T)$ be an ergodic measure preserving system and let $f,g\in L^2(X)$. If either f or g (or both) is weak mixing, then*

$$UC\text{-}\lim_n T^n f\cdot T^{2n} g = 0 \qquad \textit{in norm.}$$

Proof. We will use the van der Corput trick, in the form of Lemma 4.2.2. Let $u_n = T^n f\cdot T^{2n} g$. We have

$$\langle u_{n+h}, u_n\rangle = \int_X T^{n+h}f\cdot T^{2n+2h}g\cdot T^n\overline{f}\cdot T^{2n}\overline{g}\,\mathrm{d}\mu = \int_X \big(T^h f\cdot \overline{f}\big)\cdot T^n\big(T^{2h}g\cdot\overline{g}\big)\,\mathrm{d}\mu.$$

Using ergodicity and Theorem 2.3.1, taking a uniform Cesàro average in n we get

$$UC\text{-}\lim_n \langle u_{n+h}, u_n\rangle = \int_X T^h f\cdot\overline{f}\,\mathrm{d}\mu\int_X T^{2h}g\cdot\overline{g}\,\mathrm{d}\mu.$$

Since both sequences $h\mapsto\int_X T^h f\cdot\overline{f}\,\mathrm{d}\mu$ and $h\mapsto\int_X T^{2h}g\cdot\overline{g}\,\mathrm{d}\mu$ are bounded (by Cauchy-Schwarz inequality) and the one associated with a weak mixing function is smaller than ϵ in a set of full density (for each $\epsilon>0$) it follows that

$$C\text{-}\limsup_h \big|UC\text{-}\lim_n \langle u_{n+h}, u_n\rangle\big| < \epsilon$$

for every $\epsilon > 0$. This of course means that the limit is 0 and the conclusion follows from Lemma 4.2.2. □

We are now ready to prove Theorem 6.4.2 when the system is ergodic.

Theorem 6.4.6. *Let $\mathbf{X} = (X,\mathcal{B},\mu,T)$ be ergodic and let $A\in\mathcal{B}$ have $\mu(A) > 0$. Then*

$$UC\text{-}\lim_n \mu(A\cap T^{-n}A\cap T^{-2n}A) > 0. \tag{6.4}$$

Proof. Use Theorem 5.5.3 to decompose $1_A = f_c + f_w$ into $f_c\in H_c$ and $f_w\in H_w$. In view of Lemma 6.4.3, f_c takes values in $[0,1]$. Moreover, since $1\in H_c$ and hence $1\perp f_w$, we deduce that $\int_X f_c\,\mathrm{d}\mu = \langle f_c, 1\rangle = \langle 1_A, 1\rangle = \mu(A) > 0$. Therefore f is not a.e. 0 and so we can use Theorem 6.2.1 to deduce that

$$UC\text{-}\lim_n \int_X f_c\cdot T^n f_c\cdot T^{2n} f_c\,\mathrm{d}\mu > 0.$$

Since f_c is measurable with respect to the Kronecker factor $\mathcal{K}$, also $T^n f_c \cdot T^{2n} f_c$ is measurable with respect to $\mathcal{K}$ and therefore it is orthogonal to H_{wm}. In particular, for every $n \in \mathbb{N}$, $\big(T^n f_c \cdot T^{2n} f_c\big) \perp f_w$ and hence

$$UC\text{-}\lim_n \int_X 1_A \cdot T^n f_c \cdot T^{2n} f_c \, \mathsf{d}\mu > 0. \tag{6.5}$$

Next we use Lemma 6.4.5 3 times to deduce that

$$UC\text{-}\lim_n \int_X 1_A \cdot T^n f_w \cdot T^{2n} f_c \, \mathsf{d}\mu = 0. \tag{6.6}$$

$$UC\text{-}\lim_n \int_X 1_A \cdot T^n f_c \cdot T^{2n} f_w \, \mathsf{d}\mu = 0. \tag{6.7}$$

$$UC\text{-}\lim_n \int_X 1_A \cdot T^n f_w \cdot T^{2n} f_w \, \mathsf{d}\mu = 0. \tag{6.8}$$

Finally, adding (6.5), (6.6), (6.7) and (6.8) we obtain (6.4). □

7. Disintegration of measures and the ergodic decomposition

7.1. Conditional expectation.

Definition 7.1.1. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra and let $f \in L^1(\mathcal{B})$. The conditional expectation of f with respect to $\mathcal{D}$ is the function $\mathbb{E}[f \mid \mathcal{D}] \in L^1(\mathcal{D})$ such that for all $D \in \mathcal{D}$ we have*

$$\int_D f \, \mathsf{d}\mu = \int_D \mathbb{E}[f \mid \mathcal{D}] \, \mathsf{d}\mu$$

Informally, the conditional expectation $\mathbb{E}[f \mid \mathcal{D}]$ is the function in $L^1(X, \mathcal{D})$ that better approximates f. This sentence is made more precise below on Theorem 7.1.6. From an information theory point of view, $\mathbb{E}[f \mid \mathcal{D}](x)$ is the best guess for the value of $f(x)$ when all the information we have is $\mathcal{D}$. For instance if we have no information at all (so $\mathcal{D} = \{\emptyset, X\}$), then $\mathbb{E}[f \mid \mathcal{D}]$ is the constant function $\int_X f \, \mathsf{d}\mu$ (in probability theory denoted $\mathbb{E}[f]$), and that's the best guess one can have for $f(x)$. On the other extreme situation, when we have complete information (i.e. when $\mathcal{D} = \mathcal{B}$) then $E[f \mid \mathcal{D}] = f$ and our "guess" for $f(x)$ is $f(x)$ itself.

We need to show that the conditional expectation exists and is unique in the space $L^1(\mathcal{D})$:

Proposition 7.1.2. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra and let $f \in L^1(\mathcal{B})$. The conditional expectation $\mathbb{E}[f \mid \mathcal{D}] \in L^1(\mathcal{D})$ exists and is unique.*

Proof. To prove existence we define the complex measure $\nu : \mathcal{D} \to \mathbb{C}$ by $\nu(D) = \int_D f d\mu$ for every $D \in \mathcal{D}$ (if you are not comfortable with complex measures, split $f = f_1 - f_2 + i f_3 - i f_4$ with each f_i a non-negative real valued function and apply the proof to each f_i separately). It is easy to check that this is indeed a complex measure. Moreover if $D \in \mathcal{D}$ is such that $\mu(D) = 0$ then $\nu(D) = 0$ as well. In other words we have $\nu \ll \mu$. Therefore we can apply the Radon-Nikodym theorem to find a derivative $g = \frac{d\nu}{d\mu} \in L^1(\mathcal{D})$. This means that for every $D \in \mathcal{D}$

$$\int_D g d\mu = \nu(D) = \int_D f d\mu$$

Thus g is a conditional expectation of f with respect to $\mathcal{D}$. To prove uniqueness, assume that $g, h \in L^1(\mathcal{D})$ are both conditional expectations of f with respect to $\mathcal{D}$. Then for each $D \in \mathcal{D}$ we have $\int_D h d\mu = \int_D f d\mu = \int_D g d\mu$ which implies $\int_D h - g d\mu = 0$.

If the set $\{x \in X : h(x) - g(x) \neq 0\}$ has positive measure then without loss of generality the set $\{x \in X : h(x) - g(x) > 0\}$ has positive measure. Hence there is some $\epsilon > 0$ such that the set $D := \{x \in X : h(x) - g(x) > \epsilon\}$ has positive measure. But since both h and g are measurable in $\mathcal{D}$ we conclude that $D \in \mathcal{D}$ and hence

$$0 = \int_D h - g d\mu \geq \epsilon \mu(D) > 0$$

which is a contradiction. This shows the uniqueness of $\mathbb{E}[f \mid \mathcal{D}]$. □

We will need the following basic fact about conditional expectations:

Proposition 7.1.3. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra and let $f \in L^1(\mathcal{B})$ be real valued. The conditional expectation $\mathbb{E}[f \mid \mathcal{D}]$ satisfies*

$$\inf_{x \in X} f(x) \leq \mathbb{E}[f \mid \mathcal{D}](y) \leq \sup_{x \in X} f(x) \qquad a.s.$$

More generally, if f takes values in a convex subset of $\mathbb{C}$, then $\mathbb{E}[f \mid \mathcal{D}]$ also takes values on that convex set, but the proof is technically more cumbersome.

Proof. We prove only the first inequality, the second can be easily derived from the first one by considering the function $-f(x)$. Fix $\epsilon > 0$ and let $D := \left\{ y \in X : \mathbb{E}\left[f \mid \mathcal{D}\right](y) < \inf_{x \in X} f(x) - \epsilon \right\} \in \mathcal{D}$. We have

$$\mu(D) \left(\inf_{x \in X} f(x) - \epsilon \right) \geq \int_D \mathbb{E}\left[f \mid \mathcal{D}\right] d\mu = \int_D f d\mu \geq \mu(D) \inf_{x \in X} f(x)$$

which simplifies to $\epsilon \mu(D) \leq 0$, and thus $\mu(D) = 0$. Since $\epsilon > 0$ was arbitrary we conclude that almost surely $\inf_{x \in X} f(x) \leq \mathbb{E}[f \mid \mathcal{D}](y)$. $\square$

Lemma 7.1.4. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra and let $f \in L^1(X, \mathcal{B})$. Then almost everywhere*

$$\big| \mathbb{E}[f \mid \mathcal{D}] \big| \leq \mathbb{E}\left[|f| \mid \mathcal{D} \right] \tag{7.1}$$

Proof. By splitting f into its real and its imaginary part we may assume that f is real valued. Let $A := \{x \in X : \mathbb{E}[f \mid \mathcal{D}] > 0\}$ and $B := X \setminus A = \{x \in X : \mathbb{E}[f \mid \mathcal{D}] \leq 0\}$. Let $D \in \mathcal{D}$ be the set of points where the inequality (7.1) fails. Let $D^+ = D \cap A$ and let $D^- = D \cap B$. We have

$$\begin{aligned}
\int_D |\mathbb{E}[f \mid \mathcal{D}]| \, d\mu &= \int_{D^+} |\mathbb{E}[f \mid \mathcal{D}]| \, d\mu + \int_{D^-} |\mathbb{E}[f \mid \mathcal{D}]| \, d\mu = \left| \int_{D^+} \mathbb{E}[f \mid \mathcal{D}] d\mu \right| + \left| \int_{D^-} \mathbb{E}[f \mid \mathcal{D}] d\mu \right| \\
&= \left| \int_{D^+} f d\mu \right| + \left| \int_{D^-} f d\mu \right| \leq \int_{D^+} |f| d\mu + \int_{D^-} |f| d\mu \\
&= \int_{D^+} \mathbb{E}[|f| \mid \mathcal{D}] d\mu + \int_{D^-} \mathbb{E}[|f| \mid \mathcal{D}] d\mu = \int_D \mathbb{E}[|f| \mid \mathcal{D}] d\mu.
\end{aligned}$$

Since D is the set of points where the inequality (7.1) fails, we conclude that $\mu(D) = 0$ as desired.

$\square$

Proposition 7.1.5. *Let $(X, \mathcal{B}, \mu)$ be a probability space and let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra. The operator $\mathbb{E}[. \mid \mathcal{D}] : L^1(X, \mathcal{B}) \to L^1(X, \mathcal{D})$ is continuous.*

Proof. We show that actually the norm of the operator is 1: Let $f \in L^1(X, \mathcal{B})$. By Lemma 7.1.4 we have

$$\|\mathbb{E}[f \mid \mathcal{D}]\| = \int_X |\mathbb{E}[f \mid \mathcal{D}]| \, d\mu \leq \int_X \mathbb{E}\left[|f| \mid \mathcal{D}\right] d\mu = \int_X |f| d\mu = \|f\|.$$

$\square$

When $f \in L^2(X, \mathcal{B})$ one can use the Hilbert space structure to give a different characterization of the conditional expectation.

Theorem 7.1.6. *Let $(X, \mathcal{B}, \mu)$ be a probability space and let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra. Let $P : L^2(X, \mathcal{B}) \to L^2(X, \mathcal{D})$ be the orthogonal projection (observe that $L^2(X, \mathcal{D})$ is a closed subspace of $L^2(X, \mathcal{B})$. Then for every let $f \in L^2(X, \mathcal{B})$ we have $\mathbb{E}[f \mid \mathcal{D}] = Pf$.*

Proof. By definition of orthogonal projection, for any function $g \in L^2(X, \mathcal{D})$ we have $\langle f - Pf, g \rangle = 0$. If $D \in \mathcal{D}$ then the indicator function 1_D of D is in $L^2(X, \mathcal{D})$. Therefore

$$\int_D Pf d\mu = \int_X 1_D Pf d\mu = \langle 1_D, Pf \rangle = \langle 1_D, f \rangle = \int_X 1_D f d\mu = \int_D f d\mu$$

and hence $Pf = \mathbb{E}[f \mid \mathcal{D}]$ as desired. $\square$

7.2. **Disintegration of measures.**

Theorem 7.2.1. *Let X be a compact metric space, let $\mathcal{B}$ be the σ-algebra of Borel sets on X and let $\mu : \mathcal{B} \to [0,1]$ be a Radon probability measure. Let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra. Then for almost every $y \in X$ there exists a probability measure μ_y on $(X, \mathcal{B})$ such that for every $f \in L^1(X, \mathcal{B}, \mu)$:*

- *The function $y \mapsto \int_X f(x) d\mu_y(x)$ is in $L^1(X, \mathcal{D})$.*
- *$\mu_y([y]_{\mathcal{D}}) = 1$, where $[y]_{\mathcal{D}} = \bigcap_{y \in D \in \mathcal{D}} D$.*
-
$$\int_X f(x) d\mu(x) = \int_X \left(\int_X f(x) d\mu_y(x) \right) d\mu(y) \tag{7.2}$$

This result applies more generally than to compact metric spaces, but this restriction makes the proof technically easier.

Proof. Since X is a compact metric space, the space $C(X)$ of continuous functions from X to $\mathbb{R}$ is separable (with the topology of uniform convergence, equivalently, the supremum norm). Let $(f_n)_{n=1}^\infty$ be a countable dense set in $C(X)$. We will assume, without loss of generality, that the set $\{f_n : n \in \mathbb{N}\}$ forms a vector space over $\mathbb{Q}$. For each $n \in \mathbb{N}$, the conditional expectation $\mathbb{E}[f_n \mid \mathcal{D}]$ is defined μ-a.e. on X. Thus there is a set of full measure $Y \subset X$ such that $\mathbb{E}[f_n \mid \mathcal{D}]$ is defined on Y for all $n \in \mathbb{N}$.

For each $y \in Y$ define $L_y(f_n) = \mathbb{E}[f_n \mid \mathcal{D}](y)$. Note that, by Proposition 7.1.3 we have

$$|L_y(f_n)| = |\mathbb{E}[f_n \mid \mathcal{D}](y)| \leq \sup_{x \in X} |\mathbb{E}[f_n \mid \mathcal{D}](x)| \leq \sup_{x \in X} |f_n(x)| = \|f_n\|$$

Thus L_y can be extended to a continuous functional on $C(X)$. By the Riesz representation theorem there exists a measure μ_y on X such that $L_y(f) = \int_X f d\mu_y$.

For each $n \in \mathbb{N}$ we have that $\int_X f_n(x) d\mu_y(x) = L_y(f_n) = \mathbb{E}[f_n \mid \mathcal{D}](y)$, hence the function $y \mapsto \int_X f_n(x) d\mu_y(x)$ is in $L^1(X, \mathcal{D})$. Since $(f_n)_{n=1}^\infty$ is a dense set in $C(X)$ and $C(X)$ is a dense set in $L^1(X, \mathcal{B})$, we conclude that $(f_n)_{n=1}^\infty$ is a dense set in $L^1(X, \mathcal{B})$. It follows from Proposition 7.1.5 that the function $y \mapsto \int_X f(x) d\mu_y(x)$ is in $L^1(X, \mathcal{D})$ for any $f \in L^1(X, \mathcal{B})$.

Finally for each $n \in \mathbb{N}$ we have

$$\int_X \int_X f_n(x) d\mu_y(x) d\mu(y) = \int_X L_y(f_n) d\mu(y) = \int_X \mathbb{E}[f_n \mid \mathcal{D}] d\mu = \int_X f_n d\mu$$

and since the sequence $(f_n)_{n=1}^\infty$ is a dense set in $C(X)$ we conclude that (7.2) holds for any $f \in C(X)$.

Now given $f \in L^1(X, \mathcal{B})$, one can find $g_1, g_2, \dots \in C(X)$ such that $f = \sum_{i=1}^\infty g_i$ a.e. and $\sum_{i=1}^\infty \|g_i\|_{L^1} < \infty$. From Proposition 7.1.5 this implies that $\sum_{i=1}^\infty \left\| \mathbb{E}\left[|g_i| \mid \mathcal{D}\right] \right\| < \infty$ and hence $\sum_{i=1}^\infty \mathbb{E}\left[|g_i| \mid \mathcal{D}\right] < \infty$ a.e. For each $x \in X$ for which this series converges and $f = \sum_{i=1}^\infty g_i$ we have

$$\int_X |f| d\mu_y = \int_X \left| \sum_{i=1}^\infty g_i \right| d\mu_y \leq \sum_{i=1}^\infty \int_X |g_i| d\mu_y = \sum_{i=1}^\infty \mathbb{E}\left[|g_i| \mid \mathcal{D}\right](y) < \infty$$

We conclude that $f \in L^1(X, \mathcal{B}, \mu_y)$. Moreover, since (7.2) holds for each g_i, it is easy to deduce that (7.2) holds for f. $\square$

Given a probability space $(X, \mathcal{B}, \mu)$ where X is a compact metric space and μ is a Radon measure, and a σ-subalgebra $\mathcal{D} \subset \mathcal{B}$, we call the family of measures (μ_y) that appear in Theorem 7.2.1 the ***disintegration of μ with respect to $\mathcal{D}$***.

7.3. **The ergodic decomposition.** Given a measurable space $(X, \mathcal{B})$ (i.e. $\mathcal{B}$ is a σ-algebra on X) and a measurable map $T : X \to X$, one can have several measures on $(X, \mathcal{B})$ which are invariant under T, each giving rise of a measure preserving system. We say that a probability measure μ on $(X, \mathcal{B})$ is invariant under T or T-invariant if $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ is a measure preserving transformation. Similarly, we will say that μ is ergodic (or T-ergodic) if the measure preserving system (X, B, μ, T) is ergodic.

The ergodic decomposition theorem states that any T-invariant measure can be decomposed as a convex combination of ergodic measures.

Theorem 7.3.1 (Ergodic Decomposition)**.** *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be a measure preserving system where X is a compact metric space, $\mathcal{B}$ is the Borel σ-algebra and μ is a Radon measure. Then for μ-almost every $y \in X$ there exists a T-invariant, ergodic Radon probability measure ν_y such that for every $f \in L^1(X, \mu)$, the map $y \mapsto \int_X f d\nu_y$ is $\mathcal{B}$-measurable and invariant under T and*

$$\int_X \left(\int_X f(x) d\nu_y(x) \right) d\mu(y) = \int_X f(x) d\mu(x)$$

The last condition can be informally stated as $\mu = \int_X \nu_y d\mu(y)$, i.e., any T-invariant probability is the convex combination of the ergodic measures ν_y.

Example 7.3.2. *Let $X = \{1, 2, 3\}$ be given the discrete topology and discrete σ-algebra $\mathcal{B}$ and let μ be the uniform measure (more precisely, $\mu(\{1\}) = \mu(\{2\}) = \mu(\{3\}) = 1/3$). Let $T(1) = 2$, $T(2) = 1$ and $T(3) = 3$. The set $A = \{1, 2\}$ is invariant under T and $0 < \mu(A) < 1$, hence the system $(X, \mathcal{B}, \mu, T)$ is not ergodic.*

However, if we restrict μ to A and renormalize it, we obtain a probability measure which makes the system ergodic. More precisely, let $\nu(\{1\}) = \nu(\{2\}) = 1/2$ and $\nu(\{3\}) = 0$. Then ν is and ergodic measure, in other words, the system $(X, \mathcal{B}, \nu, T)$ is ergodic.

Also, if ν_3 is the point mass at 3 (so that $\nu_3(\{1\}) = \nu_3(\{2\}) = 0$ and $\nu_3(\{3\}) = 1$), then the system $(X, \mathcal{B}, \nu_3, T)$ is also ergodic (one can also think of ν_3 as the normalized restriction of μ to the invariant set $\{3\}$).

Finally, observe that we can write μ as the convex combination $\mu = \frac{2}{3}\nu + \frac{1}{3}\nu_3$ of the ergodic measures ν and ν_3. If we let $\nu_1 = \nu_2 = \nu$, then we can write informaly $\mu = \int_X \nu_y d\mu(y)$.

Example 7.3.3. *Let $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ be the torus group and let $X = \mathbb{T}^2$ be the unit square with the usual topology and the Borel σ-algebra $\mathcal{B}$. Let μ be the Lebesgue measure on X and let $T(x, y) = (x + \alpha, y)$ where $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is some irrational number. Any set of the form $\mathbb{T} \times B$, where $B \subset \mathbb{T}$ is a Borel set, is invariant under T. Therefore the measure preserving system $(X, \mathcal{B}, \mu, T)$ is not ergodic.*

To try to mimic the previous example, we can take some Borel set $B \subset \mathbb{T}$ such that $0 < \mu(\mathbb{T} \times B) < 1$, and let $\nu = \frac{1}{\mu(\mathbb{T}\times B)} \mu\big|_{\mathbb{T}\times B}$. The probability ν is T-invariant but, unlike in the first example, ν is not ergodic (for any choice of B).

Regardless, it is still quite intuitive what we need to do. Let λ denote the Lebesgue/Haar measure on $\mathbb{T}$. For each $y \in \mathbb{T}$, let ν_y be the measure defined as $\nu_y(B) = \lambda\big(B \cap (\mathbb{T} \times \{y\})\big)$. It is not hard to see that ν_y is T-invariant and ergodic (it is a not completely trivial exercise to verify that it is ergodic. One can show this, for instance, using Fourier analysis). Moreover it follows from Fubini's theorem that

$$\int_X f(x, y) d\mu(x, y) = \int_{\mathbb{T}} \left(\int_{\mathbb{T}} f(x, y) d\lambda(x) \right) d\lambda(y) = \int_{\mathbb{T}} \left(\int_X f(x, z) d\nu_y(x, z) \right) d\lambda(y)$$

for any $f \in L^1(X, \mu)$. To make this decomposition compatible with the notation of Theorem 7.3.1, let $\nu_{(x,y)} = \nu_y$ for all $(x, y) \in X$. Observe that the function $(x, y) \mapsto \int_X f(u) d\nu_{(x,y)}$ does not depend on x. Thus, applying Fubini's theorem again we have

$$\begin{aligned}
\int_X \left(\int_X f(u) d\nu_{(x,y)}(u) \right) d\mu(x, y) &= \int_X \left(\int_X f(u) d\nu_y(u) \right) d\mu(x, y) \\
&= \int_{\mathbb{T}} \left(\int_X f(u) d\nu_y(u) \right) d\lambda(y) \\
&= \int_X f(x, y) d\mu(x, y)
\end{aligned}$$

Example 7.3.4. *Let again $X = \mathbb{T}^2$ with the usual topology and let μ be the Lebesgue measure. Let $T(x, y) = (x + y, y)$. Again, any set of the form $\mathbb{T} \times B$, where $B \subset \mathbb{T}$ is a Borel set, is invariant under T and hence the measure preserving system (X, μ, T) is not ergodic.*

However, unlike the previous example, not all the T-invariant measures ν_y (defined by $\nu_y(B) = \lambda(B \cap \{\mathbb{T} \times \{y\}\})$) are ergodic. Indeed, the set $A = \left(\left[0, \frac{1}{4}\right] \cup \left[\frac{1}{2}, \frac{3}{4}\right] \right) \times \left\{\frac{1}{2}\right\}$ is invariant under T but $\nu_{1/2}(A) = \frac{1}{2}$. This shows that the measure $\nu_{1/2}$ is not ergodic.

In fact the measures ν_y are ergodic exactly when y is irrational (again, this can be proved with some Fourier analysis). Since the set of irrational y have full measure on $\mathbb{T}$, the ergodic decomposition of μ is the same as the one on the previous example, using only the irrational values for y.

However, in this example there are more ergodic measures. Indeed let $\frac{n}{m} \in \mathbb{T}$ be some rational point and let $x \in \mathbb{T}$ be arbitrary. Denote $\left(x, \frac{n}{m}\right)$ by u. Then the probability measure ν_u defined by

$$\nu_u\left(\left\{\left(x, \frac{n}{m}\right)\right\}\right) = \nu_u\left(\left\{\left(x + \frac{1}{m}, \frac{n}{m}\right)\right\}\right) = \cdots = \nu_u\left(\left\{\left(x + \frac{m-1}{m}, \frac{n}{m}\right)\right\}\right) = \frac{1}{m}$$

is T-ergodic. We have now found all ergodic measures for this system, so any T-invariant measure $\tilde{\mu}$ can be decomposed as

$$\int_X f(v) d\tilde{\mu}(v) = \int_X \left(\int_X f(v) d\nu_u(v)\right) d\tilde{\mu}(u)$$

for every $f \in L^1(X, \tilde{\mu})$.

7.4. **Proof of Theorem 7.3.1.** Example 7.3.4 hints that in order to find all the ergodic measures of a given system, one should look at the invariant sets (observe, however, that not all T-invariant sets give an ergodic measure: the set $A := \{(n\pi, \pi) : n \in \mathbb{Z}\} \subset \mathbb{T}^2$ is invariant for the system of Example 7.3.4 and yet no ergodic measure has A as its support).

Proposition 7.4.1. *Let $(X, \mathcal{B}, \mu, T)$ be a probability preserving system and let*

$$\mathcal{I} := \{B \in \mathcal{B} : T^{-1}B = B\}$$

Then $\mathcal{I}$ is a σ-algebra.

Proof. Let $I \in \mathcal{I}$ and let $A = X \setminus I$. Then

$$T^{-1}A = \{x \in X : Tx \in A\} = \{x \in X : Tx \notin I\} = X \setminus \{x \in X : Tx \in I\} = X \setminus I = A$$

and hence $\mathcal{I}$ is closed under complements. Now let $(I_n)_{n=1}^{\infty}$ be a sequence of sets in $\mathcal{I}$ and let $I = \bigcup I_n$. Then

$$\begin{aligned} T^{-1}I &= \{x \in X : Tx \in I\} = \left\{x \in X : Tx \in \bigcup_{n=1}^{\infty} I_n\right\} \\ &= \bigcup_{n=1}^{\infty} \{x \in X : Tx \in I_n\} = \bigcup_{n=1}^{\infty} T^{-1}I_n = \bigcup_{n=1}^{\infty} I_n = I \end{aligned}$$

and hence $\mathcal{I}$ is closed under countable unions and therefore it is a σ-algebra. □

Henceforth we will call $\mathcal{I}$ the ***σ-algebra of invariant sets***. It turns out that the ergodic measures that appear in Theorem 7.3.1 are the measures that arise from the disintegration of μ with respect to the σ-algebra of invariant sets.

Lemma 7.4.2. *Under the conditions of Theorem 7.3.1, let $\mathcal{A} \subset \mathcal{B}$ be a σ-subalgebra and let (ν_y) be the disintegration of μ with respect to $\mathcal{A}$. Then for every $f \in C(X)$ there exists a set of full measure $Y \subset X$ such that for every $y \in Y$ we have*

$$\mathbb{E}[f \mid \mathcal{A}](x) = \mathbb{E}[f \mid \mathcal{A}](y) \qquad \text{for } \nu_y \text{ almost every } x$$

Lemma 7.4.3. *For μ-a.e. y, the measure ν_y that arises from the disintegration of a T-invariant measure μ with respect to the invariant σ-algebra $\mathcal{I}$ is T-invariant and ergodic.*

Proof. We first prove that for almost every y, ν_y is T-invariant. More precisely, we will find a set $Y \subset X$ with $\mu(Y) = 1$ such that for every $y \in Y$, the measure ν_y is T-invariant. Let $D \subset C(X)$ be a countable dense set. It suffices to show that for each $f \in D$ there exists a set $Y_f \subset X$ with $\mu(Y_f) = 1$ and such that for every $y \in Y_f$ we have $\int_X Tf - f \, \mathsf{d}\nu_y = 0$. Recall by the construction of conditional measure that $\int_X Tf - f \, \mathsf{d}\nu_y = \mathbb{E}[Tf - f \mid \mathcal{I}]$, so we need to show that $\mathbb{E}[Tf - f \mid \mathcal{I}] = 0$ μ-a.e. But this follows from the following computation, which holds for each $I \in \mathcal{I}$

$$\int_I Tf - f \, \mathsf{d}\mu = \int_X 1_I \cdot Tf - 1_I \cdot f \, \mathsf{d}\mu = \int_X T\big(1_I \cdot f\big) - 1_I \cdot f \, \mathsf{d}\mu = 0$$

We now show that almost every ν_y is ergodic. It suffices to show that for every $f \in D$, there exists a set $Y_f \subset X$ with $\mu(Y_f) = 1$ and such that for every $y \in Y_f$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f(x) = \int_X f \, \mathsf{d}\nu_y \qquad \text{for } \nu_y \text{ almost every } x$$

Combining the pointwise ergodic theorem (Theorem 2.2.3) with Theorem 7.1.6 it follows that the left hand side equals $\mathbb{E}[f \mid \mathcal{I}](x)$ for every x in a full μ-measure set, while the left hand side is $\mathbb{E}[f \mid \mathcal{I}](y)$ (where the conditional expectations are both taken with respect to the measure μ). The desired conclusion now follows from Lemma 7.4.2. □

Proof of Theorem 7.3.1. Let $\mathcal{I}$ denote the invariant σ-algebra and let $(\nu_y)_{y\in Y}$ be the disintegration of μ with respect to $\mathcal{I}$, for some $Y \in \mathcal{B}$ with $\mu(Y) = 1$. By Lemma 7.4.3 each of the measures ν_y is T-invariant and T-ergodic. By the properties of the disintegration of measures we have that for every $f \in L^1(X, \mu)$, the map $y \mapsto \int_f d\mu$ is $\mathcal{I}$-measurable, and hence it is $\mathcal{B}$-measurable and T-invariant. Moreover, it follows from the properties of the disintegration of measures that

$$\int_X \left(\int_X f(x) d\nu_y(x) \right) d\mu(y) = \int_X f(x) d\mu(x)$$

and this finishes the proof. □

7.5. **Reducing multiple recurrence to ergodic systems.** Let's once again recall the statement of the multiple recurrence theorem of Furstenberg, Theorem 3.3.4.

Theorem 3.3.4. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s. and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then for every $k \in \mathbb{N}$,*

$$UC\text{-}\lim_n \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > 0. \tag{3.3}$$

In this subsection we show how Theorem 3.3.4 can be reduced to the case where the system is ergodic, and overall more convenient.

Theorem 7.5.1. *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic m.p.s., where X is a compact metric space. Let $f \in L^\infty(X)$ be a non-negative function with $\int_X f \, \mathsf{d}\mu > 0$. Then for every $k \in \mathbb{N}$,*

$$UC\text{-}\lim_n \int_X \prod_{i=0}^{k} T^{in} f \, \mathsf{d}\mu > 0.$$

We will prove Theorem 7.5.1 in the next section. Here we will establish its equivalence to Theorem 3.3.4.

Proposition 7.5.2. *Theorems 3.3.4 and 7.5.1 are equivalent.*

Proof. First we show that Theorem 3.3.4 implies Theorem 7.5.1. Observe that for any function f in the conditions of Theorem 7.5.1 there exists $\epsilon > 0$ such that the set $A := \{x \in X : f(x) > \epsilon\}$ has positive measure. Since trivially

$$\int_X \prod_{i=0}^{k} T^{in} f \, \mathsf{d}\mu \geq \epsilon^{k+1} \mu\big(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A\big),$$

this implication follows.

To prove the converse implication, first we claim that we can assume without loss of generality that X is a compact metric space and μ is a Radon measure. Indeed, after choosing the set A we want to study, consider the smallest σ-subalgebra $\mathcal{A}$ of $\mathcal{B}$ which contains A and is invariant under T. This is a countably generated σ-algebra and therefore by Caratheodory's theorem ([13, Theorem 15.3.4]) the probability space $(X, \mathcal{A}, \mu)$ is isomorphic to $(Y, \mathcal{D}, \nu)$ where Y is a compact metric space, $\mathcal{D}$ is the Borel σ-algebra and ν is a Radon measure. Since $\mathcal{A}$ is invariant under T, we can define measure preserving transformation $S : (Y, \mathcal{D}, \nu) \to (Y, \mathcal{D}, \nu)$ and a factor map $\pi : (X, \mathcal{B}, \mu, T) \to (Y, \mathcal{D}, \nu, S)$ of measure preserving systems. If the result is established in Y, it will then hold as well in X.

Assuming now that X is a compact metric space and μ is a Radon measure, we can apply the ergodic decomposition theorem (Theorem 7.3.1) to disintegrate μ into ergodic components $(\mu_y)_{y\in Y}$ for some set

$Y \in \mathcal{B}$ with $\mu(Y) = 1$. Since $0 < \mu(A) = \int_Y \mu_y(A)\,\mathsf{d}\mu(y)$, there exists a positive measure set $B \subset Y$ such that $\mu_y(A) > 0$ for all $y \in B$. Therefore, applying Theorem 7.5.1 to the function 1_A for each $y \in B$,

$$UC\text{-}\lim_n \mu_y\left(\bigcap_{i=0}^k T^{-in}A\right) = UC\text{-}\lim_n \int_Y \prod_{i=0}^k T^{in}1_A\,\mathsf{d}\mu > 0.$$

Integrating over $y \in Y$ it follows that

$$UC\text{-}\lim_n \mu\left(\bigcap_{i=0}^k T^{-in}A\right) = UC\text{-}\lim_n \int_Y \mu_y\left(\bigcap_{i=0}^k T^{-in}A\right)\mathsf{d}\mu(y) = \int_Y UC\text{-}\lim_n \mu_y\left(\bigcap_{i=0}^k T^{-in}A\right)\mathsf{d}\mu(y) > 0,$$

where in the last equality we used the dominated convergence theorem. Observe that if we knew only that the $\liminf$ was positive for a positive measure of $y \in Y$ (but not that the limit existed), it would still be possible to obtain the desired inequality using Fatou's lemma.

□

8. Proof of Furstenberg's Multiple Recurrence theorem

The purpose of this section is to give a proof of Theorem 7.5.1. We will follow Furstenberg's original proof (or more precisely, the proof in [8]), and therefore we will not obtain the full strength of Theorem 7.5.1; instead we will show that under the same conditions we have

$$UC\text{-}\liminf_n \int_X \prod_{i=0}^k T^{in}f\,\mathsf{d}\mu > 0,$$

which is enough to obtain Szemerédi's theorem (and even the strengthening in Corollary 3.3.6).

8.1. Outline of the proof.

Definition 8.1.1 (Sz factor)**.** *A factor $\mathcal{A}$ of $(X, \mathcal{B}, \mu, T)$ is called* **Sz** *if $(X, \mathcal{A}, \mu, T)$ satisfies the conclusion of the Theorem 7.5.1.*

For instance, the factor $\{\emptyset, X\}$ is trivially Sz (the only measurable functions are the constants). Theorem 7.5.1 can be rephrased as saying that $\mathcal{B}$ itself is a Sz factor. The idea to prove Theorem 7.5.1 is to prove that any proper Sz factor is contained in a strictly larger factor which is also Sz.

Definition 8.1.2 (Extension)**.** *Let $\mathcal{A}$ and $\mathcal{D}$ be factors of $(X, \mathcal{B}, \mu, T)$. If $\mathcal{A} \subset \mathcal{D}$, then we say that $\mathcal{D}$ is* **an extension of** *$\mathcal{A}$. We say that $\mathcal{D}$ is a* **non-trivial extension** *of $\mathcal{A}$ if $\mathcal{A} \neq \mathcal{D}$ in the sense that there exists a set $D \in \mathcal{D}$ such that $\mu(D \triangle A) > 0$ for every $A \in \mathcal{A}$.*

In view of Theorem 5.5.1, either the system is weak mixing, in which case Corollary 6.1.4 implies that $\mathcal{B}$ is itself Sz, or there is a non-trivial Kronecker factor, which Theorem 6.2.1 implies is a Sz factor. In either case, we have already established that every ergodic system has a non-trivial factor which is Sz.

Analogous to this, we will introduce two special types of extensions: ***weak-mixing extensions*** and ***compact extensions***. When an extension $\mathcal{A} \subset \mathcal{D}$ is weak-mixing (resp. compact), we sometimes say that the factor $\mathcal{D}$ is weak-mixing (resp. compact) relative to $\mathcal{A}$ or simply that it is a relatively weak mixing (resp. compact) extension of $\mathcal{A}$. We postpone the definitions to later subsections, but one feature of these definitions is that a system $(X, \mathcal{B}, \mu, T)$ is weak-mixing (resp. compact) if and only if $\mathcal{B}$ is a weak-mixing (resp. compact) extension of the trivial factor $\{\emptyset, X\}$.

The following theorem contains the main steps in the proof of Theorem 7.5.1.

Theorem 8.1.3.

- *Let $\mathcal{A} \subset \mathcal{D}$ be an extension between factors of $\mathcal{B}$. If the extension is weak mixing and $\mathcal{A}$ is a Sz factor, then also $\mathcal{D}$ is a Sz factor.*
- *Let $\mathcal{A} \subset \mathcal{D}$ be an extension between factors of $\mathcal{B}$. If the extension is compact and $\mathcal{A}$ is a Sz factor, then also $\mathcal{D}$ is a Sz factor.*
- *Let $\mathcal{A}$ be a factors of $\mathcal{B}$. If the extension $\mathcal{A} \subset \mathcal{B}$ is not weak mixing, then there exists a non-trivial extension $\mathcal{D}$ of $\mathcal{A}$ which is compact.*

In the rest of this subsection we see how Theorem 8.1.3 implies Theorem 7.5.1.

First we present a useful lemma that helps explain why we consider factors of $\mathcal{B}$ instead of arbitrary σ-subalgebra:

Lemma 8.1.4. *Let $f \in L^2(\mathcal{B})$ and let $\mathcal{A} \subset \mathcal{B}$ be a factor. Then $T\,\mathbb{E}[f \mid \mathcal{A}] = \mathbb{E}[Tf \mid \mathcal{A}]$.*

Proof. Note that T is a unitary operator on $L^2(\mathcal{B})$ and the conditional expectation is the orthogonal projection onto $L^2(\mathcal{A})$. Since $\mathcal{A}$ is a factor, it is easy to see that $T\,\mathbb{E}[f \mid \mathcal{A}]$ is also in $L^2(\mathcal{A})$. We will use the Riesz representation theorem, thus it suffices to show that for every $g \in L^2(\mathcal{A})$ we have $\langle T\,\mathbb{E}[f \mid \mathcal{A}], g\rangle = \langle \mathbb{E}[Tf \mid \mathcal{A}], g\rangle$. Indeed

$$\langle T\,\mathbb{E}[f \mid \mathcal{A}], g\rangle = \left\langle \mathbb{E}[f \mid \mathcal{A}], T^{-1}g\right\rangle = \left\langle f, T^{-1}g\right\rangle = \langle Tf, g\rangle = \langle \mathbb{E}[Tf \mid \mathcal{A}], g\rangle$$

The fact that $\mathcal{A}$ is a factor, which implies that also $T^{-1}g \in L^2(\mathcal{A})$, was used in the second equality. □

Next, we need a technical lemma (alternatively one could use some version of Doob's Martingale convergence Theorem).

Lemma 8.1.5. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\{\mathcal{B}_\alpha\}$ be σ-subalgebras of $\mathcal{B}$ totally ordered by inclusion and let $\mathcal{A} = \sigma\left(\bigcup B_\alpha\right)$ be the σ-algebra generated by all $\mathcal{B}_\alpha$. Let $f \in L^\infty(\mathcal{A})$ and $\epsilon > 0$. Then there is some $\mathcal{B}_\alpha$ such that $\|f - \mathbb{E}[f \mid \mathcal{B}_\alpha]\|_{L^2} < \epsilon$.*

Proof. Let $H \subset L^2(\mathcal{B})$ be the closure of the union of the $L^2(\mathcal{B}_\alpha)$, precisely $H = \overline{\bigcup L^2(\mathcal{B}_\alpha)}$, where we view $L^2(\mathcal{B}_\alpha)$ as a subspace of $L^2(\mathcal{B})$. I claim that if $f, g \in H \cap L^\infty(\mathcal{B})$, then also $fg \in H$.

To see this assume without loss of generality that $\|f\|_{L^\infty} = \|g\|_{L^\infty} = 1$, let $\epsilon > 0$ and choose $f_\alpha, g_\alpha \in L^2(\mathcal{B}_\alpha)$ be such that both $\|f - f_\alpha\| < \epsilon$ and $\|g - g_\alpha\| < \epsilon$. Note that multiplying g_α with the characteristic function of the set $\{|g_\alpha| < 2\}$ (which is in $\mathcal{B}_\alpha$) gives a function in $L^2(\mathcal{B}_\alpha)$ closer to g than g_α, thus we can assume that $\|g_\alpha\|_{L^\infty} < 2$. We now have

$$\|fg - f_\alpha g_\alpha\|_{L^2} = \|f(g - g_\alpha) + g_\alpha(f - f_\alpha)\|_{L^2} \leq \|f(g - g_\alpha)\|_{L^2} + \|g_\alpha(f - f_\alpha)\|_{L^2}$$

and also

$$\|f(g - g_\alpha)\|_{L^2}^2 = \int_X |f|^2 |g - g_\alpha|^2 d\mu \leq \int_X |g - g_\alpha|^2 d\mu = \|g - g_\alpha\|_{L^2}^2 < \epsilon^2$$

and

$$\|g_\alpha(f - f_\alpha)\|_{L^2}^2 = \int_X |g_\alpha|^2 |f - f_\alpha|^2 d\mu \leq 4\int_X |f - f_\alpha|^2 d\mu = 4\|f - f_\alpha\|_{L^2}^2 < 4\epsilon^2$$

Putting the last three equations together we get that $\|fg - f_\alpha g_\alpha\|_{L^2} < 5\epsilon^2$, since $\epsilon > 0$ was arbitrary, this proves the claim.

Now let $\mathcal{D}$ be the family of all sets $D \in \mathcal{B}$ such that $1_D \in H$. I claim that $\mathcal{D}$ is a σ-algebra. Let $D \in \mathcal{D}$, then $1 - 1_D \in H$ and so $X \setminus D \in \mathcal{D}$. Now let $\{D_i\}_{i \in \mathbb{N}}$ be any sequence of sets in $\mathcal{D}$. Let $f_i = \prod_{j \leq i} 1_{D_j}$, note that f_i is the characteristic function of the intersection $\bigcap_{j \leq i} D_j$. By the previous claim, $f_i \in H$, and $\lim f_i$ is the characteristic function of the intersection $\bigcap_{i \in \mathbb{N}} D_i$. Since H is closed, we get that the intersection of all D_i is still in $\mathcal{D}$ and hence $\mathcal{D}$ is indeed a σ-algebra as claimed.

Moreover, since any set in any $\mathcal{B}_\alpha$ has its characteristic function in H we conclude that $\mathcal{B}_\alpha \subset \mathcal{D}$, so also $\mathcal{A} \subset \mathcal{D}$. Finally since H is a closed subspace we actually get that $L^2(\mathcal{A}) \subset H$. □

We are now ready to prove Theorem 7.5.1, conditional on Theorem 8.1.3.

Proof. Let Ω be the set of all Sz factors of the system $(X, \mathcal{B}, \mu, T)$, we want to show that $\mathcal{B} \in \Omega$. We can order Ω partially by inclusion, and we will apply Zorn's lemma to find a maximal element in Ω. Let $\{\mathcal{B}_\alpha\}$ be a totally ordered family of Ω. I claim that the σ-algebra $\mathcal{A}$ generated by $\bigcup_\alpha \mathcal{B}_\alpha$ is also Sz.

Let $f \in L^\infty(\mathcal{A})$ be such that $f \geq 0$ and $\int_X f d\mu > 0$. We need to show that, for each $k \in \mathbb{N}$,

$$UC\text{-}\liminf \int_X f(x) f(T^n x) \cdots f(T^{kn} x) d\mu(x) > 0 \tag{8.1}$$

There must exist some $c > 0$ such that the set $E = \{x \in X : f(x) > c\}$ has positive measure, otherwise $\int_X f d\mu$ would be 0. Since

$$\int_X f(x) f(T^n x) \cdots f(T^{kn} x) d\mu(x) \geq c^{k+1} \int_X 1_E(x) 1_E(T^n x) \cdots 1_E(T^{kn} x) d\mu(x),$$

it suffices to assume that f itself is the characteristic function of some set, say $f = 1_E$.

Fix $k \in \mathbb{N}$ and apply Lemma 8.1.5 to find a Sz factor $\mathcal{B}_\alpha$ such that

$$\left\| 1_E - \mathbb{E}[1_E \mid \mathcal{B}_\alpha] \right\|_{L^2} < \frac{\sqrt{\mu(E)}}{2(k+1)}$$

Call $g = \mathbb{E}[1_E \mid \mathcal{B}_\alpha]$ and $h = 1_E - g$. I claim now that $g(x) > 1 - 1/2(k+1)$ in a set of positive measure.

Indeed, if that were not true, we would have $h(x) > 1/2(k+1)$ for all $x \in E$, and hence

$$\frac{\mu(E)}{4(k+1)^2} > \|h\|_{L^2}^2 = \int_X |h|^2 d\mu \geq \int_E |h|^2 d\mu > \frac{\mu(E)}{4(k+1)^2}$$

which is a contradiction. Thus the set $H := \{x \in X : g(x) > 1 - 1/2(k+1)\}$ has positive measure. Also $H \in \mathcal{B}_\alpha$, which is a Sz factor, so

$$c := UC\text{-}\liminf \int_X 1_H(x) 1_H(T^n x) \cdots 1_H(T^{kn} x) d\mu(x) > 0$$

and hence, the set $D := \{n \in \mathbb{N} : \mu(H \cap T^{-n} H \cap \cdots \cap T^{-kn} H) > c/2\}$ is syndetic. For any $n \in D$ denote $H_n := H \cap T^{-n} H \cap \cdots \cap T^{-kn} H$. If $x \in H_n$ then $T^{ni} x \in H$ for all $i = 0, 1, \ldots, k$, and hence

$$\mathbb{E}[T^{ni} 1_E \mid \mathcal{B}_\alpha](x) = \mathbb{E}[1_E \mid \mathcal{B}_\alpha](T^{ni} x) = g(T^{ni} x) > 1 - \frac{1}{2(k+1)}$$

Since each of the functions $T^{in} 1_E$ only takes values in $\{0, 1\}$ we have

$$\prod_{i=0}^{k} T^{ni} 1_E \geq 1 - \sum_{i=0}^{k} (1 - T^{in} 1_E)$$

and taking conditional expectations we get, for $x \in H_n$,

$$\mathbb{E}\left[\prod_{i=0}^{k} T^{ni} 1_E \mid \mathcal{B}_\alpha\right](x) > 1 - \sum_{i=0}^{k} \Big(1 - \mathbb{E}[T^{in} 1_E \mid \mathcal{B}_\alpha](x)\Big) > 1 - \frac{k+1}{2(k+1)} = \frac{1}{2}$$

Finally integrating we obtain

$$\int_X \prod_{i=0}^{k} T^{ni} 1_E d\mu = \int_X \mathbb{E}\left[\prod_{i=0}^{k} T^{ni} 1_E \mid \mathcal{B}_\alpha\right] d\mu > \frac{\mu(H_n)}{2} > \frac{c}{4}$$

Since this happens for every $n \in D$ which is a syndetic set, say with gaps bounded by d, we conclude that

$$UC\text{-}\liminf \mu(E \cap T^{-n} E \cap \cdots \cap T^{-in} E) > \frac{c}{4d} > 0$$

This means that the factor $\mathcal{A}$ is indeed a Sz factor.

We are now in conditions to apply Zorn's lemma to find a maximal Sz factor $\mathcal{D}$. Assume, for the sake of a contradiction, that $\mathcal{D} \neq \mathcal{B}$. If the extension $\mathcal{D} \subset \mathcal{B}$ is weak mixing, then, by the first point of Theorem 8.1.3 also $\mathcal{B}$ is Sz. Otherwise, by the third point of Theorem 8.1.3, there exists an intermediate factor $\mathcal{D}'$ such that the extension $\mathcal{D} \subset \mathcal{D}'$ is compact. But then, by the second point of Theorem 8.1.3, $\mathcal{D}'$ would also be Sz, contradicting the maximality of $\mathcal{D}$. □

8.2. **Weak-mixing and compact extensions.** We need to set up some terminology before defining weak mixing and compact extensions. Recall that throughout this section we assume that an ergodic system $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ is given. Moreover, in this subsection we also suppose that we are given an extension $\mathcal{A} \subset \mathcal{D}$ of factors of $\mathbf{X}$.

Definition 8.2.1 (Conditional inner product)**.** *Let $f, g \in L^2(\mathcal{D})$. We define their conditional inner product by:*

$$\langle f, g \rangle_{(\mathcal{D}|\mathcal{A})} := \mathbb{E}\left[f \overline{g} \mid \mathcal{A}\right] \in L^1(\mathcal{A})$$

Note that if $\mathcal{A} = \{\emptyset, X\}$ is the trivial σ-algebra, then this degenerates to the usual L^2 inner product.

An immediate property of this inner product is that $\langle hf, g\rangle_{(\mathcal{D}|\mathcal{A})} = h\langle f, g\rangle_{(\mathcal{D}|\mathcal{A})}$ for every $f, g \in L^2(\mathcal{D})$ and $h \in L^\infty(\mathcal{A})$. Indeed

$$\langle hf, g\rangle_{(\mathcal{D}|\mathcal{A})} = \mathbb{E}\left[hf\overline{g} \mid \mathcal{A}\right] = h\,\mathbb{E}\left[f\overline{g} \mid \mathcal{A}\right] = h\langle f, g\rangle_{(\mathcal{D}|\mathcal{A})}$$

This conditional inner product gives rise to a conditional norm, and then we can define the conditional Hilbert space (or Hilbert module over $L^\infty(\mathcal{A})$).

Definition 8.2.2 (Hilbert Module)**.** *We define $L^2(\mathcal{D} \mid \mathcal{A})$ to be the subspace of $L^2(\mathcal{D})$ consisting of those functions f for which the conditional norm*

$$\|f\|_{(\mathcal{D}|\mathcal{A})} := \sqrt{\langle f, f\rangle_{(\mathcal{D}|\mathcal{A})}} = \sqrt{\mathbb{E}\left[|f|^2 \mid \mathcal{A}\right]}$$

is in $L^\infty(\mathcal{A})$.

Note that the functions f and $\sqrt{\mathbb{E}\left[|f|^2 \mid \mathcal{A}\right]}$ have the same L^2 norm. Thus, in particular, if $\|f\|^2_{L^2(\mathcal{D}|\mathcal{A})} = 0$ then $f = 0$.

The conditional Cauchy-Schwartz inequality assures us that for $f, g \in L^2(\mathcal{D} \mid \mathcal{A})$, the inner product $\langle f, g\rangle_{(\mathcal{D}|\mathcal{A})}$, which is a priori only in $L^1(\mathcal{A})$, is actually in $L^\infty(\mathcal{A})$.

Proposition 8.2.3 (Conditional Cauchy-Schwartz inequality)**.** *For any functions $f, g \in L^2(\mathcal{D} \mid \mathcal{A})$ we have*

$$\left|\langle f, g\rangle_{(\mathcal{D}|\mathcal{A})}\right| \leq \|f\|_{(\mathcal{D}|\mathcal{A})}\|g\|_{(\mathcal{D}|\mathcal{A})} \qquad a.e.$$

Proof. Most proofs of the usual Cauchy-Schwartz inequality can be relativized do this situation. Note that if $f, g \in L^2(\mathcal{D} \mid \mathcal{A})$ then the function $f\|g\|_{(\mathcal{D}|\mathcal{A})} - g\|f\|_{(\mathcal{D}|\mathcal{A})}$ is also in $L^2(\mathcal{D} \mid \mathcal{A})$. We now use the trivial inequality $x^2 \geq 0$ with this function:

$$\begin{aligned} 0 &\leq \left\| f\|g\|_{(\mathcal{D}|\mathcal{A})} - g\|f\|_{(\mathcal{D}|\mathcal{A})} \right\|^2_{(\mathcal{D}|\mathcal{A})} \\ &= \left\langle f\|g\|_{(\mathcal{D}|\mathcal{A})} - g\|f\|_{(\mathcal{D}|\mathcal{A})}, f\|g\|_{(\mathcal{D}|\mathcal{A})} - g\|f\|_{(\mathcal{D}|\mathcal{A})} \right\rangle \\ &= 2\|f\|^2_{(\mathcal{D}|\mathcal{A})}\|g\|^2_{(\mathcal{D}|\mathcal{A})} - 2\|f\|_{(\mathcal{D}|\mathcal{A})}\|g\|_{(\mathcal{D}|\mathcal{A})}\langle f, g\rangle_{(\mathcal{D}|\mathcal{A})} \end{aligned}$$

After rearranging this gives the desired inequality. ∎

Observe that this also implies a conditional Triangular inequality

Corollary 8.2.4 (Conditional Triangular Inequality)**.** *Let $f, g \in L^2(\mathcal{D} \mid \mathcal{A})$. Then*

$$\|f + g\|_{L^2(\mathcal{D}|\mathcal{A})} \leq \|f\|_{L^2(\mathcal{D}|\mathcal{A})} + \|g\|_{L^2(\mathcal{D}|\mathcal{A})} \qquad a.e.$$

Proof.

$$\begin{aligned} \|f + g\|^2_{L^2(\mathcal{D}|\mathcal{A})} &= \|f\|^2_{L^2(\mathcal{D}|\mathcal{A})} + \|g\|^2_{L^2(\mathcal{D}|\mathcal{A})} + 2\langle f, g\rangle_{L^2(\mathcal{D}|\mathcal{A})} \\ &\leq \left(\|f\|_{L^2(\mathcal{D}|\mathcal{A})} + \|g\|_{L^2(\mathcal{D}|\mathcal{A})}\right)^2 \end{aligned}$$

∎

We define the norm on $L^2(\mathcal{D} \mid \mathcal{A})$ by making $\|f\| = \left\|\|f\|_{L^2(\mathcal{D}|\mathcal{A})}\right\|_{L^\infty}$. Corollary 8.2.4 implies that this is indeed a norm. This turns $L^2(\mathcal{D} \mid \mathcal{A})$ into a complete metric space.

An alternative way to understand the Hilbert module associate with an extension of factors is via disintegration of measures. Indeed let $(\nu_x)_{x\in X}$ be a disintegration of μ over the σ-algebra $\mathcal{A}$. Observe that $\langle f, g\rangle_{(\mathcal{D}|\mathcal{A})}(x) = \int_X f\overline{g}\,\mathsf{d}\nu_x$. Therefore, the conditional norm of f can be described as the function $x \mapsto \sqrt{\int_X |f|^2\,\mathsf{d}\nu_x}$ and the Hilbert Module $L^2(\mathcal{D} \mid \mathcal{A})$ is the space of those functions $f \in L^2(\mathcal{D})$ such that $\|f\|_{L^2(\mathcal{D},\mu_x)}$ is uniformly bounded for μ-almost every $x \in X$.

We can now define weak mixing extensions.

Definition 8.2.5 (Weak mixing extension)**.**

- *A function $f \in L^2(\mathcal{D} \mid \mathcal{A})$ is* **conditionally weak mixing** *if for each $g \in L^2(\mathcal{D} \mid \mathcal{A})$ we have*

$$UC\text{-}\lim_n \left|\langle T^n f, g\rangle_{(\mathcal{D}|\mathcal{A})}\right| = 0 \qquad \text{in } L^2(\mathcal{A})$$

- *The extension is called* **weak mixing** *if every $f \in L^2(\mathcal{D} \mid \mathcal{A})$ such that $\mathbb{E}[f|\mathcal{A}] = 0$ is conditionally weak mixing.*

Example 8.2.6. *If $\mathcal{A} = \{\emptyset, X\}$ is the trivial σ-algebra, then the extension $(\mathcal{D} \mid \mathcal{A})$ is weak mixing if and only if the m.p.s. $(X, \mathcal{D}, \mu, T)$ is a weakly mixing system. Indeed, this follows from Theorem 5.1.5.*

Here is a sufficient condition for an extension to be weak-mixing, using the language of disintegration of measures. This condition is also necessary when the factor is an ergodic system, but the proof of that implication is more difficult and will be omitted since we do not need it.

Proposition 8.2.7. *Let $(\mathcal{D} \mid \mathcal{A})$ be an extension such that the system $(X \times X, \mathcal{D} \otimes D, \nu, T \times T)$ is ergodic, where $\nu := \int_X \nu_x \otimes \nu_x \, \mathsf{d}\mu(x)$. Then the extension $(\mathcal{D} \mid \mathcal{A})$ is weak mixing.*

Proof. The connection between λ and conditional expectation is given by the following computation, which holds for every $f \in L^2(\mathcal{D})$:

$$\left\| \mathbb{E}[f \mid \mathcal{A}] \right\|^2 = \int_X \left| \int_X f \, \mathsf{d}\nu_x \right|^2 \mathsf{d}\mu(x) = \int_X \int_{X \times X} (f \otimes \overline{f}) \, \mathsf{d}(\nu_x \otimes \nu_x) \, \mathsf{d}\mu(x) = \int_{X \times X} f \otimes \overline{f} \, \mathsf{d}\lambda. \tag{8.2}$$

Suppose that $(X \times X, \mathcal{D} \otimes D, \nu, T \times T)$ is an ergodic system and take $f, g \in L^2(\mathcal{D} \mid \mathcal{A})$ such that $\mathbb{E}[f \mid \mathcal{A}] = 0$. We need to show that

$$UC\text{-}\lim_n |\langle T^n f, g\rangle_{(\mathcal{D}|\mathcal{A})}|^2 = 0 \qquad \text{in } L^2(\mathcal{A}) \tag{8.3}$$

On the other hand, since $\langle T^n f, g\rangle_{(\mathcal{D}|\mathcal{A})} = \mathbb{E}[T^n f \overline{g} \mid \mathcal{A}]$ we can use the triangular inequality in $L^2(\mathcal{A})$ to deduce that (8.3) will follow from

$$UC\text{-}\lim_n \left\| \mathbb{E}[T^n f \overline{g} \mid \mathcal{A}] \right\|^2_{L^2} = 0 \tag{8.4}$$

Using (8.2) and applying the ergodic theorem in $(X \times X, \mathcal{D} \otimes D, \nu, T \times T)$ we conclude that

$$\begin{aligned} UC\text{-}\lim_n \left\| \mathbb{E}[T^n f \overline{g} \mid \mathcal{A}] \right\|^2_{L^2} &= UC\text{-}\lim_n \int_{X \times X} (\overline{g} \otimes g)(T \times T)^n (f \otimes \overline{f}) d\lambda \\ &= \int_{X \times X} (\overline{g} \otimes g) d\lambda \int_{X \times X} (f \otimes \overline{f}) d\lambda \\ &= \int_{X \times X} (\overline{g} \otimes g) d\lambda \int_X \mathbb{E}[f \mid \mathcal{A}]^2 d\mu \\ &= 0. \end{aligned}$$

□

It takes a little more effort to define compact extensions:

Definition 8.2.8 (Compact extension)**.**

- *A subset $C \subset L^2(\mathcal{D} \mid \mathcal{A})$ is* **conditionally pre-compact** *if for every $\epsilon > 0$ there are finitely many functions $f_1, ..., f_r \in L^2(\mathcal{D} \mid \mathcal{A})$ such that for any $g \in C$*

$$\left\| \min_{1 \le t \le r} \|f_t - g\|_{(\mathcal{D}|\mathcal{A})} \right\|_{L^\infty(\mathcal{A})} < \epsilon.$$

- *A function $f \in L^2(\mathcal{D} \mid \mathcal{A})$ is* **conditionally compact** *if the orbit $C = \{T^n f : n \in \mathbb{Z}\}$ is conditionally pre-compact.*
- *The extension is called compact if conditionally compact functions are dense in $L^2(\mathcal{D})$.*

We stress the subtlety that, in the definition of conditionally pre-compact set, the choice of f_i depends on each $x \in X$.

Example 8.2.9. *The first example of a compact extension is when $\mathcal{A} = \{\emptyset, X\}$ is the trivial σ-algebra and $(X, \mathcal{D}, \mu, T)$ is (isomorphic to) a rotation on a compact group, i.e. X is a compact metrizable group, $\mathcal{D}$ is the Borel σ-algebra, μ is the Haar measure and $Tx = ax$ for some $a \in X$.*

Indeed, in this case, the conditional norm coincides with the L^2 norm, and hence the extension is compact if and only in for each function $f \in L^2(\mathcal{D})$, the orbit $\{T^n f : n \in \mathbb{N}\}$ is pre-compact in the L^2 norm.

A less trivial example is the skew-product.

Example 8.2.10. *Let $X = \mathbb{T}^2$, let $\mathcal{B}$ be the Borel σ-algebra in $\mathbb{T}$ and let $\mathcal{D} = \mathcal{B} \otimes \mathcal{B}$ be the Borel σ-algebra on X. Let μ be the Haar measure on $(X, \mathcal{D})$, let $\alpha \in \mathbb{T}$ and define $T : X \to X$ by $T(x, y) = (x + \alpha, y + x)$. Let $\mathcal{A} := \{A \times \mathbb{T} : A \in \mathcal{B}\}$ be the vertical σ-algebra. Then $\mathcal{A}$ is a factor of $\mathcal{D}$ and the extension $(\mathcal{D} \mid \mathcal{A})$ is a compact extension.*

Proof. Since $T^{-1}(A \times \mathbb{T}) = (T^{-1}A) \times \mathbb{T}$ we deduce that $\mathcal{A}$ is indeed a factor. To see that the extension is compact recall that $L^2(\mathcal{D})$ has the following orthonormal basis formed by characters $e_{j,k} : (x, y) \mapsto e(jx+ky)$, with $j, k \in \mathbb{Z}$. We claim that each character $e_{j,k}$ is conditionally compact.

It follows from Fubini's theorem that, for every $f_1, f_2 \in L^2(\mathcal{D} \mid \mathcal{A})$ and any point $(x, y_0) \in X$ we have

$$\|f_1 - f_2\|^2_{(\mathcal{D}|\mathcal{A})}(x, y_0) = \int_{\mathbb{T}} \big|f_1(x, y) - f_2(x, y)\big|^2 dy \tag{8.5}$$

Now, fix $\epsilon > 0$ and let $\lambda_1, \cdots, \lambda_r \in \mathbb{C}$ with each $|\lambda_t| = 1$ be an ϵ-net of the unit circle. Let $f_t : y \mapsto \lambda_t e(ky)$. For every $(x, y_0) \in X$ and $n \in \mathbb{N}$, applying (8.5) and a simple computation we have

$$\|T^m e_{j,k} - f_t\|^2_{(\mathcal{D}|\mathcal{A})}(x, y_0) = \int_{\mathbb{T}} \left| \Big[e\big(x(j + nk) + \alpha(jn + kn(n-1)/2)\big) - \lambda_t \Big] e(ky) \right|^2 dy.$$

Thus, choosing the appropriate λ_t we conclude that $\|T^m e_{j,k} - f_t\|_{(\mathcal{D}|\mathcal{A})} < \epsilon$ and hence every character is conditionally compact. It is easy to see that finite linear combinations of conditionally compact functions are still conditionally compact, and hence we found a dense subset of $L^2(\mathcal{D} \mid \mathcal{A})$ formed by compact functions. $\square$

The notation $\lfloor x \rfloor$ for a real number x, denotes the largest integer n such that $n \leq x$.

Exercise 8.2.11. *Let $X, \mathcal{D}, \mu, T$ and $\mathcal{A}$ be as in Example 8.2.10. Show that the function $f(x, y) = e(y\lfloor 1/x \rfloor)$ is not conditionally compact with respect to $\mathcal{A}$.*

8.3. **Sz lifts through weak mixing extensions.** The purpose of this subsection is to prove the following theorem:

Theorem 8.3.1. *If $(\mathcal{D} \mid \mathcal{A})$ is a weak mixing extension of a measure preserving system $(X, \mathcal{B}, \mu, T)$ and $\mathcal{A}$ is Sz, then also $\mathcal{D}$ is Sz.*

Throughout this subsection we assume that $(\mathcal{D} \mid \mathcal{A})$ is a weak mixing extension of a measure preserving system $(X, \mathcal{B}, \mu, T)$ and $\mathcal{A}$ is Sz.

Remark 8.3.2. *If an extension is weak mixing for T, then it is also weak mixing for T^i for each $i \in \mathbb{N}$. To see this note that*

$$\frac{1}{N - M} \sum_{n=N}^{M} |\langle T^{in} f, g \rangle_{(\mathcal{D}|\mathcal{A})}| \leq i \frac{1}{Ni - Mi} \sum_{n=Mi}^{Ni} |\langle T^n f, g \rangle_{(\mathcal{D}|\mathcal{A})}|$$

The next lemma asserts that weak mixing extensions are also ergodic extensions:

Lemma 8.3.3. *Let $f \in L^2(\mathcal{D})$ be such that $\mathbb{E}[f \mid \mathcal{A}] = 0$. Then*

$$UC\text{-}\lim_n T^n f = 0 \qquad \text{in } L^2(\mathcal{D})$$

Proof. By the ergodic theorem we know that this limit is the projection of f onto the space of invariant functions. Thus it suffices to show that any T-invariant function $g \in L^2(\mathcal{D})$ is measurable with respect to $\mathcal{A}$.

Let g be a T-invariant function and let $g_0 = g - \mathbb{E}[g \mid \mathcal{A}]$. By Lemma 8.1.4, $Tg_0 = Tg - \mathbb{E}[Tg \mid \mathcal{A}] = g - \mathbb{E}[g \mid \mathcal{A}] = g_0$, so g_0 is also T-invariant. Moreover g_0 is weak mixing and so

$$0 = UC\text{-}\lim_n \big|\langle T^n g_0, g_0 \rangle_{(\mathcal{D}|\mathcal{A})}\big| = \|g_0\|^2_{(\mathcal{D}|\mathcal{A})}$$

Thus $g_0 = 0$ and hence $g = \mathbb{E}[g \mid \mathcal{A}]$. Therefore every T-invariant function is indeed measurable with respect to $\mathcal{A}$ and this proves the result. $\square$

The following lemma is the key to show that weak mixing extensions of Sz systems are Sz.

Lemma 8.3.4. *Assume that the extension is weak mixing. Let $f_1, ..., f_k \in L^\infty(\mathcal{D})$ and assume that for some $i \in \{1, \ldots, k\}$ we have $\mathbb{E}[f_i \mid \mathcal{A}] = 0$. Then*

$$UC\text{-}\lim_n \prod_{i=1}^k T^{in} f_i = 0 \qquad \textit{in } L^2(\mathcal{D})$$

To get this lemma, we need to induct on some stronger hypothesis, and hence we will instead prove the following general case:

Lemma 8.3.5. *Assume that the extension is weak mixing. Let $f_1, ..., f_k \in L^\infty(\mathcal{D})$ and let $a_1, \cdots, a_k \in \mathbb{Z}$ be distinct and non-zero. Assume that, for some $i \in \{1, \ldots, k\}$, we have $\mathbb{E}[f_i \mid \mathcal{A}] = 0$. Then*

$$UC\text{-}\lim_n \prod_{i=1}^k T^{a_i n} f_i = 0 \qquad \textit{in } L^2(\mathcal{D})$$

Proof. We can (and will) assume, without loss of generality, that $\mathbb{E}[f_k \mid \mathcal{A}] = 0$, since otherwise we can just permutate the values of a_i. We proceed by induction on k. For $k = 1$ we can use Remark 8.3.2 to deduce that the extension is weak mixing for T^{a_1}, and hence we can assume that $a_1 = 1$. But now, this case reduces to Lemma 8.3.3.

Now let $k > 1$ and let $u_n = \prod_{i=1}^k T^{a_i n} f_i$. We are going to apply the van der Corput trick (Lemma 4.2.2). We have

$$\begin{aligned}
\langle u_{n+d}, u_n \rangle &= \int_X \prod_{i=1}^k T^{a_i(n+d)} f_i \prod_{i=1}^k T^{a_i n} f_i d\mu \\
&= \int_X \prod_{i=1}^k T^{a_i n} \left(T^{a_i d} f_i . f_i\right) d\mu \\
&= \int_X T^{a_1 n} \left[T^{a_1 d} f_1 . f_1 \prod_{i=2}^k T^{(a_i - a_1)n} \left(T^{a_i d} f_i . f_i\right) \right] d\mu \\
&= \int_X T^{a_1 d} f_1 . f_1 \prod_{i=2}^k T^{(a_i - a_1)n} \left(T^{a_i d} f_i . f_i\right) d\mu
\end{aligned}$$

Let $g_d = T^{a_k d} f_k . f_k$ and $g'_d = g_d - \mathbb{E}[g_d \mid \mathcal{A}]$. By the induction hypothesis we have

$$UC\text{-}\lim_n \prod_{i=2}^{k-1} T^{(a_i - a_1)n} \left(T^{a_i d} f_i . f_i\right) \cdot T^{(a_k - a_1)n} g'_d = 0 \tag{8.6}$$

On the other hand we have the trivial bound

$$\left\| T^{a_1 d} f_1 . f_1 \prod_{i=2}^{k-1} T^{(a_i - a_1)n} \left(T^{a_i d} f_i . f_i\right) \right\|_{L^\infty} \leq \prod_{i=1}^{k-1} \|f_i\|_{L^\infty}^2 =: b$$

Taking the uniform Cesàro limit of the inner product $\langle u_{n+d}, u_n \rangle$ and using the previous estimate and (8.6) (and the triangular inequality) we have

$$\begin{aligned}
&\limsup_{N-M \to \infty} \left| \frac{1}{N-M} \sum_{n=M}^N \langle u_{n+d}, u_n \rangle \right| \\
\leq\ & \limsup_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^N b \int_X \left| T^{(a_k - a_1)n} \mathbb{E}[g_d \mid \mathcal{A}] \right| d\mu \\
=\ & b \limsup_{N-M \to \infty} \frac{1}{N-M} \sum_{n=M}^N \int_X \left| \mathbb{E}[g_d \mid \mathcal{A}] \right| d\mu \\
=\ & b \int_X \left| \mathbb{E}[g_d \mid \mathcal{A}] \right| d\mu
\end{aligned}$$

Finally we have that

$$
\begin{aligned}
& \limsup_{D\to\infty} \frac{1}{D}\sum_{d=1}^{D} \limsup_{N\to\infty} \left| \frac{1}{N}\sum_{n=1}^{N} \langle u_{n+d}, u_n\rangle \right| \\
\leq \quad & b \limsup_{D\to\infty} \frac{1}{D}\sum_{d=1}^{D} \int_X |\mathbb{E}[g_d \mid \mathcal{A}]| \, d\mu \\
= \quad & b \limsup_{D\to\infty} \int_X \frac{1}{D}\sum_{d=1}^{D} \left| \langle T^{a_k d} f_k, f_k\rangle_{L^2(\mathcal{D}|\mathcal{A})} \right| d\mu \\
= \quad & 0
\end{aligned}
$$

where in the last line we used the definition of weak mixing extension (Definition 8.2.5). □

We can now prove Theorem 8.1.3. Let $f \in L^\infty(\mathcal{D})$ be a non-negative function such that $\int_X f d\mu > 0$ and let $k \in \mathbb{N}$. We need to show that

$$
UC\text{-}\liminf_n \int_X \prod_{i=0}^{k} T^{in} f \, \mathsf{d}\mu > 0. \tag{8.7}
$$

Let $g = \mathbb{E}[f \mid \mathcal{A}]$ and note that $\int_X g d\mu = \int_X f d\mu > 0$ and $g \geq 0$ a.e., by Proposition 7.1.3. Also let $h = f - g$. Replacing each appearance of f in (8.7) with $(g + h)$ and opening up the product we end up with a sum of 2^{k+1} products. Each of those products but one contain some copy of h. In view of Lemma 8.3.4 each such term vanishes after taking the uniform Cesàro limit. Therefore we conclude that

$$
UC\text{-}\liminf_n \int_X \prod_{i=0}^{k} T^{in} f \, \mathsf{d}\mu \geq UC\text{-}\liminf_n \int_X \prod_{i=0}^{k} T^{in} g \, \mathsf{d}\mu > 0,
$$

where the last inequality follows from the fact that $\mathcal{A}$ is Sz.

8.4. **Sz lifts through compact extensions I.** We will present (almost) two proofs that the Sz property lifts through compact extensions. In this section we prove the analogous result for regular (i.e. not-uniform) Cesàro averages. We will give a different proof in the next section which works for uniform Cesàro averges. We remark that Furstenberg's original proof was neither of the ones presented here: similarly to the first proof presented here, Furstenberg made use of van der Waerden's theorem, but similarly to the second proof, Furstenberg was able to deal with uniform Cesàro averages.

The purpose of this subsection is therefore to prove the following theorem:

Theorem 8.4.1. *If $(\mathcal{D} \mid \mathcal{A})$ is a compact extension of a measure preserving system $(X, \mathcal{B}, \mu, T)$ and*

$$
C\text{-}\liminf_n \int_X \prod_{i=0}^{k} T^{ni} f \, \mathsf{d}\mu > 0 \tag{8.8}
$$

holds for every $k \in \mathbb{N}$ and every non-negative $f \in L^\infty(\mathcal{A})$ with $\int_X f \, \mathsf{d}\mu > 0$, then (8.8) also holds for every non-negative $f \in L^\infty(\mathcal{D})$ with $\int_X f \, \mathsf{d}\mu > 0$.

As mentioned above, we will need the van der Waerden Theorem. We use the convenient notation $[K] = \{0, \ldots, K-1\}$.

Theorem 8.4.2. *(van der Waerden) Let $r, k \in \mathbb{N}$. There exists some $K(r,k) \in \mathbb{N}$ such that for any coloring of $\big[K(r,k)\big]$ in r colors there exists a monochromatic arithmetic progression of length k.*

We start with a lemma:

Lemma 8.4.3. *If $(\mathcal{D} \mid \mathcal{A})$ is a compact extension, for each $f \in L^2(\mathcal{D} \mid \mathcal{A})$ and each $\epsilon > 0$ there is a subset $A \in \mathcal{A}$ such that $\mu(A) > 1 - \epsilon$ and $f 1_A$ is conditionally compact.*

Proof. Let $\epsilon > 0$ and, for each $\ell \in \mathbb{N}$, set $\epsilon_\ell := \epsilon/2^\ell$ and let f_ℓ be a conditionally compact function such that $\|f - f_\ell\|_{L^2(\mathcal{D})} < \epsilon_\ell$. Let $A_\ell := \left\{x \in X : \mathbb{E}\left[|f-f_\ell|^2 \mid \mathcal{A}\right](x) > \epsilon_\ell\right\}$ and let $A^c := \bigcup_\ell A_\ell$. Since $A_\ell \in \mathcal{A}$ we have

$$\epsilon_\ell^2 > \int_X |f-f_\ell|^2 \,\mathsf{d}\mu \geq \int_{A_\ell} |f-f_\ell|^2 \,\mathsf{d}\mu = \int_{A_\ell} \mathbb{E}\left[|f-f_\ell|^2 \mid \mathcal{A}\right] \mathsf{d}\mu \geq \epsilon_\ell \mu(A_\ell)$$

whence it follows that $\mu(A_\ell) \leq \epsilon_\ell$ for every $\ell \in \mathbb{N}$, and therefore that $\mu(A^c) < \epsilon$.

Next let $A := X \setminus A^c$. To show that $f1_A$ is conditionally compact, let $\delta > 0$, take ℓ sufficiently large so that $\epsilon_\ell < \delta/2$ and, using the fact that f_ℓ is conditionally compact, let $g_1 \dots, g_r \in L^2(\mathcal{D} \mid \mathcal{A})$ be such that for every $n \in \mathbb{N}$,

$$\left\| \min_{1\leq t\leq r} \|g_t - T^n f_\ell\|_{(\mathcal{D}|\mathcal{A})} \right\|_{L^\infty(\mathcal{A})} < \delta/2.$$

Using the triangle inequality for the conditional norm (Corollary 8.2.4) we have

$$\min_{1\leq t\leq r} \left\|g_t - T^n(f1_A)\right\|_{(\mathcal{D}|\mathcal{A})} \leq \min_{1\leq t\leq r} \left\|g_t - T^n f_\ell\right\|_{(\mathcal{D}|\mathcal{A})} + \left\|T^n f_\ell - T^n(f1_A)\right\|_{(\mathcal{D}|\mathcal{A})} \qquad \text{a.e.} \tag{8.9}$$

Since

$$\left\|T^n f_\ell - T^n(f1_A)\right\|_{(\mathcal{D}|\mathcal{A})} = \mathbb{E}\left[\left|T^n(f_\ell - f1_A)\right|^2 \mid \mathcal{A}\right] = T^n\, \mathbb{E}\left[|f_\ell - f1_A|^2 \mid \mathcal{A}\right]$$

it follows that the quantity in (8.9) is smaller than δ at every point x such that $T^n x \in A$. On the other hand, if $T^n x \notin A$, then $\left\|T^n(f1_A)\right\|_{(\mathcal{D}|\mathcal{A})}(x) = 0$. It follows that, letting $g_0 \equiv 0$,

$$\min_{0\leq t\leq r} \left\|g_t - T^n(f1_A)\right\|_{(\mathcal{D}|\mathcal{A})} < \delta \qquad \text{a.e.}$$

$\square$

Proof of Theorem 8.4.1. Let $f \in L^\infty(\mathcal{D})$ be such that $f \geq 0$ and $\int_X f d\mu > 0$; and let $k \in \mathbb{N}$. Let $\epsilon > 0$ be such that if $A \in \mathcal{A}$ has $\mu(A) > 1-\epsilon$ then $\int_A f d\mu > 0$ and, using Lemma 8.4.3, let $A \in \mathcal{A}$ be such that $\mu(A) > 1-\epsilon$ and $f1_A$ is conditionally compact. Note that if (8.8) holds for $f1_A$, since $f \geq f1_A$ then it also holds for f. Therefore we can (and will) assume, without loss of generality that f is itself conditionally compact. We will also assume that $0 \leq f \leq 1$ after multiplying f, if necessary, by a constant.

Take $\epsilon > 0$ very small (how small will be determined later, but it depends only on k and f). Applying the compactness hypothesis we can find $r \in \mathbb{N}$ and $f_1, ..., f_r$ be such that for each $n \in \mathbb{N}$ we have

$$\min_{1\leq i\leq r} \|T^n f - f_i\|_{L^2(\mathcal{D}|\mathcal{A})} < \epsilon \qquad \text{a.e..}$$

Next, applying the van der Waerden Theorem (Theorem 8.4.2) we find $K \in \mathbb{N}$ so that each r-coloring of $\{1, ..., K\}$ contains an arithmetic progression of length $k+1$. For each $n \in \mathbb{N}$ and almost every $x \in X$ consider the r-coloring of $[K]$ induced by the orbit $T^{an} f$. More precisely the coloring is the map from $a \in [K]$ to some $i \in [r]$ that satisfies $\|T^{an} f - f_i\|_{L^2(\mathcal{D}|\mathcal{A})}(x) < \epsilon$. By the choice of K we can find some $a = a(x, n)$, $s = s(x, n) \in \mathbb{N}$ and $i = i(x, n) \in [r]$ such that $a + ks \leq K$ and $\|T^{(a+ts)n} f - f_i\|_{L^2(\mathcal{D}|\mathcal{A})}(x) < \epsilon$ for each $t = 0, ..., k$. Note that for each fixed a, s, i (and n), the set of those $x \in X$ such that $\|T^{(a+ts)n} f - f_i\|_{L^2(\mathcal{D}|\mathcal{A})}(x) < \epsilon$ for all $t = 0, ..., k$ is in $\mathcal{A}$, so we can choose the functions a, r, i to be measurable on $\mathcal{A}$. Using the conditional triangular inequality (Corollary 8.2.4) we also have that

$$\forall n \in \mathbb{N} \quad \forall t \in [k] \qquad \|T^{(a+ts)n} f - T^{an} f\|_{L^2(\mathcal{D}|\mathcal{A})}(x) < 2\epsilon \qquad \text{for } \mu \text{ a.e. } x \in X$$

Recall that $\|f\|_{L^\infty} \leq 1$ and that the norm $\|\cdot\|_{L^2(\mathcal{D}|\mathcal{A})}$ is bounded by the L^∞ norm. Using the Conditional Cauchy-Schwartz inequality (Proposition 8.2.3) several times we estimate

$$\mathbb{E}\left[\prod_{t=0}^k T^{n(a+ts)} f \;\middle|\; \mathcal{A}\right](x) \geq \mathbb{E}\left[\left(T^{na} f\right)^{k+1} \;\middle|\; \mathcal{A}\right](x) - 2(k+1)\epsilon \tag{8.10}$$

Next, let $g = \mathbb{E}[f \mid \mathcal{A}]$ and note that also $\int_X g d\mu = \int_X f \,\mathsf{d}\mu > 0$ and, in view of Proposition 7.1.3, $0 \leq g \leq 1$ a.e. Using Jensen inequality for conditional expectation we can get a lower bound on the right hand side of (8.10).

$$\mathbb{E}\left[(T^{na} f)^{k+1} \middle| \mathcal{A}\right](x) \;=\; \mathbb{E}\left[f^{k+1} \middle| \mathcal{A}\right](T^{na} x) \;\geq\; \Big(\mathbb{E}\left[f \mid \mathcal{A}\right](T^{na} x)\Big)^{k+1} \;=\; \Big(g(T^{na} x)\Big)^{k+1}$$

Recall that a and s depend on both n and x. However, they take values in the finite set $[K]$. Therefore

$$\Big(g(T^{na}x)\Big)^{k+1} \geq \prod_{i=0}^{K} \Big(g(T^{ni}x)\Big)^{k+1},$$

and putting the last three equations together we get

$$\mathbb{E}\left[\prod_{t=0}^{k} T^{n(a+ts)}f \;\middle|\; \mathcal{A}\right] \geq \prod_{i=0}^{K} T^{ni}g^{k+1} - 2(k+1)\epsilon \qquad \text{a.e.} \tag{8.11}$$

Since $g \in L^\infty(\mathcal{A})$ (because $f \in L^2(\mathcal{D} \mid \mathcal{A})$), also $g^{k+1} \in L^\infty(\mathcal{A})$, and moreover $g^{k+1} \geq 0$ and $\int_X g^{k+1}\,\mathsf{d}\mu > 0$. Therefore we can apply the hypothesis to deduce that

$$C\text{-}\liminf_n \int_X \prod_{i=0}^{K} T^{in}g^{k+1}\,\mathsf{d}\mu > 0$$

For each $n \in \mathbb{N}$, let $a(n), s(n) \in [K]$ and $A_n \in \mathcal{A}$ be such that $a(n,x) = a(n)$ and $s(n,x) = s(n)$ for every $x \in A_n$, and

$$\int_{A_n} \prod_{i=0}^{K} T^{in}g^{k+1}\,\mathsf{d}\mu \geq \frac{1}{K^2}\int_X \prod_{i=0}^{K} T^{in}g^{k+1}\,\mathsf{d}\mu.$$

We then have, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\int_X \prod_{t=0}^{k} T^{n(a(n)+ts(n))}f\,\mathsf{d}\mu &\geq \int_{A_n} \prod_{t=0}^{k} T^{n(a(n)+ts(n))}f\,\mathsf{d}\mu = \int_{A_n} \mathbb{E}\left[\prod_{t=0}^{k} T^{n(a(n)+ts(n))}f \;\middle|\; \mathcal{A}\right]\mathsf{d}\mu \\
&\geq \int_{A_n} \prod_{i=0}^{K} T^{ni}g^{k+1}\,\mathsf{d}\mu - 2(k+1)\epsilon \geq \frac{1}{K^2}\int_X \prod_{i=0}^{K} T^{ni}g^{k+1}\,\mathsf{d}\mu - 2(k+1)\epsilon
\end{aligned}$$

On the other hand,

$$\int_X \prod_{t=0}^{k} T^{n(a(n)+ts(n))}f\,\mathsf{d}\mu = \int_X \prod_{t=0}^{k} T^{nts(n)}f\,\mathsf{d}\mu$$

so that, making ϵ small enough,

$$C\text{-}\liminf_n \int_X \prod_{t=0}^{k} T^{nts(n)}f\,\mathsf{d}\mu > 0.$$

Finally, since $s(n)$ takes values in the finite set $[K]$ we conclude that

$$C\text{-}\liminf_n \int_X \prod_{t=0}^{k} T^{nt}f\,\mathsf{d}\mu \geq \frac{1}{K^2} C\text{-}\liminf_n \int_X \prod_{t=0}^{k} T^{nts(n)}f\,\mathsf{d}\mu > 0.$$

□

8.5. **Sz lifts through compact extensions 2.** In this section we present another proof that the multiple recurrence lifts through compact extensions. More precisely, we will prove the following theorem.

Theorem 8.5.1. *If $(\mathcal{D} \mid \mathcal{A})$ is a compact extension of a measure preserving system $(X,\mathcal{B},\mu,T)$ and $\mathcal{A}$ is Sz, then also $\mathcal{D}$ is Sz.*

The proof we will present was obtained by Furstenberg, Katznelson and Ornstein and avoids the use of van der Waerden's theorem, while working with uniform Cesàro limits.

Let $f \in L^\infty(\mathcal{D})$ be nonnegative and such that $\int_X f d\mu > 0$. Using Lemma 8.4.3, we can assume without loss of generality that f is itself conditionally compact. Fix $\epsilon > 0$ to be a very small real number to be determined later, depending only on k and f.

Given $x \in X$ and distinct $a, b \in \mathbb{Z}$, let's call the pair $\{a,b\}$ ***good*** (with respect to x) if for every $i = 0,\dots,k$ we have $\|T^{ia}f - T^{ib}f\|_{(\mathcal{D}|\mathcal{A})}(x) < \epsilon$. A finite set $F \subset \mathbb{Z}$ is said to be ***unfortunate*** (with respect to x) if no pair $\{a,b\} \subset F$ is good. Since f is conditionally compact, large sets cannot be unfortunate:

Lemma 8.5.2. *There exists $M \in \mathbb{N}$ depending only on f, k and ϵ such that if $F \subset \mathbb{Z}$ has $|F| \geq M$, then there exists $Y \in \mathcal{A}$ with $\mu(Y) = 1$ and for every $x \in Y$, the set F is not unfortunate.*

Proof. Let $r \in \mathbb{N}$ and $f_1, ..., f_r$ be such that for each $n \in \mathbb{N}$ and almost every $x \in X$ there is some $t \in [r]$ such that

$$\|T^n f - f_t\|_{L^2(\mathcal{D}|\mathcal{A})} < \frac{\epsilon}{2} \tag{8.12}$$

Let $Y \in \mathcal{A}$ be the full measure set where (8.12) holds. Let $x \in Y$ and $F \subset \mathbb{Z}$ be a finite subset. For each $a \in F$ and $i = 0, 1, \ldots, k$ there exists some $t \in [r]$ that satisfies (8.12) with $n = ai$. Thus we can associate each $a \in F$ with a sequence $t_0, t_1, \ldots t_k$ in $[r]$. Since there are only r^{k+1} possible such sequences, if $|F| > r^{k+1}$ there must be some pair $a, b \in F$ with the same sequence. Using the conditional triangular inequality we conclude that $\|T^{ia} f - T^{ib} f\|_{L^2(\mathcal{D}|\mathcal{A})}(x) < \epsilon$ for every $i = 0, \ldots, k$, i,e, the pair $\{a, b\}$ is good. ☐

Given $F \subset \mathbb{Z}$ (and keeping $\epsilon > 0$ fixed), let

$$\begin{aligned} S(F) &= \big\{x \in X : F \text{ is unfortunate with respect to } x\big\} \\ &= \bigcap_{a \neq b \in F} \bigcup_{i \in [k]} \big\{x \in X : \|T^{ia} f - T^{ib} f\|_{(\mathcal{D}|\mathcal{A})}(x) \in [\epsilon, \infty)\big\} \in \mathcal{A} \end{aligned}$$

We want to look at maximal unfortunate sets. Thus we define:

$$Q(F) = \{x \in S(F) : x \notin S(E) \text{ for all } E \subset \mathbb{Z} \text{ with } |E| > |F|\} \in \mathcal{A}$$

Lemma 8.5.3. *Almost every $x \in X$ is in some $Q(F)$, i.e.* $\mu\left(\bigcup_{\substack{F \subset \mathbb{Z} \\ 0 < |F| < M}} Q(F)\right) = 1$

Proof. On the one hand $S(F) = X$ whenever F is a singleton. On the other hand, in view of Lemma 8.5.2, if $|F| \geq M$, the set $S(F)$ has measure 0. For each $x \in Y$ let $F \subset \mathbb{Z}$ be an arbitrary unfortunate set (w.r.t. x) with maximal cardinality (among all unfortunate sets w.r.t. x). Then $|F| \leq M$ and in particular F is finite, so that $x \in Q(F)$. ☐

Let $g = \mathbb{E}[f \mid \mathcal{A}]$ and denote by $\delta := \int_X f d\mu$. Let $A := \{x : g(x) > \delta/2\} \in \mathcal{A}$. Observe that $\delta = \int_A g d\mu + \int_{X \setminus A} g d\mu \leq \mu(A) + \frac{\delta}{2}$, hence $\mu(A) > \delta/2$. Since there are only countably many finite subsets of $\mathbb{Z}$, we can find a finite set $F \subset \mathbb{Z}$ such that $\mu\big(Q(F) \cap A\big) > 0$. Since $Q(F) \subset S(F)$, for all $x \in Q(F)$ the set F is unfortunate, and hence no pair $a \neq b \in F$ is good. It follows that there exists some $i = i(x, a, b) \in [k]$ such that $\|T^{ia} f - T^{ib} f\|_{(\mathcal{D}|\mathcal{A})}(x) > \epsilon$. Thus we can find a set of positive measure $E \subset Q(F) \cap A$ and a map $q : F^2 \to [k]$ such that for all $x \in E$ and every distinct $a, b \in F$ we have $\|T^{q(a,b)a} f - T^{q(a,b)b} f\|_{(\mathcal{D}|\mathcal{A})}(x) > \epsilon$.

Lemma 8.5.4. *Let $n \in \mathbb{Z}$ and $x \in E \cap T^{-n} E \cap T^{-2n} E \cap \cdots \cap T^{-kn} E$ (in particular assume this intersection is non-empty). Then there exists $a = a(n, x) \in F$ s.t. $\forall i \in [k]$ we have $\|T^{i(n+a)} f - f\|_{(\mathcal{D}|\mathcal{A})}(x) < \epsilon$ (i.e. the pair $\{a + n, 0\}$ is good).*

Proof. Since for each $t \in [k]$ we have $T^{tn} x \in E$, we have that for any pair $a \neq b \in F$ there exists $i = q(a, b)$ independent of t such that $\|T^{ia} f - T^{ib} f\|_{(\mathcal{D}|\mathcal{A})}(T^{tn} x) > \epsilon$ and hence $\|T^{tn+ia} f - T^{tn+ib} f\|_{(\mathcal{D}|\mathcal{A})}(x) > \epsilon$. In particular we can take $t = i$ and we have

$$\|T^{i(n+a)} f - T^{i(b+n)} f\|_{(\mathcal{D}|\mathcal{A})}(x) > \epsilon \tag{8.13}$$

On the other hand, $E \subset Q(F)$, so $x \notin S(G)$ for any $G \subset \mathbb{Z}$ with $|G| > F$. We will make $G = (F + n) \cup \{0\}$. If $0 \in F + n$, then $-n \in F$ and hence, taking $a = -n$, the conclusion of the lemma is trivially true. Otherwise $|G| > |F|$ and hence there exists some good pair $a, b \in G$.If neither a nor b equal 0, then $a = a' + n$ and $b = b' + n$ with $a', b' \in F$. But this would contradict (8.13). Thus either $a = 0$ or $b = 0$ and we are done. ☐

Now we can finish the proof. By assumption we get that 1_E satisfies (8.7). Let $c_0 > 0$ be the lim inf in (8.8) with 1_E, let $I_n = E \cap T^{-n} E \cap ... \cap T^{-kn} E$ and let $B := \{n \in \mathbb{N} : \mu(I_n) > c_0/2\}$. We have

$$\begin{aligned} c &= UC\text{-}\liminf_n \mu(I_n) = UC\text{-}\liminf_n \big[1_B(n)\mu(I_n) + 1_{B^c}(n)\mu(I_n)\big] \\ &\leq \liminf_{N - M \to \infty} \frac{1}{N - M} \big|B \cap \{M, \ldots, N\}\big| + \frac{c_0}{2} \end{aligned}$$

and thus

$$\liminf_{N-M\to\infty}\frac{\big|B\cap\{M,\dots,N\}\big|}{N-M}\geq\frac{c_0}{2}$$

For each $n\in B$ and $x\in I_n$ we can apply Lemma 8.5.4 and find $a(n,x)\in F$. Using the conditional Cauchy-Schwarz inequality and the Jensen's inequality we obtain

$$\mathbb{E}\left[\prod_{i=0}^k T^{i(n+a)}f\,\middle|\,\mathcal{A}\right](x)\geq\mathbb{E}[f^{k+1}\mid\mathcal{A}](x)-k\epsilon\geq g^{k+1}(x)-k\epsilon\geq\left(\frac{\delta}{2}\right)^{k+1}-k\epsilon=:c_1.$$

By choosing now ϵ sufficiently small we can make $c_1>0$ (observe that c_1 depends only on f and k). While a depends on x, it can only take finitely many possible values. Since the function $x\mapsto a(n,x)$ is measurable with respect to A we can find an $\mathcal{A}$-measurable subset $J_n\subset I_n$ where $a(n,\cdot)$ is constant and

$$\int_{J_n}\mathbb{E}\left[\prod_{i=0}^k T^{i(n+a)}f\,\middle|\,\mathcal{A}\right]\mathsf{d}\mu\geq\frac{1}{|F|}\int_{I_n}\mathbb{E}\left[\prod_{i=0}^k T^{i(n+a)}f\,\middle|\,\mathcal{A}\right]\mathsf{d}\mu\geq\frac{\mu(I_n)}{|F|}c_1=:c_2>0.$$

Since $J_n\in\mathcal{A}$ and $f\geq 0$ we can then integrate and deduce that

$$\int_X\prod_{i=0}^k T^{i(n+a)}f\,d\mu\geq\int_{J_n}\prod_{i=0}^k T^{i(n+a)}f\,d\mu\geq c_2.$$

Let $a_1=\min F$ and $a_2=\max F$. Let $M<N\in\mathbb{Z}$ be such that $N-M>a_2-a_1$. Then we have

$$\begin{aligned}\sum_{m=M}^N\int_X\prod_{i=0}^k T^{im}f\,d\mu &= \frac{1}{|F|}\sum_{a\in F}\sum_{n=M-a}^{N-a}\int_X\prod_{i=0}^k T^{i(n+a)}f\,d\mu\\ &\geq \frac{1}{|F|}\sum_{a\in F}\sum_{n=M-a_1}^{N-a_2}\int_X\prod_{i=0}^k T^{i(n+a)}f\,d\mu\\ &= \frac{1}{|F|}\sum_{n=M-a_1}^{N-a_2}\int_X\sum_{a\in F}\prod_{i=0}^k T^{i(n+a)}f\,d\mu\\ &\geq \frac{c_2}{|F|}|B\cap[M-a_1,N-a_2]|\end{aligned}$$

Finally, taking the uniform Cesàro limit we conclude:

$$UC\text{-}\liminf_m\int_X\prod_{i=0}^k T^{im}f\,d\mu\geq\frac{c_2}{|F|}\liminf_{N-M\to\infty}\frac{|B\cap[M-a_1,N-a_2]|}{N-M}=\frac{c_0c_2}{2|F|}>0$$

8.6. **The dichotomy between compactness and weak mixing.** In this subsection we complete the proof of Theorem 7.5.1, by establishing the third piece of Theorem 8.1.3.

Proposition 8.6.1. *Let $(X,\mathcal{B},\mu,T)$ be a measure preserving system and let $\mathcal{A}\subset\mathcal{B}$ be factor. If the extension $(\mathcal{B}\mid\mathcal{A})$ is not weak mixing, then there exists a non-trivial intermediate factor $\mathcal{A}\subset\mathcal{D}\subset\mathcal{B}$ such that the extension $(\mathcal{D}\mid\mathcal{A})$ is compact.*

The first step is to show that, if the extension $(\mathcal{B}\mid\mathcal{A})$ is not weak mixing then there exists at least one function $f\in L^2(\mathcal{B}\mid\mathcal{A})$ which is conditionally compact with respect to $\mathcal{A}$, but not in $L^2(\mathcal{A})$ (observe that any function in $L^2(\mathcal{A})$ is trivially conditionally compact).

It follows from Proposition 8.2.7 that the ***relatively independent self joining*** $(X\times X,\mathcal{B}\otimes\mathcal{B},\nu,T\times T)$ is not ergodic, where ν is the measure on $X\times X$ defined by $\nu:=\int_X\nu_x\otimes\nu_x\,\mathsf{d}\mu(x)$, and $(\nu_x)_{x\in X}$ is the disintegration of μ with respect to the factor $\mathcal{A}$. Let $H\in L^\infty(\mathcal{B}\otimes\mathcal{B},\nu)$ be a $T\times T$-invariant function which is not a constant (ν almost everywhere). Since $(X,\mathcal{B},\mu,T)$ is an ergodic system, the function $H:X\times X\to\mathbb{C}$ must depend non-trivially on both coordinates.

Since $H\in L^\infty(\mu_x\times\mu_x)$, for almost every x one can define an operator $\Phi_x:L^2(\nu_x)\to L^2(\nu_x)$ given by

$$(\Phi_x f)(z)=\int_X H(z,y)f(y)d\nu_x(y).$$

Moreover, this operator is a Hilbert-Schmidt operator, and thus compact. We can glue the Φ_x together to obtain an operator $\Phi : L^2(\mathcal{B} \mid \mathcal{A}) \to L^2(\mathcal{B} \mid \mathcal{A})$ defined (almost everywhere) by $(\Phi f)(x) = (\Phi_x f)(x)$. Since H is invariant under $T \times T$, it follows that $\Phi(Tf) = T(\Phi f)$.

Since H depends non-trivially on both coordinates, it follows that there exists some $f \in L^2(\mathcal{B} \mid \mathcal{A})$ such that Φf does not belong to $L^2(\mathcal{A})$.

Next, we will show that Φf is conditionally compact. Indeed $f \in L^2(\mathcal{B} \mid \mathcal{A})$, which implies that $B := \big\| \|f\|_{L^2(\mathcal{B}|\mathcal{A})} \big\|_{L^\infty(\mathcal{A})} < \infty$ or, equivalently, that $\|f\|_{L^2(\mu_x)} \leq B$ for almost every x. Moreover, $\|Tf\|_{L^2(\mu_x)} = \|f\|_{L^2(\mu_{Tx})} < B$, and we deduce that the set $\{T^n f : n \in \mathbb{Z}\}$ is a bounded subset of $L^2(\mu_x)$, for almost every $x \in X$. Since each $\Phi_x : L^2(\mu_x) \to L^2(\mu_x)$ is compact, we deduce that, for almost every $x \in X$, the orbit

$$\{T^n(\Phi f) : n \in \mathbb{Z}\} = \Phi\big(\{T^n f : n \in \mathbb{Z}\}\big) \subset L^2(\mathcal{B}, \mu_x)$$

is pre-compact. However, the number of functions necessary to ϵ-cover the orbit of Φf in $L^2(\mu_x)$ may depend on x.

Define, for every $\epsilon > 0$, the function $M = M_\epsilon : X \to \mathbb{N}$ such that $\big\{\Phi f, T(\Phi f), \ldots, T^{M(x)}(\Phi f)\big\}$ is ϵ-dense in the whole orbit with respect to the $L^2(\mathcal{B}, \mu_x)$ norm. In symbols, we have

$$M(x) := \min\Big\{M \in \mathbb{N} : (\forall n \in \mathbb{Z})(\exists j \in [M]) \text{ s.t. } \big\|T^n(\Phi f) - T^j(\Phi f)\big\|_{L^2(\mathcal{B},\mu_x)} < \epsilon\Big\}$$

It is easy to check that M is measurable with respect to $\mathcal{A}$, and it follows from the compactness of each Φ_x that M is almost everywhere finite. Let $M_0 \in \mathbb{N}$ be such that the set $A \in \mathcal{A}$ defined by $A := \{x \in X : M(x) \leq M_0\}$ has positive measure. We will show that the orbit of Φf in $L^2(\mathcal{B}, \mu_x)$ can be ϵ-covered by M_0 functions, for ***almost every*** $x \in X$. Indeed, for each $j = 0, \ldots, M_0$, define the function

$$g_j(x) = (\Phi f)(T^{m+j} x)$$

where $m = m(x) = \min\{i \geq 0 : T^i x \in A\}$. Observe that, by ergodicity, $m(x)$ and hence $g_j(x)$ is defined almost everywhere. Take $n \in \mathbb{Z}$ and $x \in X$ such that g_j is defined and let $m = m(x)$. Since m is measurable w.r.t. $\mathcal{A}$, we have

$$\begin{aligned}
\min_{j \in [M_0]} \big\|T^n(\Phi f) - g_j\big\|^2_{L^2(\mathcal{B},\mu_x)} &= \min_{j \in [M_0]} \big\|T^n(\Phi f) - T^{m+j}(\Phi f)\big\|^2_{L^2(\mathcal{B},\mu_x)} \\
&= \min_{j \in [M_0]} \mathbb{E}\Big[\big|T^{n-m}(\Phi f) - T^j(\Phi f)\big|^2 \;\Big|\; \mathcal{A}\Big](T^m x) \leq \epsilon
\end{aligned}$$

We just proved that Φf is indeed conditionally compact. In particular, it follows that the subspace

$$F = \{f \in L^\infty(\mathcal{B}) : f \text{ is conditionally compact}\}$$

is not contained in $L^2(\mathcal{A})$. We claim that F is in fact closed under products. Indeed, let $f, g \in F$, let $\epsilon > 0$ and choose $f_1, \ldots, f_r, g_1, \ldots, g_r \in L^2(\mathcal{B} \mid \mathcal{A})$ such that for every $n \in \mathbb{Z}$ we have

$$\Big\| \min_{1 \leq t \leq r} \|T^n f - f_t\|_{L^2(\mathcal{B}|\mathcal{A})} \Big\|_{L^\infty} < \frac{\epsilon}{\|g\|_{L^\infty}} \quad \text{and} \quad \Big\| \min_{1 \leq t \leq r} \|T^n g - g_t\|_{L^2(\mathcal{B}|\mathcal{A})} \Big\|_{L^\infty} < \frac{\epsilon}{\|f\|_{L^\infty}}$$

Observe that truncating each f_t by $\|f\|_{L^\infty}$ does not alter the above inequality, so we will simply assume that $\|f_t\|_{L^\infty} \leq \|f\|_{L^\infty}$. Therefore for each $n \in \mathbb{Z}$, choosing t and s appropriately, we have

$$\begin{aligned}
\big\|T^n(fg) - f_t g_s\big\|_{(\mathcal{B}|\mathcal{A})}(x) &\leq \big\|T^n(fg) - f_t T^n g\big\|_{(\mathcal{B}|\mathcal{A})}(x) + \big\|f_t T^n g - f_t g_s\big\|_{(\mathcal{B}|\mathcal{A})}(x) \\
&\leq \|T^n g\|_{L^\infty} \|T^n f - f_t\|_{L^2(\mathcal{B}|\mathcal{A})} + \|f_t\|_{L^\infty} \|T^n g - g_s\|_{L^2(\mathcal{B}|\mathcal{A})} \\
&\leq 2\epsilon
\end{aligned}$$

We just showed that F is an algebra. Next let

$$\mathcal{D} := \{D \in \mathcal{B} : 1_D \in \bar{F}\}$$

Since F is an algebra and $\bar{F}$ is closed, the set $\mathcal{D}$ is a σ-algebra and $F \cap L^2(\mathcal{D})$ is dense in $L^2(\mathcal{D})$.

The last step is to prove that $F \subset L^\infty(\mathcal{D})$. Let $f \in F$ and let $D = \{x \in X : f(x) < a\}$ for some $a \in \mathbb{R}$. By Weierstrass' approximation theorem, the function $1_{(-\infty,a)} \in L^2(\mathbb{R}, f_*\mu)$ (where $f_*\mu$ is the pushforward measure of μ through f) can be approximated by a polynomial (even though $1_{(-\infty,a)}$ is not continuous, it has only one point of discontinuity, so it can be altered in a set of arbitrarily small measure to become continuous). This means that 1_D can be approximated by $p \circ f$ for the same polynomial $p \in \mathbb{R}[x]$. Since F is an algebra, $p \circ f \in F$ and so $1_D \in \bar{F}$ and hence $f \in L^\infty(\mathcal{D})$

Finally, note that F is invariant under T, thus $\mathcal{D}$ is invariant under T and therefore it is a compact extension of $\mathcal{A}$. This finishes the proof of Proposition 8.6.1, and hence of the Multiple recurrence theorem of Furstenberg.

References

[1] V. Bergelson and J. Moreira. Ergodic theorem involving additive and multiplicative groups of a field and $\{x + y, xy\}$ patterns. *Ergodic Theory Dynam. Systems*, 37(3):673–692, 2017.

[2] 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.

[3] J. G. van der Corput. Diophantische Ungleichungen. I. Zur Gleichverteilung Modulo Eins. *Acta Math.*, 56(1):373–456, 1931.

[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] M. Foreman, D. J. Rudolph, and B. Weiss. The conjugacy problem in ergodic theory. *Ann. of Math. (2)*, 173(3):1529–1586, 2011.

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

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

[8] H. Furstenberg, Y. Katznelson, and D. Orstein. The ergodic theoretical proof of szemerdi's theorem. *Bull. Amer. Math. Soc.*, 7:427–552, 1979.

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

[10] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)*, 161(1):397–488, 2005.

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

[12] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. John Wiley & Sons, New York-London-Sydney, 1974.

[13] H. L. Royden. *Real analysis*. Macmillan Publishing Company, New York, third edition, 1988.

[14] I. M. Vinogradov. *The method of trigonometrical sums in the theory of numbers*. Interscience Publishers, London and New York., 1947. Translated, revised and annotated by K. F. Roth and Anne Davenport.

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

[16] H. Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.*, 77(3):313–352, 1916.