Contents

Chapter 1

Preliminaries

1.1 Measure Theory

1.1.1 Basics

We begin with basic definitions from measure theory.

Definition 1.1.1.

Algebra: on set X is a collection of subsets $\mathcal{A} \subset \mathcal{P}(X)$ such that $X \in \mathcal{A}$ and is closed under complementation and finite union.

σ-algebra: on set X is a collection of subsets $\mathcal{B} \subset \mathcal{P}(X)$ such that $X \in \mathcal{B}$ and is closed under complementation and countable union.

σ-algebra: generated by $S \subset X$, $\sigma(S)$ is the smallest σ-algebra containing S.

Monotone class: is a collection of subsets $\mathcal{M} \subset \mathcal{P}(X)$ that is closed under union of increasing sequences and intersection of decreasing sequences.

Measurable space: is a pair $(X, \mathcal{B})$, where X is a set and $\mathcal{B}$ is a σ-algebra on X.

Measure: on $(X, \mathcal{B})$ is a *countably additive* function $\mu : X \to [0, \infty)$: if $\{E_i\}$ is a countable disjoint collection in $\mathcal{B}$, then $\mu(\bigcup E_i) = \sum \mu(E_i)$.

Probability space: is a triple $(X, \mathcal{B}, \mu)$, where $X, \mathcal{B}$ is measurable space and μ is a measure on it such that $\mu(X) = 1$ (μ is called a probability measure).

Measure-preserving transformation: is a measurable function $T : X \to X$: if T such that $\mu \circ T^{-1} = \mu$.

Measure-preserving system: (MPS) 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 following lemma says that measure is countably additive on almost disjoint sets.

Lemma 1.1.2. *If $\{E_i\}$ is a countable collection in $\mathcal{B}$ such that $\mu(A_i \cap A_j) = 0$ when $i \neq j$, then $\mu(\bigcup E_i) = \sum \mu(E_i)$.*

Proof. Let $N = \bigcup_{i \neq j}(A_i \cup A_j)$, then $\mu(N) = 0$. Let $B_i = A - i \setminus N$, then

$$\mu(\bigcup B_i) = \mu(\bigcup (A_i \setminus N)) = \mu((\bigcup A_i) \setminus N) = \mu(\bigcup A_i),$$
$$\mu(\bigcup B_i) = \sum \mu(B_i) = \sum \mu(A_i).$$

□

Theorem 1.1.3 (Monotone class theorem)**.** *If $\mathcal{A}$ is an algebra, then $M(\mathcal{A}) = \sigma(\mathcal{A})$, where $M(\mathcal{A})$ is the smallest monotone class containing $\mathcal{A}$.*

Definition 1.1.4 (Push-forward measure)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, $(Y, \mathcal{C})$ and $\phi : X \to Y$ be a measurable map. Then the *push-forward measure* is a measure on Y defined by

$$\mu \circ \phi^{-1}(E) = \mu(\phi^{-1}(E)).$$

Theorem 1.1.5 (Change of variables)**.** *Let $\phi : (X, \mathcal{B}, \mu) \to (Y, \mathcal{C})$ be a measurable map and f be a real-valued measurable function on Y, then*

$$\int_Y f d(\mu \circ \phi^{-1}) = \int_X f \circ \phi d\mu$$

1.1.2 Conditional Expectation

Definition 1.1.6 (Absolutely continuous)**.** Let ν, μ be measures on $(X, \mathcal{B})$. ν *is absolutely continuous with respect to* μ if

$$\mu(A) = 0 \implies \nu(A) = 0 \quad \forall A \in \mathcal{B}.$$

Theorem 1.1.7 (Radon-Nikodym theorem)**.** *Let ν, μ be σ-finite measures on $(X, \mathcal{B})$ such that $\nu \ll \mu$. There exists a measurable function $f : X \to [0, \infty)$ such that*

$$\nu(A) = \int_A f d\mu \quad \forall A \in \mathcal{B}.$$

Definition 1.1.8 (Measure-preserving system)**.** 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.*

Definition 1.1.9 (Conditional expectation)**.** Let $(X, \mathcal{B}, \mu)$ be a probability space, $f \in L^1(X, \mathcal{B}, \mu)$, and $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. The *conditional expectation of f given $\mathcal{B}$*, denoted $\mathbb{E}(f|\mathcal{A})$, is a $\mathcal{A}$-measurable function such that

$$\int_A \mathbb{E}(f|\mathcal{A}) d\mu = \int_A f d\mu \quad \forall A \in \mathcal{A}.$$

Such a function exists and is a.e. unique due to the Radon-Nikodym theorem. We first consider non-negative f. Define

$$\nu(A) = \int_A f d\mu \text{ for } A \in \mathcal{A}.$$

Then $\nu(A)$ is a finite measure on $\mathcal{A}$ and is absolutely continuous with respect to $\mu|_{\mathcal{A}}$. By Radon-Nikodym theorem, there is a function $g \in L^1(X, \mathcal{A}, \mu)$ such that

$$\nu(A) = \int_A g d\mu \quad \forall A \in \mathcal{A}.$$

We denote $g = \mathbb{E}(f|\mathcal{A})$.

For general $f \in L^1$, we write $f = f^+ - f^-$ and get $\mathbb{E}(f|\mathcal{A}) = \mathbb{E}(f^+|\mathcal{A}) - \mathbb{E}(f^-|\mathcal{A})$.

Example. $\{A_i\}_{i=0}^n$ is a finite partition of X, and $\mathcal{A}$ is the σ-algebra generated by it. Since $\mathbb{E}(f|\mathcal{A})$ is $\mathcal{A}$-measurable, it must be a linear combination of 1_{A_i}'s,

$$\mathbb{E}(f|\mathcal{A}) = \sum_{i=0}^n a_i 1_{A_i}.$$

We can compute,

$$a_i = \frac{\int_{A_i} f d\mu}{\mu(A_i)}.$$

If $f = 1_A$ for $A \subset X$, then $a_i = \frac{\mu(A_i \cap A)}{\mu(A_i)}$. This is equivalent to the definition of 'conditional probability of A given B' for events A, B from elementary probability theory.

Theorem 1.1.10. *Conditional expectation has the following properties:*

(i) $\mathbb{E}(af + bg|\mathcal{A}) = a\mathbb{E}(f|\mathcal{A}) + b\mathbb{E}(g|\mathcal{A})$ *a.e., where* $a, b \in \mathbb{R}$.

(ii) $f \leq g$ *a.e., then* $\mathbb{E}(f|\mathcal{A}) \leq \mathbb{E}(g|\mathcal{A})$ *a.e.*

(iii) $|\mathbb{E}(f|\mathcal{A})| \leq \mathbb{E}(|f||\mathcal{A})$

(iv)

Proof. (i) For any $A \in \mathcal{A}$,

$$\int_A a\mathbb{E}(f|\mathcal{A}) + b\mathbb{E}(g|\mathcal{A})\, d\mu = \int_A af + bg\, d\mu = \int_A \mathbb{E}(af + bg|\mathcal{A}).$$

(ii) For any $A \in \mathcal{A}$,

$$\int_A \mathbb{E}(f|\mathcal{A}) - \mathbb{E}(g|\mathcal{A})\, d\mu = \int_A f - g\, d\mu \geq 0.$$

Thus, $\mathbb{E}(f|\mathcal{A}) - \mathbb{E}(g|\mathcal{A}) \geq 0$. a.e.

(iii) Follows from (ii) by the fact $f \leq |f|$.

(iv)

□

1.2 Functional Analysis

Theorem 1.2.1 (Banach-Alaoglu theorem)**.** *X is a normed space. Then the unit ball of X^* is a compact Hausdorff space.*

Definition 1.2.2 (Convex hull)**.** Let V be a vector space, and $A \subset V$, then the *convex hull of A*, $\mathrm{conv}(A)$ is the smallest subset of V containing S.

$$\mathrm{conv}(A) = \left\{ \sum_{i=1}^{n} c_i x_i \mid n \in \mathbb{N}, c_i \geq 0, \sum_{i=1}^{n} c_i = 1 \right\}.$$

Theorem 1.2.3 (Hahn-Banach separation theorem, first form)**.** *Let E be a normed linear space, and $A, B \subset E$ be nonempty disjoint convex subsets. If A is open, then there exists a continuous linear funcitonal $f : E \to \mathbb{R}$, $\alpha \in \mathbb{R}$ such that*

$$f(a) \leq \alpha \leq f(b),\ \forall a \in A,\ \forall b \in B.$$

Theorem 1.2.4 (Hahn-Banach separation theorem, second form)**.** *Let E be a normed linear space, and $A, B \subset E$ be nonempty disjoint convex subsets. If A is compact and B is closed, then there exists a continuous linear funcitonal $f : E \to \mathbb{R}$, $\alpha \in \mathbb{R}$ such that*

$$f(a) < \alpha < f(b),\ \forall a \in A,\ \forall b \in B.$$

Chapter 2

Ergodic Theory

2.1 Recurrence

Recurrence is a basic and deep property of all measure-preserving system.

Lemma 2.1.1. *Let A be a measurable set with positive measure, then there is an n such that $\mu(A \cap T^{-n}A) > 0$.*

Proof. Take $k > 1/\mu(A)$, and consider the sets $A, T^{-1}A, \ldots, T^{-k}A$. $T^{-i}A$ are not almost pairwise disjoint, since if they were, then by using lemma 1.1.2,

$$1 = \mu(X) \geq \mu(\bigcup_{i=1}^{k} T^{-i}A) = \sum_{i=1}^{k} \mu(A) > 1.$$

So, there are $1 \leq i < j \leq n$ such that $\mu(T^{-i}A \cap T^{-j}A) > 0$. Then $\mu(T^{-i}(A \cap T^{-(i-j)}A)) = \mu(A \cap T^{-(i-j)}A) > 0$. □

Theorem 2.1.2 (Poincare recurrence theorem)**.** *Let A be a measurable set with positive measure, then almost every point of A returns to A, i.e., there is a set $E \subset A$, $\mu(E) = 0$ such that is if $x \in A \setminus E$, then there exist $n \in \mathbb{N}$, $T^n x \in A$. Furthermore, the points return infinitely often, i.e., there are infinitely many n such that $T^n x \in A$.*

Proof. Consider the set of points of A that do not return to A, $E = A \setminus \bigcup_{n \in \mathbb{N}} T^{-n}A$. Then, $E \cap T^{-n}E \subset A \cap T^{-n} = \phi$ for all $n \in \mathbb{N}$. So, by the previous lemma, $\mu(E) = 0$, and every point of E returns to A at least once.

To show they return infinitely often, consider the set of points of A that return only finitely many points. Note, if $x \in A$ returns to A k times then there are $n_1 < \cdots < n_k$ such that $T^{n_k} \in E$. So, $\bigcup_{n_1 < \cdots < n_k} T^{-n_k}E$, where the union is over all k-tuples, $\{n_1 < \cdots < n_k\}$ consists of the points of A that return exactly k many times. Thus, the set $\bigcup_{k \in \mathbb{N}} \bigcup_{n_1 < \cdots < n_k} T^{-n_k}E$ are the set of points of A that return finitely many times, and has measure 0 since $\mu(E) = 0$. □

2.2 Ergodicity

2.2.1 Ergodicity

Definition 2.2.1 (Invariant set)**.** A measurable set A is an *invariant set* if $T^{-1}A = A$.

Definition 2.2.2 (Invariant function)**.** A measurable function f is an *invariant function* if $f \circ T = f$.

Lemma 2.2.3. *f is bounded, $\mathcal{B}$-measurable and A is a T-invariant set, then $\int_A f \circ T d\mu = \int_A f d\mu$.*

Proof.

$$\int_A f \circ T d\mu = \int_X (f \circ T) 1_{T^{-1}A} d\mu = \int_X (f 1_A) \circ T d\mu = \int_X f 1_A d(\mu \circ T^{-1}) = \int_A f d\mu.$$

□

Note, if A is an invariant set, then $X \setminus A$ is also an invariant set.

Definition 2.2.4 (Ergodic system)**.** A measurable-preserving system $(X, \mathcal{B}, \mu, T)$ is *ergodic* if there are no non-trivial invariant sets, i.e., if A is an invariant set, then $\mu(A) = 0$ or 1.

Ergodicity is an irreducibility condition: if T is non-ergodic, then it can be reduced to to simpler transformations $T|_A$ and $T|_{X\setminus A}$, since A and $X \setminus A$ do not interact, i.e., orbits in one of them do not enter the other.

There are several ways of defining ergodicity.

Theorem 2.2.5. *$(X, \mathcal{B}, \mu, T)$ is a MPS, then the following are equivalent,*

(i) T is ergodic.

(ii) if $T^{-1}A = A$ mod-μ [i.e. $\mu(T^{-1}A \triangle A) = 0$], then $\mu(A) = 0$ or 1.

(iii) if $f : X \to \mathbb{R}$ is T-invariant, then f is constant a.e.

(iv) if $f \in L^1$ such that $f \circ T = f$ a.e., then f is constant a.e.

Proof.

(i) $\iff$ *(iii):* If T is not ergodic, there is a non-trivial invariant set A. Then 1_A is a measurable function such that $1_A \circ T = 1_{T^{-1}A} = 1_A$ which is not constant a.e.

Now, assume T is ergodic and $f \circ T = f$ a.e. For $k \in \mathbb{Z}$ and $n \in \mathbb{N}$, define

$$X(k, n) = \{x \in X \mid \frac{k}{2^n} \leq f(x) \leq \frac{k+1}{2^n}\}.$$

Since, $T^{-1}X(k, n) \triangle X(k, n) \subset x \mid f(x) \neq f(Tx)$, it must have measure 0 or 1. For a fixed n, $X(k, n)$ are disjoint and its union covers X, $\mu(X(k_n, n)) = 1$ for exactly one k_n. Now consider the set $\bigcap_n X(k_n, n)$. It has measure 1 and we can show that f is constant on it.

(ii) $\iff$ *(iv):* Same as previous.

(iii) $\iff$ *(iv):* If f is a T-invariant function, we construct integrable functions $f\chi_n$, where χ_n is the characteristic function on $\{|f| < n\}$. Then $f\chi_n$ are constant a.e., and since they have non-trivial intersection of domains, for all n the constants are same. As $f\chi_n \to f$ pointwise, we get that f is constant a.e.

If $f \in L^1$ such that $f \circ T = f$ a.e., let $g(x) = \limsup f(T^n(x))$, then g is measurable and $g \circ T = T$. So g is constant a.e. We are done if we show $f = g$ a.e.
Note, $f(T^{n+1}x) = f(T^n x)\ \forall n \implies f(T^n x) = f(x)\ \forall n \implies g(x) = f(x)$,
and $f(T^{n+1}x) = f(T^n x) \iff T^n x \in \{fT = f\} \iff x \in T^{-n}\{fT = f\}$.
So,$\{g = f\} \supset \bigcap_n T^{-n}\{fT = f\}$, which being the intersection of sets with measure 1, has measure 1.

□

Theorem 2.2.6. *$(X, \mathcal{B}, \mu, T)$ is a MPS, then the following are equivalent,*

(i) T is ergodic.

(ii) if $A \in \mathcal{B}, \mu(A) > 0$, then $\bigcup_{n=m}^{\infty} T^{-n}A = X$ mod-μ $\forall m$.

(iii) if $A, B \in \mathcal{B}$, $\mu(A), \mu(B) > 0$, then $\mu(T^{-n}A \cap B) > 0$ for infinitely many n.

Proof.

(i) $\implies$ *(ii):* Let $A' = \bigcup_{n=m}^{\infty} T^{-n}A$, then $T^{-1}A' = \bigcup_{n=m+1}^{\infty} T^{-n}A \subset A'$. And since $\mu(T^{-1}A) = \mu(A)$, we have $T^{-1}A' = A'$ mod-μ. By ergodicity, $A' = X$ mod-μ.

(ii) $\implies$ *(iii):* We have $\mu(\bigcup_{n=1}^{\infty} T^{-n}A) = 1$. So, $0 < \mu(B) = \mu(B \cap \bigcup_n T^{-n}A) = \mu(\bigcup_n (B \cap T^{-n}A))$. Hence, for some n_0, $\mu(B \cap T^{-n_0}A) > 0$. We can find another such n by taking the union from n_0 instead of 1 in the first line.

(iii) $\implies$ *(i):* Assume T is not ergodic and let A be a non-trivial invariant subset, then $A \cap (X \setminus A) = \phi \implies (T^{-n}A) \cap X \setminus A = \phi \ \forall n \implies \mu(T^{-n}A \cap (X \setminus A)) = 0 \ \forall n$, which is a contradiction.

□

Example. On S^1 with the Lebesgue measure, the rotation $T_\alpha(z) = ze^{2\pi i\alpha}$ is ergodic if and only if α is irrational.

Proof. If $\alpha = p/q \in \mathbb{Q}$, $f(z) = z^q$ is a such that it is non-constant a.e, T_α-invariant function. Hence, T_α is not ergodic.

If $\alpha \in \mathbb{Q}^c$: Let f be a T_α-invariant function in L^1. Consider the Fourier series, $f = \sum a_n \chi_n$, where $\chi_n(z) = z^n$ are the characters on S^1. Then,

$$f \circ T_\alpha(z) = \sum a_n e^{2\pi i n\alpha} z^n = \sum a_n z^n = f(z).$$

By uniqueness of Fourier coefficients, for $n \neq 0$, $a^n e^{2\pi i n\alpha} = a_n$. Which implies that $a_n = 0$, when $n \neq 0$, and that f is not constant a.e. □

Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system and let $A \in \mathcal{B}$ with positive measure. By ergodicity, $\bigcup_{n=1}^{\infty} T^{-n}A = X$ mod-μ, which says that a.e. $x \in X$ returns to A. Kac's formula gives us the average return time.

Define the function $r_A(x) : X \to \mathbb{N}$ by $r_A(x) = \min\{r \geq 1 \mid T^n x \in A\}$. Since $n_A((-\infty, k]) = \bigcup_{i=0}^{k} T^{-i}A$ is measurable for every $k \in \mathbb{N}$, r_A is measurable.

Theorem 2.2.7 (Kac's return time formula)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a MPS such that T is invertible. Then*

$$\int_A r_A d\mu = 1.$$

Proof. Let $A_n = A \cap \{r_A = n\}$, then on A, $r_A = \sum_{n=1}^{\infty} n1_{A_n}$. So,

$$\int_A r_A d\mu = \sum_{n=1}^{\infty} n\mu(A_n) = \sum_{n=1}^{\infty} \sum_{m=1}^{n} \mu(T^m A_n).$$

We show $\{T^m A_n \mid n \in \mathbb{N},\ m = 1, \ldots, n\}$ are disjoint and their union has full measure. Then, the result follows from countable additivity of μ.

To show: $\mu(\bigcup_{n=1}^{\infty} \bigcup_{m=1}^{n} T^m A_n) = 1$.

Proof. By ergodicity, a.e. $x \in X$, there is a smallest $m \geq 1$ such that $y := T^{-m}x \in A$. Let $n = r_A(y)$. Then, if $m > n$, $T^n y = T^{-(m-n)}x \in A$, but $m - n < m$, which is a contradiction of the minimality. Thus, $m \leq n$, and $x \in T^m A_n$. □

To show: $\{T^m A_n \mid n \in \mathbb{N},\ m = 1, \ldots, n\}$ are disjoint.

Proof. We show if $x \in T^{m'} A_{n'}$, then $(m', n') = (m, n)$. First, $T^{-m'}x \in A_{n'} \subset A$. So, by minimality, $m \leq m'$. If $m < m'$, since $T^{-m}x = T^{m'-m}(T^{-m'}x) \in A_n \subset A$, $n' = r_A(T^{-m'}x) \leq m' - m < m'$, which is a contradiction. Thus, $m = m'$. Now, if $x \in T^m A_n \cap T^m A_{n'}$, observe if $n \neq n'$, $A_n \cap A_{n'} = \phi \implies T^m A_n \cap T^m A_{n'} = \phi$. Hence, $n = n'$. □

□

2.2.2 Ergodic decomposition

We have seen that ergodic systems are non-decomposable. In this section, we look at how non-ergodic systems can be decomposed into ergodic ones. To do so, we will use measure disintegration.

Definition 2.2.8 (Probability kernel)**.** Let $(X, \mathcal{B})$ and $(Y, \mathcal{C})$ be two measurable spaces and $\{\theta_x\}_{x\in X}$ be a family of probability measures on Y, then $\{\theta_x\}_{x\in X}$ is called a *probability kernel from* $(X, \mathcal{B})$ *to* $(Y, \mathcal{C})$ if for each $E \in \mathcal{C}$, the map $x \mapsto \theta_x(B)$ is $\mathcal{B}$-measurable.

Example. Every measurable map $\phi : (X, \mathcal{B}) \to (Y, \mathcal{C})$ induces a kernel by $x \mapsto \delta_{\phi(x)}$, since for $E \in \mathcal{C}$, the map $x \mapsto \delta_{\phi(x)}(E) = 1_E(\phi(x))$ is a composition of measurable functions.

Definition 2.2.9 (Measure integration)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space, $(Y, \mathcal{C})$ be a measurable space and $\{\theta_x\}_{x\in X}$ be a kernel from $(X, \mathcal{B})$ to $(Y, \mathcal{C})$, then we can define a probability measure on Y,

$$\nu(E) = \int \nu_x(E) d\mu(x).$$

ν is a measure by an application of monotone convergence theorem: for positive measurable functions $\{f_n\}_{n=1}^{\infty}$, $\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n$.

We now look at the inverse problem and study how we can decompose a given measure as an integral of other measures. We can do this on some special spaces.

Definition 2.2.10 (Standard measurable space)**.** A measurable space $(X, \mathcal{B})$ is *standard* if there exists a complete and separable metric on X for which $\mathcal{B}$ is the Borel σ-algebra.

We will use the following result from advanced measure theory, that classifies standard measurable spaces, by cardinality.

Theorem 2.2.11. *There are three standard measurable spaces, up to measurable isomorphism: finite discrete space, countable discrete space and* $[0, 1]$ *with the usual Borel* σ*-algebra.*

Example. These standard measurable space are isomorphic: $[0, 1]$ with the Borel σ-algebra, $S^{\mathbb{Z}}$ (S is a finite set) with the product Borel σ-algebra.

Definition 2.2.12 (Measure disintegration)**.** Let $(X, \mathcal{B}, \mu)$ be a measure space and $\mathcal{F} \subset \mathcal{B}$ a sub-σ-algebra, then a kernel from $(X, \mathcal{F})$ to $(X, \mathcal{B})$, $\{\theta_x\}_{x\in X}$ is called the *disintegration of* μ *over* $\mathcal{F}$ if

1. $\mu(E) = \int \theta_x(E) d\mu(x), \ \forall E \in \mathcal{B}$.
2. if $f : X \to \mathbb{C}$ is a bounded $\mathcal{B}$-measurable function, then

$$\mathbb{E}(f|\mathcal{F})(x) = \int f(t) d\theta_x(t) \ \mu\text{-a.e.} \tag{2.1}$$

We begin by examining the simplest case when $\mathcal{F} = \sigma\{A_1, \ldots, A_n\}$ is a finite σ-algebra, where $\{A_i\}_{i=1}^n$ are measurable sets that partition A. Denote $A(x)$ to be the unique A_i containing x, and we can define the kernel $\theta_x = \frac{\mu|_{A(x)}}{\mu(A(x))}$. Then,

$$\int \theta_x(E) d\mu(x) = \int \frac{\mu|_{A_x}(E)}{\mu(A(x))} d\mu(x) = \sum_{i=1}^{n} \int \frac{\mu(E \cap A_i)}{\mu(A_i)} d\mu|_{A_i}(x) = \mu(E).$$

So, θ_x is a disintegration of μ over $\mathcal{F}$.

Writing $\theta_x(E) = \mathbb{E}(1_E|\mathcal{F})(x)$, we can extend this to when $\mu(E) = 0$. But since, conditional expectation is well-defined a.e., this works when there are countably many E. We deal with this technicality in the next theorem, by first finding a pre-measure on a countable algebra.

Theorem 2.2.13. *When $(X, \mathcal{B}, \mu)$ is a measure space and $\mathcal{F} \subset \mathcal{B}$ a sub-σ-algebra, there is a disintegration of μ over $\mathcal{F}$. Further, the kernel is unique: if θ'_x is another kernel, then $\theta_x = \theta'_x$ μ-a.e.*

Proof. We provide a proof for when $X = S^{\mathbb{Z}}$, $S = \{1, \dots, n\}$ is a finite set, with the Borel σ-algebra on the product topology.

Let $\pi_n : S^{\mathbb{Z}} \to S^{2n+1}$ be the projection into $-n^{\text{th}}$ to n^{th} coordinates, with the discrete topology on S^{2n+1}. Define $\mathcal{A}_n := \{\pi_n^{-1}(E) \mid E \subset A^{2n+1}\}$. This is an algebra and $\mathcal{A}_n \subset \mathcal{A}_{n+1}$. So, $\mathcal{A} := \bigcup_{n=1}^{\infty} \mathcal{A}_n$ is also an algebra. Now, we can define

$$\theta_x(E) = \mathbb{E}(1_E|\mathcal{F})(x), \text{ for } E \in \mathcal{A}.$$

Since $\mathcal{A}$ is countable, it is well-defined on a set of full measure and $0 \leq \theta_x(E) \leq 1$, call it X_0.

Claim: θ_x is a pre-measure on $\mathcal{A}$.

Proof. Clearly, $\theta_x(X) = 1$, and is finitely additive. Instead of showing countable additivity, we show that finite additivity is enough, since no element in $\mathcal{A}$ can be written as a countable union. Let $A \in \mathcal{A}$ such that $A = \bigsqcup_{i=1}^{\infty} A_i$, $A_i \in \mathcal{A}$. It follows that infinitely many A_i are empty from the fact A is compact and A_i are disjoint. $\square$

Now, we can extend θ_x to a probability measure, which we also call θ_x, on $\sigma(\mathcal{A})$, by Caratheodory extension theorem. And $\sigma(\mathcal{A}) = \mathcal{B}$. Indeed, we can check that $\mathcal{A}$ forms a basis on the product topology.

Claim: The map $X_0 \ni x \mapsto \theta_x(E)$ is $\mathcal{F}$-measurable and $\theta_x(E) = \mathbb{E}(1_E|\mathcal{F})(x)$, for $E \in \mathcal{B}$.

Proof. Let $\mathcal{C} = \{B \in \mathcal{B} \mid x \mapsto \theta_x(B) \text{ is measurable and } \theta_x(B) = \mathbb{E}(1_B|\mathcal{F})(x)\}$. From the definition of θ_x, $\mathcal{A} \subset \mathcal{C}$. By the properties of measures and by linearity of conditional expectation, we can see $\mathcal{C}$ is an algebra. And by continuity of measures and conditional expectation, we can also check $\mathcal{C}$ is a monotone class. So, by monotone class theorem, $\sigma(\mathcal{C}) = M(\mathcal{C}) = \mathcal{C}$. Thus, $\mathcal{B} = \sigma(\mathcal{A}) \subset \sigma(\mathcal{C}) = \mathcal{C}$. $\square$

Now, we extend $\{\theta_x\}_{x \in X_0}$ to the whole of X, by mapping $x \in X_0^c$ to a fixed probability measure. Hence, $\{\theta_x\}_{x \in X}$ is a kernel from $(X, \mathcal{F})$ to $(X, \mathcal{B})$. And,

$$\int \theta_x(E) d\mu(x) = \int \mathbb{E}(1_E|\mathcal{F})(x) d\mu(x) = \mu(E).$$

Note,

$$\mathbb{E}(1_E|\mathcal{F})(x) = \int 1_E(t) d\theta_x(t).$$

So, by linearity and continuity of integral and conditional expectation, we can replace 1_E by any bounded $\mathcal{B}$-measurable functions $f : X \to \mathbb{R}$ to get (2.1).

Finally, the uniqueness of kernel follows from (2.1), by taking f to be characteristic functions. $\square$

Theorem 2.2.14 (Ergodic decomposition)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a standard measure preserving system, and $\mathcal{F} \subset \mathcal{B}$ the sub-σ-algebra of the T-invariant sets. If $\{\theta_x\}_{x \in X}$ is the disintegration of μ over $\mathcal{F}$, then θ_x is T-invariant and ergodic for μ-a.e.*

Proof.

Claim: $\theta'_x = \theta_x \circ T^{-1}$ is a disintegration of μ over $\mathcal{F}$.

Proof. Clearly, θ_x is a kernel. And,

$$\int \theta'_x(E)d\mu(x) = \int \theta_x(T^{-1}E)d\mu(x) = \mu(T^{-1}E) = \mu(E).$$

Lastly,

$$\int f(t)d\theta'_x(t) = \int f(t)d(\theta_x \circ T^{-1}(t)) = \int f \circ T(t)d\theta_x(t) = \mathbb{E}(f \circ T|\mathcal{F})(x) = \mathbb{E}(f|\mathcal{F})(x) \text{ a.e.}$$

The second equality is due to a change of variable, and the last equality follows, since for any $A \in \mathcal{F}$,

$$\int_A \mathbb{E}(f \circ T|\mathcal{F})d\mu = \int_A f \circ T d\mu = \int_A f d\mu = \int_A \mathbb{E}(f|\mathcal{F})d\mu.$$

□

Thus, by uniqueness of disintegration, $\theta_x \circ T^{-1} = \theta_x$, which means that θ_x is T-invariant. To show θ_x is ergodic, take $f : X \to \mathbb{R}$ a bounded, $\mathcal{B}$-measurable, T-invariant function. This means, $\mathcal{F}$-measurable, since for U open in $\mathbb{C}$,

$$f \circ T = f \implies (f \circ T)^{-1}U = T^{-1}f^{-1}U = f^{-1}U \implies f^{-1}U \in \mathcal{F}.$$

So, $\mathbb{E}(f|\mathcal{F}) = f$ a.e. Now,

$$\begin{aligned}\operatorname{Var}_x(f) &= \int \left[f(t) - \int f(s)d\theta_x(s)\right]^2 d\theta_x(t) = \int f(t)^2 d\theta_x(t) - \left(\int f(t)d\theta_x(t)\right)^2 \\ &= \mathbb{E}(f^2|\mathcal{F}) - \mathbb{E}(f|\mathcal{F})^2 = f^2 - f^2 = 0.\end{aligned}$$

Hence, the first integrand is 0. So, f is constant a.e. and θ_x is ergodic. □

2.3 Ergodic Theorems

2.3.1 Mean ergodic theorem

Lemma 2.3.1. *If T is a contraction, then $T^*v = v \iff Tv = v$.*

Proof. $\implies$: If $T^*v = v$, then

$$\begin{aligned}\|Tv - v\|^2 &= \langle Tv - v, Tv - v\rangle \\ &= \|Tv\|^2 - \langle Tv, v\rangle - \langle v, Tv\rangle + \|v\|^2 \\ &\le 2\|v\|^2 - 2\|v\|^2 = 0.\end{aligned}$$

So, $Tv = v$.

The other direction is similar. □

Theorem 2.3.2 (Von Neumann's mean ergodic theorem for Hilbert space)**.** *$\mathcal{H}$ is a Hilbert space and T is a contraction [i.e. T is a bounded operator and $\|T\| \le 1$]. Let $\mathcal{M} = \{v \in \mathcal{H} \mid Tv = v\}$ and $\pi : \mathcal{H} \to \mathcal{M}$ be the orthogonal projection. Then*

$$S_n(v) := \frac{1}{n}\sum_{k=0}^{n-1} T^k(v) \to \pi(v) \quad \forall v \in \mathcal{H}.$$

Proof. We write $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$.

Claim: Let $\mathcal{N} = \{v - Tv \mid v \in \mathcal{H}\}$, then $\mathcal{M}^\perp = \overline{\mathcal{N}}$.

Proof. Since $\overline{\mathcal{N}}, \mathcal{M}$ are closed, we can show $\mathcal{N}^\perp = \mathcal{M}$ (since $\overline{\mathcal{N}}^\perp = \mathcal{N}^\perp$).
Let $v \in \mathcal{N}^\perp$, then

$$\langle v, w - Tw \rangle = 0 \implies \langle v, w \rangle = \langle v, Tw \rangle = \langle T^* v, w \rangle .$$

Since this is for every $w \in \mathcal{H}$, we get $T^* v = v$, and by the the lemma, $Tv = v$. Hence, $v \in \mathcal{M}$.

Now, let $v \in \mathcal{M}$, then $Tv = v$ and $T^* v = v$. So,

$$\langle v, Tw - w \rangle = \langle T^* v, w \rangle - \langle v, w \rangle = 0$$

Hence, $v \in \mathcal{N}^\perp$. □

By linearity of the operators, it is enough to show the statement holds on $\mathcal{M}$ and $\mathcal{M}^\perp$ separately. When $v \in \mathcal{M}$,

$$S_n(v) = v \to v$$

When $v \in \mathcal{M}^\perp = \overline{\mathcal{N}}$, by continuity, we only show when $v = w - Tw \in \mathcal{N}$

$$\begin{aligned} \|S_n(w - Tw)\| &= \left\| \frac{1}{n} \sum_{k=0}^{n-1} T^k (w - Tw) \right\| = \left\| \frac{1}{n} (w - T^n w) \right\| \\ &\leq \frac{1}{n} \|w\| + \|T^n w\| \leq \frac{2}{n} \|w\| \to 0 = \pi(v) \end{aligned}$$

□

Let $(X, \mathcal{B}, \mu, T)$ be a MPS. We define an operator (for $1 \leq p \leq \infty$)

$$U_T : L^p(X, \mu) \to L^p(X, \mu), \quad U_T(f) = f \circ T.$$

Then U_T is an isomtery, since (because T is μ-invariant)

$$\|U_T f\|_p = \left(\int |f(T(x))|^p d\mu(x) \right)^{1/p} = \left(\int |f(x)|^p d\mu(x) \right)^{1/p} = \|f\|_p , \quad (1 \leq p < \infty)$$

For $p = \infty$, it follows from the fact that $\|f\|_\infty = \lim_{p \to \infty} \|f\|$.

Corollary 2.3.3 (Mean ergodic theorem for dynamical systems)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a MPS and $M_T \subset \mathcal{B}$ be the σ-algebra of T-invariant subsets of $\mathcal{B}$, i.e. $M_T = \sigma\{E \in \mathcal{B} \mid T^{-1} E = E\}$, then for $f \in L^2(X, \mathcal{B}, \mu)$,*

$$S_n(U_T)(f) = \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \to \mathbb{E}(f | M_T) \text{ in } \|\cdot\|_2 .$$

If the system is ergodic, then

$$S_n(U_T)(f) \to \int f d\mu \text{ in } \|\cdot\|_2 .$$

Proof. We use the above theorem on $\mathcal{H} = L^2(X, \mathcal{B}, \mu)$ and U_T as the contraction.

Let $\pi_T : L^2(X, \mathcal{B}, \mu) \to L^2(X, M_T, \mu)$ be the orthogonal projection, then $\pi_T(f) = \mathbb{E}(f | M_T)$. Indeed, $\pi_T(f)$ is clearly M_T-measurable and for $f \in L^2(X, \mathcal{B}, \mu)$, $E \in M_T$,

$$\int_E \pi_T(f) d\mu = \langle \pi_T(f), 1_E \rangle = \langle f, 1_E \rangle = \int_E f d\mu.$$

Finally, we show that $\mathcal{M} := \{f \in L^2(X, \mathcal{B}, \mu) \mid U_T(f) = f\} = L^2(X, M_T, \mu)$. First, take $1_A \in L^2(X, M_T, \mu)$, where $A \in M_T$. But, $U_T(1_A) = 1_{T^{-1}A} = 1_A$. So $1_A \in \mathcal{M}$. By linearity and continuity of U_T, $L^2(X, M_T, \mu) \subset \mathcal{M}$. For the reverse inclusion, we have to show that $f \in \mathcal{M}$ is M_T-measurable, or that for all $a \in \mathbb{R}$, $f^{-1}(-\infty, a)$ is T-invariant. This follows from $T^{-1}f^{-1}(-\infty, a) = (f \circ T)^{-1}(-\infty, a) = f^{-1}(-\infty, a)$.

When the system is ergodic, $M_T = \{\phi, X\}$. Since, $\mathbb{E}(f|M_T)$ is M_T-measurable, $\mathbb{E}(f|M_T)$ is a linear multiple of 1_X, $\mathbb{E}(f|M_T) = a1_X$. So, $a = \int \mathbb{E}(f|M_T)d\mu = \int f d\mu$. ☐

We now look at the generalization of the theorem to Banach spaces. Let $(X, \|\cdot\|)$ be a Banach space, and $T : X \to X$ a continuous linear map such that $\sup_{n\in\mathbb{N}} \|T^n\| < \infty$. We call such operators, *power bounded operators*. Note, for such T,

$$\lim_{n\to\infty} \frac{T^n(x)}{n} = 0, \text{ and } \sup_{n\in\mathbb{N}} \|S_n\| < \infty.$$

We define the average

$$S_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} T^k(x).$$

Consider the following subspaces,

$$N = \{x \in X \mid T(x) = x\} = \ker(I - T),$$
$$M = \{x \in X \mid \lim_{n\to\infty} S_n(x) \text{ exists}\}.$$

And define the map,

$$P : M \to X, \; P(x) = \lim_{n\to\infty} S_n(x).$$

Lemma 2.3.4. *P has the following properties,*

(i) $P \in \mathcal{B}(X)$

(ii) $P(x) = PT(x) = TP(x)$.

(iii) $x \in \overline{Im(I - T)} \implies P(x) = 0$

Proof. (i)

$$\|P(x)\| = \lim_{n\to\infty} \|S_n\| \|x\| \implies \|P\| < \infty,$$

since $\sup_n \|S_n\| < \infty$.

(ii)

$$S_n(Tx) = \frac{1}{n}\sum_{k=0}^{n-1} T^k(Tx) = \frac{1}{n}\sum_{k=0}^{n-1} T^k(x) + \frac{T^n x - x}{n} \implies PT(x) = P(x),$$

since $\frac{T^n(x)}{n} \to 0$. And

$$S_n(Tx) = \frac{1}{n}\sum_{k=0}^{n-1} T^k(Tx) = T\left(\frac{1}{n}\sum_{k=0}^{n-1} T^k(x)\right) \implies PT(x) = TP(x),$$

by linearity and continuity of T.

(iii) If $x = y - Ty \in \mathrm{Im}(I - T)$, for some $y \in X$. Then, $P(x) = P(y) - PT(y) = 0$. By continuity of P, we can extend to x in the closure.

☐

Theorem 2.3.5. *The following hold,* *(i) M is a T-invariant closed subspace of X.*

(ii) $P : M \to N$ is a projection, i.e., $P^2 = P$, and $M = N \oplus \overline{Im(I-T)}$.

Proof. *M is T-invariant:* We have to show if $x \in M$, then $Tx \in M$. This follows from $PT(x) = P(x)$.

M is closed: Consider a sequence $x_n \in M$ such that $x_n \to x$. To show $\lim S_n(x)$ exists, we show $S_n(x)$ is Cauchy. Given $\epsilon > 0$, choose k such that $\|x - x_k\| < \epsilon$. Since $S_n(x_k)$ is Cauchy, choose m, n such that $\|(S_n - S_m)(x_k)\| < \epsilon$. Then,

$$\begin{aligned}\|(S_n - S_m)(x)\| &\leq \|S_n(x - x_k)\| + \|(S_n - S_m)(x_k)\| + \|S_m(x_k - x)\| \\ &\leq (2 \sup_{n \in \mathbb{N}} \|S_n\| + 1)\epsilon \to 0.\end{aligned}$$

$P^2 = P$: $\mathrm{Im}(T) \subset N$ follows from the fact that $TP = P$. Since elements of N are T fixed points and hence S_n fixed points, this also shows that $P^2 = P$.

$M = N \oplus \overline{Im(I-T)}$: For $x \in M$, $x = P(x) + (I-P)(x)$. Since, $(I - S_n)(x) \in \mathrm{Im}(I-T)$, we have $(I-P)(x) \in \overline{\mathrm{Im}(I-T)}$. And if $x \in N \cap \overline{\mathrm{Im}(I-T)}$, then from the lemma, $P(x) = x = 0$. ☐

Theorem 2.3.6. *For $x, y \in X$, the following are equivalent,*

(i) $\lim_{n\to\infty} S_n x = y$.

(ii) there is a subsequence (n_k) such that $S_{n_k}(x) \rightharpoonup y$.

(iii) $y \in N \cap \overline{conv}^{weak}\{T^n x \mid n \geq 0\}$.

(iv) $y \in N \cap \overline{conv}^{\|\cdot\|}\{T^n x \mid n \geq 0\}$.

Proof. *(i) $\implies$ (ii):* This is due to the fact that norm-convergent sequences are also weakly-convergent.

(iii) $\implies$ (iv): This is a consequence of convex sets having same closure in norm topology and weak topology.

(ii) $\implies$ (iii): Clearly, $y \in \overline{\mathrm{conv}}^{\mathrm{weak}}\{T^n x \mid n \geq 0\}$. Showing $y \in N$ is similar to showing $TP = P$, but using weak-continuity of T instead.

(iv) $\implies$ (i): Since $y \in \overline{\mathrm{conv}}^{\|\cdot\|}\{T^n x \mid n \geq 0\}$, there is a sequence $\{y_n\} \in \mathrm{conv}^{\|\cdot\|}\{T^n x \mid n \geq 0\}$, and $y_n = \sum_{i=1}^{k_n} c_{n_i} T^{n_i}(x)$. Now,

$$\begin{aligned}y_n = x - (x - y_n) &= x - \left(\sum_{i=1}^{k_n} c_{n_i}\right)(I - T^{n_i})(x) \\ &= x - (I-T)(x'_n),\end{aligned}$$

for some $x'_n \in X$, since $I - T^{n_i} = (I-T)(I + T + T^2 + \cdots + T^{n_i - 1})$. Hence, $y_n = x - x'_n$, where $x'_n \in \mathrm{Im}(I-T)$. Taking limits, $y = x - x'$, where $x' \in \overline{\mathrm{Im}(I-T)}$. Thus, $\lim S_n y = y = \lim S_n x$, as $\lim S_n x' = 0$, by the lemma.

☐

Theorem 2.3.7 (Mean ergodic theorem for Banach spaces)**.** *Let X be a reflexive Banach space, and $T \in \mathcal{B}(X)$ such that $\sup_{n\in\mathbb{N}} \|T^n\| < \infty$, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} T^k(x) \text{ exists for all } x \in X.$$

Proof. We know, in a reflexive Banach space, every bounded sequence has a weakly convergent subsequence. The result follows, by the previous theorem. □

2.3.2 Pointwise ergodic theorem

$(X, \mathcal{B}, \mu, T)$ is a measure-preserving system. Define the average operator on $L^1(\mu)$,

$$S_n f = \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k.$$

Let $\mathcal{F}$ be the σ-algebra of T-invariant sets.

Theorem 2.3.8 (Birkhoff's pointwise ergodic theorem)**.** *For $f \in L^1(\mu)$, the following holds,*

$$S_n f(x) \to \mathbb{E}(f|\mathcal{F})(x) \text{ a.e.} \tag{2.2}$$

If the system is ergodic, then $S_n f(x) \to \int f d\mu$ a.e.

For the proof, we first find a dense set of L^1 where the statement holds and then attempt to extend to L^1.

Lemma 2.3.9. *There is a dense set $S \subset L^1$ such that for $f \in L^1$,* (2.2) *holds.*

Proof. Since L^2 is dense in L^1, it is sufficient to find a dense set of L^2. Define $S_1 = \{f \in L^2 \mid fT = f\}$ and $S_2 = \{g - gT \mid g \in L^\infty\}$. Clearly, (2.2) holds on S_1. On S_2, $\frac{1}{n} S_n(g - gT) = \frac{1}{n}(g - gT^n) \to 0$, and $\int_F \mathbb{E}(g - gT|\mathcal{F}) = \int_F g - gT = 0$.

Claim: $\overline{S_2} = \overline{\{g - gT \mid g \in L^2\}}$ in L^2.

Proof. Since $L^\infty \subset L^2$, we have $\overline{S_2} \subset \overline{\{g - gT \mid g \in L^2\}}$. Now, let $g \in L^2$. By density of L^∞ in L^2, there is a $g' \in L^\infty$ such that $\|g - g'\|_2 < \epsilon/2$. Then, $\|g - gT - (g' - g'T)\| < \|g - gT\| + \|g' - g'T\| < \epsilon$. Hence, $g - gT \in \overline{S_2}$ and $\overline{\{g - gT \mid g \in L^2\}} \subset \overline{S_2}$. □

We saw in the proof of the mean ergodic theorem that, $L^2 = S_1 \oplus \overline{S_2}$. But, $S_1 \oplus \overline{S_2} \subset \overline{S_1 \oplus S_2}$. Thus, $S_1 \oplus S_2$ is the required dense set. □

To extend to all of L^1, we need the maximal inequality. But before that, we look at its discrete version.

If $\hat{f} : \mathbb{N} \to [0, \infty)$, define the average over $I \subset \mathbb{N}$,

$$S_I f = \frac{1}{|I|} \sum_{i \in I} \hat{f}(i).$$

Theorem 2.3.10 (Discrete maximal inequality)**.** *Let $\hat{f} : \mathbb{N} \to [0, \infty)$, $I \subset \mathbb{N}$ be finite interval and $J \subset I$ be subset such that $\forall j \in J$, $I_j \subset I$ be sub-intervals of I with left-endpoint j. If $S_{I_j} \hat{f} > t$, then*

$$S_I \hat{f} > t \frac{|J|}{|I|}.$$

Proof. First, if I_j are disjoint, then $\{I_j\}, \{I \setminus \bigcup_j I_j\}$ partition I. So,

$$S_I \hat{f} = \frac{|I \setminus \bigcup_j I_j|}{|I|} S_{I \setminus \cup I_j} \hat{f} + \sum \frac{|I_j|}{|I|} S_{I_j} \hat{f} \geq \frac{|\bigcup I_j|}{|I|} S_{I_j} \hat{f} \geq \frac{|J|}{|I|} t.$$

If I_j are not disjoint, we find a sub-collection $J_0 \subset J$ such that $\{I_j\}_{j \in J_0}$ are disjoint and $J \subset \bigcup_{j \in J_0} I_j$. Then the above inequalities will still go through, with union over J_0 instead of J. Let $j_1 = \min J$ and define $j_k = \min(J \setminus \bigcup_{i=1}^{k-1} I_{j_i})$. $\square$

Theorem 2.3.11 (Maximal inequality)**.** *$f \in L^1$, $f \geq 0$, then $\forall t > 0$,*

$$\mu\{x \in X \mid \sup_n S_n f(x) > t\} \leq \frac{1}{t} \int f d\mu.$$

Proof. Let $A = \{\sup_n S_n f > t\}$. Fix $x \in X$, define $\hat{f}_x : \mathbb{N} \to [0, \infty)$, $\hat{f}_x(i) = f(T^i x)$. Now, we try to use the discrete maximal inequality on $\hat{f}_x$. If $T^j x \in A$, then $S_{n_j} T^j x > t$ for some n_j. Let $I_j = [j, j + n_j - 1]$, then

$$S_{I_j} \hat{f} = \frac{1}{n_j} \sum_{i=j}^{j+n_j-1} f(T^i x) = \frac{1}{n_j} \sum_{i=0}^{n_j-1} f T^i (T^j x) = S_{n_j} T^j x > t.$$

Consider $I = [0, M-1]$, $J = \{j \in I \mid T^j x \in A, I_j \subset I\}$. By the discrete maximal inequality on $\{I_j\}_{j \in J}$,

$$S_I \hat{f} > t \frac{|J|}{M}.$$

Let $A_R = \{\sup_{0 \leq n \leq R} S_n f > t\}$, then $J \supset \{j \in [0, M-R] =: I' \mid T^j x \in A_R\}$ and

$$|J| \geq \sum_{j=0}^{M-R} 1_{A_R}(T^j x) = \sum_{j=0}^{M-R} \hat{1}_{A_R}(j) = (M-R) S_{I_R} \hat{1}_{A_R}.$$

Thus, $S_I \hat{f} > t \frac{|J|}{M} \geq t \frac{M-R}{M} S_{I_R} \hat{1}_{A_R}$. Now,

$$\begin{aligned}
\int f d\mu &= \frac{1}{M} \sum_{i=0}^{M-1} \int f T^i(x) d\mu(x) = \int S_I \hat{f}_x d\mu(x) > t \frac{M-R}{M} \int S_{I'} \hat{1}_{A_R, x} d\mu(x) \\
&= t \left(1 - \frac{R}{M}\right) \mu(A_R) \xrightarrow{M \to \infty} t\mu(A_R) \xrightarrow{R \to \infty} t\mu(A)
\end{aligned}$$

The last equality follows from,

$$\int S_{I'} \hat{1}_{A_R, x} d\mu(x) = \int \frac{1}{M-R} \sum_{i=0}^{M-R} \hat{1}_{A_R, x}(i) d\mu(x) = \frac{1}{M-R} \sum_{i=0}^{M-R} \int 1_{A_R}(T^i x) d\mu(x) = \int 1_{A_R} d\mu.$$

$\square$

Now we can complete the proof of the pointwise ergodic theorem.

Proof of pointwise ergodic theorem. We write $S(f) = \mathbb{E}(f|\mathcal{F})$, for convenience. Let $f \in L^1$, and take $g \in S$. Then, a.e. $x \in X$,

$$|S_n f - Sf| \leq |S_n f - S_n g| + |S_n g - Sf| \leq S_n |f - g| + |S_n g - Sf|.$$

But, $S_n g \to Sg$ a.e., implies $|S_n g - Sf| \to |Sg - Sf| \le S|g - f|$, by property of conditional expectation. So,

$$\limsup_{n\to\infty} |S_n f - Sf| \le \limsup_{n\to\infty} S_n|f - g| + S|g - f|.$$

Note, if $\limsup |S_n f - Sf| > \epsilon$, then either $\limsup S_n|f - g| > \epsilon/2$ or $S|g - f| > \epsilon/2$. Therefore,

$$\mu\{\limsup_{n\to\infty} |S_n f - Sf| > \epsilon\} \le \mu\{\limsup_{n\to\infty} S_n|f - g| > \frac{\epsilon}{2}\} + \mu\{S|g - f| > \frac{\epsilon}{2}\} \le \frac{2}{\epsilon}\|f - g\|_1 + \frac{2}{\epsilon}\|f - g\|_1.$$

The first term comes from the maximal inequality and the second term from Markov's inequality. Since the right hand side can be made arbitrarily small, we have $\mu\{\limsup |S_n f - Sf| > \epsilon\} = 0$. Thus, $\limsup |S_n f - Sf| = 0$ a.e., and $S_n f \to Sf$ a.e. □

2.4 Mixing

Definition 2.4.1 (Strongly mixing). $(X, \mathcal{B}, \mu, T)$ is *strong mixing* if for all $A, B \in \mathcal{B}$,

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

It is clear that strongly ergodic systems are ergodic.

Theorem 2.4.2. *$(X, \mathcal{B}, \mu, T)$ is strongly mixing if and only if for every $f, g \in L^2$,*

$$\lim_{n\to\infty} \int f \cdot (gT^n) d\mu = \int f d\mu \cdot \int g d\mu.$$

Furthermore, the above is true for every $f, g \in L^2$ if it holds for a dense set of L^2.

Proof. We first see how the second statement implies the first. For $A, B \in \mathcal{B}$, take $f = 1_A, g = 1_B$. Then $1_A \cdot 1_B(T^n) = 1_A \cdot 1_{T^{-n}B} = 1_{A\cap T^{-n}B}$. So, the statement holds for simple functions, and by the density of simple functions in L^2, statement holds for the entire L^2.

Now, suppose $S \subset L^2$ is a dense subset such the limit holds on S. For $f, g \in L^2$, let $f', g' \in S$, such that $\|f' - f\| < \epsilon, \|g' - g\| < \epsilon$. Then,

$$\begin{aligned}
\left|\int f \cdot gT^n - \int f \cdot \int g\right| &\le \left|\int (f - f' + f') \cdot (g - g' + g')T^n - \int f \int g\right| \\
&\le \left|\int (f - f') \cdot (g - g')T^n\right| + \left|\int (f - f') \cdot g'T^n\right| + \left|\int f' \cdot (g - g')T^n\right| \\
&\quad + \left|\int f' \cdot g'T^n - \int f' \cdot \int g'\right| + \left|\int f' \cdot \int g' - \int f \cdot \int g'\right| \\
&\quad + \left|\int f \cdot \int g' - \int f \cdot \int g\right| \\
&\le \epsilon^2 + \epsilon\|g'\| + \epsilon\|f'\| + \epsilon + \epsilon\|g'\| + \epsilon\|f\| \to 0.
\end{aligned}$$

The last inequality is follows applications of Cauchy-Schwarz inequality. □

The above technique will be used repeatedly to extend similar statements from a dense subset to the entire space.

We also have a similar characterization for ergodicity.

Theorem 2.4.3. *$(X, \mathcal{B}, \mu, T)$ is ergodic if and only if $A, B \in \mathcal{B}$,*

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

Proof. $\Longleftarrow$: If $\mu(A), \mu(B) > 0$, then $\mu(A \cap T^{-n}B) > 0$, for infinitely many n. If not, then the limit would go to 0, which is not possible.

$\Longrightarrow$: For $A, B \in \mathcal{B}$,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mu(A\cap T^{-n}B) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \int 1_A \cdot 1_B(T^k) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \langle 1_A, 1_B(T^k)\rangle \\ &= \left\langle 1_A, \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} 1_B(T^k)\right\rangle = \left\langle 1_A, \lim_{n\to\infty} S_n 1_B\right\rangle \\ &= \left\langle 1_A, \int 1_B\right\rangle = \mu(A)\mu(B).\end{aligned}$$

The penultimate equality is due to mean ergodic theorem.

□

As before, we can similarly extend to integrals of functions instead of measures of sets.

Definition 2.4.4 (Weakly mixing)**.** $(X, \mathcal{B}, \mu, T)$ is *weakly mixing* if for all $A, B \in \mathcal{B}$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} |\mu(A\cap T^{-k}B) - \mu(A)\mu(B)| = 0.$$

We can again extend this to functions instead of sets: for all $f, g \in L^2$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \left|\int f\cdot gT^k - \int f \int g\right| = 0.$$

We require the concept of density, before studying weakly mixing further.

Definition 2.4.5 (Upper density)**.** The density of subset $I \subset \mathbb{N}$ is

$$d(I) = \lim_{n\to\infty} \frac{|I \cap \{1, \dots, n\}|}{n}.$$

Definition 2.4.6 (Convergence in density)**.** a_n is said to *converge in density* to a, $\text{d-}\lim_{n\to\infty} a_n = a$ or $a_n \xrightarrow{d} a$, if

$$d(\{n \mid |a_n - a| > \epsilon\}) = 0.$$

The definition of weakly mixing can now be reformulated as $\mu(A\cap T^{-n}B) \xrightarrow{d} \mu(A)\mu(B)$.

Notice, usual convergence implies convergence in density, because the above set will always have finite elements and hence, zero density. So, we have strong mixing implies weak mixing.

But the converse is not true. The sequence $a_n = 1$, when n is prime and 0 otherwise, will converge to 0, as the primes have 0 density, but does not converge in norm.

Lemma 2.4.7. *For bounded sequences a_n, the following are equivalent,*

(i) $\text{d-}\lim_{n\to\infty} a_n = a$.

(ii) *there is a subset $J \subset \mathbb{N}$ with $d(J) = 0$ such that* $\lim_{n\to\infty,\ n\notin J} a_n = 0$.

(iii) $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} |a_k - a| = 0$.

(iv) $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} (a_k - a)^2 = 0$.

Proof. *(ii)* $\implies$ *(i)*: Follows from the fact that, $\{i \in [1,n] \mid |a_i - a| \geq \epsilon\} \subset \{i \in [1,n] \cap J^c \mid |a_i - a| \geq \epsilon\} \cup \{i \in [1,n] \cap J \mid |a_i - a| \geq \epsilon\}$. The first set is finite and the second has density zero.

(i) $\implies$ *(iii)*: For $\epsilon > 0$, the set $\{n \mid |a_n - a| \geq \epsilon\}$ has density zero. And since the sequences are bounded, $|a_n - a| < K$ for some $K > 0$. Thus,

$$\frac{1}{n}\sum_{k=0}^{n-1} |a_n - a| < \epsilon + \frac{K}{n}\{i \in [1,n] \mid |a_i - a| \geq \epsilon\} \to \epsilon.$$

(iii) $\implies$ *(ii)*: Define $J_k = \{n \mid |a_n - a| \geq 1/k\}$. Then $J_k \subset J_{k+1}$ and each J_k has density zero, since $\frac{1}{n}\sum |a_n - a| \geq \frac{1}{kn}|J_k \cap [1,n]|$. This also means that, there is a sequence $\{l_k\}$ such that $|J_k \cap [1,n]|/n < 1/k \ \forall n \geq l_k$. Now, define $J = \bigcup_k (J_k \cap [l_k, l_{k+1}))$, then J also has density zero. Indeed, $J \cap [1,n] \subset J_k \cap [1,n]$, where $l_k \leq n < l_{k+1}$. Finally, if $n > l_k$ and $n \notin J$, then $n \notin J_k$, and hence $|a_n - a| < 1/k$.

□

Now, we can see weakly mixing implies ergodic. If it was not ergodic, then the series *(ii)* cannot converge to zero, by the previous theorem on ergodicity.Using *(ii)*, we can also easily show convergence in density respects sums and products.

Corollary 2.4.8. *Strong mixing implies weak mixing, and weak mixing implies ergodicity.*

Theorem 2.4.9. *$(X, \mathcal{B}, \mu, T)$ is weakly mixing if and only if $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T)$ is ergodic.*

Proof. $\impliedby$: We can see the condition implies ergodicity. If T is not ergodic $T \times T$ cannot be ergodic, since if f is a non-constant T-invariant measurable function, then (f, f) is a non-constant $T \times T$-invariant measurable function.

Let $f, g \in L^2(\mu)$, and define $\tilde{f}, \tilde{g} \in L^2(\mu \times \mu)$ by $\tilde{f}(x,y) = f(x)f(y), \tilde{g}(x,y) = g(x)g(y)$. Since $T \times T$ is ergodic,

$$\frac{1}{n}\sum_{k=0}^{n-1} \int \tilde{f} \cdot \tilde{g}(T \times T)^k \to \int \tilde{f} \int \tilde{g}.$$

By Fubini's theorem, the left integral is $(\int f \cdot gT^k)^2$ and the right integral is $(\int f \int g)^2$. Thus, $\frac{1}{n}\sum_{k=0}^{n-1}(\int f \cdot gT^k)^2 \to (\int f \int g)^2$. Note, since T is ergodic, $\frac{1}{n}\sum_{k=0}^{n-1}\int f \cdot gT^k \to \int f \int g$. Finally,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\left(\int f \cdot gT^k - \int f \cdot \int g\right)^2 &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\Bigg[\left(\int f \cdot gT^k\right)^2 + \left(\int f \int g\right)^2 \\ &\quad - 2\left(\int f \cdot gT^k\right)\left(\int f \int g\right)\Bigg] = 0.\end{aligned}$$

$\implies$: We prove for $A_1 \times A_2, B_1 \times B_2 \in \mathcal{B} \times \mathcal{B}$.

$$\begin{aligned}&(\mu \times \mu)\left((A_1 \times A_2) \cap (T \times T)^{-k}(B_1 \times B_2)\right) = (\mu \times \mu)\left((A_1 \cap T^{-k}B_1) \cap (A_2 \cap T^{-k}B_2)\right) \\ &\quad = \mu(A_1 \cap T^{-k}B_1)\mu(A_2 \cap T^{-k}B_2) \xrightarrow{d} \mu(A_1)\mu(A_2)\mu(B_1)\mu(B_2) = \mu(A_1 \times A_2)\mu(B_1 \times B_2).\end{aligned}$$

We know, functions of the form $1_{A_1}(x).1_{A_2}(y) = 1_{A_1 \times A_2}(x,y),\ A_1, A_2 \in \mathcal{B}$ form an orthonormal basis for $L^2(\mu \times \mu)$. Thus, by the density argument, the above convergence can be extended to all sets in the product σ-algebra, and we have $T \times T$ is ergodic.

□

Now, we discuss some multiplier properties of weak mixing systems.

Theorem 2.4.10. *$(X, \mathcal{B}, \mu, T)$ is weak mixing if and only if $X \times Y$ is ergodic for every ergodic system $(Y, \mathcal{C}, \nu, S)$.*

Proof. $\Longleftarrow$ *:* Take $Y = \{*\}$, then $X \times \{*\} \cong X$ is ergodic. So, $X \times X$ is ergodic, and X is weak mixing

$\Longrightarrow$ *:* Again by density argument, it is enough to consider cylindrical sets $A \times C, B \times D \in \mathcal{B} \times \mathcal{C}$,

$$\frac{1}{n}\sum_{k=0}^{n-1}(\mu \times \nu)\left[(T \times S)^{-k}(A \times C) \cap (B \times D)\right] = \frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}A \cap B)\nu(S^{-k}C \cap D)$$

$$= \frac{1}{n}\sum_{k=0}^{n-1}\Big[\{\mu(T^{-k}A \cap B) - \mu(A)\mu(B)\}\nu(S^{-k}C \cap D) + \mu(A)\mu(B)[\nu(S^{-k}C \cap D)]\Big]$$

$$\to \mu(A)\mu(B)\nu(C)\nu(D), \text{ as } n \to \infty.$$

The first term goes to 0 and second term converges, due to ergodicity of μ and ν respectively. $\square$

Corollary 2.4.11. *(i) X_1, X_2 is weak mixing $\Longrightarrow$ $X_1 \times X_2$ is weak mixing.*

(ii) T is weak mixing $\Longrightarrow$ T^n is weak mixing for all $n \in \mathbb{N}$

(iii) if T is invertible, T is weak mixing $\Longleftrightarrow$ T^{-1} is weak mixing

Proof. (i) Take ergodic system Y, then $X_2 \times Y$ is ergodic, which implies $X_1 \times X_2 \times Y$ is also ergodic and hence, $X_1 \times X_2$ is weak mixing.

(ii) We first show, T is weak mixing $\Longrightarrow$ T^n is ergodic for all $n \in \mathbb{N}$.
Assume T^n is not ergodic, and take a non-constant a.e., T^n-invariant function $f \in L^2(\mu)$. Define a measure-preserving system, $(Y, \mathcal{C}, \nu, S)$ where $Y = \{0, 1, \ldots, n-1\}$, $\mathcal{C} = \mathcal{P}(Y)$, $\nu(A) = \frac{|A|}{n}$ and $S : Y \to Y$, $S(i) = i + 1 \mod n$. It is easy to see that S is an ergodic system. So, $T \times S$ is ergodic. Define $F \in L^2(\mu \times \nu)$, $F(x, i) = f(T^{n-i}x)$. Then, F is $T \times S$-invariant: $F((T \times S)(x, i)) = F(Tx, i + 1) = f(T^{n-i}x) = F(x, i)$. As, F is non-constant a.e., this contradicts the ergodicity of $T \times S$. Thus, T^n is ergodic.
Now, T is weak mixing $\Longrightarrow$ $T \times T$ is weak mixing $\Longrightarrow$ $T^n \times T^n$ is ergodic $\Longrightarrow$ T^n is weak mixing.

(iii) Follows from the fact that T is ergodic if and only if T^{-1} is ergodic, which implies $T \times T$ is ergodic if and only if $T^{-1} \times T^{-1}$ is ergodic.

$\square$

We now look at the relationship between properties of a measure-preserving system $(X, \mathcal{B}, \mu, T)$ and the point-spectrum of U_T. Recall, $\lambda \in \mathbb{C}$ is an *eigenvalue* of U_T if $U_T f = \lambda f$, for some $0 \neq f \in L^2(\mu)$, and f is the called *eigenfunction*. The vector space of $f \in L^2(\mu)$ satisfying the equation is called the *eigenspace*, E_λ and the *point-spectrum* of U_T is defined by $\sigma_p(U_T) = \{\lambda \in \mathbb{C} \mid U_T f = \lambda f,\ f \in L^2(\mu)\}$. An eigenvalue is said to be *simple* if the corresponding eigenspace is one-dimensional.

Theorem 2.4.12. *If T is ergodic, then σ_p is a subgroup of S^1, and every eigenvalue of U_T is simple. T is ergodic if and only if 1 is a simple eigenvalue of U_T.*

Proof. $\sigma_p \subset S^1$*:* If $\lambda \in \sigma_p$, then $U_T f = \lambda f$ for some non-zero $f \in L^2$. So, $\|f \circ T\| = \|f\| = |\lambda|\,\|f\|$, and $\lambda \in S^1$.

σ_p is closed under product: $\lambda_1, \lambda_2 \in \sigma_p$, and $U_T f = \lambda f$ and $U_T g = \lambda g$ for non-zero $f, g \in L^2$. Then $U_T(f \cdot g) = \lambda_1 \lambda_2 (f \cdot g)$, and $\lambda_1 \lambda_2 \in \sigma_p$.

σ_p is closed under inverse: If $U_T f = \lambda f$, then $|f \circ T| = |f|$. So, $|f|$ is T-invariant and hence, non-zero constant a.e. Thus, since μ is probability measure, $\left\|f^{-1}\right\|_2 < \infty$ and $f^{-1} \in L^2$. Now, $U_T(f \cdot f^{-1}) = U_T(1) = U_T(f) U_T(f^{-1}) \implies U_T(f^{-1}) = \lambda^{-1} f^{-1}$. Hence, $\lambda^{-1} \in \sigma_p$.

eigenvalues are simple: Consider $f, g \in E_\lambda$, then $U_T(f/g) = (\lambda f)/(\lambda g) = f/g$ and $f/g = c$ for some $c \in \mathbb{C}$. Thus, E_λ is one-dimensional.

$\Longleftrightarrow$: Follows from the observation that 1 is a simple eigenvalue of U_T is equivalent to the statement that T-invariant functions are constant a.e.

□

Theorem 2.4.13. *If T is ergodic, then it is weak mixing if and only if $\sigma_p = \{1\}$.*

Proof. *$\Longrightarrow$:* Let $1 \neq \lambda \in \sigma_p$ and non-zero $f \in L^2$ such that $U_T f = \lambda f$. Then $\int f = \int U_T f = \lambda \int f$. As $\lambda \neq 1$, $\int f = 0$. Now,

$$\frac{1}{n} \sum_{k=0}^{n-1} \left| \int f T^k \cdot g - \int f \int g \right| = \frac{1}{n} \sum_{k=0}^{n-1} \left| \langle U_T^k f, \overline{g} \rangle \right| = \langle f, \overline{g} \rangle .$$

The last equality is due to $|\lambda| = 1$. Since, T is weak mixing, the above quantity goes to 0 as $n \to \infty$. Thus, we get $\langle f, g \rangle = 0$ for all $g \in L^2$. So, $f = 0$, which is a contradiction.

$\Longleftarrow$: We show $T \times T$ is ergodic. Let $k \in L^2(\mu \times \mu)$ be a $T \times T$-invariant function. Define $k^*(y, x) = \overline{k(x, y)}$. Then k^* is also $T \times T$-invariant. We can assume $k = k^*$, since we can write any k as sum of such functions, $k = \frac{1}{2}(k + k^*) + \frac{1}{2i} i(k - k^*)$. Define $A : L^2(\mu) \to L^2(\mu)$, $Af(x) = \int k(x, y) f(y) d\mu(y)$.

A is a well-defined bounded operator:

$$\|Af\|_2^2 \leq \int \left(\int |k(x, y) f(y)| d\mu(y) \right)^2 d\mu(x) \leq \|k\|_2^2 \|f\|_2^2 .$$

The last inequality is due to Cauchy-Schwarz inequality. Thus, $Af \in L^2$ and A is bounded.

A is self-adjoint:

$$\langle f, A^* g \rangle = \langle Af, g \rangle = \iint k(x, y) f(y) \overline{g}(x) dy dx = \iint \overline{F^*(y, x) g(x)} dx\, f(y) dy = \langle f, Ag \rangle$$

A is compact: We show that A is a Hilbert-Schmidt operator and use the fact that Hilbert-Schmidt operators are compact.

$AU_T = U_T A$:

$$\begin{aligned} AU_T(f)(x) = Af(Tx) &= \int F(x, y) f(Ty) dy = \int F(Tx, Ty) f(Ty) dy \\ &= \int F(Tx, y) f(y) dy = U_T A(f)(x) \end{aligned}$$

Consider the eigenspaces E_λ corresponding to non-zero eigenvalues λ of A. Then $A|_{E_\lambda} = \lambda.\mathrm{id}$ is compact. We know, $\mathrm{id} : E_\lambda \to E_\lambda$ is compact if and only if E_λ is finite-dimensional. We also have $U_T(E_\lambda) \subset E_\lambda$: if $f \in E_\lambda$, then $AU_T(f) = U_T A(f) = \lambda U_T(f)$, and $U_T(f) \in E_\lambda$. Thus we get $U_T : E_\lambda \to E_\lambda$ is an isomtery between finite-dimensional space, which makes it unitary (as isometries are injective) and hence, diagonalizable. But, $\sigma_p(U_T) = \{1\}$ and by previous theorem, 1 is a simple eigenvalue. Thus, E_λ are one-dimensional and with A and U_T commuting, every eigenfunction of A is also an eigenfunction of U_T. But, 1 is the only eigenfunction of U_T and hence of A. Spectral theorem for self-adjoint compact operator states that $Af = \sum_n \lambda_n \langle f, e_n \rangle e_n$, where λ_n are eigenvalues with eigenfunctions e_n. So, $Af = \lambda \langle f, 1 \rangle 1$, where λ is the eigenvalue of 1. Finally,

$$\begin{aligned} Af(x) &= \lambda \int f(y)dy = \iint k(x,y)f(y)dy \\ \implies &\iint [k(x,y) - \lambda]f(y)dy = 0 \\ \implies &\iint [k(x,y) - \lambda]f(y)g(x) = 0, \text{ for all } f, g \in L^2 \end{aligned}$$

Hence, k is constant a.e., and T is weak-mixing.

□

Chapter 3

Multiple Recurrence Theorem

3.1 Multiple Recurrence Theorem

An interesting generalization of the Poincare recurrence is the multiple recurrence theorem, which states any positive measure set has a positive measure subset that returns to it in an arithmetic progression of arbitrary length. It has many applications, an important one being that it connects ergodic theory and combinatorial number theory.

Theorem 3.1.1 (Multiple Recurrence Theorem)**.** *For a measure-preserving system $(X, \mathcal{B}, \mu, T)$, and a set $A \in \mathcal{B}$, with $\mu(A) > 0$, and for any $k \in \mathbb{N}$, there is some $n \geq 1$ such that*

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

When $k = 1$, this reduces to Poincare recurrence.

3.2 Szemeredi's theorem

We now use this theorem to prove Szemeredi's theorem, which states that any set of integers with positive density contains arithmetic progressions of arbitrary length.

Definition 3.2.1 (Upper density)**.** The *upper density* of a set $A \subset \mathbb{Z}$ is defined as

$$d(A) = \limsup_{N \to \infty} \frac{1}{2N+1} |S \cap \{-n, -n+1, \ldots, n-1, n\}|$$

Theorem 3.2.2 (Szemeredi's Theorem)**.** *If $A \subset \mathbb{Z}$ has positive upper density, then A contains arithmetic progressions of arbitrary length, that is, for all $k \in \mathbb{N}$, there exists $a, b \in \mathbb{Z}$, $b \neq 0$ such that $a, a+b, a+2b, \ldots, a+kb \in A$.*

Proof. □

3.3 Outline of proof of Multiple Recurrence

The rest of the chapter will deal with understanding the proof of the multiple recurrence theorem. We prove a stronger statement, and call this property the *MR-property*:

Theorem 3.3.1. *For a measure-preserving system $(X, \mathcal{B}, \mu, T)$, and a set $A \in \mathcal{B}$, with $\mu(A) > 0$, and for any $k \in \mathbb{N}$, we have*

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n} A \cap T^{-2n} A \cap \ldots T^{-kn} A) > 0$$

We first show that MR-property is satisfied by some systems, such as, weak-mixing systems and Kronecker systems. The main ingredient is understanding how the property lifts by weak-mixing and compact extensions, and understanding the relationship between the two extensions. We then lift the MR-property from Kronecker systems to arbitrary measure-preserving systems, using appropriate extensions.

Definition 3.3.2 (Extensions and factors)**.** Let $(X, \mathcal{B}_X, \mu, T)$ and $(Y, \mathcal{B}_Y, \nu, S)$ be two MPS. Y is a *factor* of X if there are sets $X' \in \mathcal{B}_X$, $Y' \in \mathcal{B}_Y$ with $\mu(X') = 1$, $\nu(Y') = 1$, $TX' \subset X'$, $SY' \subset Y'$, and a measure-preserving map $\phi : X' \to Y'$, such that $\phi \circ T = S \circ \phi$.
And X is called an *extension* of Y.

Essentially, we have a measure-preserving map $\phi : X \to Y$ such that $\phi \circ T = S \circ \phi$ a.e. We next show that factors of a system are in 1-1 correspondence with T-invariant sub-σ-algebras ($\mathcal{A} \subset \mathcal{B}_X$ is T-invariant if $T^{-1}\mathcal{A} = \mathcal{A} \mod \mu$). Given a factor $\phi : X \to Y$, we have T-invariant sub-σ-algebras, $\phi^{-1}\mathcal{B}_Y$. For the other direction,

Theorem 3.3.3. *If $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system and $\mathcal{A} \subset \mathcal{B}_X$ is a T-invariant sub-σ-algebra, then*

Proof. □

We prove the multiple recurrence theorem in the following steps, which we elaborate on in the subsequent sections.

Step 1.

(i) Every MPS has an invertible extension

(ii) A MPS has MR-property if and only if its invertible extension does.

(iii) Every invertible extension has a factor, which is a standard probability space.

Now, given any MPS $(X, \mathcal{B}, \mu, T)$, consider its invertible extension and then a standard probability factor of it. If the standard probability factor satisfies MR-property, then $(X, \mathcal{B}, \mu, T)$ also does. Thus, it is sufficient to prove that any standard measure-preserving system $(X, \mathcal{B}, \mu, T)$ satisfies MR-property.

Step 2.

(i) Weak-mixing and compact systems satisfy MR-property.

(ii) A system is not weak-mixing if and only if it has a non-trivial compact factor.

(iii) If $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \ldots$ is an increasing chain of factors of X that satisfy MR-property, then $\sigma(\cup_{n\geq 1}\mathcal{A}_n)$ also satisfies it.

So, if X is weak-mixing, we are done. If it is not weak-mixing, then consider the family of factors, $\mathcal{A} \subset \mathcal{B}$ such that $(X, \mathcal{A}, \mu, T)$ satisfy MR-property. We know, it is non-empty, since there is a compact factor that satisfies it. By Zorn's lemma, we can show there is a maximal sub-σ-algebra of this family, call it, $\mathcal{B}_\infty$.

Step 3

(i) If X is a weak-mixing extension of Y, which satisfies MR-property, then X also satisfies MR-property.

(ii) If X is a compact extension of Y, which satisfies MR-property, then X also satisfies MR-property.

(iii) If $X \to Y$ is not a weak-mixing extension, then there exists an intermediate factor of $X \to Z$, such that $Z \to Y$ is a compact extension.

Hence, if $(X, \mathcal{B}, \mu, T) \to (X, \mathcal{B}_\infty, \mu, T)$ is a weak-mixing extension, we are done. And if it is not, there is a non-trivial compact extension $(X, \mathcal{C}) \to (X, \mathcal{B}_\infty)$. So $(X, \mathcal{C})$ satisfies MR-property, which contradicts the maximality of $\mathcal{B}_\infty$. Thus, the extension must be weak-mixing, and this completes the proof.

3.3.1 Step 1

Definition 3.3.4 (Invertible extension)**.** Any $(X, \mathcal{B}, \mu, T)$ has an invertible extension, $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu}, \tilde{T})$, where

- $\tilde{X} = \{x \in X^{\mathbb{Z}} \mid Tx_k = x_{k+1} \ \forall k \in \mathbb{Z}\}$
- $(\tilde{T}x)_k = x_{k+1}$ for all $k \in \mathbb{Z}$ and $x \in \tilde{X}$
- $\tilde{\mathcal{B}}$ is the product σ-algebra
- $\tilde{\mu}$ is the product measure

$\pi_0 : \tilde{X} \to X$ is called the invertible extension, where $\pi_0 : X^{\mathbb{Z}} \to X$ is the 0-th projection.

It is easy to show that the above is a MPS. Many properties like ergodicity, weak-mixing and MR-property are preserved under invertible extensions.

Theorem 3.3.5. *$(X, \mathcal{B}, \mu, T)$ has MR-property if and only if its invertible extension does.*

Proof. Note, $\mu = \tilde{\mu} \circ \pi_0^{-1}$. Let $\tilde{A} \in \tilde{\mathcal{B}}$ with positive measure, and let $A = \pi_0(\tilde{A})$. Then $\pi_0^{-1}(A \cap \ldots T^{-kn} A) = (\tilde{A} \cap \ldots \tilde{T}^{-kn} \tilde{A})$. This follows from $x \in \pi_0^{-1}(A \cap \ldots T^{-kn} A) \iff T^{in} x_0 = x_{in} \in A$, for $0 \le i \le k$ and $x_{in} \in A \iff \tilde{T}^{in} x \in \tilde{A}$. Now, the result is clear for one direction. For the converse, start with $A \in \mathcal{B}$ with positive measure and let $\tilde{A} = \pi_0^{-1}(A)$. □

Theorem 3.3.6. *Every invertible system $(X, \mathcal{B}, \mu, T)$ has a factor, that is a standard probability space.*

Proof. Consider the system, $\{0,1\}^{\mathbb{Z}}$, with the product σ-algebra and measure, and the shift operator. Fix $A \in \mathbb{B}$ of positive measure and define a map $\phi : X \to \{0,1\}^{\mathbb{Z}}$, by $\phi(x) = \chi_A(T^n x)$. ϕ is clearly a factor. □

3.3.2 Step 2

To define compact systems, we recall some definitions from point-set topology. A set is *precompact* if its closure is compact and is *totally bounded* if for all $\epsilon > 0$, the set can be covered by a finite number of ϵ-balls. The two are equivalent in a complete metric space.

Definition 3.3.7 (Almost periodic functions)**.** $f \in L^2(X)$ is *almost periodic* if its orbit $\{T^n f\}_{n \in \mathbb{Z}}$ is precompact in $L^2(X)$.

Definition 3.3.8 (Compact systems)**.** $(X, \mathcal{B}, \mu, T)$ is a *compact system* if every $f \in L^2(X)$ is almost periodic.

Theorem 3.3.9. *Compact systems satisfy MR-property.*

Proof. □

Theorem 3.3.10. *$(X, \mathcal{B}, \mu, T)$ is not weak-mixing if and only if it has a non-trivial compact factor.*

Proof. □

Theorem 3.3.11. *Let $(X, \mathcal{B}, \mu, T)$ be a standard invertible MPS and $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \ldots$ is an increasing chain of factors that satisfy MR-property, then $\sigma(\cup_{n \ge 1} \mathcal{A}_n)$ also satisfies it.*

Proof. Let $\mathcal{A} = \sigma(\bigcup A_i)$ and $A \in \mathcal{A}$ with positive measure. Then for any $\epsilon > 0$, there exists some $A_1 \in A_n$ for some $n \in \mathbb{N}$ such that $\mu(A \triangle A_1) < \epsilon$. Fix $k \in \mathbb{N}$ and let $\eta = \frac{1}{2(k+1)}$ and $\epsilon = \frac{1}{4}\eta\mu(A)$. Define $A_0 = \{y \in A_1 \mid \mu_x(A) \ge 1 - \eta\}$.

Claim: $\mu(A_0) > \frac{1}{2}\mu(A)$.

Proof.

$$\frac{1}{4}\eta\mu(A) = \epsilon > \mu(A_1 \setminus A) = \int_{A_1} \mu_x(A_1 \setminus A) d\mu \geq \int_{A_1 \setminus A_0} (1 - \mu_x(A)) d\mu \geq \eta\mu(A_1 \setminus A_0).$$

Hence, $\mu(A_1 \setminus A_0) \leq \frac{1}{4}\mu(A)$, which implies $\mu(A_0) = \mu(A_1) - \mu(A_1 \setminus A_0) > \frac{1}{2}\mu(A)$ (we can show $\mu(A_1) \geq \frac{3}{4}\mu(A)$). □

Claim: $\mu(A \cap \dots T^{-kn} A) \geq \frac{1}{2}\mu(A_0 \cap \dots T^{-kn} A_0)$.

Proof. If $x \in A_0 \cap \dots T^{-kn} A_0$, then for $0 \leq i \leq k$,

$$\mu_x(T^{-in}(A)) = \mu_{T^{in}x}(A) \geq 1 - \eta.$$

Thus,

$$\mu_x(A \cap \dots T^{-kn} A) \geq 1 - (k+1)\eta = \frac{1}{2}.$$

The claim follows by integrating over $A_0 \cap \dots T^{-kn} A_0$. □

Finally, since A_0 satisfies MR-property, by the above claim, A also satisfies.

□

3.3.3 Weak-mixing extension

Definition 3.3.12 (Weak-mixing extension)**.** Let $\phi : (X, \mathcal{B}_X, \mu, T) \to (Y, \mathcal{B}_Y, \nu, S)$ be an extension. Define a measure $\mu \times_Y \mu$ on $(X \times X, \mathcal{B} \times \mathcal{B})$ given by disintegration with respect to $\mathcal{A} := \phi^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_X$,

$$(\mu \times_Y \mu)_x = \mu_x \times \mu_x.$$

Then $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu}, \tilde{T}) := (X \times X, \mathcal{B} \times \mathcal{B}, \mu \times_Y \mu, T \times T)$ is a MPS and the extension is called *weak-mixing* if this system is ergodic.

Note, if we take the trivial system $(X, \{\phi, X\}, \mu, T)$, then any weak-mixing extension over it is simply a weak-mixing system. This follows from the fact that the measure disintegration over the trivial σ-algebra is the measure itself, i.e., $\mu_x = \mu$.

Theorem 3.3.13. *If X is a weak-mixing extension of Y, which satisfies MR-property, then X also satisfies MR-property.*

Corollary 3.3.14. *If X is weak-mixing, then it satisfies MR-property.*

Proof. Follows from the remark above. □

We fix some notations for the remainder for the section. We have $\phi : (X, \mathcal{B}_X, \mu, T) \to (Y, \mathcal{B}_Y, \nu, S)$ a weak-mixing extension. Let $\mathcal{A} := \phi^{-1}(\mathcal{B}_Y)$. For $f, g \in L^\infty(X)$, define $f \otimes g \in L^\infty(X \times X)$,

$$f \otimes g(x_1, x_2) = f(x_1)g(x_2).$$

We start with a useful identity.

Lemma 3.3.15. *If $f, g \in L^\infty_\mu(X)$, then*

$$\int f \otimes g d\tilde{\mu} = \int \mathbb{E}(f|\mathcal{A})\, \mathbb{E}(g|\mathcal{A})\, d\mu. \tag{3.1}$$

Proof.

$$\begin{aligned}\int f\otimes g(x_1,x_2)d\tilde{\mu}(x_1,x_2) &= \iint f\otimes g(x_1,x_2)d\mu_y(x_1,x_2)d\mu(y)\\ &= \iiint f(x_1)g(x_2)d\mu_y(x_1)d\mu_y(x_2)d\mu(y)\\ &= \int \mathbb{E}\left(f|\mathcal{A}\right)(y)\mathbb{E}\left(g|\mathcal{A}\right)(y)d\mu(y).\end{aligned}$$

□

Lemma 3.3.16. *If $f,g\in L^\infty_\mu(X)$, then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\|[\mathbb{E}\left(fT^n g|\mathcal{A}\right)-\mathbb{E}\left(f|\mathcal{A}\right)\mathbb{E}\left(T^n g|\mathcal{A}\right)]\|_{L^2_\mu}=0.$$

Proof. First let f be such that $\mathbb{E}\left(f|\mathcal{A}\right)=0$, then

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int\mathbb{E}\left(fT^n g|\mathcal{A}\right)^2 d\mu &= \lim_{n\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int(f\otimes f)(T^n g\otimes T^n g)d\tilde{\mu}\\ &= \int(f\otimes f)\left[\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\tilde{T}^n(g\otimes g)\right]d\tilde{\mu}\\ &= \int(f\otimes f)d\tilde{\mu}\int(g\otimes g)d\tilde{\mu}=0.\end{aligned}$$

The last line is using Birkhoff's ergodic theorem (by ergodicity of $\tilde{T}$). Now, for any $f\in L^\infty(X)$, consider $f-\mathbb{E}\left(f|\mathcal{A}\right)$. Then $\mathbb{E}\left(f-\mathbb{E}\left(f|\mathcal{A}\right)|\mathcal{A}\right)=0$, and

$$\mathbb{E}\left((f-\mathbb{E}\left(f|\mathcal{A}\right))\cdot T^n g|\mathcal{A}\right)=\mathbb{E}\left(fT^n g|\mathcal{A}\right)-\mathbb{E}\left(f|\mathcal{A}\right)\mathbb{E}\left(T^n g|\mathcal{A}\right).$$

□

Lemma 3.3.17 (van der Corput lemma)**.** *If $\{x_n\}$ be a bounded sequence in a Hilbert space H, and*

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

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

Proof. We normalize so that $\|x_n\|=1$, which we can write as $x_n=O(1)$. This gives,

$$\frac{1}{N}\sum_{n=0}^{N-1}x_n=\frac{1}{N}\sum_{n=0}^{N-1}x_{n+h}+O\left(\frac{|h|}{N}\right).$$

Averaging over $0\le h\le H-1$ for some $H\ge 1$,

$$\frac{1}{N}\sum_{n=0}^{N-1}x_n=\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{H}\sum_{h=0}^{H-1}x_{n+h}+O\left(\frac{H}{N}\right).$$

By triangle inequality and squaring (using $(a+b)^2 \le 2(a^2+b^2)$),

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} x_n \right\|^2 \le O\left(\frac{1}{N} \sum_{n=0}^{N-1} \left\| \frac{1}{H} \sum_{h=0}^{H-1} x_{n+h} \right\|^2 \right) + O\left(\frac{H^2}{N^2}\right)$$
$$= O\left(\frac{1}{H^2} \sum_{0\le h,h'\le H} \frac{1}{N} \sum_{n=0}^{N-1} \langle x_{n+h}, x_{n+h'} \rangle \right) + O\left(\frac{H^2}{N^2}\right).$$

Keeping H fixed,

$$\limsup_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} x_n \right\|^2 \le O\left(\frac{1}{H^2} \sum_{0\le h,h'\le H} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle x_{n+h}, x_{n+h'} \rangle \right).$$

Rewriting, $\sum_{n=0}^{N-1} \langle x_{n+h}, x_{n+h'} \rangle = \sum_{n=0}^{N-1} \left\langle x_{n+|h-h'|}, x_n \right\rangle$ gives

$$\limsup_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} x_n \right\|^2 \le O\left(\frac{1}{H} \sum_{h=0}^{H} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle x_{n+h}, x_n \rangle \right).$$

The result follows by taking the limit as $H\to\infty$. □

Lemma 3.3.18. *For $f_1,\ldots,f_k \in L^\infty(X)$,*

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} \left(T^n f_1 \cdots T^{kn} f_k - T^n \mathbb{E}\left(f_1|\mathcal{A}\right) \cdots T^{kn}\mathbb{E}\left(f_k|\mathcal{A}\right)\right) \right\|_2 = 0.$$

Proof. Proof is by induction on k. When $k=1$, the result follows from the mean ergodic theorem: $\frac{1}{N}\sum_{n=0}^{N-1} T^n f \to \int f$ and $\frac{1}{N}\sum_{n=0}^{N-1} T^n \mathbb{E}\left(f|\mathcal{A}\right) \to \int \mathbb{E}\left(f|\mathcal{A}\right) = \int f$ in $\|\cdot\|_2$.

For the inductive step, we can assume $\mathbb{E}\left(f_j|\mathcal{A}\right) = 0$ for some j. Using

$$\prod_{i=1}^k a_i - \prod_{i=1}^k b_i = \sum_{j=1}^k \left[\left(\prod_{i=1}^k a_i \right) (a_j - b_j) \left(\prod_{i=j+1}^k b_i \right) \right],$$

and with $a_i = T^{in} f_i$ and $b_i = T^{in}\mathbb{E}\left(f_i|\mathcal{A}\right)$, we can reduce the original term to a sum with each summand having the same form, but with a function $T^{in} f_i - T^{in}\mathbb{E}\left(f_i|\mathcal{A}\right)$ that satisfies our assumption.

The result follows if we can use van der Corput lemma, with $x_n = \prod_{i=1}^k T^{in} f_i$.

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle x_n, x_{n+h} \rangle &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \prod_{i=1}^k \left(T^{in} f_i \cdot T^{i(n+h)} f_i \right) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \prod_{i=1}^k T^{in} \left(f_i \cdot T^{ih} f_i \right) \\
(T\text{-invariance of } \mu) &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f_1 T^h f_1 \prod_{i=2}^k T^{(i-1)n} \left(f_i \cdot T^{ih} f_i \right) \\
(\text{induction hypothesis}) &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f_1 T^h f_1 \prod_{i=2}^k \mathbb{E}\left(T^{(i-1)n} \left(f_i \cdot T^{ih} f_i \right) \middle| \mathcal{A} \right) \\
&\le \prod_{i\ne j} \|f_i\|_\infty \left\| \mathbb{E}\left(f_j T^{jh} f_j \middle| \mathcal{A}\right) \right\|_2
\end{aligned}$$

By the previous lemma, the last term converges in density to 0. This satisfies the hypothesis for van der Corput lemma. □

Lemma 3.3.19. *$\tilde{X}$ is a weak-mixing extension of Y.*

Proof. We need to show $(\tilde{X} \times \tilde{X}, \mathcal{B}_{\tilde{X}} \times \mathcal{B}_{\tilde{X}}, \tilde{\mu} \times_Y \tilde{\mu}, \tilde{T} \times \tilde{T})$ is ergodic. We work on a dense set of $L^\infty(\tilde{X} \times \tilde{X})$. Let $F = f_1 \otimes f_2$ and $G = g_1 \otimes g_2$ for $f_i, g_i \in L^\infty_{\tilde{\mu}}(\tilde{X})$. Here, $\mathcal{A} :=$

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int F(\tilde{T} \times \tilde{T})^n G \, d(\tilde{\mu} \times_Y \tilde{\mu}) &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int (f_1 T^n g_1) \otimes (f_2 T^n g_2) d(\tilde{\mu} \times_Y \tilde{\mu}) \\
\text{(using 3.1)} &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \mathbb{E}\left(f_1 T^n g_1 | \mathcal{A}\right) \mathbb{E}\left(f_2 T^n g_2 | \mathcal{A}\right) d\mu \\
\text{(by previous lemma)} &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \prod_{i=1}^{2} \mathbb{E}\left(f_i | \mathcal{A}\right) T^n \mathbb{E}\left(g_i | \mathcal{A}\right) d\mu \\
\text{(by ergodicity of } \tilde{T}) &= \int \prod_{i=1}^{2} \mathbb{E}\left(f_i | \mathcal{A}\right) d\mu \int \prod_{i=1}^{2} T^n \mathbb{E}\left(T^n g_i | \mathcal{A}\right) d\mu \\
&= \int F d\tilde{\mu} \int G d\tilde{\mu}.
\end{aligned}$$

□

Lemma 3.3.20. *For $f_0, \ldots, f_k \in L^\infty(X)$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left\| \mathbb{E}\left(\prod_{i=0}^{k} T^{in} f_i \middle| \mathcal{A} \right) - \prod_{i=0}^{k} \mathbb{E}\left(T^{in} f_i | \mathcal{A}\right) \right\|_{L^2_\mu} = 0.$$

Proof. We prove by induction on k. First, assume $\mathbb{E}\left(f_k | \mathcal{A}\right) = 0$. By the previous lemma, $\tilde{X}$ is a weak-mixing extension over Y. Using lemma 17,

$$\lim_{N\to\infty} \left\| \left(\frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^{k} \tilde{T}^{in} (f_i \otimes f_i) \right) \right\|_{L^2_{\tilde{\mu}}} = 0,$$

from the fact that $\mathbb{E}\left(f_k \otimes f_k | \mathcal{A}\right) = 0$ (by application of 3.1). Note,

$$\prod_{i=1}^{k} \tilde{T}^{in} (f_i \otimes f_i) = \prod_{i=1}^{k} T^{in} f_i \otimes \prod_{i=1}^{k} T^{in} f_i,$$

which gives, using 3.1 and Cauchy-Schwarz inequality,

$$\int \left(\frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}\left(T^{in} f_i | \mathcal{A}\right) \right) d\mu = \int \left(\frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^{k} \tilde{T}^{in} (f_i \otimes f_i) \right) d\tilde{\mu} \to 0.$$

Now, for any $f_k \in L^\infty(X)$, consider instead $(f_k - \mathbb{E}\left(f_k | \mathcal{A}\right)) + \mathbb{E}\left(f_k | \mathcal{A}\right)$. The proof for $f_k - \mathbb{E}\left(f_k | \mathcal{A}\right)$ is as above. For $\mathbb{E}\left(f_k | \mathcal{A}\right)$,

$$\begin{aligned}
&\frac{1}{N} \sum_{n=0}^{N-1} \left\| \mathbb{E}\left(\prod_{i=0}^{k-1} T^{in} f_i \cdot T^{kn} \mathbb{E}\left(f_k | \mathcal{A}\right) \middle| \mathcal{A} \right) d\mu - \prod_{i=0}^{k-1} \mathbb{E}\left(T^{in} f_i | \mathcal{A}\right) \cdot T^{kn} \mathbb{E}\left(f_k | \mathcal{A}\right) \right\|_2^2 \\
&\quad = \frac{1}{N} \sum_{n=0}^{N-1} \left\| T^{kn} \mathbb{E}\left(f_k | \mathcal{A}\right) \right\|_\infty^2 \left\| \mathbb{E}\left(\prod_{i=0}^{k-1} T^{in} f_i \middle| \mathcal{A} \right) - \prod_{i=0}^{k-1} \mathbb{E}\left(T^{in} f_i | \mathcal{A}\right) \right\|_2^2 \to 0,
\end{aligned}$$

using the induction hypothesis. □

Corollary 3.3.21. *For $B_0, \ldots, B_k \in \mathcal{B}_X$,*

$$\frac{1}{N}\sum_{n=0}^{N-1}\int [\mu_x(B_0 \cap \ldots B_k) - \mu_x(B_0)\ldots\mu_x(B_k)]^2\, d\mu \to 0.$$

We are now ready for the main result.

Theorem 3.3.22. *If Y satisfies MR-property, then so does X.*

Proof. Let $B \in \mathcal{B}_X$ with positive measure. Choose $a > 0$ such that the set $A = \{x \mid \mu_x(B)\}$ has positive measure. Then,

$$\begin{aligned}
\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mu(B\cap\ldots T^{-kn}B) &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int\mu_x(B\cap\ldots T^{-kn}B)d\mu \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int\mu_x(B)\ldots\mu_x(T^{-kn}B)d\mu \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\int\mu_x(B)\ldots\mu_{T^{kn}x}(B)d\mu \\
&\geq \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}a^{k+1}\mu(A\cap\ldots T^{-kn}A) > 0.
\end{aligned}$$

The penultimate inequality comes from restricting the integral to $A \cap \ldots T^{-kn}A$, on which we have $\mu_{T^{in}x}(B) > 0$. ☐

3.3.4 Compact extension

Definition 3.3.23 (Relative almost periodic functions)**.** Let $(X, \mathcal{B}_X, \mu, T) \to (Y, \mathcal{B}_Y, \nu, S)$ be an extension. $f \in L^2_\mu(X)$ is *almost periodic* relative to Y if for every $\epsilon > 0$, there is an $r \in \mathbb{Z}$ and function $g_1, \ldots, g_r$ such that

$$\min_{i=1,\ldots,r}\|T^k f - g_i\|_{L^2_{\mu_x}} < \epsilon.$$

for all $n \in \mathbb{N}$ and a.e $x \in X$.

Definition 3.3.24 (Compact extension)**.** $X \to Y$ is a *compact extension*, if the set of functions almost periodic relative to Y is dense in $L^2_\mu(X)$, where μ_x is the measure disintegration .

Theorem 3.3.25. *If X is a compact extension of Y, which satisfies MR-property, then X also satisfies MR-property.*

3.3.5 Step 3.

This is the final theorem to complete the proof.

Theorem 3.3.26. *If $X \to Y$ is not a weak-mixing extension, then there exists an intermediate factor of $X \to Z$, such that $Z \to Y$ is a compact extension.*