

Ergodic Theory

Ronggang Shi, noted by afsd

Contents

1	Week 1	3
2	Week 2	4
3	Week 3	7
4	Week 4	8
5	Week 5	10
6	Week 6	12
7	Week 7	14
8	Week 8	15
9	Week 9	16
10	Week 10	18
11	Week 11	20
12	Week 12	22
13	Week 13	24
14	Week 14	25
15	Week 15	27

Introduction

This course is named “Introduction to Ergodic Theory” and takes place in spring, 2023. The students are assumed to be familiar with real analysis (measure, integral and Littlewood maximal function), functional analysis (Fourier analysis of L^2 space) and basic topology. Others like conditional expectation will be included in the course. Some parts of the theory are included in the exercises, but not in this note.

This course includes the following: basic facts about recurrence, ergodicity and mixing; ergodic decomposition; measure-theoretic entropy and topological entropy; expanding systems.

This note does not include proofs, and is intended to serve as a review before the final exam.

The final exam is on June 9, 2023.

Week 1

A p.m.s. (X, \mathcal{B}, μ, f) refers to a probability space (X, \mathcal{B}, μ) and a measure-preserving map $f : X \rightarrow X$.

Theorem 1.1. (Poincare recurrence)

For a p.m.s. (X, \mathcal{B}, μ, f) and $A \in \mathcal{B}$, we have $\mu(\{x \in A : f^n x \notin A, \forall n \geq 1\}) = 0$.

Moreover, $\mu(\{x \in A : f^n x \in A \text{ i.o.}\}) = \mu(A)$.

Proof. For the first statement, set $N = \{x \in A : f^n x \notin A, \forall n \geq 1\}$.

Then $f^{-n}N$ are pairwise disjoint, and hence $\mu(N) = 0$.

For the second statement, replace f by f^n . □

Proposition 1.2. For a p.m.s. (X, \mathcal{B}, μ, f) , TFAE:

- (1) f is ergodic.
- (2) $\forall B \in \mathcal{B}$ such that $\mu(B \Delta f^{-1}(B)) = 0$, we have $\mu(B) = 0$ or 1 .
- (3) $\forall A \in \mathcal{B}$ such that $\mu(A) = 0$, we have $\mu(\bigcup_{n \geq 1} f^{-n}A) = 0$.
- (4) $\forall A, B \in \mathcal{B}$ such that $\mu(A) > 0, \mu(B) > 0$, there is some $n \geq 1$ such that $\mu(f^{-n}A \cap B) > 0$.
- (5) Any f -invariant measurable function is constant.

Proof. (1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (1). And (2) \Leftrightarrow (5). □

Week 2

U_f is the Koopman operator on L^2 (and sometimes L^1).

Theorem 2.1. For a p.m.s (X, \mathcal{B}, μ, f) , f is ergodic iff 1 is a simple eigenvalue of U_f .

Remark 2.1. 1 is always a eigenvalue of U_f , since constant functions serve as the eigenvectors.

Example 2.1. (Full) Bernoulli Shift.

$(X^{\mathbb{Z}}, \lambda^{\otimes \mathbb{Z}}, \sigma)$, where X is a finite set, λ is a probability on X , and σ is the bilateral shift.

It is an ergodic system, which is proved by the following argument:

Set $\mathcal{S} := \{ \prod_{k \in \mathbb{Z}} A_k : A_k = X \text{ except for a finite number of } k \}$. The semi-algebra \mathcal{S} generates the product σ -algebra \mathcal{B} .

Let $\mathcal{A}(\mathcal{S})$ be the algebra generated by \mathcal{S} . For $A \in \mathcal{S}$,

$$\mu(\sigma^{-j}(A) \cap A) = \mu(A)^2 \text{ when } |j| \text{ is sufficiently large.}$$

Suppose $B \in \mathcal{B}$ and $\sigma^{-1}(B) = B, \mu(B) < 1$. Approximate B by $A \in \mathcal{A}(\mathcal{S})$, i.e. $\mu(A \Delta B) < \varepsilon$.

Note that $\mu(A \Delta B) = \mu(\sigma^{-j}(A) \Delta \sigma^{-j} B) = \mu(\sigma^{-j}(A) \Delta B)$. Take $|j|$ sufficiently large,

$$\mu(B) \leq \mu(A)^2 + 2\varepsilon \leq (\mu(B) + \varepsilon)^2 + 2\varepsilon.$$

Example 2.2. Suppose (X, \mathcal{B}, μ, f) is an extension of (Y, \mathcal{C}, ν, g) . If (μ, f) is ergodic, then (ν, g) is ergodic.

$$\begin{array}{ccc} (X, \mu) & \xrightarrow{f} & (X, \mu) \\ h \downarrow & & \downarrow h \\ (Y, \nu) & \xrightarrow{g} & (Y, \nu) \end{array}$$

Just prove by definition.

Example 2.3. The one-sided Bernoulli shift is a factor of the full Bernoulli shift, and thus is ergodic.

Example 2.4. Suppose $X = \{0, 1, \dots, m-1\}$ in the one-sided Bernoulli shift and $\vec{p} = (p_0, p_1, \dots, p_{m-1})$ is a probability vector, viewed as a probability on X . Then \mathbb{R}/\mathbb{Z} with a appropriate measure is a factor of this system and thus ergodic.

Example 2.5. Suppose $\gamma \in M_{nn}(\mathbb{Z})$, $\det(\gamma) \neq 0$ and acts on $\mathbb{T}^n := \mathbb{R}^n/\mathbb{Z}^n$.

Then $(\mathbb{T}^n, \text{Lebesgue measure}, \gamma)$ is ergodic iff every root of unity is not an eigenvalue of γ .

To prove it, just recall that $L^2(\mathbb{T}^n)$ has an orthonormal basis $e_v(x) = e^{2\pi i \langle v, x \rangle}$, $v \in \mathbb{Z}^n$. A direct calculation shows that $U_\gamma e_v = e_{\gamma^t v}$.

If some root of unity $e^{2\pi i n/m}$ is an eigenvalue of γ , then 1 is an eigenvalue of $(\gamma^m)^t$. Suppose $(\gamma^m)^t v = v$, $v \in \mathbb{Z}^n \setminus \{0\}$, $m \geq 1$. Now $\varphi := e_v + e_{\gamma^t v} + \dots + e_{(\gamma^{m-1})^t v}$ is a non-constant eigenvector(eigenfunction) of U_γ .

Conversely, if γ is not ergodic, there is a non-constant $\varphi \in L^2$ such that $U_\gamma \varphi = \varphi$. Expand $\varphi = \sum_{v \in \mathbb{Z}^n} b_v e_v$, and we know that $b_v = b_{\gamma^t v}$. There is some $v \neq 0$ such that $b_v \neq 0$. However, $\sum_{v \in \mathbb{Z}^n} |b_v|^2 < \infty$, so the set $\{(\gamma^m)^t v : m \in \mathbb{Z}\}$ must be finite.

Example 2.6. The system $(\mathbb{T}^n, \text{Lebesgue measure}, R_\alpha)$, where R_α is translation by $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ is ergodic iff $\{1, \alpha_1, \dots, \alpha_n\}$ is \mathbb{Q} -linearly independent.

As above, $U_f e_v = c_v e_v$ where $c_v = e^{2\pi i \langle v, \alpha \rangle}$, so the system is ergodic iff $c_v \neq 1$ when $v \neq 0$.

Lemma 2.2. Set $B = \{\varphi - U_f \varphi : \varphi \in L_\mu^2\}$, $I = \{\varphi \in L_\mu^2(X) : U_f \varphi = \varphi\}$. Then $B^\perp = I$.

Proof. For $\varphi \in I, \psi \in L_\mu^2$,

$$\begin{aligned} \langle \varphi, \psi - U_f \psi \rangle &= \langle \varphi, \psi \rangle - \langle \varphi, U_f \psi \rangle \\ &= \langle \varphi, \psi \rangle - \langle U_f \varphi, U_f \psi \rangle \\ &= 0. \end{aligned}$$

Conversely, if $\varphi \perp B, \varphi \in L_\mu^2$,

$$\begin{aligned} \|\varphi - U_f \varphi\|^2 &= \langle \varphi - U_f \varphi, \varphi - U_f \varphi \rangle \\ &= \langle -U_f \varphi, \varphi - U_f \varphi \rangle \\ &= \langle -U_f \varphi, \varphi \rangle + \langle U_f \varphi, U_f \varphi \rangle \\ &= \langle -U_f \varphi, \varphi \rangle + \langle \varphi, \varphi \rangle \\ &= \langle \varphi - U_f \varphi, \varphi \rangle \\ &= 0. \end{aligned}$$

□

Theorem 2.3. (von Neumann Mean Ergodic Theorem)

For a p.m.s. (X, \mathcal{B}, μ, f) and $\varphi \in L^2_\mu(X)$, $A_N \varphi$ converges to $P_I \varphi$ in L^2 norm.

Here, $A_N = \frac{1}{N} \sum_{n=0}^{N-1} U_f^n$, and P_I is the orthogonal projection onto I .

Proof. Write $L^2_\mu = I \oplus \overline{B}$. For $\varphi \in L^2_\mu$, $\varphi = \varphi_1 + \varphi_2$ where $\varphi_1 = P_I \varphi, \varphi_2 \in \overline{B}$.

Obviously $A_N \varphi_1 = \varphi_1, \forall N \geq 0$.

For $\varphi_3 = U_f \psi - \psi \in B$, $A_N \varphi_3 = A_N(U_f \psi - \psi) = \frac{1}{N}(U_f^N \psi - \psi) \rightarrow 0$.

Approximate φ_2 by φ_3 . □

Week 3

Theorem 3.1. For a p.m.s. (X, \mathcal{B}, μ, f) and $\varphi \in L^1_\mu$, $A_N \varphi$ converges to an f -invariant function in L^1_μ .

Proof. Approximate. □

Theorem 3.2. (Birkhoff Ergodic Theorem)

For a p.m.s. (X, \mathcal{B}, μ, f) and $\varphi \in L^1_\mu$, $A_N \varphi$ converges to an f -invariant function almost everywhere.

Remark 3.1. By taking a subsequence, the two limit functions above coincide almost everywhere.

Proof. The idea is to mimic the Hardy-Littlewood maximal function. The main tool is the following theorem.

The complete proof is omitted. □

Theorem 3.3. (Maximal Ergodic Theorem)

For a p.m.s. (X, \mathcal{B}, μ, f) and a real-valued function $\varphi \in L^1_\mu$, set $E_\alpha = \{x \in X : \sup_{n \geq 1} A_n \varphi(x) > \alpha\}$, $\forall \alpha \in \mathbb{R}$. Then $\alpha \mu(E_\alpha) \leq 3 \|\varphi\|_1$.

The maximal ergodic theorem is proved with the help of the following lemma.

Lemma 3.4. For $b \in l^1(\mathbb{Z})$, $\alpha > 0$, set $b^*(a) = \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} b(a+i)$ and $E_\alpha^b = \{a \in \mathbb{Z} : b^*(a) > \alpha\}$. Then $\alpha \cdot \#E_\alpha^b \leq 3 \|b\|_1$.

Week 4

First we define subadditivity.

A sequence of numbers $\{a_n\}$ where $a_n \in [-\infty, \infty)$ is subadditive if

$$a_{m+n} \leq a_m + a_n, \forall m, n > 0.$$

A sequence of measurable functions $\{\varphi_n\}$ where $\varphi : X \rightarrow [-\infty, \infty)$ is subadditive with respect to the map $f : X \rightarrow X$ if

$$\varphi_{m+n}(x) \leq \varphi_m(x) + \varphi_n(f^m x), \forall m, n > 0, \forall x \in X.$$

Theorem 4.1. (Kingman, 1968)

Given a p.m.s. (X, \mathcal{B}, μ, f) . Suppose $\varphi_n : X \rightarrow [-\infty, \infty)$ is subadditive with respect to f and the positive part $\varphi_1^+ \in L_\mu^1(X)$. Then $\lim_{n \rightarrow \infty} \frac{1}{n} \varphi_n(x)$ exists for μ -a.e. $x \in X$, and the limit function $\tilde{\varphi}$ satisfies:

- (i) $\tilde{\varphi}$ is f -invariant.
- (ii) $\tilde{\varphi}^+ \in L_\mu^1$.
- (iii) $\int_X \tilde{\varphi} d\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \int_X \varphi_n d\mu = \int_{n \geq 1} \frac{1}{n} \int_X \varphi_n d\mu$.

For an Euclidean space \mathbb{R}^m , we can construct its k -alternating space $\Omega^k(\mathbb{R}^m)$. When \mathbb{R}^m is equipped with a specific inner product (usually the Euclidean inner product), there is a natural transformation from \mathbb{R}^m to its dual, i.e. a 1-form. Then $\Omega^k(\mathbb{R}^m)$ is also an inner product space:

$$\langle v_1 \wedge \cdots \wedge v_k, w_1 \wedge \cdots \wedge w_k \rangle = \det(\langle v_i, w_j \rangle)_{(i,j)}.$$

Here we identify the element $v \in \mathbb{R}^m$ with its natural image $v^*(\cdot) = \langle v, \cdot \rangle \in (\mathbb{R}^m)^*$.

For a matrix $g \in \text{GL}_m(\mathbb{R})$, we denote its (inner product space) operator norm by $\|g\|$ and the operator norm induced by g on $\Omega^k(\mathbb{R}^m)$ by $\|g\|_{\Omega^k}$.

A direct calculation shows that:

Lemma 4.2. Suppose $g \in \text{GL}_m(\mathbb{R})$ and the eigenvalues of $(g^t g)^{1/2}$ are listed with multiplicities as $a_1 \geq a_2 \geq \cdots a_m > 0$. Then $\|g\|_{\Omega^k} = a_1 a_2 \cdots a_k$.

Theorem 4.3. (Oseledets, 1965)

Given a p.m.s. (X, \mathcal{B}, μ, f) and measurable map $A : X \rightarrow \mathrm{GL}_m(\mathbb{R})$ such that $\log \|A(x)\|, \log \|A^{-1}(x)\| \in L^1_\mu$. Set $A(n, x) = A(f^{n-1}x) \cdots A(fx)A(x)$.

Let $\psi_1(n, x) \leq \cdots \leq \psi_m(n, x)$ denote the eigenvalues of $(A^t(n, x)A(n, x))^{1/2}$. Then $\lim_{n \rightarrow \infty} \frac{1}{n} \log \psi_i(n, x)$ exists and is f -invariant.

Proof. Note that the operator norm of $A(n, x)$ satisfies the subadditivity condition, after taking logarithm.

Note that the induced operator norm on $\Omega^k(\mathbb{R}^m)$ satisfies the condition after taking logarithm too. \square

Week 5

Recall that for a p.m.s. (X, \mathcal{B}, μ, f) , ergodicity can be formulated as follows:

$$\frac{1}{N} \sum_{n=0}^{N-1} \mu(f^{-n}A \cap B) \rightarrow \mu(A)\mu(B), \forall A, B \in \mathcal{B}.$$

The system is weak-mixing if

$$\frac{1}{N} \sum_{n=0}^{N-1} |\mu(f^{-n}(A) \cap B) - \mu(A)\mu(B)| \rightarrow 0, N \rightarrow \infty, \forall A, B \in \mathcal{B}.$$

The system is (strong-)mixing if

$$\lim_{n \rightarrow \infty} \mu(f^{-n}(A) \cap B) = \mu(A)\mu(B), \forall A, B \in \mathcal{B}.$$

The system is k -fold mixing (or mixing of order k), if

$$\mu(A_0 \cap f^{-n_1}A_1 \cap \cdots f^{-n_k}A_k) \rightarrow \mu(A_0) \cdots \mu(A_k), \text{ when } n_i - n_{i-1} \rightarrow \infty.$$

It is still an open question whether mixing implies mixing of all orders.

Theorem 5.1. For a p.m.s. (X, \mathcal{B}, μ, f) , TFAE:

- (1) The system is weak-mixing.
- (2) For every $A, B \in \mathcal{B}$, there is a set $J \subset \mathbb{N}$ with zero density such that

$$\lim_{n \rightarrow \infty, n \notin J} \mu(f^{-n}A \cap B) = \mu(A)\mu(B).$$

- (3) For every $A, B \in \mathcal{B}$,

$$\frac{1}{N} \sum_{n=0}^{N-1} |\mu(f^{-n}(A) \cap B) - \mu(A)\mu(B)|^2 \rightarrow 0, N \rightarrow \infty.$$

The proof is based on the following lemma.

Lemma 5.2. For a sequence $\{a_n\}_{n \geq 1}$ where $0 \leq a_n \leq R < \infty$, TFAE:

- (1) $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} a_n = 0$.
- (2) There is a $J \subset \mathbb{N}^*$ with zero density such that $\lim_{n \rightarrow \infty, n \notin J} a_n = 0$.

Theorem 5.3. TFAE:

- (1) (X, \mathcal{B}, μ, f) is weak-mixing.
- (2) $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \times \mu, f \times f)$ is weak-mixing.
- (3) $(X \times X, \mathcal{B} \otimes \mathcal{B}, \mu \times \mu, f \times f)$ is ergodic.
- (4) For any ergodic p.m.s (Y, ν, g) , $(X \times Y, \mu \times \nu, f \times g)$ is ergodic.

Proof. Just prove that $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1)$ and $(1) \Rightarrow (4) \Rightarrow (3)$.

A lemma about approximation is needed. □

Lemma 5.4. Given a p.m.s. (Y, \mathcal{C}, ν, g) and a semi-algebra \mathcal{S} such that $\sigma(\mathcal{S}) = \mathcal{C}$. Consider the statement:

$$\frac{1}{N} \sum_{n=0}^{N-1} |\nu(g^{-n}C \cap D) - \nu(C)\nu(D)| \rightarrow 0.$$

If the statement holds true for all $C, D \in \mathcal{S}$, then it holds true for all $C, D \in \mathcal{C}$.

Now turn to the description of weak-mixing by the Koopman operator.

Theorem 5.5. For a p.m.s. (X, \mathcal{B}, μ, f) , it is weak-mixing iff 1 is the unique eigenvalue of U_f and it is a simple eigenvalue.

Or equivalently, U_f has no non-constant eigenfunctions in L^2_μ .

Week 6

Suppose (X, d) is a compact metric space. Then $P(X)$, equipped with w^* -topology, is a compact metrizable space.

If $f : X \rightarrow X$ is continuous, then $f_* : P(X) \rightarrow P(X)$ is continuous. Set $P_f(X) = \{\mu \in P(X) : f_*\mu = \mu\}$.

Lemma 6.1. $P_f(X)$ is non-empty.

Proof. For any $x \in X$, set $\mu_N = \frac{1}{N} \sum_{n=0}^{N-1} f_*^n \delta_x = \frac{1}{N} \sum_{n=0}^{N-1} \delta_{f^n x}$.

Any accumulation point of $\{\mu_N\}$ lies in $P_f(X)$. □

Theorem 6.2. If $f : X \rightarrow X$ is continuous, then $\mu \in P_f(X)$ is ergodic iff μ is an extreme point of $P_f(X)$.

Proof. If f is non-ergodic, then there is a measurable set A such that $f^{-1}(A) = A$, $0 < \mu(A) < 1$.

Write $\mu = \mu(A) \cdot \frac{1}{\mu(A)} \mu|_A + \mu(A^c) \cdot \frac{1}{\mu(A^c)} \mu|_{A^c}$.

Conversely, suppose $\mu = t\mu_1 + (1-t)\mu_2$, $\mu_i \in P_f(X)$, $t \in (0, 1)$.

Then μ_1 is absolutely continuous with respect to μ , so $d\mu_1 = \varphi d\mu$, $\varphi : X \rightarrow [0, \infty)$ and $\int_X \varphi d\mu = 1$.

Let $B = \{\varphi < 1\}$ be a measurable set. Then $\mu_1(B) = \mu_1(f^{-1}B)$ implies that $\int_B \varphi d\mu = \int_{f^{-1}B} \varphi d\mu$.

So $\int_{B \setminus f^{-1}B} \varphi d\mu = \int_{f^{-1}B \setminus B} \varphi d\mu$.

But $\varphi < 1$ on $B \setminus f^{-1}B$ and $\varphi \geq 1$ on $f^{-1}B \setminus B$, so both sets must be μ -null sets.

This means that B is f -invariant, $\mu(B) = 0$ or 1 . The identity $\int_X \varphi d\mu = 1$ eliminates the possibility of $\mu(B) = 1$.

Finally, $\mu(B) = 0$ obviously implies that $\varphi = 1$ μ -a.e. and $\mu_1 = \mu$. □

Theorem 6.3. If $\mu, \nu \in P_f(X)$ are different ergodic measures, then $\mu \perp \nu$.

Proof. Since $\mu \neq \nu$, there is a $\varphi \in C(X)$ such that $\mu(\varphi) \neq \nu(\varphi)$.

By Birkhoff's ergodic theorem, the (a.e. pointwise) limit of $\frac{1}{N} \sum_{n=0}^{N-1} \varphi(f^n x)$ is $\mu(\varphi)$ and $\nu(\varphi)$.

Be careful that the above a.e. convergence is under two measures μ, ν respectively. So, let A denote the subset where the above identity holds for μ , then $\mu(A) = 1$ and $\nu(A) = 0$. \square

Now prepare for Ergodic Decomposition. First Conditional Expectation.

Theorem 6.4. Suppose (X, \mathcal{B}, μ) is a probability space and $\mathcal{A} \subseteq \mathcal{B}$ is a σ -subalgebra. Then there is a unique map:

$$\mathbb{E}(\cdot|\mathcal{A}) : L^1(X, \mathcal{B}, \mu) \rightarrow L^1(X, \mathcal{A}, \mu)$$

satisfying the following conditions:

- (1) $\mathbb{E}(\varphi|\mathcal{A}) = \varphi, \forall \varphi \in L^1(X, \mathcal{A}, \mu)$.
- (2) $\int_A \mathbb{E}(\varphi|\mathcal{A}) d\mu = \int_A \varphi d\mu, \forall A \in \mathcal{A}, \varphi \in L^1(X, \mathcal{B}, \mu)$.

Moreover, the map also satisfies:

- (3) $\mathbb{E}(\cdot|\mathcal{A})$ is a positive operator of norm 1. Here the positive structure comes from the obvious positive cone.
- (4) If $\psi \in L^\infty(X, \mathcal{A}, \mu)$, then $\mathbb{E}(\psi\varphi|\mathcal{A}) = \mathbb{E}(\varphi|\mathcal{A})\psi$.
- (5) If $\mathcal{A}' \subset \mathcal{A}$ is a σ -subalgebra, then $\mathbb{E}(\cdot|\mathcal{A}') = \mathbb{E}(\mathbb{E}(\cdot|\mathcal{A})|\mathcal{A}')$.
- (6) $|\mathbb{E}(\varphi|\mathcal{A})| \leq \mathbb{E}(|\varphi||\mathcal{A})$.

Theorem 6.5. (Martingale Convergence Theorem)

Given a probability space (X, \mathcal{B}, μ) and $\varphi \in L^1(X, \mathcal{B}, \mu)$. Suppose (\mathcal{A}_n) is an increasing sequence of σ -algebras and $\mathcal{A} = \sigma(\bigcup_{n \geq 1} \mathcal{A}_n)$. Then $\mathbb{E}(\varphi|\mathcal{A}_n) \rightarrow \mathbb{E}(\varphi|\mathcal{A})$ μ -a.e. and in L^1 .

The theorem also holds for a decreasing sequence of σ -algebras.

Lemma 6.6. (Doob's Inequality)

If $\varphi \in L^1(X, \mathcal{B}, \mu)$ and $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \dots \subseteq \mathcal{A}_N \subseteq \mathcal{B}$. Set $E = \{x \in X : \max_{1 \leq i \leq N} \mathbb{E}(\varphi|\mathcal{A}_i) > \lambda\}$. Then $\mu(E) \leq \frac{1}{\lambda} \|\varphi\|_1$.

If $\{\mathcal{A}_n\}$ is increasing or decreasing, then the conclusion holds for $E = \{x : \max_{n \geq 1} \mathbb{E}(\varphi|\mathcal{A}_n) > \lambda\}$.

Week 7

For a countable measurable partition ξ of X , the conditional expectation with respect to $\sigma(\xi)$ is easy to calculate:

$$\mathbb{E}(\varphi|\sigma(\xi))(x) = \int_A \varphi \frac{d\mu}{\mu(A)}, x \in A \in \xi.$$

We denote the set of all p -integrable measurable functions by \mathcal{L}^p , i.e. the usual L^p without the equivalence relation of modulo null sets. We also denote the atom of x with respect to \mathcal{A} by $[x]_{\mathcal{A}}$.

Theorem 7.1. (Disintegration of a measure)

X is a compact metric space, $\mu \in P(X)$, $\mathcal{A} \subseteq \mathcal{B}_X$ where \mathcal{B}_X is the Borel σ -algebra of X .

Then there is a full-measure subset $X' \in \mathcal{A}$ and a measurable map $X' \rightarrow P(X)$, $x \mapsto \mu_x^{\mathcal{A}}$ such that:

- (1) $\mathbb{E}(\varphi|\mathcal{A})(x) = \int_X \varphi d\mu_x^{\mathcal{A}}, \forall \varphi \in \mathcal{L}^1(X, \mathcal{B}, \mu)$.
- (2) If \mathcal{A} is countably generated, then $\mu_x^{\mathcal{A}}([x]_{\mathcal{A}}) = 1, \forall x \in X'$. And $\mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}}$ iff $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$.
- (3) If (1) holds for a countable dense set of $C(X)$, the required map is uniquely determined μ -a.e.
- (4) If $\tilde{\mathcal{A}}$ is another σ -subalgebra of \mathcal{B}_X and $\mathcal{A} = \tilde{\mathcal{A}} \mod \mu$, then $\mu_x^{\mathcal{A}} = \mu_x^{\tilde{\mathcal{A}}}, \mu$ -a.e.

Week 8

Lemma 8.1. Suppose X is a compact metric space, $\mu \in P(X)$, and \mathcal{A} is a σ -subalgebra of \mathcal{B}_X . Then there exists a countably generated $\mathcal{A}' \subseteq \mathcal{A}$ such that $\mathcal{A}' = \mathcal{A} \mod \mu$.

Proposition 8.2. X, μ as above. Suppose $\mathcal{A}' \subseteq \mathcal{A}$ are countably generated σ -subalgebras of \mathcal{B}_X . Then for μ -a.e. $z \in X$ and for $\mu_z^{\mathcal{A}'}$ -a.e. $x \in X$, we have $(\mu_z^{\mathcal{A}'})_x^{\mathcal{A}} = \mu_x^{\mathcal{A}}$.

Corollary 8.3. If \mathcal{A}_n is increasing or decreasing to \mathcal{A} , then $\mu_x^{\mathcal{A}_n} \rightarrow \mu_x^{\mathcal{A}}$ in the w^* -topology for μ -a.e. $x \in X$.

Week 9

Given a probability space (X, \mathcal{B}, μ) and a finite or countable partition ξ , the information function is defined by $I_\mu(\xi)(x) = -\log \mu([x]_\xi)$.

The entropy of the partition ξ is defined by

$$H_\mu(\xi) = \int_X I_\mu(\xi) d\mu = - \sum_{A \in \xi} \mu(A) \log \mu(A).$$

Here $0 \log 0 = 0$.

Define $\varphi(x) = \begin{cases} x \log x, & x > 0. \\ 0, & x = 0 \end{cases}$ Then φ is continuous on $[0, \infty)$ and convex.

Theorem 9.1. Suppose ξ is a partition of (X, \mathcal{B}, μ) and $\sharp \xi = k$. Then $H_\mu(\xi) \leq \log k$, and the equality holds iff $\mu(A) = \frac{1}{k}, \forall A \in \xi$.

For partitions ξ, η , we define $\xi \vee \eta = \{P \cap Q : P \in \xi, Q \in \eta\}$.

We define that ξ is coarser than η , denoted by $\xi < \eta$ if for any $Q \in \eta$, there exists $P \in \xi$ such that $Q \subseteq P \pmod{\mu}$.

Lemma 9.2. If $\xi < \eta$, then $\xi \vee \eta = \eta \pmod{\mu}$.

The conditional entropy of ξ with respect to η is

$$H_\mu(\xi|\eta) = \sum_{B \in \eta} \mu(B) H_{\frac{\mu|_B}{\mu(B)}}(\xi).$$

The conditional information function of ξ with respect to η is

$$I_\mu(\xi|\eta)(x) = -\log \frac{\mu([x]_{\xi \vee \eta})}{\mu([x]_\eta)}, \text{ if } \mu([x]_{\xi \vee \eta}) \neq 0.$$

Proposition 9.3. Suppose ξ, η, ζ are partitions of (X, \mathcal{B}, μ) .

- (1) $H_\mu(\xi|\eta) = \int_X I_\mu(\xi|\eta) d\mu$.
- (2) $I_\mu(\xi \vee \eta|\zeta) = I_\mu(\eta|\zeta) + I_\mu(\xi|\eta \vee \zeta)$.
- $H_\mu(\xi \vee \eta|\zeta) = H_\mu(\eta|\zeta) + H_\mu(\xi|\eta \vee \zeta)$.
- (3) If $\xi < \eta$, then $H_\mu(\xi|\zeta) \leq H_\mu(\eta|\zeta)$, and $H_\mu(\zeta|\xi) \geq H_\mu(\zeta|\eta)$.
- (4) $\xi < \eta$ iff $H_\mu(\xi|\eta) = 0$.

Corollary 9.4. (1) $H_\mu(\xi \vee \eta) \leq H_\mu(\eta) + H_\mu(\xi)$.

- (2) If $H_\mu(\eta) < \infty$, then $H_\mu(\xi|\eta) = H_\mu(\xi \vee \eta) - H_\mu(\eta)$.

Lemma 9.5. Given $k \geq 1, \varepsilon > 0$. There exists $\delta > 0$ such that for all partitions $\xi = \{P_1, \dots, P_k\}, \eta = \{Q_1, \dots, Q_k\}$ satisfying $\mu(P_i \Delta Q_i) \leq \delta$, we have $H_\mu(\xi|\eta) < \varepsilon$.

Lemma 9.6. $\phi : (X, \mathcal{B}, \mu) \rightarrow (Y, \mathcal{C}, \nu)$ is a measure-preserving map between probability spaces, and ξ, η are partitions of Y . Then

$$I_\nu(\xi|\eta) \circ \phi = I_\mu(\phi^{-1}\xi|\phi^{-1}\eta),$$

$$H_\nu(\xi|\eta) = H_\mu(\phi^{-1}\xi|\phi^{-1}\eta).$$

Week 10

Let ξ^n denote $\xi \vee f^{-1}\xi \vee \dots \vee f^{-n+1}\xi$.

Lemma 10.1. $\{H_\mu(\xi^n)\}_{n \geq 1}$ is subadditive.

So we define the entropy (X, μ, f) with respect to ξ to be $h_\mu(f, \xi) := \lim_{n \rightarrow \infty} \frac{1}{n} H_\mu(\xi^n)$.

The entropy of (X, μ, f) is $h_\mu(f) := \sup\{h_\mu(f, \xi) : H_\mu(\xi) < \infty\}$.

We will only consider partitions with finite entropy.

Example 10.1. Bernoulli shift.

$X = Z^\mathbb{N}$, where $Z = \{0, 1, \dots, m-1\}$, and the measure is given by a probability vector $\vec{p} = (p_0, \dots, p_{m-1})$.

The partition ξ consists of m sets of the form $\{i\} \times Z \times Z \times \dots$.

A direct calculation shows that $H_\mu(\xi^n) = -n \sum_{i=0}^{m-1} p_i \log p_i$, and $h_\mu(\sigma, \xi) = -\sum_{i=0}^{m-1} p_i \log p_i$.

Note that $\bigcup_{n \geq 1} \xi^n$ generates the product σ -algebra, and we will see later that this implies that $h_\mu(\sigma) = h_\mu(\sigma, \xi)$.

Proposition 10.2. $h_\mu(f, \xi) = \lim_{n \rightarrow \infty} H_\mu(\xi | f^{-1}\xi^n)$.

Proposition 10.3. (1) $h_\mu(f, \xi) \leq H_\mu(\xi)$.

(2) $h_\mu(f, \xi \vee \eta) \leq h_\mu(f, \xi) + h_\mu(f, \eta)$.

(3) $h_\mu(f, \xi \vee \eta) \leq h_\mu(f, \eta) + H_\mu(\xi | \eta)$.

Proposition 10.4. (1) $h_\mu(f, \xi) = h_\mu(f, \xi^k) = h_\mu(f, f^{-1}\xi)$.

(2) If f is bimeasurable, then $h_\mu(f, \xi) = h_\mu(f^{-1}, \xi) = h_\mu(f, \bigvee_{i=-k}^k f^{-i}\xi)$.

(3) $h_\mu(f^k) = k h_\mu(f)$.

(4) If f is bimeasurable, then $h_\mu(f) = h_\mu(f^{-1})$.

Proposition 10.5. $h_\mu(f) = \sup\{h_\mu(f, \xi) : \# \xi < \infty\}$.

Proof. It follows from the following lemma. □

Lemma 10.6. If $H_\mu(\xi) < \infty$, then for any $\varepsilon > 0$, there exists $\eta < \xi$ such that $\# \eta < \infty$ and $H_\mu(\xi | \eta) < \varepsilon$.

Proposition 10.7. If (Y, \mathcal{C}, ν, g) is a factor of (X, \mathcal{B}, μ, f) , then $h_\mu(f) \geq h_\nu(g)$.

Theorem 10.8. (Kolmogorov-Sinai)

Given a sequence of partitions $\xi_1 < \xi_2 < \dots$ and $H_\mu(\xi_i) < \infty$. If $\sigma(\bigcup_{n \geq 1} \xi_n) = \mathcal{B} \text{ mod } \mu$, then $\lim_{n \rightarrow \infty} h_\mu(f, \xi_n) = h_\mu(f)$.

We call ξ a one-sided generator if $H_\mu(\xi) < \infty$ and $\sigma(\bigcup_{n \geq 0} f^{-n}\xi) = \mathcal{B} \text{ mod } \mu$.

When f is bimeasurable, ξ is a two-sided generator if $H_\mu(\xi) < \infty$ and $\sigma(\bigcup_{n \in \mathbb{Z}} f^{-n}\xi) = \mathcal{B} \text{ mod } \mu$.

Theorem 10.9. (Kolmogorov-Sinai)

- (1) If ξ is a one-sided generator, then $h_\mu(f) = h_\mu(f, \xi)$.
- (2) If f is bimeasurable, ξ is a two-sided generator, then $h_\mu(f) = h_\mu(f, \xi)$.
- (3) If f is bimeasurable, ξ is a one-sided generator, then $h_\mu(f) = 0$.

The third statement needs the following lemma, which is obvious intuitively and similar to the first Kolmogorov-Sinai Theorem.

Lemma 10.10. If $\xi_n < \xi_{n+1}$, $H_\mu(\xi_n) < \infty$, and $\sigma(\bigcup_{n \geq 1} \xi_n) = \mathcal{B} \text{ mod } \mu$, then for any η with finite entropy, we have $\lim_{n \rightarrow \infty} H_\mu(\eta | \xi_n) = 0$.

Proposition 10.11. (X, d) is a compact metric space, $f : X \rightarrow X$ is measurable, $\mu \in P_f(X)$. Suppose $\xi_1 < \xi_2 < \dots$ and $H_\mu(\xi_n) < \infty$. If $\text{diam}([x]_{\xi_n}) \rightarrow 0$ for μ -a.e. $x \in X$, then $h_\mu(f) = \lim_{n \rightarrow \infty} h_\mu(f, \xi_n)$.

Corollary 10.12. X, d, f, μ as above. ξ is a partition with finite entropy. If $\text{diam}([x]_{\xi_n}) \rightarrow 0$ for μ -a.e. $x \in X$, then $h_\mu(f) = h_\mu(f, \xi)$.

Week 11

Theorem 11.1. (Shannon-McMillen-Breiman)

Given a partition ξ with finite entropy. Then for μ -a.e. $x \in X$, $\frac{1}{n}I_\mu(\xi^n)(x)$ converges.

Let $h_\mu(f, \xi)(x)$ denote the limit. Then:

- (1) $\frac{1}{n}I_\mu(\xi^n)(x)$ converges in L^1 to $h_\mu(f, \xi)(x)$.
- (2) $h_\mu(f, \xi)$ is f -invariant.

The proof of the theorem requires a reformulation of conditional information function as conditional expectation.

Now comes to the topological entropy.

Always assume that (X, d) is a compact metric space.

Definition 11.2. For an open cover \mathcal{U} of X , define

$$N(\mathcal{U}) := \int \{\#\gamma : \gamma \subseteq \mathcal{U} \text{ is a subcover}\}$$

For open covers \mathcal{U}, \mathcal{V} , we can define similarly $\mathcal{U} \vee \mathcal{V}$.

It is obvious that $N(\mathcal{U} \vee \mathcal{V}) \leq N(\mathcal{U})N(\mathcal{V})$.

Define $\mathcal{U}^n = \mathcal{U} \vee f^{-1}\mathcal{U} \vee \dots \vee f^{-n+1}\mathcal{U}$. Then $\log N(\mathcal{U}^n)$ is subadditive.

Define $h(f, \mathcal{U}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log N(\mathcal{U}^n)$ to be the entropy of (X, f) with respect to \mathcal{U} .

The entropy of (X, f) is $h(f) := \sup\{h(f, \mathcal{U}) : \mathcal{U} \text{ open cover}\}$.

$\mathcal{U} < \mathcal{V}$ if $\forall V \in \mathcal{V}, \exists U \in \mathcal{U}$ such that $V \subseteq U$.

Definition 11.3. The Bowen metric on X is defined by

$$d_n(x, y) := \max_{0 \leq i \leq n-1} d(f^i x, f^i y).$$

The Bowen ball is define by $B_n(x, r) := \{y \in X : d_n(x, y) < r\}$.

(1) $G \subseteq X$ is (n, ε) -generating if $\{B_n(x, \varepsilon)\}_{x \in G}$ is an open cover of X .

$g_n(f, \varepsilon) := \inf\{\#\mathcal{G} : \mathcal{G} \text{ } (n, \varepsilon)\text{-generating}\}$.

$g(f, \varepsilon) := \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log g_n(f, \varepsilon)$.

$g(f) := \lim_{\varepsilon \rightarrow 0+} g(f, \varepsilon)$.

(2) $S \subseteq X$ is (n, ε) -separated if x, y are different elements of S , then $d_n(x, y) > \varepsilon$.

$$s_n(f, \varepsilon) := \sup\{\#S : S \text{ } (n, \varepsilon)\text{-separated}\}.$$

$$s(f, \varepsilon) := \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log s_n(f, \varepsilon).$$

$$s(f) := \lim_{\varepsilon \rightarrow 0+} s(f, \varepsilon).$$

Week 12

Lemma 12.1. $g_n(f, \varepsilon) \leq s_n(f, \varepsilon) \leq g_n(f, \frac{\varepsilon}{2})$.

From the lemma, we know that $g(f) = s(f)$.

Moreover,

Theorem 12.2. $g(f) = s(f) = h(f)$.

From the proof of theorem, it is shown that:

Corollary 12.3.

$$\begin{aligned} h(f) &= \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log g_n(f, \varepsilon) \\ &= \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log s_n(f, \varepsilon). \end{aligned}$$

Proposition 12.4. If \mathcal{U} is an open cover of X , then $h(f, \mathcal{U}) = h(f, \mathcal{U}^k)$, $\forall k \geq 1$.

Lemma 12.5. (1) \mathcal{U}^{n+k-1} is a subcover of $(\mathcal{U}^k)^n$.

(2) For any subcover \mathcal{V} of $(\mathcal{U}^k)^n$, there is a subcover \mathcal{V}' of \mathcal{U}^{n+k-1} such that $\# \mathcal{V}' \leq \# \mathcal{V}$.

Corollary 12.6. If f is a homeomorphism, then $h(f, \mathcal{U}) = h(f, \bigvee_{j=-k}^k f^j \mathcal{U})$.

Proposition 12.7. $\lim_{\text{diam } \mathcal{U} \rightarrow 0} h(f, \mathcal{U}) = h(f)$. Here $\text{diam } \mathcal{U}$ is $\sup\{\text{diam } U : U \in \mathcal{U}\}$.

Actually, for any sequences of open covers \mathcal{U}_k with $\text{diam } \mathcal{U}_k \rightarrow 0$, it holds that $\lim_{k \rightarrow \infty} h(f, \mathcal{U}_k) = h(f)$.

Proposition 12.8. $h(f^k) = kh(f)$.

Now comes to a good class of system.

Say f is expansive, if there is $\varepsilon_0 > 0$ such that $d(f^n x, f^n y) < \varepsilon_0, \forall n \geq 0$ implies $x = y$.

If f is a homeomorphism, say f is two-sided expansive if there is $\varepsilon_0 > 0$ such that $d(f^n x, f^n y) < \varepsilon_0, \forall n \in \mathbb{Z}$ implies $x = y$.

Example 12.1. (1) Multiplying by $m \in \mathbb{N}^*$ on \mathbb{T} is expansive.

(2) Multiplying by $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ on \mathbb{T}^2 is two-sided expansive but not expansive.

ε_0 is called a constant of expansivity in both cases.

Theorem 12.9. Suppose $\varepsilon_0 > 0$ is a constant of expansivity for f .

(1) $h(f) = h(f, \mathcal{U})$ when $\text{diam } \mathcal{U} < \varepsilon_0$.

(2) $h(f) = g(f, \varepsilon) = s(f, \varepsilon)$ when $\varepsilon < \varepsilon_0/2$.

Particularly, $h(f) < \infty$.

Week 13

Theorem 13.1. (Variational Principle)

Suppose (X, d) is a compact metric space and $f : X \rightarrow X$ is continuous. Then $h(f) = \sup\{h_\mu(f) : \mu \in P_f(X)\}$.

When the entropy map $\mu \mapsto h_\mu(f)$ is upper semi-continuous, there is a $\mu \in P_f(X)$ such that $h_\mu(f) = h(f)$.

Say $x \in X$ is a periodic point of f , if $f^n x = x$ for some $n \geq 1$.

Proposition 13.2. If f is expansive, then $\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \#\text{Fix}(f^n) \leq h(f)$. Here $\text{Fix}(f^n) = \{x \in X : f^n x = x\}$.

Example 13.1. Subshift of finite type.

$Z_m = \{0, 1, \dots, m-1\}$, $\Sigma = Z_m^{\mathbb{N}}$. Suppose A is an $m \times m$ matrix with entries 0, 1, and for any i , there is some j such that $A_{ij} = 1$. Then A gives rise to a subshift of finite type as follows:

$\Sigma_A = \{(x_0, x_1, \dots) \in \Sigma : A_{x_i x_{i+1}} = 1, \forall i \geq 0\}$, and we call a word (w_0, \dots, w_n) A -admissible if $A_{w_i w_{i+1}} = 1, \forall 0 \leq i \leq n$.

Also set $[w_0, \dots, w_n] = \{x \in \Sigma_A : x_i = w_i, 0 \leq i \leq n\} \neq \emptyset$ when (w_0, \dots, w_n) is A -admissible.

Theorem 13.3. Suppose (Σ_A, σ_A) is a subshift of finite type given by $m \times m$ matrix A . Then $h(\sigma_A) = \log \max\{|\lambda_i| : \lambda_i \text{ eigenvalue of } A\}$.

Lemma 13.4. If $a, b \in Z_m$, set $L_n(a, b) = \#\{[w_0 \dots w_n] : w_0 = a, w_n = b, w_0, \dots, w_n \text{ } A\text{-admissible}\}$. Then $L_n(a, b) = A_{a,b}^n$.

Say A is aperiodic, if there is some $n \geq 1$ such that $A_{i,j}^n > 0, \forall i, j$.

Theorem 13.5. If A is aperiodic, then $\lim_{n \rightarrow \infty} \frac{1}{n} \log \#\text{Fix}(\sigma_A^n) = h(\sigma_A)$.

Week 14

Suppose (X, d) is a compact metric space, $f : X \rightarrow X$ is continuous. Say f is expanding if there exists $\sigma > 1, r > 0$ such that $\forall p \in X$,

- (1) $\forall x, y \in B(p, r), d(fx, fy) \geq \sigma d(x, y)$.
- (2) $f(B(p, r))$ contains an open neighbourhood of $\overline{B(fp, r)}$.

On Riemannian manifolds, the second condition is implied by the first.

Denote the closure of all periodic points of f by Λ . We will prove that Λ is similar to X in dynamical sense.

Lemma 14.1. If $f : X \rightarrow X$ is expanding, and a subset $K \subseteq X$ is compact such that $f^{-1}(K) = K$, then $f|_K$ is expanding.

Now comes to inverse branch. By (2) of the definition of expanding maps, we can define an injective continuous map $h_p : B(fp, r) \rightarrow B(p, r)$ such that $f \circ h_p = \text{id}$. Note that h_p is contractive with constant σ^{-1} .

Lemma 14.2. If f is (σ, r) -expanding, then $f^{-1}(B(y, r)) = \bigcup_{x \in f^{-1}y} h_x(B(y, r))$.

Lemma 14.3. If f is (σ, r) -expanding, then $\forall \epsilon \leq r$, we have $f^n(B_{n+1}(p, \epsilon)) = B(f^n p, \epsilon)$. Here B_n is the Bowen ball.

Corollary 14.4. If f is (σ, r) -expanding, then f is r -expansive.

To obtain a periodic orbit, we can use pseudo orbit to approximate.

Say $\{x_n\} \subseteq X$ is a δ -pseudo orbit if $d(fx_n, x_{n+1}) < \delta, \forall n \geq 0$.

Say $\{x_n\}$ is a δ -pseudo periodic orbit if in addition there is some k such that $x_{n+k} = x_n, \forall n \geq 0$.

Theorem 14.5. (Shadow Lemma)

If f is (σ, r) -expanding, then for $0 < \epsilon < \frac{r}{2}$, there is some $\delta > 0$ ($\delta = \frac{1-\sigma^{-1}}{2}\epsilon$) such that for any δ -pseudo orbit $\{x_n\}$, there is a unique $x \in X$ such that $d(f^n, x_n) < \epsilon, \forall n \geq 0$.

Moreover, when $\{x_n\}$ is a δ -pseudo k -periodic orbit, the chosen x is k -periodic.

We say $\{x_n\}_{n \leq 0}$ is a pre-orbit of x_0 if $f(x_n) = x_{n+1}, \forall n < 0$.

Lemma 14.6. If $d(x, y) < r$, $\{x_n\}_{n \leq 0}$ is a pre-orbit of $x_0 = x$, then there is a pre-orbit $\{y_n\}_{n \leq 0}$ of y such that $d(x_n, y_n) \rightarrow 0, n \rightarrow -\infty$.

Theorem 14.7. Let Λ be the closure of all periodic points of f . Then $f(\Lambda) = \Lambda$ and $f|_{\Lambda}$ is expanding.

Week 15

Theorem 15.1. X, f, Λ as before. Then $X = \bigcup_{k \geq 0} f^{-k} \Lambda$.

Corollary 15.2. (1) For any $\mu \in P_f(X)$, we have $\mu(\Lambda) = 1$.
 (2) $h(f) = h(f|_\Lambda)$.

Theorem 15.3. Suppose f is expanding, $X = \Lambda$ and f is topologically exact. Then for any $\varepsilon > 0, s \in \mathbb{N}^*$, there is some $k \in \mathbb{N}^*$ such that $\forall x_1, \dots, x_s \in X, n_1, \dots, n_s \in \mathbb{N}^*, k_1, \dots, k_s \geq k$, there is a point $p \in X$ such that: denote $m_0 = 0, m_j = \sum_{i < j} (n_k + k_i)$,

- (1) $d(f^{m_j+i} p, f^i x_j) < \varepsilon, \forall 1 \leq j \leq s, 0 \leq i \leq n_j - 1$.
- (2) $f^{m_s} p = p$.

Corollary 15.4. X, f as above. For any $\mu \in P_f(X)$, there is a sequence of periodic $\nu_n \in P_f(X)$ such that $\nu \rightarrow \mu$ in w^* -topology.

Here, periodic measures mean measures of the form

$$\frac{1}{k} \sum_{0 \leq i < k} \delta_{f^i p}$$

where p is k -periodic and δ is the Dirac measure.