

Basic concepts with proofs in *Dynamical Systems*, Topological Dynamics and Ergodic Theory

Liu Zhizhou

August 7, 2022

This is a lecture notes of the book [BS02], but full of details.

Contents

1	Topological Dynamics	1
1.1	Properties of Limit Sets	2
1.2	Recurrent and Non-wandering Sets	3
1.3	Topological Entropy	4
1.4	Calculate Topological Entropy on Some Examples	11
2	Ergodic Theory	15
2.1	Measure Theory Preliminaries	15
2.2	Interaction of measures and topologies	19
2.3	Space of Functions	20
2.4	Recurrence	21
2.5	Ergodicity and Mixing	23
2.6	Examples of Ergodicity and Mixing	28
2.7	Important Ergodic Theorems	29
2.8	Invariant Measure for Continuous Map	31

1 Topological Dynamics

Definition 1.0.1: Basic Definition in Topological Dynamics

We introduce the definitions for dynamical systems, homomorphism, semiconjugacy/conjugacy, isometry.

1. A *topological dynamical system* (X, f) is a topological space X and a continuous map $f : X \rightarrow X$. In the context, we always assume X is compact and metric (with metric d) for simplicity.
2. A continuous map f is called *homomorphism* if it is bijective and the inverse is continuous.

3. Let $f : X \rightarrow X$ and $g : Y \rightarrow Y$ be topological systems. A *topological semiconjugacy* from g to f is a surjective continuous map $\pi : Y \rightarrow X$ such that $f \circ \pi = \pi \circ g$. Then (X, f) is a *factor* of (Y, g) and (Y, g) is an *extension* of (X, f) . If π is homomorphism, then it is called a *topological conjugacy*.

$$\begin{array}{ccc} Y & \xrightarrow{g} & Y \\ \downarrow \pi & & \downarrow \pi \\ X & \xrightarrow{f} & X \end{array}$$

4. If (X, d) and (Y, d') are metric spaces, then $f : X \rightarrow Y$ is an *isometry* if $d'(f(x_1), f(x_2)) = d(x_1, x_2)$, for all $x_1, x_2 \in X$. (The same definition in normed space)

Be careful that $f^n(x)$ always represents $f \circ \dots \circ f$ for n times, or in terminology, we say that f iterates n times.

1.1 Properties of Limit Sets

Definition 1.1.1: ω and α limit points and sets

Here we give a definition for ω and α limit points and sets. Let $f : X \rightarrow X$ be a topological dynamical system. Let $x \in X$.

1. A point $y \in Y$ is an ω limit point of x if there is a sequence of natural numbers $n_k \rightarrow \infty$ (as $k \rightarrow \infty$) such that $f^{n_k}(x) \rightarrow y$.
2. The ω limit set of x is the set $\omega(x)$ or denoted $\omega_f(x)$ of all ω limit points of x .
3. If f is invertible, the α limit point and set of x is just change f^{n_k} from the above definition to f^{-n_k} . In other words, the α limit sets of x for f is the ω limit sets of x for f^{-1} .

Note that the limit sets we defined here is just like the limit sets in sequence of numbers.

Proposition 1.1.2: Equivalent definition

Let $f : X \rightarrow X$ be a topological dynamical system. Let $x \in X$. Then

$$\omega(x) = \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{i \geq n} f^i(x)}.$$

Consequently, if f is invertible, then

$$\alpha(x) = \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{i \geq n} f^{-i}(x)}.$$

Proof. We need to prove two sides of inclusion. Firstly, assume $y \in \omega(x)$, then exists

$\{n_k\}$ infinite sequence with $n_k \rightarrow \infty$ such that $f^{n_k}(x) \rightarrow y$. Our aim is to prove $y \in \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{i \geq n} f^i(x)}$, i.e. for all $n \in \mathbb{N}$, $y \in \overline{\bigcup_{i \geq n} f^i(x)}$.

To show this, note that $y \in \overline{A}$ for some set A if and only if for any $\epsilon > 0$, there exist $z \in A$ such that $d(y, z) < \epsilon$, i.e. $A \cap B(y, \epsilon) \neq \emptyset$ for any $\epsilon > 0$.

Now for any $\epsilon > 0$, since $f^{n_k}(x) \rightarrow y$, we have for any $n \in \mathbb{N}$, there exist $n_k \geq n$ such that $f^{n_k}(x) \in B(y, \epsilon)$. Thus, $\forall n \in \mathbb{N}$, we have $\bigcup_{i \geq n} f^i(x) \cap B(y, \epsilon) \neq \emptyset$, which means $y \in \overline{\bigcup_{i \geq n} f^i(x)}$ for all $n \in \mathbb{N}$.

Conversely, assume $y \in \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{i \geq n} f^i(x)}$. Then $y \in \overline{\bigcup_{i \geq n} f^i(x)}$ for all $n \in \mathbb{N}$. So that $\forall n \in \mathbb{N}$, we have $\bigcup_{i \geq n} f^i(x) \cap B(y, \epsilon) \neq \emptyset$. Hence there must exist $\{n_k\}$ with $n_k \rightarrow \infty$ such that $f^{n_k}(x) \rightarrow y$. \square

The observation we used in the proof of the proposition is used to express the feeling of points being approximated by sequence in the topological space. And this proposition makes the connection between 'limit' in the sense of sequence and the 'limit' in the sense of closed sets. This proposition is somewhat trivial because we have similar results for limit points in sequence of numbers.

Proposition 1.1.3: closed and f invariant of limit sets

Both ω and α limit sets are closed and f invariant.

Proof. By the equivalent definition of ω and α limit sets, since it is the intersection of closed sets, it is automatically closed. For the invariant property, choose $y \in \omega(x)$, then exists $\{n_k\}$ such that $f^{n_k}(x) \rightarrow y$. Consider $f(y)$. Since f is continuous, we should have $f^{n_k+1}(x) \rightarrow f(y)$, which shows $f(y) \in \omega(x)$. Here the sequence is $n_k + 1$.

For α limit sets, choose $y \in \alpha(x)$, then exists $\{n_k\}$ such that $f^{-n_k}(x) \rightarrow y$. Consider $f(y)$ also. Since f is continuous, we should have $f^{-n_k+1}(x) \rightarrow f(y)$, which shows $f(y) \in \alpha(x)$. Here the sequence is $n_k - 1$. \square

1.2 Recurrent and Non-wandering Sets

Definition 1.2.1: Recurrent and Non-wandering Points and Sets

Let $f : X \rightarrow X$ be a topological dynamical system. Let $x \in X$.

1. A point x is called *(positively) recurrent* if $x \in \omega(x)$;
2. The set $R(f)$ is defined to be the set that consists of all recurrent points;
3. A point x is *non-wandering* if for any neighborhood U of x , there exists $n \in \mathbb{N}$ such that $f^n(U) \cap U \neq \emptyset$;
4. The set $NW(f)$ is defined to be the set that consists of all non-wandering points.

What does $f(U) \cap U$ mean? Take $x \in f(U) \cap U$, then $x \in U$ belongs to the image of U . That is, the image of a point in U after 1 iteration. Generally, the $f^n(U) \cap U$ is the set

of points after n iterations beginning with a point in U . Likewise, what does $f^{-n}(U) \cap U$ mean? It means the set of points that will lie in U after n iterations.

Proposition 1.2.2

We have the following properties:

1. The set $R(f)$ is f invariant;
2. The set $NW(f)$ is closed and f invariant;
3. $NW(f)$ contains $\omega(x)$ and $\alpha(x)$ for all $x \in X$.

Proof. To show the first property, we need to show that if x is a recurrent point, i.e. $x \in \omega(x)$, then $f(x) \in \omega(f(x))$. If there exists $\{n_k\}$ with $n_k \rightarrow \infty$ such that $f^{n_k}(x) \rightarrow x$, then the same sequence ensures $f^{n_k}(f(x)) = f(f^{n_k}(x)) \rightarrow f(x)$, since f is continuous.

To prove $NW(f)$ is closed, choose $\{x_n\}$ with $x_n \rightarrow x$ where $x_n \in NW(f)$. For any neighborhood U contains x , there exists r such that $B(x, r) \subset U$. And for any $\epsilon > 0$, there exists N such that $x_N \in B(x, \epsilon)$. Then there must exists $x_N \in B(x, r) \subset U$. Since x_N is non-wandering, there exists $n \in \mathbb{N}$ such that $f^n(U) \cap U \neq \emptyset$.

And $NW(f)$ is f invariant because if $x \in NW(f)$, then there exists n such that $f^n(U) \cap U \neq \emptyset$, i.e. there exists $y \in U$ such that $f^n(y) \in U$. Consider $f(x)$, for any neighborhood U of $f(x)$, then $f^{-1}(U)$ will also be open neighborhood of x . Thus there exists n and $y \in f^{-1}(U)$ such that $f^n(y) \in f^{-1}(U)$. So $f^n(f(y))$ will be in U and $f(y) \in U$, which shows $f^n(U) \cap U \neq \emptyset$.

Finally, for any fixed x , assume $y \in \omega(x)$, there must exists $\{n_k\}$ with $n_k \rightarrow \infty$ such that $f^{n_k}(x) \in B(y, \epsilon)$ for any $\epsilon > 0$. Then $f^{n_2-n_1} \circ f^{n_1}(x) = f^{n_2}(x) \in B(y, \epsilon)$. This shows that for any neighborhood U contains y there exists N such that $f^N(U) \cap U \neq \emptyset$.

For $y \in \alpha(x)$, the proof is similar. Just note that $f^{n_2-n_1} \circ f^{-n_2}(x) = f^{-n_1}(x) \in B(y, \epsilon)$. \square

From the properties and the proof above, we can think of recurrence points as points, say x that will have a subsequence of $f^n(x)$ that converges to itself. And non-wandering points are worse. It only grants that there exists sequence of points, say y_k , in neighborhood of x such that $f^{n_k}(y_k) \rightarrow x$ for some $n_k \rightarrow \infty$. From this, the fact that $R(f) \subset NW(f)$ is obvious. And it is also obvious that all periodic points (i.e. exists $N \in \mathbb{N}$ such that $f^N(x) = x$) and all stationary points (i.e. $f(x) = x$) are in $R(f) \subset NW(f)$.

In fact, we have the following proposition.

Proposition 1.2.3

$$\overline{R(f)} \subset NW(f).$$

1.3 Topological Entropy

Topological entropy is the exponential growth rate of the number of essentially different orbit segments of length n . It measures the complexity of the orbit structure of a dynamical system. And we shall show that it is a topological invariant under conjugacy.

We now define a metric $d_n(x, y)$ to measure the maximum distance between x and y after iterates for n times. It is defined to be

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

It is easy to see that $d_n \geq d_{n-1}$ and $d_1 = d$. Moreover, we have the following proposition.

Proposition 1.3.1

Let (X, d) be a compact metric space. Show that the metrics d_i all induce the same topology on X .

Proof. Let $B^n(x, r) = \{y \in X : d_n(x, y) < r\}$. To show d_n and d induce the same topology, it is equivalent to show that: (i) for any $\epsilon > 0$, there exists $\delta > 0$ such that $B^1(x, \delta) \subset B^n(x, \epsilon)$; and (ii) for any $\epsilon' > 0$, there exists $\delta' > 0$ such that $B^n(x, \delta') \subset B^1(x, \epsilon')$.

For (i), choose $\delta = \epsilon$. Then for $y \in B^n(x, \delta)$, we have $d_n(x, y) \geq d(x, y)$ and $d_n(x, y) < \delta = \epsilon$, so $d(x, y) < \epsilon$, i.e. $y \in B^1(x, \epsilon)$.

For (ii), since f is continuous, f^n must also be continuous. For the $\epsilon' > 0$, there exists $\delta_n > 0$ such that once $d(x, y) < \delta_n$, then $d(f^n(x), f^n(y)) < \epsilon'$. Let $\delta' = \min_n \delta_n$. Then for $y \in B^1(x, \delta')$, then we must have $y \in B^n(x, \delta')$, since we have $d_n(x, y) = d(f^n(x), f^n(y)) < \epsilon'$ for all n . Therefore, $B^1(x, \delta') \subset B^n(x, \epsilon')$. \square

We then define three numbers respectively.

Definition 1.3.2: spanning number, separated number, covering number

Fix $\epsilon > 0$.

1. A subset $A \subset X$ is (n, ϵ) *spanning* if for every $x \in X$ there is $y \in A$ such that $d_n(x, y) < \epsilon$. We denote the minimum cardinality of A by $\text{span}(n, \epsilon, f)$;
2. A subset $A \subset X$ is (n, ϵ) *separated* if any two distinct points in A are at least ϵ apart in the metric d_n . Any (n, ϵ) separated set is finite. Let $\text{sep}(n, \epsilon, f)$ denote the maximum cardinality of A .
3. Let $\text{cov}(n, \epsilon, f)$ be the minimum cardinality of a covering of X by sets of d_n -diameter less than ϵ .

By compactness, there are finite (n, ϵ) spanning sets and covering sets, so that the corresponding numbers are finite. The separated number is also finite, since otherwise, by compactness, there is a convergent subsequence of the separated sequence $\{x_n\}$, this contradicts with $d(x_i, x_j) \geq \epsilon$.

Remember that $d_n(x, y)$ is the maximum distance after n times iterations. Therefore, for instance, $\text{cov}(n, \epsilon, f)$ measures the complexity after n iterations. For fixed ϵ , the more chaotic the map f is, the smaller the covering number is.

In other words, think of $f^n(x)$ as continuous time model, i.e. $f^t(x), t \in \mathbb{R}_+$. Then for any fixed point x , $f^t(x)$ is like a 'noodle' with length t . Then, for any element x in the spanning set A , there is noodle of radius ϵ . And the spanning number is finite means that

finite many noodles can span the compact space X . Also, separate number counts the noodles in a different way. Instead of counting the minimum number that generates the whole space X , it counts for the maximum number of distinguishable noodles. And the only difference between spanning and covering is that in spanning, the noodles must be in shape of cylinder with radius (of every slice) less than ϵ , while in covering, the noodles can be of arbitrary shape with diameter (of every slice) less than ϵ .

Intuitively, since circles are of diameter 2ϵ , we must have $\text{cov}(n, 2\epsilon, f) \leq \text{span}(n, \epsilon, f)$. And also, the ϵ -separated noodles is enough to generate the space, so $\text{span}(n, \epsilon, f) \leq \text{sep}(n, \epsilon, f)$. Rigorous proof and more information are presented in the following proposition.

Proposition 1.3.3

$$\text{cov}(n, 2\epsilon, f) \leq \text{span}(n, \epsilon, f) \leq \text{sep}(n, \epsilon, f) \leq \text{cov}(n, \epsilon, f).$$

Proof. Suppose A is an (n, ϵ) spanning set of minimum cardinality. Then the open balls of radius ϵ centered at the points of A cover X . By compactness, there exists $\epsilon_1 < \epsilon$ such that the balls of radius ϵ_1 centered at the points of A also cover X . Their diameter is $2\epsilon_1 < 2\epsilon$. Here we choose ϵ_1 this way because the definition of covering sets needs sets of d_n -diameter less than ϵ . Hence, we have the same number of covering set as the spanning set but this covering set might not be the minimum. Thus $\text{cov}(n, 2\epsilon, f) \leq \text{span}(n, \epsilon, f)$.

Suppose B is an (n, ϵ) separated set of maximum cardinality. Then for any $x \in X$, $d(x, y) \leq \epsilon$ for some $y \in B$, i.e. $\bigcup_{y \in B} B(y, \epsilon)$ spans X . However, this spanning set may not be the minimum, and thus $\text{span}(n, \epsilon, f) \leq \text{sep}(n, \epsilon, f)$.

Finally, suppose \mathfrak{C} is an (n, ϵ) covering set of maximum cardinality. If $\text{sep}(n, \epsilon, f) < \text{cov}(n, \epsilon, f)$, then there exists $\text{sep}(n, \epsilon, f)$ ϵ -separated points. However, note that at least a pair of separated points will lay in the same covering set by Pigeonhole Principle. This will leads to a contradiction since this pair of separated points will have distance strictly less than ϵ . \square

Note that A, B consists of points but CC is the collection of sets in the above proof. The entropy is defined as following. First, let

$$h_\epsilon(f) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log(\text{cov}(n, \epsilon, f)).$$

We have the following lemma.

Lemma 1.3.4

The limit $\lim_{n \rightarrow \infty} \frac{1}{n} \log(\text{cov}(n, \epsilon, f)) = h_\epsilon(f)$ exists and is finite.

Proof. Let U have d_m -diameter less than ϵ , i.e. $\forall x, y \in U, \max_{i=0, \dots, m-1} d(f^i(x), f^i(y)) < \epsilon$. Let V have d_n -diameter less than ϵ , i.e. $\forall x, y \in V, \max_{i=0, \dots, n-1} d(f^i(x), f^i(y)) < \epsilon$. Then $\forall x, y \in U \cap f^{-m}(V)$, we have $x, y \in U$ and $f^m(x), f^m(y) \in V$, so that $U \cap f^{-m}(V)$ has d_{m+n} -diameter less than ϵ . Hence

$$\text{cov}(m+n, \epsilon, f) \leq \text{cov}(m, \epsilon, f) \cdot \text{cov}(n, \epsilon, f)$$

since the number of $U \cap f^{-m}(V)$ kind set is $\text{cov}(m, \epsilon, f) \cdot \text{cov}(n, \epsilon, f)$ and the union must covers X (for any $x \in X$ we can find a pair of U, V such that $x \in U$ and $f^m(x) \in V$ by the definition of covering set). Now let $a_n = \log(\text{cov}(n, \epsilon, f)) \geq 0$ is sub-additive. Recall we have a lemma¹ in sequence of numbers says that if $\{a_n\}$ is sub-additive and non-negative then

$$\lim_{n \rightarrow \infty} \frac{a_n}{n} = \inf_{n \geq 0} \frac{a_n}{n}.$$

Therefore in our situation,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(\text{cov}(n, \epsilon, f)) = \inf_{n \geq 0} \frac{1}{n} \log(\text{cov}(n, \epsilon, f)),$$

also note that $1/n \log(\text{cov}(n, \epsilon, f))$ is bounded from below (one of the lower bound is 0, obviously), we have the limit exists and is finite. \square

Hence, we have

$$h_\epsilon(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \log(\text{cov}(n, \epsilon, f)).$$

Since $\text{cov}(n, \epsilon, f)$ increases monotonically as ϵ decreases (we need more noodles to cover the space if the diameter became thinner), $h_\epsilon(f)$ does as well. Thus, the limit

$$h_{\text{top}} = h(f) = \lim_{\epsilon \rightarrow 0^+} h_\epsilon(f)$$

exists (in the sense of $\mathbb{R} \cup \infty$). And we define it to be the topological entropy as below.

Definition 1.3.5: topological entropy

Let (X, d) be a compact metric space, and $f : X \rightarrow X$ a continuous map, the *topological entropy* of f , denoted by $h_{\text{top}}(f)$ or simply $h(f)$, is defined as

$$h_{\text{top}}(f) = h(f) = \lim_{\epsilon \rightarrow 0^+} h_\epsilon(f) = \lim_{\epsilon \rightarrow 0^+} \lim_{n \rightarrow \infty} \frac{1}{n} \log(\text{cov}(n, \epsilon, f)).$$

By previous inequality (Proposition 1.3.3), we have

$$\frac{1}{n} \log(\text{cov}(n, 2\epsilon, f)) \leq \frac{1}{n} \log(\text{span}(n, \epsilon, f)) \leq \frac{1}{n} \log(\text{sep}(n, \epsilon, f)) \leq \frac{1}{n} \log(\text{cov}(n, \epsilon, f))$$

and take $\limsup_{n \rightarrow \infty}$ or $\liminf_{n \rightarrow \infty}$ then take $\lim_{\epsilon \rightarrow 0^+}$, we will find

$$h(f) = \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \frac{1}{n} \log(\text{cov}(n, \epsilon, f)) = \lim_{\epsilon \rightarrow 0^+} \liminf_{n \rightarrow \infty} \frac{1}{n} \log(\text{cov}(n, \epsilon, f)) \quad (1)$$

$$= \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \frac{1}{n} \log(\text{span}(n, \epsilon, f)) = \lim_{\epsilon \rightarrow 0^+} \liminf_{n \rightarrow \infty} \frac{1}{n} \log(\text{span}(n, \epsilon, f)) \quad (2)$$

$$= \lim_{\epsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \frac{1}{n} \log(\text{sep}(n, \epsilon, f)) = \lim_{\epsilon \rightarrow 0^+} \liminf_{n \rightarrow \infty} \frac{1}{n} \log(\text{sep}(n, \epsilon, f)) \quad (3)$$

We summarize this in the following proposition.

¹Statement of the lemma: Let $\{a_n\}$ be a sub-additive sequence of non-negative real numbers, i.e. $0 \leq a_{m+n} \leq a_m + a_n$ for all $m, n \geq 0$. Then $\lim_{n \rightarrow \infty} a_n/n = \inf_{n \geq 0} a_n/n$. And the proof of this lemma is first fix n and consider $m = kn + r$, then using $\epsilon - N$ language to conclude.

Lemma 1.3.6: Equivalent definition of topological entropy

Let (X, d) be a compact metric space, and $f : X \rightarrow X$ a continuous map. Equation (1)-(3) are equivalent definition of the topological entropy $h(f)$.

Here we comes to see some properties of the number, topological entropy.

Proposition 1.3.7

The topological entropy of an isometry is zero.

Proof. Note that if f is an isometry, then $d_n(x, y) = d(f^n(x), f^n(y)) = d(x, y)$ for all $n \in \mathbb{N}$. Thus, once we fixed ϵ , $h_\epsilon(f)$ will be zero for all $n \in \mathbb{N}$ by definition $h_\epsilon(f) = \lim_{n \rightarrow \infty} 1/n \log(\text{cov}(n, \epsilon, f))$, which completes the proof since $\lim_{\epsilon \rightarrow 0^+} h_\epsilon(f) = \lim_{\epsilon \rightarrow 0^+} 0 = 0$. \square

Lemma 1.3.8: Independent of metric

The topological entropy of a continuous map $f : X \rightarrow X$ does not depend on the choice of a particular metric generating the topology of X .

Proof. Let d, d' both generate the topology of X . Take $\epsilon > 0$, let

$$\delta(\epsilon) = \sup\{d'(x, y) : d(x, y) \leq \epsilon\}.$$

We must have $\delta(\epsilon) > 0$ otherwise x, y will be the same point. We will also have $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$, otherwise there will be a sequence $\{x_n\}$ converges to x in metric d , but do not converge to x in metric d' , which is not possible since they generate the same topology. Now let U be a set with d_n -diameter less than ϵ , then it will have d'_n -diameter at most $\delta(\epsilon)$. This is directly followed from the definition of $\delta(\epsilon)$. From this, we will have $\text{cov}'(n, \delta(\epsilon), f) \leq \text{cov}(n, \epsilon, f)$, where cov and cov' correspond to the metrics d and d' , respectively. Hence,

$$\lim_{\delta \rightarrow 0^+} \lim_{n \rightarrow \infty} \frac{1}{n} \log(\text{cov}'(n, \delta, f)) \leq \lim_{\epsilon \rightarrow 0^+} \lim_{n \rightarrow \infty} \frac{1}{n} \log(\text{cov}(n, \epsilon, f)).$$

Interchanging d and d' gives the opposite inequality. \square

Corollary 1.3.9: Invariant property of topological entropy

Topological entropy is an invariant of topological conjugacy.

Proof. Suppose $f : X \rightarrow X$ and $g : Y \rightarrow Y$ are topologically conjugate dynamical systems, with conjugacy $\pi : Y \rightarrow X$, see Definition 1.0.1. Let d be a metric on X . Then $d'(y_1, y_2)$ defined to be $d(\pi(y_1), \pi(y_2))$ is a metric on Y generating the topology of Y since π is a homomorphism. Thus, since π is isometry, then $d_n(x, y) = d'_n(\pi(x), \pi(y))$. With the help of the homomorphism π , we have $\text{cov}(n, \epsilon, f) = \text{cov}'(n, \epsilon, g)$, where cov and cov' correspond to the metrics d and d' , respectively. And the entropy is independent of metric by Lemma 1.3.8, it follows that $h(f) = h(g)$. \square

Proposition 1.3.10

Let $f : X \rightarrow X$ be a continuous map of a compact metric space X . Then

1. $h(f^m) = m \cdot h(f)$ for $m \in \mathbb{N}$.
2. If f is invertible, then $h(f^{-1}) = h(f)$. Thus $h(f^m) = |m| \cdot h(f)$ for all $m \in \mathbb{Z}$.
3. If A_i , $i = 1, \dots, k$ are closed (not necessarily disjoint) forward f -invariant subsets of X , whose union is X , then

$$h(f) = \max_{1 \leq i \leq k} h(f|A_i).$$

In particular, if A is a closed forward invariant subset of X , then $h(f|A) \leq h(f)$.

Proof. For the first one, note that

$$\max_{0 \leq i < n} d(f^{mi}(x), f^{mi}(y)) \leq \max_{0 \leq j < mn} d(f^j(x), f^j(y)).$$

So that given ϵ , if U has d_{mn} -diameter less than ϵ under f , then it must have d_n -diameter less than ϵ under f^m . Therefore, $\text{cov}(n, \epsilon, f^m) \leq \text{cov}(mn, \epsilon, f)$. Fixed m , take logarithm, divided by n , take $\lim_{n \rightarrow \infty}$ and then $\lim_{\epsilon \rightarrow 0+}$, we will obtain $h(f^m) \leq m \cdot h(f)$. Conversely, for any $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that $d(x, y) < \delta(\epsilon)$ implies $d(f^i(x), f^i(y)) < \epsilon$ for $i = 0, \dots, m$, i.e. $d_m(x, y) < \epsilon$. Then $\text{cov}(n, \delta(\epsilon), f^m) \geq \text{cov}(mn, \epsilon, f)$, since if U has d_n -diameter over less than $\delta(\epsilon)$, then it must be d_{mn} -diameter over less than ϵ , i.e. $d_n(x, y) = d(f^n(x), f^n(y)) < \delta(\epsilon)$ implies $d(f^{mn}(x), f^{mn}(y)) = d_{mn}(x, y) < \epsilon$. Hence, $h(f^m) \geq m \cdot h(f)$.

For the second one, just note that the n -th image of an (n, ϵ) -separated set for f , say $f^n(B)$, where B is the (n, ϵ) -separated set, is (n, ϵ) -separated for f^{-1} . Since for any x, y in $f^n(B)$, $f^{-1}(x), f^{-1}(y)$ will be in B so that they are ϵ -separated. Hence, $h(f^{-1}) \geq h(f)$. Conversely, the n -th image of an (n, ϵ) -separated set for f^{-1} is (n, ϵ) -separated for f , too.

For the third one, it is clear that

$$\text{cov}(n, \epsilon, f) \leq \sum_{i=1}^k \text{cov}(n, \epsilon, f|A_i) \leq k \max_{1 \leq i \leq k} \text{cov}(n, \epsilon, f|A_i).$$

And note that $1/n \log k$ will vanish as $n \rightarrow 0$, we will obtain $h(f) \leq \max_{1 \leq i \leq k} h(f|A_i)$. On the other hand, just note that any (n, ϵ) -separated set in A_i is (n, ϵ) -separated in X , so $\max_{1 \leq i \leq k} h(f|A_i) \leq h(f)$. \square

Proposition 1.3.11

Let (X, d^X) and (Y, d^Y) be compact metric spaces, and $f : X \rightarrow X$, $g : Y \rightarrow Y$ continuous maps. Then

1. $h(f \times g) = h(f) + h(g)$.

2. If f is a factor of g (or equivalently, g is an extension of f , see Definition 1.0.1), then $h(g) \geq h(f)$.

Proof. For the first part, the metric $d((x, y), (x', y')) = \max\{d^X(x, x'), d^Y(y, y')\}$ generates the product topology on $X \times Y$, and

$$d_n((x, y), (x', y')) = \max\{d_n^X(x, x'), d_n^Y(y, y')\}.$$

If $U \subset X$ and $V \subset Y$ have the respective n -diameter less than ϵ , then $U \times V$ has d_n -diameter less than ϵ . Hence,

$$\text{cov}(n, \epsilon, f \times g) \leq \text{cov}(n, \epsilon, f) \cdot \text{cov}(n, \epsilon, g),$$

so $h(f \times g) \leq h(f) + h(g)$. On the other hand, if $A \subset X$ and $B \subset Y$ are both (n, ϵ) -separated, then $A \times B$ is (n, ϵ) -separated by the definition of the deduced metric. Hence

$$\text{sep}(n, \epsilon, f \times g) \geq \text{sep}(n, \epsilon, f) \cdot \text{sep}(n, \epsilon, g),$$

so $h(f \times g) \geq h(f) + h(g)$ by Lemma 1.3.6, the equivalent definition of the entropy.

For the second part, by the definition of f is a factor of g , there exists a continuous projection $\pi : Y \rightarrow X$ such that $f \circ \pi = \pi \circ g$. For any $\epsilon > 0$, exists $\delta_i(\epsilon) > 0$ such that

$$d^Y(g^i(y_1), g^i(y_2)) < \delta(\epsilon) \implies d^X(\pi \circ g^i(y_1), \pi \circ g^i(y_2)) < \epsilon,$$

where $i = 1, \dots, n$. Let $\delta(\epsilon) = \min_{1 \leq i \leq n} \delta_i(\epsilon)$. Then

$$d_n^Y(y_1, y_2) < \delta(\epsilon) \implies d_n^X(\pi(y_1), \pi(y_2)) < \epsilon$$

since $\pi \circ g^i = f^i \circ \pi$. Therefore, for $V \subset Y$ with d_n^Y -diameter less than $\delta(\epsilon)$, $\pi(V)$ has d_n^X -diameter less than ϵ . We will have

$$\text{cov}(n, \epsilon, f) \leq \text{cov}(n, \delta(\epsilon), g).$$

Then take logarithm, divided by n , take $\lim_{n \rightarrow \infty}$ and then $\lim_{\epsilon \rightarrow 0+}$, the proof is complete, since $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$ by the continuity of π . \square

Proposition 1.3.12

Let (X, d) be a compact metric space, and $f : X \rightarrow X$ an expansive homomorphism with expansiveness constant δ . Then $h(f) = h_\epsilon(f)$ for any $\epsilon < \delta$.

Proof. Fix γ and ϵ with $0 < \gamma < \epsilon < \delta$. We will show that $h_\gamma(f) = h_\epsilon(f)$. By monotonicity, it suffices to show that $h_\gamma(f) \leq h_\epsilon(f)$.

By expansiveness, for distinct points x and y , there is some $i \in \mathbb{Z}$ such that $d(f^i(x), f^i(y)) \geq \delta > \epsilon$. By compactness, there is $k = k(\gamma, \epsilon) \in \mathbb{N}$ such that if $d(x, y) \geq \gamma$, then $d(f^i(x), f^i(y)) > \epsilon$ for some $|i| \leq k$. Otherwise, if there is no such k , then we can find x_n, y_n with $d(x, y) \geq \gamma$ but $d(f^i(x_n), f^i(y_n)) \leq \epsilon$ for any $i \in \mathbb{Z}$. Take convergent sub-sequence from $\{x_n\}$ and $\{y_n\}$, and assume the sub-sequence converges to x and y . We would get $d(x, y) \geq \gamma$ and $d(f^i(x), f^i(y)) \leq \epsilon < \delta$, contradicting with expansiveness. Hence, if $A \subset X$ is (n, γ) -separated, then $f^{-k}(A)$ is $(n + 2k, \epsilon)$ -separated by preceded argument. We have

$$\text{sep}(n, \gamma, f) \leq \text{sep}(n + 2k, \epsilon, f).$$

Then take logarithm, divided by n , take $\liminf_{n \rightarrow \infty}$, we get $h_\gamma(f) \leq h_\epsilon(f)$. \square

The meaning of this proposition is that there is not necessary to concern about the behavior after $\epsilon < \delta$, since the map f will eventually make the distance between points no less than δ as $n \rightarrow \infty$.

1.4 Calculate Topological Entropy on Some Examples

The details of the example can be found in Chapter 1 of [BS02].

Proposition 1.4.1: Entropy of full one- or two- sided shift

The topological entropy of full one- or two- sided m shift is $\log m$.

Proof. First consider (Σ_+, σ) . The metric is $d(x, y) = 2^{-l}$, where $l = \min\{|i| : x_i \neq y_i\}$. We claim that the collection

$$\mathfrak{C} = \{C_{w_1, \dots, w_{M+n}}^{1, \dots, M+n} : w_1, \dots, w_{M+n} \in \{1, \dots, m\}\}$$

is the minimal (n, ϵ) covering set. We first show that it is (n, ϵ) covering. Consider $\epsilon = 2^{-M}$. We can think of M sufficiently large since we will finally let $M \rightarrow \infty$. For any $x \in \Sigma_+$, we have $x \in C_{w_1, \dots, w_{M+n}}^{1, \dots, M+n} \in \mathfrak{C}$. Then for any words w in the cylinder, x and w coincides with the first $M+n$ symbols. So after i shifts, $\sigma^i(x), \sigma^i(w)$ coincides with the first $M+n-i$ symbols, i.e. $d(\sigma^i(x), \sigma^i(w)) \leq 2^{-(M+n-i)}$. Therefore

$$d_n(x, w) \leq 2^{-(M+1)} < 2^{-M},$$

which shows \mathfrak{C} is a (n, ϵ) covering set. Moreover, it is a minimal covering set. Since if we remove one of the cylinders in \mathfrak{C} , say $C_{w_1, \dots, w_{M+n}}^{1, \dots, M+n}$. Then let $w = (w_1, \dots, w_{M+n})$. It is clear that this w is not in any of the cylinders in \mathfrak{C} , that is, w will different the elements in any cylinder in \mathfrak{C} at least one symbol, i.e. for any elements x , in any cylinders in \mathfrak{C} , $d(x, w) > 2^{-(M+n)}$. So we must have $d(\sigma^n(x), \sigma^n(w)) > 2^{-M}$, which shows that w can not be covered by the covering set after removing itself. Note that $\text{card } \mathfrak{C} = m^{M+n}$,

$$h(\sigma) = \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{M+n}{n} \log m = \log m.$$

Then consider (Σ, σ) . The idea is the same. We claim that the collection

$$\mathfrak{C} = \{C_{w_{-M}, \dots, w_{M+n}}^{-M, \dots, M+n} : w_{-M}, \dots, w_{M+n} \in \{1, \dots, m\}\}$$

is the minimal (n, ϵ) covering set. Also consider $\epsilon = 2^{-M}$. We can think of M sufficiently large since we will finally let $M \rightarrow \infty$. For any $x \in \Sigma$, we have $x \in C_{w_{-M}, \dots, w_{M+n}}^{-M, \dots, M+n} \in \mathfrak{C}$. Then for any words w in the cylinder, x and w coincides with the coordinates $\{-M, \dots, M+n\}$. So after i shifts, $\sigma^i(x), \sigma^i(w)$ coincides with the coordinates $\{-M+i, M+n-i\}$, i.e. $d(\sigma^i(x), \sigma^i(w)) \leq 2^{-(M+n-i)}$ since $|-M+i| > |M+n-i|$ for large M . Therefore

$$d_n(x, w) \leq 2^{-(M+1)} < 2^{-M},$$

which shows \mathfrak{C} is a (n, ϵ) covering set. Similarly, this is the minimal covering set because the removed element won't be in the cover of the remained cylinders after i shift for some $0 \leq i < n$. Thus the minimal covering number is m^{2M+n} . Consequently,

$$h(\sigma) = \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{2M+n}{n} \log m = \log m.$$

□

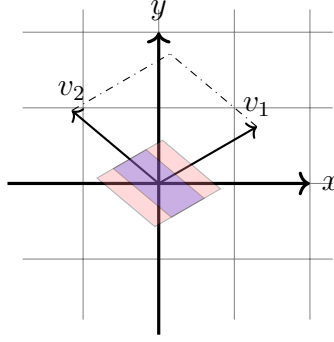


Figure 1: The red parallelogram is a ‘ball’ centered at origin with radius ϵ respect to metric d ; the blue parallelogram is a ‘ball’ centered at origin with radius ϵ respect to metric d_2 .

Proposition 1.4.2: Entropy of hyperbolic toral automorphism

Let \tilde{A} be a 2×2 integer matrix with determinant 1 and eigenvalues λ, λ^{-1} , with $|\lambda| > 1$; and let $A : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ be the associated hyperbolic toral automorphism. Then $h(A) = \log |\lambda|$.

Proof. The natural projection $\pi : \mathbb{R}^2 \rightarrow \mathbb{R}^2/\mathbb{Z}^2 = \mathbb{T}^2$ is a local homomorphism, and $\pi A = A\pi$. Any metric \tilde{d} on \mathbb{R}^2 invariant under integer translation, i.e. $\tilde{d}(x+a, y+b) = \tilde{d}(x, y)$ for all $a, b \in \mathbb{Z}$, induces a metric d on \mathbb{T} , where we define $d(x, y) = \tilde{d}(\pi^{-1}(x), \pi^{-1}(y))$. For these metrics, π is a local isometry.

Let v_1, v_2 be eigenvectors of \tilde{A} with length 1 corresponding to the eigenvalues λ, λ^{-1} . For $x, y \in \mathbb{R}^2$, write $x - y = a_1 v_1 + a_2 v_2$ and define $\tilde{d}(x, y) = \max\{|a_1|, |a_2|\}$. It is clear that \tilde{d} is invariant under integer translation. A \tilde{d} -ball of radius ϵ is a parallelogram whose sides are of (Euclidean) length 2ϵ and parallel to v_1 and v_2 , by setting $|a_1| = |a_2| = \epsilon$.

In the metric \tilde{d}_{n+1} , a ball of radius ϵ is a parallelogram with side length $2\epsilon|\lambda|^{-n}$ in the v_1 -direction and 2ϵ in the v_2 -direction. Indeed, $d_{n+1}(x, y) = \max_{1 \leq i \leq n} \{d(A^i x, A^i y)\} < \epsilon$. Also note that $A^i(x - y) = A^i(a_1 v_1 + a_2 v_2) = \lambda^i a_1 v_1 + \lambda^{-i} a_2 v_2$ if we assume $x - y = a_1 v_1 + a_2 v_2$. Hence $\max_{1 \leq i \leq n} \{|\lambda^i a_1|, |\lambda^{-i} a_2|\} = \max\{|\lambda^n a_1|, |a_2|\} < \epsilon$. We can then obtain the side length. In particular, the Euclidean area of a \tilde{d} -ball of radius ϵ is not greater than $4\epsilon^2|\lambda|^{-n}$. Since the induced metric d on \mathbb{T}^2 is locally isometric to \tilde{d} , we conclude that for sufficiently small ϵ , the Euclidean area of a d_n -ball of radius ϵ in \mathbb{T}^2 is at most $4\epsilon^2|\lambda|^{-n}$. It follows that the minimal number of balls d_n -radius ϵ needed to cover \mathbb{T}^2 is at least $|\lambda|^n/4\epsilon^2$, which is the area of \mathbb{T}^2 , 1, divided by area of $d_n - \epsilon$ ball. In other words,

$$\text{span}(n, \epsilon, A) \geq \frac{|\lambda|^n}{4\epsilon^2}.$$

We have obtained a lower bound of $h(A)$ which is

$$\lim_{\epsilon \rightarrow 0^+} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{|\lambda|^n}{4\epsilon^2} = \lim_{\epsilon \rightarrow 0^+} \log |\lambda| = \log |\lambda|.$$

For the upper bound, since the closed \tilde{d}_n -balls are parallelograms, there is a tiling of the plane by ϵ -balls whose interiors are disjoint. The Euclidean area of such a ball is $C\epsilon^2|\lambda|^{-n}$

where C depends on the angle between v_1 and v_2 . For small ϵ , any ϵ -ball that intersects the unit square $[0, 1] \times [0, 1]$ is entirely contained in the larger square $[-1, 2] \times [-1, 2]$. Therefore the number of balls that intersect the unit square does not exceed the area of the larger square divided by the area of a \tilde{d}_n ball of radius ϵ . It follows that

$$\text{span}(n, \epsilon, A) \leq \frac{9\lambda^n}{C\epsilon^2},$$

thus $h(A) \leq \log |\lambda|$. \square

Proposition 1.4.3: Entropy of E_m

The topological entropy of the expanding map $E_m : S^1 \rightarrow S^1$ is $\log m$.

Proof. The distance on S^1 is $d(x, y) = \min\{|x - y|, 1 - |x - y|\}$. Recall that if $d(x, y) < 1/2m$, then $d(E_mx, E_my) = md(x, y)$. Consider x, y is sufficiently close. We can think so since we have to let $\epsilon \rightarrow 0^+$ when calculating entropy. With this assumption, we obtain

$$d_{n+1}(x, y) = \max_{0 \leq i \leq n} d(E_m^i x, E_m^i y) = m^n d(x, y).$$

Therefore, the d_{n+1} ball with radius ϵ would have natural length $2\epsilon/m^n = 2m^{-n}\epsilon$. Then consider the closed balls with disjoint interior. Since the geometrical figure is clear (they are all actually intervals), it is clear that the minimal number of $(n + 1, 2\epsilon)$ balls needed to cover S^1 will be

$$\frac{1}{2m^{-n}\epsilon} \leq \text{cov}(n + 1, 2\epsilon, E_m) < \frac{1}{2m^{-n}\epsilon} + 1.$$

Divided by $n + 1$, then taking $\liminf_{n \rightarrow \infty}$ and $\lim_{\epsilon \rightarrow 0^+}$, we get $h(E_m) = \log m$. \square

Proposition 1.4.4: Entropy of solenoid

The topological entropy of the solenoid map $F : S \rightarrow S$ is $\log 2$.

Proof. Recall that F is topologically conjugate to the automorphism $\alpha : \Phi \rightarrow \Phi$, where

$$\Phi = \{(\phi)_{i=0}^\infty : \phi_i \in [0, 1), \phi_i = 2\phi_{i+1} \pmod{1}\},$$

and α is coordinate-wise multiplication by 2 ($\pmod{1}$). Thus $h(F) = h(\alpha)$. Define distance on Φ to be

$$d((\phi), (\phi')) = \sum_{n=1}^{\infty} \frac{1}{2^n} |\phi_n - \phi'_n|,$$

where $|\cdot|$ is the distance on $S^1 = [0, 1] \pmod{1}$.

The map $\pi : \Phi \rightarrow S^1$, $(\phi) \mapsto \phi_0$ is a semiconjugacy from α to E_2 . Hence, $h(\alpha) \geq h(E_2) = \log 2$ (Proposition 1.4.3). We now show that $h(\alpha) \leq \log 2$ by constructing an (n, ϵ) -spanning set.

Fix $\epsilon > 0$ and choose $k \in \mathbb{N}$ such that $2^{-k} < \epsilon/2$. For $n \in \mathbb{N}$, let $A_n \subset \Phi$ consists of 2^{n+2k} sequences, $\psi^j = (\psi_i^j)$, where $\psi_i^j = j \cdot 2^{-(n+k+i)} \pmod{1}$, $j = 0, \dots, 2^{n+2k} - 1$. We claim that A_n is (n, ϵ) -spanning. Let $\phi = (\phi_i)$ be a point in Φ . Choose $j \in \{0, \dots, 2^{n+2k} - 1\}$

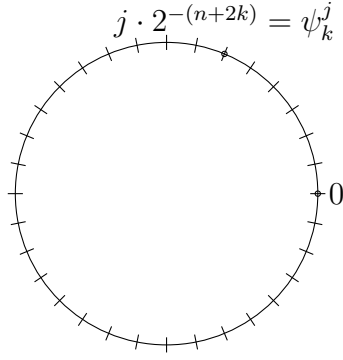


Figure 2: We can separate S^1 into 2^{n+2k} pieces with each of length $2^{-(n+2k)}$, here $n = 1, k = 2$.

so that $|\phi_k - j \cdot 2^{-(n+2k)}| \leq 2^{-(n+2k+1)}$. It is easy to see the reason with the help of Figure 2. Rigorously speaking, since ϕ_k lies in one of the intervals of

$$[j \cdot 2^{-(n+2k)}, (j+1) \cdot 2^{-(n+2k)}), \quad j \in \{0, \dots, 2^{-(n+2k)} - 2\},$$

there must exist such j . Then $|\phi_i - \psi_i^j| \leq 2^{k-i} 2^{-(n+2k+1)}$ since we have $E_m x$ preserves the distance when $d(E_m x, E_m y) < 1/2$. It follows that

$$d(\alpha^m \phi, \alpha^m \psi^j) = \sum_{i=1}^{\infty} \frac{|2^m \phi_i - 2^m \psi_i^j|}{2^i} < \sum_{i=0}^k \frac{2^m |\phi_i - \psi_i^j|}{2^i} + \frac{1}{2^k} < 2^m \sum_{i=0}^k \frac{2^{k-i} 2^{-(n+2k+1)}}{2^i} + \frac{1}{2^k} < \frac{1}{2^{k-1}} < \epsilon,$$

where we use the distance on S^1 is no more than 1. Thus $d_n(\phi, \psi^j) < \epsilon$, so A_n is (n, ϵ) -spanning. Hence,

$$h(\alpha) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log 2^{n+2k} = \log 2.$$

□

Proposition 1.4.5: Solenoid map is expansive

The solenoid map is expansive. Consequently, $h(F) = h_\epsilon(F)$ for ϵ less than the expansiveness constant (Proposition 1.3.12), which in fact is $1/3$.

Proof. [Wil55] Since the solenoid map F is topologically conjugate to the automorphism $\alpha : \Phi \rightarrow \Phi$. It suffices to show α is expansive, i.e. there is a $\delta > 0$ such that for any distinct points $\phi, \psi \in \Phi$ there exists $n \in \mathbb{Z}$ such that $d(\alpha^n(\phi), \alpha^n(\psi)) > \delta$. Recall the metric in Φ we used is

$$d(\phi, \psi) = \sum_{n=0}^{\infty} \frac{1}{2^n} |\phi - \psi|.$$

There are two cases for distinct points ϕ, ψ . The first case is that $\phi_i \neq \psi_i$ for each $i \in \mathbb{N}$. The second case is that $\phi_i = \psi_i$ for $i = 0, \dots, n-1$, but not equal in $i \geq n$. We can only have those two cases since once $\phi_n = \psi_n$ for some n , then $\phi_i = \psi_i$ for $i \leq n$ by the definition of points in Φ . And we cannot conclude $\phi_i = \psi_i$ for $i > n$ since E_m is not invertible, as there might have two preimages.

Now consider case 1, i.e. $\phi_i \neq \psi_i$ for each $i \in \mathbb{N}$. Since α is coordinate-wise multiplication by 2 (mod 1), consider only the first coordinate ϕ_0 and ψ_0 . If $|\phi_0 - \psi_0| < 1/4$, then $|\alpha(\phi_0) - \alpha(\psi_0)| = 2|\phi_0 - \psi_0|$. It is expanding, so we must have some $n \in \mathbb{N} \subset \mathbb{Z}$ such that $|\alpha^n(\phi_0) - \alpha^n(\psi_0)| > 1/4$. Then $d(\alpha^n(\phi), \alpha^n(\psi)) > 1/4$.

For the second case, if we iterate backward n times, then

$$\alpha^{-n}(\phi) = (\phi_n, \phi_{n+1}, \dots), \quad \alpha^{-n}(\psi) = (\psi_n, \psi_{n+1}, \dots)$$

and $\phi_n \neq \psi_n$. We should have $|\phi_n - \psi_n| = 1/2$ since $\phi_{n-1} = 2\phi_n = 2\psi_n = \psi_{n-1} \pmod{1}$. Then $d(\alpha^{-n}(\phi), \alpha^{-n}(\psi)) > 1$.

For both cases, α is expansive, so the proof completes. \square

2 Ergodic Theory

In Topological dynamics, we are interested in every point in the dynamics and the qualitative behavior of it. However, in Ergodic theory, we are interested in the statistical properties of a dynamical system and the behavior of a full-measure set.

2.1 Measure Theory Preliminaries

Definition 2.1.1: σ -algebra

A collection of set $\mathfrak{A} \subset \{A : A \subset X\}$ is called σ -algebra if it satisfies:

1. $\mathfrak{A} \neq \emptyset$;
2. If $A \in \mathfrak{A}$, then $X - A \in \mathfrak{A}$;
3. If $\{A_n\}_{n \in \mathbb{N}} \subset \mathfrak{A}$, then

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathfrak{A}.$$

Proposition 2.1.2

If $\{A_n\}_{n \in \mathbb{N}} \subset \mathfrak{A}$, then

$$\bigcap_{n \in \mathbb{N}} A_n \in \mathfrak{A}.$$

Proof. Use the second and third assumption in the definition of σ -algebra. \square

Any non-empty family of sets in 2^X , \mathcal{A} , generates a σ -algebra. We will call it $\mathfrak{A}(\mathcal{A})$, the σ -algebra generated by \mathcal{A} . (We haven't defined what is 'generate', I think it means the span with operators intersections, unions and complements.)

Definition 2.1.3

Given a set X compact and metric, \mathfrak{A} a σ -algebra on X . We say a function

$\mu : \mathfrak{A} \rightarrow \mathbb{R}_0^+$ is a *measure* if it is σ -additive, i.e.

$$\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n),$$

for any disjoint $A_n \in \mathfrak{A}$.

Proposition 2.1.4

$$\mu(\emptyset) = 0.$$

Definition 2.1.5

1. If $\mu(A) = 0$, then we call A a *null-set*. If $\mu(X - A) = 0$, then A is a *full-measure* set.
2. We say that a measure μ is *complete*, if $A \subset \mathfrak{A}$ such that $\mu(A) = 0$, then for all $E \subset A$ we have $E \in \mathfrak{A}$.
3. And if $A \in \mathfrak{A}$, then A is said to be *measurable*.
4. Given \mathfrak{A} a σ -algebra, with μ measure. $\overline{\mathfrak{A}}$ is a completion of \mathfrak{A} if it is the smallest complete σ -algebra containing \mathfrak{A} . In other words, $\overline{\mathfrak{A}}$ is the σ -algebra generated by \mathfrak{A} and all the subsets of null-sets of μ .

Definition 2.1.6: Measure space

A *measure space* is the triple (X, \mathfrak{A}, μ) , where X any set, \mathfrak{A} is a σ -algebra, μ is a measure.

In the context, we will always assume \mathfrak{A} is complete and \mathfrak{A} is σ -finite, i.e.

$$X = \bigcup_{n \in \mathbb{N}} X_n \quad \mu(X_n) < \infty.$$

Definition 2.1.7

Let (X, \mathfrak{A}, μ) and (Y, \mathfrak{B}, ν) be a measure space.

1. If $\mu(X) = 1$, then we called the corresponding measure space a *probability space* and μ is a probability measure. If $\mu'(X) < \infty$ then $\nu(A) = \mu'(A)/\mu'(X)$ will be a probability measure.
2. The *product measure* is $(X \times Y, \mathfrak{C}, \mu \times \nu)$ where \mathfrak{C} is the completion of the σ -algebra generated by the sets of the form $A \times B$ where $A \in \mathfrak{A}$, $B \in \mathfrak{B}$, and $(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$.
3. Let (X, \mathfrak{A}, μ) and (Y, \mathfrak{B}, ν) be a measure space. $T : X \rightarrow Y$ is a *measurable map* if for all $B \in \mathfrak{B}$, $T^{-1}(B) \in \mathfrak{A}$.

4. A measurable map $T : X \rightarrow Y$ is *non-singular* if $\nu(E) = 0$ then $\mu(T^{-1}(E)) = 0$. The pre-image of a null-set is a null-set.

Definition 2.1.8: Measure-preserving and Invariant

1. A measurable map $T : X \rightarrow Y$ is *measure-preserving* if

$$\mu(T^{-1}(A)) = \nu(A)$$

for all $A \in \mathfrak{B}$. In particular, for measure $T : X \rightarrow X$, then the definition of measure preserving will be like

$$\mu(T^{-1}(A)) = \mu(A)$$

for all $A \in \mathfrak{A}$.

2. When a transformation preserves a measure μ , then we say that μ is *T-invariant*.

The meaning of preserving measure in practical is that when the dynamic transformation T acts on some subset of X , then the volume of the subset won't change. See the informal explanation in the beginning of section 2.5.

Example 2.1.9: Lebesgue measure is invariant under E_2

The Lebesgue measure on the unit interval $[0, 1]$ is invariant under the E_2 transformation, where E_2 is the doubling map defined on S^1 , i.e. $E_2 : x \mapsto 2x \pmod{1}$.

Proof. It suffices to prove that the measure of any interval is the same as the measure of the preimage of the interval under E_2 . Indeed, consider a measurable subset A of $[0, 1]$. The first one of Littlewood's three principles (cf. Chapter 1, Section 4.3 in [SS05]) says that every set is nearly a finite union of intervals. Formally, we have the following theorem (cf. Chapter 1, Theorem 3.4 in [SS05]):

Suppose E is measurable subset of \mathbb{R}^d with finite measure. Then for every $\epsilon > 0$ there exists a finite union $F = \bigcup_{j=1}^N Q_j$ of closed cubes such that $m(E \triangle F) \leq \epsilon$.

In our case, we can assume without loss of generality that $I = \bigcup_{j=1}^N I_j$ with I_j almost disjoint intervals that satisfies $m(A \triangle I) \leq \epsilon$. If $m(E_2^{-1}I_j) = m(I_j)$ for all j then

$$m(E_2^{-1}(I)) = m(E_2^{-1}(\bigcup_{j=1}^N I_j)) = m(\bigcup_{j=1}^N E_2^{-1}(I_j)) = \sum_{j=1}^N m(E_2^{-1}(I_j)) = \sum_{j=1}^N m(I_j) = m(I).$$

Therefore

$$m(E_2^{-1}(A)) = m(E_2^{-1}(I)) + m(E_2^{-1}(A - I)) = m(I) + m(E_2^{-1}(A - I)) = m(A) - m(A - I) + m(E_2^{-1}(A - I)).$$

Now consider $I = [a, b]$, with $m(I) = b - a$. We have $E_m^{-1}(I) = [\frac{a}{2}, \frac{b}{2}] \cup [\frac{a+1}{2}, \frac{b+1}{2}]$ so that $E_m^{-1}(I) = \frac{b-a}{2} + \frac{b-a}{2} = b - a$. The proof is complete. \square

Lemma 2.1.10: Invariant of integral

T is measure preserving if and only if for each $f \in L^1(X, \mathfrak{A}, \mu)$, we have

$$\int f d\mu = \int f \circ T d\mu.$$

Proof. First note that for $A \in \mathfrak{A}$, $\mathbf{1}_A \in L^1(X, \mathfrak{A}, \mu)$ and $\mathbf{1}_A \circ T = \mathbf{1}_{T^{-1}(A)}$. The latter is because they both equal 1 if $Tx \in A$ and equal 0 otherwise.

Assume T is measure preserving, then $\mu(A) = \mu(T^{-1}(A))$ for $A \in \mathfrak{A}$, so that

$$\int \mathbf{1}_A d\mu = \mu(A) = \mu(T^{-1}(A)) = \int \mathbf{1}_{T^{-1}(A)} d\mu = \int \mathbf{1}_A \circ T d\mu.$$

Hence we prove the result for simple functions. And then for any $f \in L^1(X, \mathfrak{A}, \mu)$, we can find an increasing sequence of simple functions f_n with $f_n \rightarrow f$ as $n \rightarrow \infty$. For each n we have

$$\int f_n d\mu = \int f_n \circ T d\mu,$$

then applying Monotone Convergence Theorem to both sides, we get

$$\int f d\mu = \int f \circ T d\mu.$$

Then for general real-valued f , consider $f = f^+ - f^-$, we have

$$\int f d\mu = \int f^+ - f^- d\mu = \int f^+ \circ T - f^- \circ T d\mu = \int f \circ T d\mu,$$

which completes the proof of the first part.

Conversely, just let $f = \mathbf{1}_A$, then

$$\mu(A) = \int \mathbf{1}_A d\mu = \int \mathbf{1}_A \circ T d\mu = \int \mathbf{1}_{T^{-1}(A)} d\mu = \mu(T^{-1}(A)).$$

□

This lemma is very useful. Let us see an example using it.

Example 2.1.11

Let T be a measure-preserving transformation of (X, \mathfrak{A}, μ) , and let $f \in L^1(X, \mu)$ satisfy $f(T(x)) \leq f(x)$ for a.e. x . Then $f(T(x)) = f(x)$ for a.e. x .

We will see later that f is called essentially T -invariant function.

Proof. Assume that the set of x such that $f(T(x)) < f(x)$ has positive measure. Therefore we must have

$$\int f T d\mu < \int f d\mu,$$

which contradicts with the lemma.

□

Lemma 2.1.12

If T is a invertible, measurable transformation and T^{-1} measurable and non-singular then $\{T^n\}_{n \in \mathbb{Z}}$ forms a group of measurable transformation.

We need the assumption ‘non-singular’ because in principle, an injective transformation could be singular. For instance, we could have an *atom*, i.e. $x \in X$ and $\mu(\{x\}) > 0$, and also $\mu(\{Tx\}) = 0$. Then we would have the singular property for $E = \{Tx\}$.

Definition 2.1.13: Isomorphic, automorphism and extension

1. Two measure spaces, (X, \mathfrak{A}, μ) , (Y, \mathfrak{B}, ν) are *isomorphic* if there exists full set $X' \subset X$, full set $Y' \subset Y$ and $T : X' \rightarrow Y'$ bijection such that T, T^{-1} are measurable and measure preserving.
2. An isomorphism $T : (X, \mathfrak{A}, \mu) \rightarrow (X, \mathfrak{A}, \mu)$ is called *automorphism*.
3. Let T and S be measure preserving transformation of (X, \mathfrak{A}, μ) and (Y, \mathfrak{B}, μ) . T is an *extension* of S , if there exists full set $X' \subset X$ and full set $Y' \subset Y$, and exists a measure preserving map $\psi : X' \rightarrow Y'$ satisfying $\psi \circ T = S \circ \psi$. If ψ

is an isomorphism, then T, S are *isomorphic*.

$$\begin{array}{ccc} X' & \xrightarrow{T} & X' \\ \downarrow \psi & & \downarrow \psi \\ Y' & \xrightarrow{S} & Y' \end{array}$$

Later we will define a measure μ on (Σ^n, σ) . In dynamical systems, any transformation isomorphic to σ is call *Bernoulli*.

In particular, the product of two measure sets is an extension of either of the two sets.

2.2 Interaction of measures and topologies

Definition 2.2.1: Borel σ -algebra and Borel measure

1. Let X be a topological space, (X, d) a compact metric space. The *Borel σ -algebra* is the smallest σ -algebra containing all open sets of X .
2. If \mathfrak{B} is the Borel σ -algebra, then a measure μ called a *Borel measure* if $\mu(K) < \infty$ for all K compact.

Proposition 2.2.2

A Borel measure is always regular, i.e. for all $A \in \mathfrak{B}$, $\mu(A) = \inf\{\mu(U) : U \text{ open, } A \subset U\}$, $\mu(A) = \sup\{\mu(K) : K \text{ compact, } K \subset A\}$. In particular, for all $A \in \mathfrak{B}$, there exists $K \subset A \subset U$ such that for all $\epsilon > 0$, $\mu(U - K) < \epsilon$.

Definition 2.2.3

A finite measure space is called a *Lebesgue space* if it is isomorphic to the union of an interval $[0, a]$ (with the Lebesgue measure) and at most countably many atoms.

Lemma 2.2.4

For a separable complete metric space X , define finite Borel measure on X , \mathfrak{A} be the completion of the Borel σ -algebra w.r.t μ . Then (X, \mathfrak{A}, μ) is a Lebesgue space. In particular, $[0, 1]^2$ in the Lebesgue measure is isomorphic to $[0, 1]$.

Definition 2.2.5

A Lebesgue space with no atoms is called *non-atomic* and it is always isomorphic to $([0, 1], m)$, where m is the Lebesgue measure.

Definition 2.2.6

Let (X, \mathfrak{B}, μ) be a measure space. A property holds *mod 0* in X or μ -almost everywhere or μ -a.e. x if the property holds on a full measure set.

Example 2.2.7: Not proven yet

The hyperbolic automorphism $A : \mathbb{T}^2 \rightarrow \mathbb{T}^2$, with $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ is dense almost every orbit.

Definition 2.2.8

Let (X, \mathfrak{B}, μ) be a measure space. Two measurable functions are *equivalent* if they coincide mod 0.

Example 2.2.9

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases} \text{ on } [0, 1] \text{ is equivalent to } 0.$$

Proof. Since they are equal besides a set of measure 0. □

2.3 Space of Functions

Definition 2.3.1: L^p space

For $p \in (0, \infty)$ the space $L^p(X, \mu)$ is the set of equivalent classes mod 0 of measure functions $f : X \rightarrow \mathbb{C}$ such that

$$\int_X |f|^p d\mu < \infty$$

For $p > 1$, the norm is defined to be $\|f\|_p = (\int_X |f|^p d\mu)^{1/p}$.

Recall that $L^2(X, \mu)$ is a Hilbert space $(f, g)_2 = \int_X f \bar{g} d\mu$. $L^\infty(X, \mu)$ is a space of essentially bounded functions and $\|f\|_\infty = \text{ess sup}_X |f|$. Essentially means we do not consider set of measure 0.

Proposition 2.3.2

If μ finite, then $L^\infty \subset L^p$ for all $p > 0$.

Definition 2.3.3

The continuous functions $f : X \rightarrow \mathbb{C}$ with compact support, denoted by $C_0(X, \mu)$ if there exists $K \subset X$ compact set such that $f(x) = 0$ outside K .

Proposition 2.3.4

$C_0(X, \mu)$ is dense in $L^p(X, \mu)$ for all $p > 0$.

2.4 Recurrence

Theorem 2.4.1: Poincare Recurrence Theorem

Let (X, \mathfrak{B}, μ) be a probability measure space, $T : X \rightarrow X$ be measure preserving. If A measurable set, then for μ -a.e. $x \in A$ there exists $n \in \mathbb{N}$ such that $T^n(x) \in A$. Moreover, a.e. $x \in A$, there exists infinitely many $k \in \mathbb{N}$ such that $T^k(x) \in A$.

Proof. Let $B = \{x \in A : T^n(x) \notin A, \forall n \in \mathbb{N}\}$, which is the points that never return to A . Note that

$$\begin{aligned} B &= \{x \in A : x \notin T^{-n}(A), \forall n \in \mathbb{N}\} = \{x \in A : x \in [T^{-n}(A)]^c, \forall n \in \mathbb{N}\} \\ &= \{x \in A : x \in \bigcap_{n \in \mathbb{N}} [T^{-n}(A)]^c\} = \left\{x \in A : x \in \left[\bigcup_{n \in \mathbb{N}} T^{-n}(A) \right]^c\right\} \\ &= A - \bigcup_{n \in \mathbb{N}} T^{-n}(A). \end{aligned}$$

Claim that all preimages of B are 2 by 2 disjoint. It is enough to prove $B \cap T^{-j}(B) = \emptyset$ for all $j \in \mathbb{N}$, since

$$T^{-k}(B) \cap T^{-j}(B) = T^{-k}(B \cap T^{-j+k}(B)).$$

Now note that

$$\begin{aligned} B \cap T^{-j}(B) &= \left(A - \bigcup_{n \in \mathbb{N}} T^{-n}(A) \right) \cap T^{-j} \left(A - \bigcup_{n \in \mathbb{N}} T^{-n}(A) \right) \\ &= \left(A - \bigcup_{n \in \mathbb{N}} T^{-n}(A) \right) \cap T^{-j}(A) = \emptyset. \end{aligned}$$

However, we then have

$$\mu\left(\bigcup_{j \in \mathbb{N}} T^{-j}(B)\right) = \sum_{j \in \mathbb{N}} \mu(T^{-j}(B)) = \sum_{j \in \mathbb{N}} \mu(B) \leq \mu(X) = 1.$$

So we must have $\mu(B) = 0$. In other words, for $x \in A - B$, or almost every x , there exists $n \in \mathbb{N}$ such that $T^n(x) \in A$.

Now define $C = \{x \in A : \exists n \in \mathbb{N}, T^k(x) \notin A \forall k \geq n\}$, the set of points in A that return to A only finite number of times. We want to prove $\mu(C) = 0$. We will complete proof similarly. Note that in fact as what we done previously

$$C = \bigcup_{n \in \mathbb{N}} \left[A - \bigcup_{k \geq n} T^{-k}(A) \right].$$

Let $C_n = A - \bigcup_{k \geq n} T^{-k}(A)$. Then it suffices to show $\mu(C_n) = 0$. It is clear that $C_n \in \mathfrak{B}$. Rewrite C_n as

$$C_n = A - \bigcup_{k \geq n} T^{-k}(A) \subset \bigcup_{k \geq 0} T^{-k}(A) - \bigcup_{k \geq n} T^{-k}(A).$$

So that

$$\mu(C_n) \leq \mu\left(\bigcup_{k \geq 0} T^{-k}(A)\right) - \mu\left(\bigcup_{k \geq n} T^{-k}(A)\right) = 0$$

since $T^{-n}\left(\bigcup_{k \geq 0} T^{-k}(A)\right) = \left(\bigcup_{k \geq n} T^{-k}(A)\right)$. This ends the proof. \square

We give an example to show that Poincare Recurrence Theorem fails if $\mu(X) = \infty$. Take $X = \mathbb{R}$, μ be the Lebesgue measure. $T(x) = x + 1$ of course measure preserving. However for example points in $A = [0, 1/2]$ will never return, i.e. $T^{-j}(A) \cap A = \emptyset$ for $j \geq 1$.

We have previously defined another type of (forward) recurrence, which is $x \in \omega(x)$. We will see later that for a (finite) Borel measure there is a relationship between two notions.

Definition 2.4.2: Support of measure

Let X be a compact metric space, μ a Borel measure on X . The *support of μ* is the set

$$\text{supp}(\mu) = \{x \in X : \mu(B(x, \epsilon)) > 0, \forall \epsilon > 0\}.$$

Proposition 2.4.3

The support measure $\text{supp}(\mu)$ is closed.

Proof. Consider a convergent sequence $x_n \in \text{supp}(\mu)$ with $x_n \rightarrow x$. For any $\epsilon > 0$, there exists N such that $x_N \in B(x, \epsilon)$. Since $B(x, \epsilon)$ is open, there is a open ball centered at x_N , $B(x_N, \delta) \subset B(x, \epsilon)$. By assumption, $\mu(B(x_N, \delta)) > 0$, and thus $\mu(B(x, \epsilon)) > 0$, i.e. $x \in \text{supp}(\mu)$, which shows $\text{supp}(\mu)$ is closed. \square

Proposition 2.4.4

1. $\text{supp}(\mu)$ is the intersection of all closed full measure set.
2. $\text{supp}(\mu)$ is the complement of the union of all open null measure set.

Proof. It is enough to prove the first one, just noting that

$$\left(\bigcup_i U_i\right)^c = \left(\bigcup_i F_i^c\right)^c = \bigcap_i F_i.$$

To prove it, first, we have to prove for each $x \in \text{supp}(\mu)$, $x \in F$ once F is a closed full measure set. If not, $x \in F^c$ with F^c open and null measure, it contradict with $x \in \text{supp}(\mu)$. Second, we have to prove if $x \in F$ for any F closed full measure set, then $x \in \text{supp}(\mu)$. If not, then for some $x_0 \in F_i$, $\mu(B(x_0, \epsilon_0)) = 0$ for some ϵ . x_0 will not be in the closed full measure set of $F_i - B(x_0, \epsilon_0)$, contradiction. \square

Proposition 2.4.5

X compact metric space, μ Borel probability measure on X . $T : X \rightarrow X$ continuous transformation which is measure preserving. Then a.e. x is recurrent ($x \in \omega(x)$). This implies $\text{supp}(\mu) \subset \overline{R(T)}$.

Proof. Since X is a compact metric space, so there exists an open countable basis $\{U_i\}_{i \in \mathbb{N}}$, i.e. for any $U \subset X$ open set, there exists a subsequence $\{i_n\}$ such that $U = \bigcup_{i_n} U_{i_n}$. By the Poincare Recurrence Theorem there exists a full measure set \tilde{U}_i in U_i such that for all $x \in \tilde{U}_i$, there exists n_i with $T^{n_i}(x) \in U_i$. Let $X_i = (X - U_i) \cup \tilde{U}_i$. Then $\mu(X_i) = 1$, i.e. X_i is a full measure set of X . Consequently, denote $\tilde{X} = \bigcap_{i=1}^{\infty} X_i$, then it also has full measure by sub-additivity. We want to prove that all the point in \tilde{X} is recurrent. For each $x \in \tilde{X}$, let $i_n(x)$ be a subsequence of i depends on x such that $U_{i_{n+1}(x)} \subset U_{i_n(x)}$ and $\bigcap_{n \in \mathbb{N}} U_{i_n(x)} = \{x\}$. So there exists $N_{i_n(x)} \in \mathbb{N}$ such that $T^{N_{i_n(x)}}(x) \in U_{i_n(x)}$. Let $n \rightarrow \infty$, $T^{N_{i_n(x)}}(x) \rightarrow x$, which shows $x \in \tilde{X}$ is recurrent.

Note that

$$1 = \mu(\tilde{X}) = \mu(R(T)) = \mu(\overline{R(T)}),$$

so $\overline{R(T)}$ has full measure and closed. Then by Proposition 2.4.4, we know that $\text{supp}(\mu) \subset \overline{R(T)}$. \square

2.5 Ergodicity and Mixing

Before definition, we shall introduce a informal explanation of how those abstract definitions came [Wik22].

Consider a measure-preserving dynamical system, written $(X, \mathfrak{A}, \mu, T)$. The set X is understood to be the total space to be filled, for example, the smoke-filled room, a long river, etc. The measure μ is understood to define the natural volume of the space X and of its subspace. The collection of subspaces is denoted by \mathfrak{A} (formally, it is the collection of measurable sets which satisfies certain conditions, cf. Definition 2.1.1) so that for any

given set $A \subset X$, the volume of it is $\mu(A)$ (we will explain later when the measure is probability measure).

The time evolution of the system is describe by a map $T : X \rightarrow X$. Given some subset $A \subset X$, its map $T(A)$ will in general be a deformed version of A : it is squashed or stretched, folded or cut into pieces, for example, the horseshoe map. The set $T(A)$ must have the same volume as A . Such map is natural in practical since stretching and folding will not change the volume of a object (noodles, for example in horseshoe). And such map is called *measure preserving* map. However, a formal difficulty arises when one tries to reconcile the volume of sets with the need to preserve their size under T . That is because, in general (when the map is not invertible), several different points in the domain of a function would map into the same point in its range, i.e. there may be $x \neq y$ but $T(x) = T(y)$. But if we consider the inverse map $T^{-1} : \mathfrak{A} \rightarrow \mathfrak{A}$, one can find that this inverse map can capture all the points since there won't be a point in the domain that have two images. In other words, it will not lose track of where things came from. Therefore, the proper definition of a measure-preserving map is $\mu(A) = \mu(T^{-1}(A))$.

Our interest now is the time evolution of the system. If a set $A \in \mathfrak{A}$ eventually comes to fill all of X over a long period of time(that is, if $T^n(A)$ approaches all of X for large n), the system is said to be *ergodic*. If every set A behaves in this way, the system is a *conservative system*. On the contrary, if some subsets wander away, so that $T^n(A)$ won't come back to A thus won't approach all of X , then it is a *dissipative system*. An example would be water running downhill: once it's run down, it will never come back up again. The lake that forms at the bottom of this river can, however, well-mixed. The ergodic decomposition theorem states that every ergodic system can be split into two parts: the conservative part and the dissipative part.

Mixing is a stronger statement than ergodicity. Mixing asks for this ergodic property to hold between any two sets A and B , not just between A and X . That is, given two sets $A, B \in \mathfrak{A}$, a system is said to be topologically mixing if there is an integer N such that, for all $n > N$, one has $T^n(A) \cap B \neq \emptyset$.

The above discussion appeals to a physical sense of a volume in 3-D. Now we go to the abstract space with the probability measure, $\mu(X) = 1$, or equivalently, finite measure (However, the following definitions are available in measure space, rather than finite measure space).

Take the fish in the lake as an example. Suppose X is the whole lake, A is a subspace of the lake, the measure $\mu(A)$ is the proportion of fish in the subset A . And T a measure preserving transformation. Here measure preserving means that no fish would be dead or born. Then the condition of mixing says that as time goes by, the fish in A must be well-mixed in the lake. However, the ergodicity only ask the fish to approach all of the lake when time goes to infinity, the fish might swan in group without mixing (the N may not exist such that for all $n > N$, $T^n(A) \cap B \neq \emptyset$, but we will have the trace of almost every fish will be dense in the lake, see Proposition 2.5.9).

Definition 2.5.1: Essentially T -invariant of both functions and sets

$T : X \rightarrow X$ measure preserving transformation. $f : X \rightarrow \mathbb{R}$ is a measurable function is *essentially T -invariant* if $T_*f = f$ a.e, in other words,

$$\mu(\{x \in X : f(T(x)) \neq f(x)\}) = 0.$$

A measurable set A is *essentially T -invariant* if its characteristic function $\mathbf{1}_A$ is essentially T -invariant. Equivalently, if $\mu(T^{-1}(A) \triangle A) = 0$.

Proof of the equivalence. The proof is trivial. Just note that

$$\begin{aligned}
 & \{x \in X : \mathbf{1}_A(T(x)) \neq \mathbf{1}_A(x)\} \\
 &= \{x \in X : \mathbf{1}_A(T(x)) = 1, \mathbf{1}_A(x) = 0\} \cup \{x \in X : \mathbf{1}_A(T(x)) = 0, \mathbf{1}_A(x) = 1\} \\
 &= \{x \in X : T(x) \in A, x \notin A\} \cup \{x \in X : T(x) \notin A, x \in A\} \\
 &= (T^{-1}(A) - A) \cup (A - T^{-1}(A)) \\
 &= T^{-1}(A) \triangle A.
 \end{aligned}$$

□

Proposition 2.5.2: Another equivalent definition

A measurable function f is essentially T -invariant if and only if $f(T^n(x)) = f(x)$ a.e..
A measurable set A is measure preserving if and only if $\mu(T^{-n}(A) \triangle A) = 0$.

Proof. The second statement is a consequence of the first statement. To prove the first one, note that

$$\begin{aligned}
 0 &= \mu(\{x \in X : f(T(x)) \neq f(x)\}) = \mu(T^{-1}\{x \in X : f(T(x)) \neq f(x)\}) \\
 &= \mu(\{y \in X : f(T^2(y)) \neq f(T(y))\}).
 \end{aligned}$$

This means that $fT^2 = fT$ a.e. If \mathcal{A} is complete then the a.e. property is transitive, so that $fT^2 = fT = f$ a.e. Then by induction, one can show that $fT^n = fT^{n-1} = \dots = f$ a.e. □

Definition 2.5.3: Ergodicity

A measure preserving transformation $T : X \rightarrow X$ is *ergodic* if every essentially T -invariant set is full or null measure, i.e. if a subset $A \subset X$ satisfies $\mu(T^{-1}(A) \triangle A) = 0$, then either $\mu(A) = 0$ or $\mu(A) = \mu(X)$. Equivalently, T is ergodic if any essentially T -invariant measurable function is constant mod 0.

Proof of the equivalence. One side is trivial. If we assume any essentially T -invariant measurable function is constant mod 0, then for any essentially T -invariant set A , the characteristic function $\mathbf{1}_A$ is constant mod 0, since the definition says that $\mathbf{1}_A$ is essentially T -invariant function. That is, A is either of full measure or null measure.

For the other side, assume that any essentially T -invariant set is full or null measure. Then consider an essentially invariant function ϕ , we want to prove it is constant mod 0. Consider the set $\{\phi(x) > c\}$. This set is obviously essentially T -invariant, since ϕ is an essentially invariant function (the difference after the action T is of measure zero). Therefore, $\mu(\{\phi(x) > c\})$ is of either full or null measure. Symmetrically, $\{\phi(x) < c\}$ is also of full or null measure. So that $\mu(\{\phi(x) = c\})$ is of full or null measure for all c . Since for each c , the set is disjoint, there is only one c that has full measure. This shows that ϕ is constant mod 0. □

Proposition 2.5.4

Let T be a measure-preserving transformation on a finite measure space (X, \mathfrak{A}, μ) , and let $p \in (0, \infty]$. Then T is ergodic if and only if every essentially invariant function $f \in L^p(X, \mu)$ is constant mod 0.

Proof. One side is trivial. Assume T is ergodic, then by definition, every essentially invariant function is constant mod 0.

To prove the converse, let f be an essentially invariant measurable function on X , we want to show that f is constant mod 0. Consider for every $M > 0$ the function,

$$f_M(x) = \begin{cases} f(x), & |f(x)| \leq M \\ 0, & |f(x)| > M \end{cases}.$$

This function is clearly bounded by M and essentially invariant. Therefore, $f_M(x) \in L^\infty(X, \mu) \subset L^p(X, \mu)$ (cf. Proposition 2.3.2). By assumption $f_M(x)$ is constant mod 0, for all $M > 0$. It follows that f is constant mod 0. \square

This proposition loose the definition of ergodicity from the set of all essentially invariant measure function to all essentially invariant L^p measurable function, for some $p \in (0, \infty]$.

Proposition 2.5.5

Let (X, \mathfrak{A}, μ) be a measure space, and suppose that $f : X \rightarrow \mathbb{R}$ is essentially invariant for a measurable transformation T on X . Then there is a strictly invariant measurable function \tilde{f} such that $f(x) = \tilde{f}(x)$ mod 0.

Proof. Consider the full measure set

$$A = \{x \in X : f(T^n(x)) = f(x)\}.$$

We define \tilde{f} to be

$$\tilde{f}(x) = \begin{cases} f(x), & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

Then $\tilde{f}(x) = f(x)$ mod 0 clearly. And $\tilde{f}(T(x)) = f(T(x)) = f(x)$ since $T(x) \in A$, so that \tilde{f} is strictly T -invariant. \square

The strict T -invariant function is exactly the original function but defined on the full measure set that makes f be invariant.

Definition 2.5.6: (Strong) mixing

A measure-preserving transformation T on a probability space (X, \mathfrak{A}, μ) is called *(strong) mixing* if

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

for any two measurable sets $A, B \in \mathfrak{A}$.

The definition is actually meaning that the fish that goes from A to B when the time is large is exactly determined by the initial amount of fish in A and B . The assumption of measure-preserving of T means that the total number of fish in any subset of the lake won't change.

Definition 2.5.7: Weak mixing

A measure-preserving transformation T of a probability space (X, \mathfrak{A}, μ) is called *weak mixing* if for all $A, B \in \mathfrak{A}$,

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

The definition means that the fish in the lake may not be mixed such well, but when we average on time, they can be viewed as mixed well.

Proposition 2.5.8

Mixing implies weak mixing, and weak mixing implies ergodicity.

Proof. Mixing implies weak mixing is followed immediately by Stolz's Theorem. And if T is weak mixing, assume A is essentially invariant measurable set, i.e. $\mu(T^{-n}(A) \triangle A) = 0$, then $\mu(T^{-i}(A) \cap A) = \mu(A)$, so that $\mu(A) = \mu(A)^2$ and thus $\mu(A) = 0$ or 1 . \square

Proposition 2.5.9: Topological Consequences

Let X be a compact metric space, $T : X \rightarrow X$ a continuous map, and μ a T -invariant Borel measure on X . Then

1. If T is ergodic, then the orbit of μ -almost every point is dense in $\text{supp}(\mu)$.
2. If T is mixing, then T is topologically mixing on $\text{supp}(\mu)$.

Proof. For the first one, suppose T is ergodic. Let U be a non-empty open set in $\text{supp}(\mu)$. By the definition of support (cf. Definition 2.4.2), $\mu(U) > 0$. Consider the set that whose forward orbits would be in U , namely $B = \bigcup_{i \in \mathbb{N}} T^{-i}(U)$. We claim that B is of full measure. To see this, note that B is an essentially T -invariant set and by ergodicity, B is either of full or null measure, however, B contains U , so we must have B is of full measure. The reason that B is an essentially T -invariant set is that since $T^{-1}(B) \subset B$, then

$$\mu(T^{-1}(B) \triangle B) = \mu(B) - \mu(T^{-1}(B)) = 0$$

for T is measure preserving. Now we have B is of full measure and B is the set whose forward orbits are in U . Since $\text{supp}(\mu)$ is a compact metric space, there exists a countable open basis $\{U_k\}_{k \in \mathbb{N}}$. For each k , we have the corresponding B_k . Now consider $B^* = \bigcap_{k \in \mathbb{N}} B_k$. Since each B_k is of full measure, then B^* is of full measure. And each point in B^* has forward orbits in every U_k , which shows that the forward orbits are dense in $\text{supp}(\mu)$.

For the second one, assume T is mixing, then by definition, given any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for $n > N$,

$$|\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)| < \epsilon,$$

for any $A, B \subset \text{supp}(\mu)$. Since $A, B \subset \text{supp}(\mu)$, we have $\mu(A)\mu(B) > 0$, which implies $\mu(T^{-n}(A) \cap B) > 0$ for $n > N$ for some large N . Therefore $T^{-n}(A) \cap B \neq \emptyset$. For $x \in T^{-n}(A) \cap B$, $T^n x \in A \cap T^n(B)$, i.e. T is topologically mixing on $\text{supp}(\mu)$. \square

2.6 Examples of Ergodicity and Mixing

Proposition 2.6.1: Circle rotation

The circle rotation R_α is ergodic with respect to Lebesgue measure if and only if α is irrational.

Proof. If α is rational then R_α can not be ergodic since the orbit of x , say $\{R_\alpha^n(x)\}_{n=1}^N$, is finite but the union of ϵ -ball centered at each point in the orbit is strictly T -invariant. But this set is neither full nor null measure.

Suppose α is irrational. By Proposition 2.5.4, it is enough to prove that any bounded R_α -invariant function $f : S^1 \rightarrow \mathbb{C}$ is constant mod 0. Since $f \in L^2$, the Fourier series $\sum_{n=-\infty}^{\infty} a_n e^{2n\pi i x}$ of f converges to f in L^2 norm. And also $f \circ R_\alpha = \sum_{n=-\infty}^{\infty} a_n e^{2n\pi i(x+\alpha)}$. If f is R_α -invariant function, by the uniqueness of Fourier coefficients, $a_n = a_n e^{2n\pi i \alpha}$ for all $n \in \mathbb{Z}$. Since $e^{2n\pi i \alpha} \neq 1$ for $n \neq 0$, we have $a_n = 0$ for $n \neq 0$. Thus f is constant mod 0. \square

Proposition 2.6.2: Expanding endomorphism

An expanding endomorphism $E_m : S^1 \rightarrow S^1$ is mixing with respect to Lebesgue measure.

Proof. Since any measurable subsets of S^1 can be approximated by a finite union of intervals of the form $[p/m^i, (p+1)/m^i]$, $p \in \{0, \dots, m^i - 1\}$ (this is because firstly Littlewood's Principle says any measurable subsets can be approximated by a finite union of disjoint intervals and secondly each intervals can be approximated by disjoint intervals of such form), so it is sufficient to consider two intervals

$$A = \left[\frac{p}{m^i}, \frac{p+1}{m^i} \right], \quad B = \left[\frac{q}{m^j}, \frac{q+1}{m^j} \right],$$

where $p \in \{0, \dots, m^i - 1\}$, $q \in \{0, \dots, m^j - 1\}$. Then $E_m^{-1}(B)$ will have m preimages and each has size $1/m^{j+1}$:

$$E_m^{-1}(B) = \bigcup_{k=0}^{m-1} \left[\frac{km^j + q}{m^{j+1}}, \frac{km^j + q + 1}{m^{j+1}} \right].$$

By induction, we have $E_m^{-n}(B)$ is the union of m^n uniformly spaced intervals of length $1/m^{j+n}$. Thus for $n > i$, the intersection $A \cap E_m^{-n}(B)$ consists of m^{n-i} intervals of length $1/m^{n+j}$. Thus

$$\mu(A \cap E_m^{-n}(B)) = \frac{m^{n-i}}{m^{n+j}} = m^{-(i+j)} = \mu(A)\mu(B).$$

□

Proposition 2.6.3: Space of symbols

The shift $\sigma : \Sigma_2 \rightarrow \Sigma_2$ preserves the measure

$$\mu(C_{x_1, x_2, \dots, x_n}^{w_1, w_2, \dots, w_n}) = p(x_1)p(x_2) \dots p(x_n),$$

where $p(\{0\}) + p(\{1\}) = 1$. And also μ is mixing.

Proof. To prove σ preserves μ , note that $\sigma^{-1}C_{x_1, x_2, \dots, x_n}^{w_1, w_2, \dots, w_n} = C_{x_1, x_2, \dots, x_n}^{w_1+1, w_2+1, \dots, w_n+1}$, so it is clear that σ preserves the measure. And consider $A = C_{x_1, x_2, \dots, x_n}^{w_1, w_2, \dots, w_n}$, $B = C_{y_1, y_2, \dots, y_m}^{v_1, v_2, \dots, v_m}$, when N sufficiently large, to be more specific, when $N > \max\{0, w_n - v_1\}$, the positions are disjoint, so $A \cap \sigma^{-N}(B) = C_{x_1, x_2, \dots, x_n, y_1, \dots, y_m}^{w_1, w_2, \dots, w_n, v_1+N, \dots, v_m+N}$. Therefore

$$\mu(A \cap \sigma^{-N}(B)) = p(x_1) \dots p(x_n) \cdot p(y_1) \dots p(y_m) = \mu(A)\mu(B).$$

□

2.7 Important Ergodic Theorems

Theorem 2.7.1: Von Neumann Ergodic Theorem

Let T be a measure-preserving transformation of a finite measure space (X, \mathfrak{A}, μ) . For $f \in L^2(X, \mathfrak{A}, \mu)$, set

$$f_N^+ = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)).$$

Then f_N^+ converges to $L^2(X, \mathfrak{A}, \mu)$ to a T -invariant function \bar{f} .

If T is furthermore assumed to be ergodic, then \bar{f} would be constant mod 0.

Theorem 2.7.2: Birkhoff Ergodic Theorem

Let T be a measure-preserving transformation in a finite measure space (X, \mathfrak{A}, μ) , and let $f \in L^1(X, \mathfrak{A}, \mu)$. Then the limit

$$\bar{f}(x) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x))$$

exists for a.e. $x \in X$, is μ -integrable and T -invariant, and satisfies

$$\int_X \bar{f}(x) d\mu = \int_X f(x) d\mu.$$

Also, if T is furthermore assumed to be ergodic, we have the following corollary.

Corollary 2.7.3

A measure-preserving transformation T of a finite measure space (X, \mathfrak{A}, μ) is ergodic if and only if for each $f \in L^1(X, \mathfrak{A}, \mu)$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \frac{1}{\mu(X)} \int_X f(x) d\mu \quad a.e.x,$$

i.e., if and only if the time average equals the space average for every L^1 function.

Proof. If T is ergodic, then $\bar{f}(x)$ is constant mod 0. As a consequence, for every $f \in L^1(X, \mathfrak{A}, \mu)$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \bar{f}(x) = \frac{1}{\mu(X)} \cdot \int_X f(x) d\mu \quad a.e.x.$$

Conversely, assume f is an invariant function, then

$$f(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \frac{1}{\mu(X)} \int_X f(x) d\mu \quad a.e.x$$

which shows $f(x)$ is constant a.e. x . Then T is ergodic. □

Corollary 2.7.4

A measure-preserving transformation T of a finite measure space (X, \mathfrak{A}, μ) is ergodic if and only if for every $A \in \mathfrak{A}$, for a.e. $x \in X$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_A(T^k(x)) = \frac{\mu(A)}{\mu(X)}.$$

Proof. If T is ergodic, take $f = \mathbf{1}_A$ in the previous corollary, then the proof is complete. For the converse, assume A is T -invariant, multiply both sides by $\mathbf{1}_A$, and note that $\mathbf{1}_A(T^k(x)) \cdot \mathbf{1}_A = \mathbf{1}_{T^{-k}(A) \cap A}$, then integral on both sides, applying dominated convergent theorem, we obtain

$$\mu(A) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}(A) \cap A) = \frac{\mu(A)^2}{\mu(X)}$$

implies $\mu(A) = 0$ or 1 . □

Corollary 2.7.5

Let T be a measure-preserving transformation of a probability measure space

(X, \mathfrak{A}, μ) . Prove that T is ergodic if and only if

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

for any $A, B \in \mathfrak{A}$.

Proof. For one side, assume $A, B \in \mathfrak{A}$, and set $f(x) = \mathbf{1}_A$, then $\bar{f} = \frac{\mu(A)}{\mu(X)} = \mu(A)$ a.e. x . By Birkhoff's Ergodic Theorem, we have

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_A(T^k(x)) \rightarrow \mu(A).$$

Multiply both sides by $\mathbf{1}_B$ and note that $\mathbf{1}_A(T^k(x)) \cdot \mathbf{1}_B = \mathbf{1}_{T^{-k}(A) \cap B}$. The latter is because $\mathbf{1}_A(T^k(x)) \cdot \mathbf{1}_B = 1$ if and only if $T^k(x) \in A$ and $x \in B$ if and only if $x \in T^{-k}(A) \cap B$ if and only if $\mathbf{1}_{T^{-k}(A) \cap B} = 1$. Therefore we obtain

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{T^{-k}(A) \cap B} \rightarrow \mu(A) \mathbf{1}_B.$$

Integral both sides over X , by the dominated convergence theorem, we have

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

Conversely, let $A = B$ and assume A is T -invariant, we have

$$\frac{1}{n} \sum_{k=0}^{n-1} \mu(A) = \mu(A) \rightarrow \mu(A)^2.$$

Therefore we must have $\mu(A) = 0$ or 1 , namely T is ergodic. \square

2.8 Invariant Measure for Continuous Map

Assume in this section that X is a compact metric space, then all continuous function are integrable.

Theorem 2.8.1: Krylov-Bogolubov

Let X be a compact metric space and $T : X \rightarrow X$ a continuous map. Then there is a T -invariant Borel probability measure μ on X .

Proof. We have proved in class that $\int \phi \circ T d\mu = \int \phi d\mu$ for all $\phi \in C^0(X)$. Then set $\phi = \mathbf{1}_A$ a.e. and note that $\mathbf{1}_A \circ T = \mathbf{1}_{T^{-1}(A)}$, we obtain $\mu(T^{-1}(A)) = \mu(A)$. \square

Definition 2.8.2: convergent in weak* topology

Let $\mathcal{M} = \mathcal{M}(X)$ denote the set of all Borel probability measures on X . A sequence of measures $\mu_n \in \mathcal{M}$ converges in the weak* topology to a measure $\mu \in \mathcal{M}$ if $\int_X f d\mu_n \rightarrow \int_X f d\mu$ for every $f \in C(X)$.

Proposition 2.8.3

\mathcal{M} is a compact metric space.

Proof. Assume $\mathcal{F} = \{\phi_0, \phi_1, \dots\}$ be a dense set of $C^0(X)$. The metric in \mathcal{M} is defined as following. For any $\mu, \nu \in \mathcal{M}$,

$$d(\mu, \nu) = \sum_{n=0}^{\infty} \frac{|\int \phi_n d\mu - \int \phi_n d\nu|}{2^n}.$$

This is indeed a measure. It is clear that $d(\mu, \nu) = d(\nu, \mu)$ and $d(\mu, \nu) \geq 0$. If $d(\mu, \nu) = 0$, then we have $\int \phi_n d\mu = \int \phi_n d\nu$ for all ϕ_n . Combined with \mathcal{F} is dense, we have $\mu(A) = \int \mathbf{1}_A d\mu = \int \mathbf{1}_A d\nu = \nu(A)$, for every A , i.e. $\mu = \nu$. Also note that for $\mu, \nu, \pi \in \mathcal{M}$,

$$\left| \int \phi_n d\mu - \int \phi_n d\pi \right| \geq \left| \int \phi_n d\mu - \int \phi_n d\nu \right| + \left| \int \phi_n d\nu - \int \phi_n d\pi \right|$$

we then have $d(\mu, \pi) \geq d(\mu, \nu) + d(\nu, \pi)$. Then apply diagonal method. \square

Proposition 2.8.4

\mathcal{M} is a convex set.

Proof. We need to check that for any $\mu, \nu \in \mathcal{M}$, we have $\pi = t\mu + (1-t)\nu \in \mathcal{M}$ for any $t \in [0, 1]$. For any set $A \in \mathfrak{A}$, $\pi(A) = t\mu(A) + (1-t)\nu(A) \geq 0$. Also if $\{A_n\}_{n \in \mathbb{N}}$ is a series of disjoint sets, then

$$\pi \left(\bigcup_{n \in \mathbb{N}} A_n \right) = t\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) + (1-t) \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} [t\mu + (1-t)\nu](A_n) = \sum_{n \in \mathbb{N}} \pi(A_n).$$

And also $\pi(X) = t\mu(X) + (1-t)\nu(X) = 1$. \square

Let $\mathcal{M}_T \subset \mathcal{M}$ denote the set of all T -invariant Borel probability measures on X .

Proposition 2.8.5

Ergodic T -invariant measure are precisely the extreme points of \mathcal{M}_T

Proof. If μ is not ergodic, then there is a T -invariant measurable subset $A \subset X$ with $0 < \mu(A) < 1$. Let $\mu_A(B) = \mu(B \cap A)/\mu(A)$ and $\mu_{X-A}(B) = \mu(B \cap (X-A))/\mu(X-A)$ for any measurable set B . We first check that both μ_A and μ_{X-A} are T -invariant measure.

By definition, it suffices to prove $\mu(T^{-1}(B) \cap A) = \mu(B \cap A)$ and $\mu(T^{-1}(B) \cap (X - A)) = \mu(B \cap (X - A))$. For the former, we have

$$\mu(T^{-1}(B) \cap A) = \mu(T^{-1}(B) \cap T^{-1}(A)) = \mu(T^{-1}(A \cap B)) = \mu(A \cap B).$$

For the latter, just note that if A is T -invariant, then $X - A$ is also T -invariant and the proof is the same. Indeed, we have the fact that $T^{-1}(A) = A$, which shows if $Tx \in A$, then $x \in A$ and if $x \in A$ then $Tx \in A$. Therefore, if $x \in T^{-1}(X - A)$, then $Tx \notin A$, so $x \notin A$, $x \in X - A$. Conversely, if $x \in X - A$, $x \notin A$, then $Tx \notin A$, $Tx \in X - A$. \square

Proposition 2.8.6

Let X and Y be compact metric spaces and $T : X \rightarrow Y$ a continuous map. Then T induces a natural map $\mathcal{M}(X) \rightarrow \mathcal{M}(Y)$, and that this map is continuous in the weak* topology.

Proof. The natural map induced by T is $\pi_T : \mu \mapsto \mu \circ T^{-1}$. That is $\pi_T(\mu)(A) = \mu(T^{-1}(A))$. And we simply denote the induced measure by ν . If $\int f d\mu_n \rightarrow \int f d\mu$ for any f as $n \rightarrow \infty$, then for any simple function, $\mathbf{1}_A$,

$$\begin{aligned} \int \mathbf{1}_A d\nu_n &= \int \mathbf{1}_A d\mu_n(T^{-1}) = \mu_n(T^{-1}(A)) = \int \mathbf{1}_{T^{-1}(A)} d\mu_n \\ &\rightarrow \int \mathbf{1}_{T^{-1}(A)} d\mu = \mu(T^{-1}(A)) = \int \mathbf{1}_A d\mu(T^{-1}) = \int \mathbf{1}_A d\nu \quad (n \rightarrow \infty). \end{aligned}$$

Since this is true for simple function, it is true for any continuous function, i.e. the map is continuous. \square

References

- [BS02] Michael Brin and Garrett Stuck. *Introduction to dynamical systems*. Cambridge university press, 2002.
- [SS05] Elias M Stein and Rami Shakarchi. *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton lectures in analysis. Princeton Univ. Press, Princeton, NJ, 2005.
- [Wik22] Wikipedia contributors. Ergodicity — Wikipedia, the free encyclopedia. <https://en.wikipedia.org/w/index.php?title=Ergodicity&oldid=1086512997>, 2022. [Online; accessed 20-May-2022].
- [Wil55] Robert F Williams. A note on unstable homeomorphisms. In *Proc. Amer. Math. Soc.*, volume 6, pages 308–309, 1955.