Chapter 4

Ergodic theory

4.1 The Poincaré recurrence theorem

Ergodic theory deals with dynamics in a measure space. Let us begin with a definition.

Definition 4.1.1: Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two measure spaces, and $f\colon X \to Y$ a measurable map. We shall denote by $f_*\mu$ the measure on $(Y, \mathcal{B})$ given by $f_*\mu(B) = \mu\big(f^{-1}(B)\big)$ for all $B \in \mathcal{B}$. We say that f is *measure-preserving* if $f_*\mu = \nu$. If $(X, \mathcal{A}, \mu) = (Y, \mathcal{B}, \nu)$ and f is measure-preserving, we say that μ is *f-invariant*, or that f is an *endomorphism* of $(X, \mathcal{A}, \mu)$.

The first result in ergodic theory is Poincaré's recurrence theorem, which states that in a probability Borel measure space almost every point is recurrent. This is a consequence of the following

Proposition 4.1.1: *Let $f\colon X \to X$ be an endomorphism of a probability space $(X, \mathcal{A}, \mu)$. Given $A \in \mathcal{A}$, let $\tilde{A} \subseteq A$ be the set of points $x \in A$ such that $f^j(x) \in A$ for infinitely many $j \in \mathbb{N}$, that is*

$$\tilde{A} = A \cap \bigcap_{n=0}^{\infty} \bigcup_{j \geq n} f^{-j}(A).$$

Then $\tilde{A} \in \mathcal{A}$ and $\mu(\tilde{A}) = \mu(A)$.

Proof: Let $C_n = \{x \in A \mid f^j(x) \notin A \text{ for all } j \geq n\}$. We clearly have

$$\tilde{A} = A \setminus \bigcup_{n=1}^{\infty} C_n;$$

therefore it suffices to prove that $C_n \in \mathcal{A}$ and $\mu(C_n) = 0$ for all $n \geq 1$. Now, we have

$$C_n = A \setminus \bigcup_{j \geq n} f^{-j}(A), \tag{4.1.1}$$

and thus $C_n \in \mathcal{A}$. Furthermore, (4.1.1) implies

$$C_n \subseteq \bigcup_{j \geq 0} f^{-j}(A) \setminus \bigcup_{j \geq n} f^{-j}(A),$$

and thus

$$\mu(C_n) \leq \mu\left(\bigcup_{j \geq 0} f^{-j}(A)\right) - \mu\left(\bigcup_{j \geq n} f^{-j}(A)\right).$$

But since

$$\bigcup_{j \geq n} f^{-j}(A) = f^{-n}\left(\bigcup_{j \geq 0} f^{-j}(A)\right)$$

and f is measure-preserving, it follows that

$$\mu\left(\bigcup_{j \geq 0} f^{-j}(A)\right) = \mu\left(\bigcup_{j \geq n} f^{-j}(A)\right),$$

and we are done. □

As a consequence we have

Theorem 4.1.2: (Poincaré's recurrence theorem) *Let X be a separable metric space, $f\colon X \to X$ a Borel-measurable map, and μ an f-invariant probability Borel measure on X. Then μ-almost every point of X is f-recurrent.*

Proof: Since X is separable, we can find (exercise) a countable basis of open sets $\{U_n\}_{n\in\mathbb{N}}$ such that

$$\lim_{n\to+\infty} \operatorname{diam}(U_n) = 0$$

and

$$\forall m \geq 0 \qquad \bigcup_{n\geq m} U_n = X.$$

Set $\tilde{U}_n = \{x \in U_n \mid f^j(x) \in U_n \text{ for infinitely many } j > 0\}$. Proposition 4.1.1 says that each $\tilde{U}_n$ is a Borel set and $\mu(U_n \setminus \tilde{U}_n) = 0$. Put

$$\tilde{X} = \bigcap_{m=0}^{\infty} \bigcup_{n\geq m} \tilde{U}_n,$$

so that $x \in \tilde{X}$ if and only if there are infinitely many $n \geq 0$ so that $f^j(x) \in U_n$ for infinitely many $j > 0$. We have

$$\mu(X \setminus \tilde{X}) = \mu\left(\bigcup_{m=0}^{\infty}\left(X \setminus \bigcup_{n\geq m} \tilde{U}_n\right)\right) = \mu\left(\bigcup_{m=0}^{\infty}\left(\bigcup_{n\geq m} U_n \setminus \bigcup_{n\geq m} \tilde{U}_n\right)\right) \leq \mu\left(\bigcup_{m=0}^{\infty}\bigcup_{n\geq m} (U_n \setminus \tilde{U}_n)\right) = 0.$$

So it suffices to show that every point of $\tilde{X}$ is recurrent. Given $\varepsilon > 0$, choose $m \geq 0$ so that $\operatorname{diam}(U_n) \leq \varepsilon$ if $n \geq m$. If $x \in \tilde{X}$, there must exist $n \geq m$ such that $x \in \tilde{U}_n$, and thus $f^j(x) \in U_n$ for infinitely many $j \geq 0$. But this implies $d\big(x, f^j(x)\big) < \varepsilon$ for infinitely many $j \geq 0$, and thus (being ε arbitrary) x is recurrent. □

Example 4.1.1. If $f\colon X \to X$ has a periodic point $p \in X$ of exact period n, then

$$\mu = \frac{1}{n}\sum_{j=0}^{n-1} \delta_{f^j(p)}$$

is an f-invariant Borel probability measure, where δ_x is the Dirac measure concentrated at $x \in X$.

Example 4.1.2. The Lebesgue measure of S^1 is R_α-invariant for any $\alpha \in \mathbb{R}$.

Exercise 4.1.1. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces, and $f\colon X \to Y$ a measurable map. Prove if there exists a generating sub-algebra $\mathcal{B}_0 \subseteq \mathcal{B}$ such that $\mu\big(f^{-1}(B)\big) = \nu(B)$ for all $B \in \mathcal{B}_0$ then f is measure-preserving.

Example 4.1.3. The Lebesgue measure of S^1 is E_m-invariant for any $m \in \mathbb{Z}$, because the inverse image of an interval of length ℓ small enough is the union of $|m|$ disjoint intervals of length $\ell/|m|$, and we can apply the previous exercise. Poincaré's recurrence theorem applied to E_{10} then says that there exists a Borel set $X \subset S^1$ of full measure such that for all $j \geq 1$ the decimal expansion of each $x \in X$ contains the sequence of its first j digits infinitely many times.

Example 4.1.4. *The Gauss transformation.* Let $\phi\colon [0,1] \to [0,1]$ be given by

$$\phi(x) = \begin{cases} \dfrac{1}{x} - \left\lfloor \dfrac{1}{x} \right\rfloor & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

This map has important connections with the theory of continuous fractions. We claim that ϕ preserves the Borel probability measure μ given by

$$\mu(A) = \frac{1}{\log 2}\int_A \frac{1}{1+x}\,dx$$

for every Borel set $A \subseteq [0,1]$. Using the Exercise 4.1.1, it suffices to prove that ϕ preserves the measures of finite unions of intervals, and hence it suffices to check that $\mu\big(\phi^{-1}([a,b])\big) = \mu([a,b])$ for every interval $[a,b] \subseteq [0,1]$. Since μ is absolutely continuous with respect to the Lebesgue measure, it suffices to check that $\mu\big(\phi^{-1}([0,b])\big) = \mu([0,b])$ for every $b \in [0,1]$. But indeed we have

$$\phi^{-1}([0,b]) = \bigcup_{n=1}^{\infty}\left[\frac{1}{b+n}, \frac{1}{n}\right],$$

and so

$$\begin{aligned}
\mu\big(\phi^{-1}([0,b])\big) &= \sum_{n=1}^{\infty}\mu\left(\left[\frac{1}{b+n},\frac{1}{n}\right]\right) = \sum_{n=1}^{\infty}\frac{1}{\log 2}\int_{1/(b+n)}^{1/n}\frac{1}{1+t}\,dt = \sum_{n=1}^{\infty}\frac{1}{\log 2}\log\frac{1+\frac{1}{n}}{1+\frac{1}{b+n}}\\
&= \frac{1}{\log 2}\sum_{m=1}^{\infty}\left(\log\frac{n+1}{b+n+1} - \log\frac{n}{b+n}\right) = \frac{1}{\log 2}\log(b+1)\\
&= \frac{1}{\log 2}\int_0^b \frac{1}{1+t}\,dt = \mu([0,b]).
\end{aligned}$$

Example 4.1.5. Assume we have given a probability measure on the set $\{0,\dots,N-1\}$, that is a sequence of non-negative numbers $p_0,\dots,p_{N-1} \geq 0$ so that $p_0+\dots+p_{N-1} = 1$. The associated *Bernoulli* (or *product*) *measure* on Ω_N and Ω_N^+ is the probability Borel measure μ defined on the cylinders by

$$\mu(C^{m_1\dots m_r}_{a_1\dots a_r}) = p_{a_1}\cdots p_{a_r}.$$

It is clearly (full) left shift-invariant. A slightly more involved construction yields another kind of invariant measure. A *stochastic* matrix is a matrix $S \in M_{N,N}(\mathbb{R}^+)$, with row and column indexes running from 0 to $N-1$, such that $\sum_{i=0}^{N-1} s_{ij} = 1$ for $j = 0,\dots,N-1$. If S is a stochastic matrix, it is possible to prove that there exists a vector $p = (p_0,\dots,p_{N-1}) \in \mathbb{R}^N$ with non-negative coordinates such that $Sp = p$ and $p_0+\dots+p_{N-1} = 1$; furthermore, p is unique if S is transitive (in the sense of Definition 1.7.9). Then the *Markov measure* $\mu_{S,p}$ on Ω_N is defined on cylinders by

$$\mu_{S,p}(C^{m_1\dots m_r}_{a_1\dots a_r}) = s_{a_1a_2}\cdots s_{a_{r-1}a_r}p_{a_r}.$$

Again, it is easy to check that it is shift-invariant.

Exercise 4.1.2. Let $f\colon X \to X$ be an endomorphism of a measure space $(X,\mathcal{A},\mu)$. Given $A \in \mathcal{A}$ with non-zero measure, put $\tilde{A} = \{x \in A \mid f^j(x) \in A$ for infinitely many $j \geq 1\}$, and define $\tilde{f}\colon \tilde{A} \to A$ by setting $\tilde{f}(x) = f^j(x)$, where j is the least positive index such that $f^j(x) \in A$ ($\tilde{f}$ is sometimes called the *first-return map* associated to f). Prove that $\tilde{f}(\tilde{A}) \subseteq \tilde{A}$, that $\mu\big(\tilde{A}\setminus\tilde{f}(\tilde{A})\big) = 0$, and that $\mu|_{\tilde{A}}$ is $\tilde{f}$-invariant.

Exercise 4.1.3. (Hopf) Let X be a locally compact separable metric space and $f\colon X \to X$ a Borel-measurable map. If μ is a locally finite f-invariant Borel measure, prove that μ-almost every $x \in X$ is recurrent or has empty ω-limit set. (*Hint:* since μ is locally finite, Lusin's theorem yields a sequence $\{X_i\}$ of open sets with finite measure whose union has full measure. For every X_i let $\tilde{X}_i$ be defined as in the previous exercise. Applying Theorem 4.1.2 to $\tilde{f}\colon \tilde{X}_i \to \tilde{X}_i$ deduce that almost every point of $\tilde{X}_i$ is recurrent. Conclude that almost every point of X_i is recurrent or has ω-limit set disjoint from X_j.)

Definition 4.1.2: Let μ be a Borel measure on a topological space X. The *support* of μ is the closed set

$$\operatorname{supp}\mu = X \setminus \bigcup\{A \mid A \text{ open in } X,\ \mu(A) = 0\}.$$

Lemma 4.1.3: *Let μ be a Borel measure on a topological space X. Then:*

(i) *if X is second countable then $\mu(X \setminus \operatorname{supp}\mu) = 0$;*
(ii) *any set of full measure is dense in* $\operatorname{supp}\mu$*;*
(iii) *if $f: X \to X$ is continuous and μ is f-invariant then* $\operatorname{supp}\mu$ *is f-invariant;*
(iv) *if X is a separable metric space, $f: X \to X$ is continuous, and μ is an f-invariant probability measure, then* $\operatorname{supp}\mu$ *is contained in the closure of recurrent points, and hence* $\operatorname{supp}\mu \subseteq NW(f)$.

Proof: (i) Indeed, we can write $X \setminus \operatorname{supp}\mu$ as countable union of measure zero open sets.

(ii) Assume that C is not dense in $\operatorname{supp}\mu$; this means that there exist $x \in \operatorname{supp}\mu$ and an open neighbourhood A of x disjoint from $\overline{C}$. Since $A \cap \operatorname{supp}\mu \neq 0$, we must have $\mu(A) > 0$; therefore $\mu(X \setminus C) \geq \mu(A) > 0$, and so C is not of full measure.

(iii) Since f is continous and μ is f-invariant, we clearly have $f^{-1}(X \setminus \operatorname{supp}\mu) \subseteq X \setminus \operatorname{supp}\mu$. Therefore $f^{-1}(\operatorname{supp}\mu) \supseteq \operatorname{supp}\mu$, and hence $\operatorname{supp}\mu$ is f-invariant.

(iv) Theorem 4.1.2 says that the set of recurrent points has full measure; hence it is dense in $\operatorname{supp}\mu$, and we are done. □

4.2 Existence of invariant measures

A natural question now is whether invariant measures exist. We have a positive answer in the case of continuous self-maps of compact metric spaces. To prove it let us recall a couple of definitions and theorems.

Definition 4.2.1: Let $\mathcal{A}$ be a σ-algebra on a set X. We shall denote by $\mathcal{M}_{\mathcal{A}}(X)$ the set of probability measures on $\mathcal{A}$. If X is a topological space and $\mathcal{A}$ is the σ-algebra of Borel subsets, we shall write $\mathcal{M}(X)$ instead of $\mathcal{M}_{\mathcal{A}}(X)$. If $f: X \to X$ is a $\mathcal{A}$-measurable map, we shall denote by $\mathcal{M}^f_{\mathcal{A}}(X)$ the set of f-invariant probability measures on $\mathcal{A}$, and by $\mathcal{M}^f(X)$ the set of f-invariant Borel probability measures.

The Borel measures on compact Hausdorff spaces have a nice relationship with the dual of the space of continuous functions:

Theorem 4.2.1: (Riesz representation theorem) *Let X be a compact Hausdorff space. Then:*

(i) *for every bounded linear functional T on $C^0(X)$ there exists a unique pair of mutually singular finite Borel measures μ and ν such that $T(\varphi) = \int_X \varphi\,d\mu - \int_X \varphi\,d\nu$ for all $\varphi \in C^0(X)$.*
(ii) *a bounded linear functional T on $C^0(X)$ is positive (that is, $T(\varphi) \geq 0$ if $\varphi \geq 0$) if and only if there exists a unique finite Borel measure μ such that $T(\varphi) = \int_X \varphi\,d\mu$ for all $\varphi \in C^0(X)$.*

As a consequence, if X is a compact Hausdorff space we can realize $\mathcal{M}(X)$ as the subset of the unit ball of the dual of $C^0(X)$ consisting of the positive functionals. Now, the dual of a normed vector space is endowed with the weak-$*$ topology, i.e., the topology of pointwise convergence (a sequence $\{T_j\}$ of functionals on the normed space V converges to a functional T if and only if $T_j(v) \to T(v)$ for all $v \in V$), and we have the famous

Theorem 4.2.2: (Banach-Alaoglu) *The unit ball (with respect to the dual norm) of the dual of a normed vector space is weak-$*$ compact.*

This is actually a consequence of Tychonoff's theorem on the product of compact spaces. Indeed, if B is the unit ball in the normed vector space V, we can realize the unit ball of the dual as a closed (with respect to the product topology) subset of $[-1,1]^B$, which is compact by Tychonoff's theorem. Since the topology of pointwise convergence coincides with the product topology, we are done.

Since the set of positive linear functionals on $C^0(X)$ is clearly closed with respect to the weak-$*$ topology, as a corollary we have

Corollary 4.2.3: *Let X be a compact Hausdorff space. Then $\mathcal{M}(X)$ is compact in the weak-$*$ topology.*

Exercise 4.2.1. Let X be a compact metric space, and take a countable subset $\{g_j\}_{j\in\mathbb{N}} \subset C^0(X)$ dense in the unit ball of $C^0(X)$. Given μ, $\nu \in \mathcal{M}(X)$ put

$$d(\mu,\nu) = \sum_{j=0}^{\infty} \frac{1}{2^j} \left| \int_X g_j \, d\mu - \int_X g_j \, d\nu \right| .$$

Prove that d is a metric on $\mathcal{M}(X)$ inducing the weak-$*$ topology.

We are now able to prove the existence of invariant Borel measures:

Theorem 4.2.4: (Krylov-Bogolubov) *Let $f\colon X \to X$ be a continuous self-map of a compact metric space X. Then there exists an f-invariant probability Borel measure.*

Proof: Let $f_*\colon \mathcal{M}(X) \to \mathcal{M}(X)$ be defined by $(f_*\mu)(A) = \mu\bigl(f^{-1}(A)\bigr)$ for any Borel subset $A \subseteq X$. In particular we have

$$\forall \varphi \in C^0(X) \qquad \int_X (\varphi \circ f)\, d\mu = \int_X \varphi \, d(f_*\mu),$$

and thus f_* is continuous in the weak-$*$ topology. Our aim is to find a fixed point of f_*.

Take $\mu_0 \in \mathcal{M}(X)$, and let $\mu_n \in \mathcal{M}(X)$ be defined by

$$\mu_n = \frac{1}{n+1} \sum_{m=0}^{n} (f_*)^m \mu_0 .$$

Since, by Corollary 4.2.3 and Exercise 4.2.1, $\mathcal{M}(X)$ is a compact metric space, we can extract a subsequence $\{\mu_{n_k}\}$ converging to $\mu \in \mathcal{M}(X)$. Now,

$$f_*\mu_{n_k} = \frac{1}{n_k+1} \sum_{m=0}^{n_k} (f_*)^{m+1}\mu_0 = \frac{1}{n_k+1} \sum_{m=0}^{n_k} (f_*)^m \mu_0 - \frac{1}{n_k+1}\mu_0 + \frac{1}{n_k+1}(f_*)^{n_k+1}\mu_0 .$$

The last two terms converge to 0 when k goes to infinity; hence

$$f_*\mu = \lim_{k\to\infty} f_*\mu_{n_k} = \lim_{k\to+\infty} \frac{1}{n_k+1} \sum_{m=0}^{n_k} (f_*)^m \mu_0 = \lim_{k\to\infty} \mu_{n_k} = \mu,$$

and we are done. □

Exercise 4.2.2. Let X be a compact Hausdorff space, and $f\colon X \to X$ a continuous map. Prove that a bounded positive linear functional T on $C^0(X)$ is of the form $T(\varphi) = \int_X \varphi \, d\mu$ for a f-invariant finite Borel measure $\mu \in \mathcal{M}^f(X)$ if and only if $T(\varphi \circ f) = T(\varphi)$ for all $\varphi \in C^0(X)$.

Exercise 4.2.3. Let $f\colon [0,1] \to [0,1]$ be given by

$$f(x) = \begin{cases} \frac{1}{2}x & \text{if } 0 < x \le 1, \\ 1 & \text{if } x = 0. \end{cases}$$

Prove that f is Borel measurable and that there are no f-invariant probability Borel measures on $[0,1]$.

4.3 Ergodic measures

In general, there may exist several probability measures invariant under the action of a given continuous map (remember the Example 4.1.1). We would like to single out measures related to the whole dynamics of the map, and not only to part of it.

Definition 4.3.1: Let $f\colon X \to X$ be a continuous map of a topological space. An f-invariant Borel measure μ is *ergodic* if every completely f-invariant Borel set $A \subseteq X$ is either of zero or of full measure, that is, either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$.

There is a particular case when this happen:

Definition 4.3.2: We shall say that a continuous map $f\colon X \to X$ of a topological space is *uniquely ergodic* if there exists one and only one f-invariant Borel probability measure.

The unique invariant measure of a uniquely ergodic map is ergodic. To prove it we need a definition:

Definition 4.3.3: Let $(X, \mathcal{A}, \mu)$ be a measure space, and $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$. Then the *conditional measure* associated to A is the probability measure μ_A defined by

$$\forall B \in A \qquad \mu_A(B) = \frac{\mu(B \cap A)}{\mu(A)}.$$

We shall sometimes write $\mu(B \mid A)$ instead of $\mu_A(B)$.

Proposition 4.3.1: *Let $f\colon X \to X$ a uniquely ergodic continuous map. Then the unique f-invariant Borel probability measure μ is ergodic.*

Proof: If μ is not ergodic we can find a completely f-invariant Borel set A so that $0 < \mu(A), \mu(X \setminus A) < 1$. Then μ_A and $\mu_{X\setminus A}$ are f-invariant Borel probability measures, and they are distinct because $\mu_A(A) = 1$ whereas $\mu_{X\setminus A}(A) = 0$. □

Lemma 4.3.2: *Let μ be a finite f-invariant measure for a continuous self-map f of a topological space X. Then μ is ergodic if and only if every $\varphi \in L^1(X, \mu)$ which is f-invariant it is constant μ-almost everywhere.*

Proof: If μ is not ergodic, then there is completely f-invariant Borel set A with $\mu(A)$, $\mu(X \setminus A) \neq 0$. Then the characteristic function χ_A of A is an element of $L^1(X, \mu)$ which is f-invariant and not constant μ-almost everywhere.

Conversely, assume μ ergodic, and let $f \in L^1(X, \mu)$ be f-invariant. Then for every $c \in \mathbb{R}$ the set $A_c = \{x \in X \mid f(x) \leq c\}$ is completely f-invariant; therefore either $\mu(A_c) = 0$ or $\mu(A_c) = \mu(X)$, and, since $X = \bigcup_{c\in\mathbb{Q}} A_c$, we must have $\mu(A_c) = \mu(X)$ for at least some $c \in \mathbb{R}$. Analogously, since $\varnothing = \bigcap_{c\in\mathbb{Q}} A_c$, we must have $\mu(A_c) = 0$ for at least some $c \in \mathbb{R}$. Let

$$c_0 = \inf\{c \in \mathbb{R} \mid \mu(A_c) = \mu(X)\} > -\infty.$$

Then

$$\mu\left(\{x \in X \mid f(x) < c_0\}\right) = \mu\left(\bigcup_{\substack{c<c_0\\ c\in\mathbb{Q}}} A_c\right) = 0,$$

and thus

$$\mu(\{x \in X \mid f(x) = c_0\}) = \mu(A_c) - \mu\left(\{x \in X \mid f(x) < c_0\}\right) = \mu(X),$$

that is $f \equiv c_0$ μ-almost everywhere. □

The arguments used in the previous section also yields the existence of ergodic measures. To see this, we need the following

Lemma 4.3.3: *Let $f\colon X \to X$ be a continuous map of a topological space X. If $\mu \in \mathcal{M}^f(X)$ is not ergodic then there exist μ_1, $\mu_2 \in \mathcal{M}^f(X)$ and $0 < t < 1$ so that $\mu_1 \neq \mu_2$ and $\mu = t\mu_1 + (1-t)\mu_2$.*

Proof: Let $A \subset X$ be a completely f-invariant Borel set such that $0 < \mu(A) < 1$, and set $\mu_1 = \mu_A$, $\mu_2 = \mu_{X\setminus A}$ and $t = \mu(A)$. Then $\mu_1 \neq \mu_2$ and $\mu = t\mu_1 + (1-t)\mu_2$. □

Definition 4.3.4: Let V be a vector space, and $K \subset V$ a closed convex subset. An *extreme point* of K is a point $v \in K$ such that if we can write $v = tv_1 + (1-t)v_2$ with $v_1, v_2 \in K$ and $0 < t < 1$ then necessarily $v_1 = v_2 = v$.

So extreme points of $\mathcal{M}^f(X)$ are ergodic measures (the converse is true too; see Exercise 4.3.1). Thus to prove the existence of ergodic measures it suffices to prove the existence of extreme points:

Theorem 4.3.4: *Let $f: X \to X$ be a continuous self-map of a compact metric space X. Then there exists an ergodic f-invariant probability Borel measure.*

Proof: Let $\{\varphi_j\}_{j\in\mathbb{N}}$ be a countable dense set in $C^0(X)$, and define $\mathcal{M}_j$ by setting $\mathcal{M}_0 = \mathcal{M}^f(X)$ and

$$\mathcal{M}_{j+1} = \left\{\mu \in \mathcal{M}_j \;\middle|\; \int_X \varphi_j \, d\mu = \max_{\nu\in\mathcal{M}_j} \int_X \varphi_j \, d\nu\right\}.$$

Since $\nu \mapsto \int_X \varphi_j \, d\nu$ is continuous (and recalling Theorem 4.2.4) every $\mathcal{M}_j$ is a not empty convex compact set; hence their intersection $\mathcal{G}$ is not empty. To conclude the proof it suffices to show that every $\mu \in \mathcal{G}$ is an extreme point of $\mathcal{M}^f(X)$.

Take $\mu \in \mathcal{G}$, and write $\mu = t\mu_1 + (1-t)\mu_2$ with $0 < t < 1$ and $\mu_1, \mu_2 \in \mathcal{M}^f(X)$. Then

$$\forall \varphi \in C^0(X) \qquad \int_X \varphi \, d\mu = t \int_X \varphi \, d\mu_1 + (1-t) \int_X \varphi \, d\mu_2.$$

Since $\mu \in \mathcal{M}_1$, this implies

$$t \int_X \varphi_0 \, d\mu_1 + (1-t) \int_X \varphi_0 \, d\mu_2 = \int_X \varphi_0 \, d\mu \geq \max\left\{\int_X \varphi_0 \, d\mu_1, \int_X \varphi_0 \, d\mu_2\right\},$$

which is possible if and only if $\int_X \varphi_0 \, d\mu = \int_X \varphi_0 \, d\mu_1 = \int_X \varphi_0 \, d\mu_2$; in particular, $\mu_1, \mu_2 \in \mathcal{M}_1$. Repeating this argument, by induction we see that

$$\int_X \varphi_j \, d\mu = \int_X \varphi_j \, d\mu_1 = \int_X \varphi_j \, d\mu_2$$

for every $j \in \mathbb{N}$. But $\{\varphi_j\}$ is dense in $C^0(X)$; hence $\int_X \varphi \, d\mu = \int_X \varphi \, d\mu_1 = \int_X \varphi \, d\mu_2$ for all $\varphi \in C^0(X)$. The uniqueness statement in Theorem 4.2.1.(ii) then implies $\mu_1 = \mu_2 = \mu$, and so μ is an extreme point of $\mathcal{M}^f(X)$, as desired. □

Actually, it is possible to use convex analysis (more precisely, Choquet's theorem on extreme points) to prove a sort of decomposition of any invariant measure as integral or ergodic ones:

Theorem 4.3.5: *Every invariant Borel probability measure μ for a continuous self-map f of a compact metric space X can be decomposed into an integral of ergodic invariant Borel probability measures in the following sense: there is a partition (modulo null sets) of X into completely f-invariant subsets $\{X_\alpha\}_{\alpha\in A}$, where A is a measure space endowed with a probability measure $d\alpha$ (actually, A can be taken a Lebesgue space), so that every X_α carries an ergodic f-invariant measure μ_α and we have*

$$\forall \varphi \in C^0(X) \qquad \int_X \varphi \, d\mu = \int_A \int_{X_\alpha} \varphi \, d\mu_\alpha \, d\alpha.$$

Two different ergodic measures are related to completely different parts of the dynamics:

Proposition 4.3.6: *Two distinct ergodic measures for a continuous self-map f of a topological space X are mutually singular.*

Proof: Let μ and μ_1 be two f-invariant ergodic measures, and assume that μ_1 is not absolutely continuous with respect to μ. Then there is a Borel set A such that $\mu(A) = 0$ but $\mu_1(A) > 0$. Let

$$\tilde{A} = \bigcap_{n\geq 0} \bigcup_{j\geq n} f^{-j}(A)$$

be the set of points whose orbit intersects A infinitely many times. Then $\tilde{A}$ is a completely f-invariant subset such that $\mu(\tilde{A}) = 0$ and $\mu_1(\tilde{A}) > 0$ (compare Proposition 4.1.1). By ergodicity, it follows that $X \setminus \tilde{A}$ has full μ-measure while $\tilde{A}$ has full μ_1-measure, and hence μ and μ_1 are mutually singular.

If, on the other hand, $\mu_1 << \mu$, the Radon-Nikodym derivative $d\mu_1/d\mu$ is f-invariant, and thus μ-almost everywhere constant (by Lemma 4.3.2), which implies $\mu_1 = \mu$. □

Remark 4.3.1. Notice that the last argument in the previous proof shows that if μ_1 is f-invariant, μ is ergodic and $\mu_1 << \mu$ then $\mu_1 = \mu$.

Exercise 4.3.1. Prove that a measure $f \in \mathcal{M}^f(X)$ is ergodic if and only if it is an extreme point of $\mathcal{M}^f(X)$.

There is a stronger notion of ergodicity:

Definition 4.3.5: An f-invariant measure μ is *mixing* if

$$\lim_{n\to\infty} \mu\big(f^{-n}(A) \cap B\big) = \mu(A)\mu(B) \tag{4.3.1}$$

for every pair of measurable sets A and B.

Lemma 4.3.7: *Any mixing measure is ergodic.*

Proof: Indeed, if $A \subseteq X$ is completely f-invariant we have

$$\mu(A)\mu(X \setminus A) = \lim_{n\to\infty} \mu\big(f^{-n}(A) \cap X \setminus A\big) = \mu\big(A \cap (X \setminus A)\big) = 0,$$

and hence $\mu(A) = 0$ or $\mu(X \setminus A) = 0$. □

Remark 4.3.2. In the next section we shall show that an f-invariant probability measure μ is ergodic if and only if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mu\big(f^{-m}(A) \cap B\big) = \mu(A)\mu(B)$$

for every pair of measurable sets A and B.

Ergodicity is the statistical equivalent of topological transitivity, unique ergodicity of minimality, and mixing of topological mixing. We shall give a better explanation of this in the next section, but meanwhile we can prove the following:

Proposition 4.3.8: *Let $f: X \to X$ be a homeomorphism of a locally compact Hausdorff space with a countable basis of open sets and no isolated points. Assume that there exists an ergodic measure μ with* $\operatorname{supp}\mu = X$. *Then f is topologically transitive.*

Proof: Since $\operatorname{supp}\mu = X$, every open set has positive measure. But, since μ is ergodic, there cannot be two not empty disjoint completely f-invariant open subsets of X and, by Corollary 1.4.5, f is topologically transitive. □

Remark 4.3.3. If f is a homeomorphism, it is easy to check that the support of an f-invariant measure is completely f-invariant. Then if an ergodic measure μ gives no mass to isolated points the previous argument shows that f is topologically transitive when restricted to the support of μ. In the next section we shall show that this is true even when f is not a homeomorphism.

Remark 4.3.4. Not every f-invariant measure for a topologically transitive map is ergodic. For instance, if μ_1 and μ_2 are Bernoulli measures on Ω_2, they are ergodic for the shift (they are mixing; see Example 4.3.5); on the other hand, $\mu = (\mu_1 + \mu_2)/2$ is still shift-invariant and of full support, but it is not ergodic, because of Proposition 4.3.6.

Proposition 4.3.9: *Let $f: X \to X$ be a uniquely ergodic continuous self-map of a compact metric space X, and μ the unique element of $\mathcal{M}^f(X)$. Then f is minimal on* $\operatorname{supp}\mu$.

Proof: Take $x \in \operatorname{supp}\mu$, and let $\Lambda \subseteq \operatorname{supp}\mu$ be the closure of the orbit of x. Since Λ is clearly f-invariant, Theorem 4.2.4 yields an $f|_\Lambda$-invariant Borel measure ν. Define then a Borel measure $\tilde\nu$ on X by setting $\tilde\nu(A) = \nu(A \cap \Lambda)$ for every Borel set A. Since Λ is f-invariant we have

$$\tilde\nu\big(f^{-1}(A)\big) = \nu\big(f^{-1}(A) \cap \Lambda\big) = \nu\big((f|_\Lambda)^{-1}(A \cap \Lambda)\big) = \nu(A \cap \Lambda) = \tilde\nu(A),$$

and hence $\tilde\nu$ is f-invariant. By unique ergodicity, $\tilde\nu = \mu$; but since $\operatorname{supp}\tilde\nu \subseteq \Lambda$, this implies $\Lambda = \operatorname{supp}\mu$, and we are done. □

Exercise 4.3.2. Let $f: X \to X$ be a continuous map of a topological space X. Assume that there exists a mixing measure μ. Prove that f is topologically mixing on $\operatorname{supp}\mu$.

We shall now describe some examples of ergodic and uniquely ergodic maps. We begin with some generalities on topological groups.

Definition 4.3.6: A Borel measure on a topological group G invariant under all left (respectively, right) translations will be said *left* (respectively, *right*) *invariant.*

We have the following important (and deep)

Theorem 4.3.10: (Haar) *Let G be a locally compact topological group. Then there exists a unique (up to multiplicative constants) locally finite left invariant Borel measure on G.*

Definition 4.3.7: Let G be a compact topological group. The unique (by the previous theorem) Borel left invariant probability measure on G is called the *Haar measure* of G.

EXAMPLE 4.3.1. The usual Lebesgue measure on the torus $\mathbb{T}^n$ (the one induced by the Lebesgue measure of $\mathbb{R}^n$), suitably normalized, is the Haar measure of $\mathbb{T}^n$.

Lemma 4.3.11: *The Haar measure μ of a compact group G is also right invariant, and invariant under continuous homomorphisms.*

Proof: Given $x \in G$ let $\nu = (R_x)_*\mu$, where R_x is the right translation by x. Then ν is left invariant, because for all $y \in G$ and Borel sets A we have

$$\nu\big(L_y^{-1}(A)\big) = \mu\big(R_x^{-1}L_y^{-1}(A)\big) = \mu\big(L_y^{-1}R_x^{-1}(A)\big) = \mu\big(R_x^{-1}(A)\big) = \nu(A),$$

where L_y is the left translation by y. Since we clearly have $\nu(G) = 1$, the uniqueness of the Haar measure yields $\nu = \mu$, and thus μ is right invariant.

Analogously, if $H: G \to G$ is a continuous homomorphism, setting $\nu = H_*\mu$ the same argument shows that ν is left-invariant with $\nu(G) = 1$, and hence $\nu = \mu$, as desired. □

Exercise 4.3.3. Prove that the support of the Haar measure of a compact group G is the whole group G.

Theorem 4.3.12: *Let G be a compact abelian group, μ its Haar measure, and $L_x: G \to G$ a left translation. Then the following properties are equivalent:*

(i) *L_x is topologically transitive;*
(ii) *L_x is minimal;*
(iii) *μ is L_x-ergodic;*
(iv) *L_x is uniquely ergodic;*
(v) *the orbit of x is dense.*

Proof: (i) ⇒ (ii) Lemma 1.4.2.

(ii) ⇒ (v) Obvious.

(v) ⇒ (iv) Take $\nu \in \mathcal{M}^{L_x}(G)$; it suffices to prove that ν is the Haar measure of G. Given $y \in G$, choose a sequence $\{n_j\} \subset \mathbb{N}^*$ such that $L_x^{n_j-1}(x) = n_j x$ converges to y. We shall prove that

$$\forall \varphi \in C^0(G) \qquad \lim_{j\to\infty} \int_G \varphi \circ L_{n_j x}\, d\nu = \int_G \varphi \circ L_y \, d\nu. \tag{4.3.2}$$

By uniform continuity of φ, for every $\varepsilon > 0$ there is a neighbourhood U of the identity element e such that $|\varphi(y_1) - \varphi(y_2)| \le \varepsilon$ for every $y_2 \in y_1 + U$. Now,

$$|(\varphi \circ L_{n_j x})(z) - (\varphi \circ L_y)(z)| = |\varphi(z + n_j x) - \varphi(z + y)|;$$

if j is large enough we have $n_j x - y \in U$, and thus $\|\varphi \circ L_{n_j x} - \varphi \circ L_y\|_\infty \le \varepsilon$, and (4.3.2) follows. But then

$$\int_G \varphi\, d(L_y)_*\nu = \int_G \varphi \circ L_y\, d\nu = \lim_{j\to\infty} \int_G \varphi \circ L_{n_j x}\, d\nu = \lim_{j\to\infty} \int_G \varphi \, d(L_x^{n_j})_*\nu = \int_G \varphi\, d\nu.$$

But this means that ν is left invariant, and hence it is the Haar measure of G, as claimed.

(iv) $\Rightarrow$ (iii) Proposition 4.3.1.

(iii) $\Rightarrow$ (i) If G is a finite group, then the Haar measure is the normalized counting measure. If x has period strictly less than the cardinality of G, then the orbit of x is a completely L_x-invariant subset with measure strictly greater than zero and strictly less than one, against the ergodicity assumption. So the orbit of x must be the whole of G, and L_x is topologically transitive.

If G has an isolated point, then all points are isolated, and thus G (being compact) must be finite. So we can assume that G has no isolated points, and in this case the assertion follows from Proposition 4.3.11, because it is easy to see that $\operatorname{supp}\mu = G$ (exercise). □

Corollary 4.3.13: (i) *Irrational rotations of S^1 are uniquely ergodic.*
(ii) *The Lebesgue measure of $\mathbb{T}^n$ is ergodic for a translation $T_\gamma: \mathbb{T}^n \to T^n$ if and only if $1, \gamma_1, \ldots, \gamma_n$ are rationally independent if and only if T_γ is uniquely ergodic.*

Proof: It follows from Propositions 1.4.1, 1.4.7 and 4.3.12. □

Example 4.3.2. The Lebesgue measure μ of S^1 is ergodic for the maps $E_m: S^1 \to S^1$ with $|m| > 1$. Indeed, let $\varphi \in L^1(S^1, \mu)$ be f-invariant, and let

$$\varphi(x) = \sum_{k=-\infty}^{+\infty} \varphi_k \exp(2\pi i k x)$$

its expansion in Fourier series. Since $\varphi \circ E_m = \varphi$ almost everywhere, it follows that $\varphi_{mk} = \varphi_k$ for all $k \in \mathbb{Z}$. But $|\varphi_k| \to 0$ as $|k| \to +\infty$; hence necessarily $\varphi_k = 0$ for $k \neq 0$, and thus φ is constant. The assertion then follows from Lemma 4.3.2. □

We end this section with a list of examples of mixing measures, together with an (unproved) proposition useful in proving that a measure is mixing.

Definition 4.3.8: A collection $\mathcal{C}$ of measurable sets in a measure space $(X, \mathcal{A}, \mu)$ is *dense* if for every $A \in \mathcal{A}$ and any $\varepsilon > 0$ there is $A' \in \mathcal{C}$ so that $\mu(A \triangle A') < \varepsilon$, where $A \triangle A' = (A \setminus A') \cup (A' \setminus A)$ is the symmetric difference. More generally, a collection $\mathcal{C} \subset \mathcal{A}$ is said *sufficient* if the family of finite disjoint unions of elements of $\mathcal{C}$ is dense.

Proposition 4.3.14: *Let $f: X \to X$ be an endomorphism of a measure space $(X, \mathcal{A}, \mu)$. Then:*
(i) *if (4.3.1) holds for all A and B belonging to a sufficient collection of measurable sets, then μ is mixing;*
(ii) *μ is mixing if and only if a given complete system Φ of functions in $L^2(X, \mu)$ and any φ, $\psi \in \Phi$ one has*

$$\lim_{n \to +\infty} \int_X (\varphi \circ f^n) \overline{\psi}\, d\mu = \int_X \varphi\, d\mu \cdot \int_X \overline{\psi}\, d\mu. \tag{4.3.3}$$

Example 4.3.3. The Lebesgue measure of the torus is never mixing with respect to translations T_γ, while it is mixing for hyperbolic automorphisms of $\mathbb{T}^2$.

Example 4.3.4. The Lebesgue measure of S^1 is mixing with respect to the maps E_m with $|m| > 1$.

Example 4.3.5. The Bernoulli measure is mixing with respect to the full shift both in Ω_N and in Ω_N^+.

4.4 Birkhoff's theorem

The main result of this section is the following

Theorem 4.4.1: (Birkhoff) *Let $f: X \to X$ be an endomorphism of a probability space $(X, \mathcal{A}, \mu)$. Then for every $\varphi \in L^1(X, \mu)$ the following limit*

$$\varphi_f(x) = \lim_{n \to +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi\bigl(f^j(x)\bigr)$$

exists μ-almost everywhere. Furthemore, $\varphi_f \in L^1(X,\mu)$ is f-invariant and

$$\int_X \varphi_f \, d\mu = \int_X \varphi \, d\mu. \tag{4.4.1}$$

Proof: Let $\mathcal{I} = \{A \in \mathcal{A} \mid f^{-1}(A) = A\}$ be the invariant σ-algebra associated to f. Notice that a function $\psi: X \to \mathbb{R}$ is measurable with respect to $\mathcal{I}$ if and only if it is f-invariant and $\mathcal{A}$-measurable.

Given $\psi \in L^1(X,\mu)$ set

$$\Psi_n = \max_{1\le k\le n} \sum_{j=0}^{k-1} \psi \circ f^j \in L^1(X,\mu);$$

since $\Psi_{n+1} \ge \Psi_n$ for μ-almost every $x \in X$ we have either $\Psi_n(x) \to +\infty$ or the sequence $\{\Psi_n(x)\}$ is bounded. Furthermore,

$$\Psi_n \circ f = \Psi_{n+1} - \psi + \min\{0, \Psi_n \circ f\},$$

and hence the set $A_\psi = \{x \in X \mid \Psi_n(x) \to +\infty\}$ belongs to $\mathcal{I}$. Furthermore,

$$\Psi_{n+1} - \Psi_n \circ f = \psi - \min\{0, \Psi_n \circ f\} \searrow \psi \quad \text{on } A_\psi.$$

So, by the Dominated Convergence Theorem,

$$0 \le \int_{A_\psi} (\Psi_{n+1} - \Psi_n) \, d\mu = \int_{A_\psi} (\Psi_{n+1} - \Psi_n \circ f) \, d\mu \to \int_{A_\psi} \psi \, d\mu. \tag{4.4.2}$$

Now let $\psi_{\mathcal{I}} \in L^1(X,\mu)$ be the (invariant) Radon-Nykodim derivative of $\psi\mu|_{\mathcal{I}}$ with respect to $\mu|_{\mathcal{I}}$, that is the unique f-invariant function in $L^1(X,\mu|_{\mathcal{I}}) \subset L^1(X,\mu)$ such that

$$\int_A \psi \, d\mu = \int_A \psi_{\mathcal{I}} \, d\mu$$

for all $A \in \mathcal{I}$. In particular, (4.4.2) implies that if $\psi_{\mathcal{I}} < 0$ then necessarily $\mu(A_\psi) = 0$.

Another general remark is that we have

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \psi \circ f^k \le \limsup_{n\to\infty} \frac{\Psi_n}{n} = 0 \tag{4.4.3}$$

for μ-almost every $x \notin A_\psi$.

Now take $\varepsilon > 0$ and put $\psi = \varphi - \varphi_{\mathcal{I}} - \varepsilon$. Then $\psi_{\mathcal{I}} = -\varepsilon$, which means that $\mu(A_\psi) = 0$. But then (4.4.3) yields

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k \le \varphi_{\mathcal{I}} + \varepsilon$$

μ-almost everywhere. Replacing φ by $-\varphi$ yields

$$\liminf_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k \ge \varphi_{\mathcal{I}} - \varepsilon$$

μ-almost everywhere. It follows that $\varphi_f = \varphi_{\mathcal{I}}$ μ-almost everywhere, and in particular

$$\int_X \varphi_f \, d\mu = \int_X \varphi_{\mathcal{I}} \, d\mu = \int_X \varphi \, d\mu.$$

□

Definition 4.4.1: The function φ_f is said the *time-average* (or *orbital average*) of φ. In particular, if $\varphi = \chi_A$ is the characteristic function of a subset A, then

$$(\chi_A)_f(x) = \lim_{n\to+\infty} \frac{1}{n}\#\{j \in \{0,\ldots,n-1\} \mid f^j(x) \in A\}$$

is denoted by τ_A and called the *average time* spent by x in A.

In particular, Birkhoff's theorem implies that

$$\int_X \tau_A \, d\mu = \mu(A)$$

for all measurable sets A.

Corollary 4.4.2: *Let $f\colon X \to X$ be an endomorphism of a probability space $(X, \mathcal{A}, \mu)$, and $1 \le p \le \infty$. Then for every $\varphi \in L^p(X,\mu)$ the time-average φ_f belongs to $L^p(X,\mu)$ with $\|\varphi_f\|_p \le \|\varphi\|_p$, and*

$$\lim_{n\to\infty} \left\| \varphi_f - \frac{1}{n}\sum_{j=0}^{n-1} \varphi\circ f^j \right\|_p = 0 \tag{4.4.4}$$

for $1 \le p < \infty$.

Proof: First of all, applying Birkhoff's theorem to $|\varphi|$ we get

$$|\varphi_f(x)| \le \lim_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} \bigl|\varphi\bigl(f^j(x)\bigr)\bigr| = |\varphi|_f(x) \tag{4.4.5}$$

for μ-almost all $x \in X$. In particular, if $\varphi \in L^\infty(X,\mu)$ then $\varphi_f \in L^\infty(X,\mu)$ and $\|\varphi_f\|_\infty \le \|\varphi\|_\infty$.

If $1 \le p < \infty$ we have

$$\int_X \left(\frac{1}{n}\sum_{j=0}^{n-1}|\varphi\circ f^j|\right)^p d\mu = \left\|\frac{1}{n}\sum_{j=0}^{n-1}|\varphi\circ f^j|\right\|_p^p \le \left(\frac{1}{n}\sum_{j=0}^{n-1}\|\varphi\circ f^j\|_p\right)^p = \left(\frac{1}{n}\sum_{j=0}^{n-1}\|\varphi\|_p\right)^p = \|\varphi\|_p^p,$$

where we used the fact that $\|\varphi\circ f^j\|_p = \|\varphi\|_p$ because f is measure-preserving. Then (4.4.5) and Fatou's lemma imply that $\varphi_f \in L^p(X,\mu)$ and $\|\varphi_f\|_p \le \|\varphi\|_p$.

To prove (4.4.4), assume for a moment that $\varphi \in L^\infty(X,\mu) \subset L^1(X,\mu)$. Then

$$\left|\varphi_f(x) - \frac{1}{n}\sum_{j=0}^{n-1}\varphi\circ f^j(x)\right|^p \le \left|\|\varphi_f\|_\infty + \frac{1}{n}\sum_{j=0}^{n-1}\|\varphi\circ f^j\|_\infty\right|^p \le 2^p\|\varphi\|_\infty^p,$$

and (4.4.4) follows from the Dominated Convergence Theorem.

Take now $\varphi \in L^p(X,\mu)$. Given $\varepsilon > 0$, choose $\tilde\varphi \in L^\infty(X,\mu)$ so that $\|\varphi - \tilde\varphi\|_p \le \varepsilon/3$, and choose $N > 0$ such that

$$\forall n \ge N \qquad \left\|\tilde\varphi_f - \frac{1}{n}\sum_{j=0}^{n-1}\tilde\varphi\circ f^j\right\|_p \le \varepsilon/3.$$

Then $\|\varphi_f - \tilde\varphi_f\|_p = \|(\varphi-\tilde\varphi)_f\|_p \le \|\varphi - \tilde\varphi\|_p \le \varepsilon/3$ and

$$\left\|\frac{1}{n}\sum_{j=0}^{n-1}(\varphi-\tilde\varphi)\circ f^j\right\|_p \le \frac{1}{n}\sum_{j=0}^{n-1}\|\varphi-\tilde\varphi\|_p \le \varepsilon/3;$$

summing up we get

$$\forall n \ge N \qquad \left\|\varphi_f - \frac{1}{n}\sum_{j=0}^{n-1}\varphi\circ f^j\right\|_p \le \varepsilon,$$

and we are done. □

Corollary 4.4.3: *Let $f\colon X \to X$ be an endomorphism of a probability space $(X, \mathcal{A}, \mu)$. Then for every pair of measurable sets A, $B \in \mathcal{A}$ the limit*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu\big(f^{-j}(A) \cap B\big)$$

exists.

Proof: We have

$$\mu\big(f^{-j}(A) \cap B\big) = \int_X \chi_{f^{-j}(A)} \chi_B \, d\mu = \int_X (\chi_A \circ f^j) \chi_B \, d\mu.$$

Since $\chi_A \in L^1(X, \mu)$, we can apply the previous corollary to conclude that the sequence

$$\frac{1}{n} \sum_{j=0}^{n-1} \chi_A \circ f^j$$

converges in $L^1(X, \mu)$. Thus

$$\frac{1}{n} \sum_{j=0}^{n-1} \mu\big(f^{-j}(A) \cap B\big) = \int_X \left(\frac{1}{n} \sum_{j=0}^{n-1} \chi_A \circ f^j \right) \chi_B \, d\mu$$

also converges. □

Corollary 4.4.4: *Let $f\colon X \to X$ be an endomorphism of a probability space $(X, \mathcal{A}, \mu)$. Then the following properties are equivalent:*

(i) *μ is ergodic;*
(ii) *for every A, $B \in \mathcal{A}$ we have*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu\big(f^{-j}(A) \cap B\big) = \mu(A) \cdot \mu(B);$$

(iii) *for every $\varphi \in L^1(X, \mu)$ we have*

$$\varphi_f(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi\big(f^j(x)\big) \equiv \int_X \varphi \, d\mu$$

μ-almost everywhere; in other words, the time-average is almost always equal to the space average;
(iv) *$\tau_A \equiv \mu(A)$ μ-almost everywhere for all $A \in \mathcal{A}$.*

Proof: (i) ⇒ (iii) Since φ_f is f-invariant, it must be constant μ-almost everywhere (Lemma 4.3.2), and the assertion follows from (4.4.1).

(iii) ⇒ (iv) Obvious.

(iv) ⇒ (ii) By the Dominated Convergence Theorem

$$\begin{aligned} \mu(A)\mu(B) &= \int_X \tau_A \chi_B \, d\mu = \int_X \left(\lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A \circ f^j \right) \chi_B \, d\mu = \lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \int_X (\chi_A \circ f^j) \chi_B \, d\mu \\ &= \lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu\big(f^{-j}(A) \cap B\big). \end{aligned}$$

(ii) ⇒ (i). If A is completely f-invariant we have

$$\mu(A)\mu(X \setminus A) = \lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu\big(f^{-j}(A) \cap X \setminus A\big) = \lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu\big(A \cap X \setminus A\big) = 0,$$

and μ is ergodic. □

Remark 4.4.1. Corollary 4.4.4.(iv) says that if μ is ergodic then the average time spent in A by μ-almost every point is exactly equal to the measure of A.

Corollary 4.4.5: *Let $f: X \to X$ be a self-map of a separable metric space X, and μ an ergodic Borel probability measure. Then the orbit of μ-almost every $x \in \operatorname{supp}\mu$ is dense in $\operatorname{supp}\mu$. In particular, f is topologically transitive on $\operatorname{supp}\mu$.*

Proof: Let $\{U_m\}_{m\in\mathbb{N}}$ be a countable basis of open sets for the induced topology on $\operatorname{supp}\mu$; clearly, $\mu(U_m) > 0$ for all $m \in \mathbb{N}$. Applying Corollary 4.4.4.(iv) simultaneously to the characteristic functions of U_m we obtain a set $A \subseteq \operatorname{supp}\mu$ of full measure such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_{U_m}\big(f^j(x)\big) = \mu(U_m) > 0$$

for all $x \in A$ and all $m \in \mathbb{N}$. But then the orbit of every $x \in A$ must intersect all U_m's, and hence is dense.□

Example 4.4.1. Let $f: [0,1] \to \mathbb{R}$ be continuous, and $0 < x < 1$. Let $0.a_1a_2a_3\ldots$ be the decimal expansion of x, and set $x_n = 0.a_na_{n+1}\ldots$. Then

$$\lim_{n\to+\infty} \frac{f(x_1) + \cdots + f(x_n)}{n} = \int_0^1 f(x)\,dx$$

for almost every $x \in (0,1)$. Indeed, the limit on the left is $f_{E_{10}}$, and E_{10} is ergodic.

One might expect that if μ is ergodic and $\varphi \in C^0(X)$ then the time-average might be constant everywhere instead of just constant μ-almost everywhere. This is not the case, as shown by the following

Theorem 4.4.6: *Let $f: X \to X$ be a continuous self-map of a compact metric space X. Then the following properties are equivalent:*

(i) *f is uniquely ergodic;*

(ii) *for every $\varphi \in C^0(X)$ the limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi\big(f^j(x)\big)$$

exists for every $x \in X$ and does not depend on x;

(iii) *for every $\varphi \in C^0(X)$ the sequence of functions*

$$\frac{1}{n} \sum_{j=0}^{n-1} \varphi\big(f^j(x)\big) \tag{4.4.6}$$

converges uniformly to a constant.

Proof: (i) ⇒ (iii) If (iii) does not hold, there is $\varphi \in C^0(X)$ so that the sequence (4.4.6) does not converge uniformly to a constant; in particular, it does not converge uniformly to $\int_X \varphi\,d\mu$, where μ is the unique element of $\mathcal{M}^f(X)$. Then there exists $\varepsilon > 0$, a sequence $n_k \to +\infty$, and a sequence $\{x_k\} \subset X$ such that

$$\forall k \in \mathbb{N} \qquad \left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi\big(f^j(x_k)\big) - \int_X \varphi\,d\mu \right| \ge \varepsilon.$$

Theorem 4.2.1 implies that for every $k \in \mathbb{N}$ there exists $\mu_k \in \mathcal{M}(X)$ such that

$$\forall \psi \in C^0(X) \qquad \int_X \psi\,d\mu_k = \frac{1}{n_k} \sum_{j=0}^{n_k-1} \psi\big(f^j(x_k)\big).$$

Since $\mathcal{M}(X)$ is weakly-* compact, we can assume that the sequence $\{\mu_k\}$ converges to $\nu \in \mathcal{M}(X)$. We claim that $\nu \in \mathcal{M}^f(X)$. Indeed,

$$\int_X (\psi \circ f)\, d\nu = \lim_{k\to\infty} \int_X (\psi \circ f)\, d\mu_k = \lim_{k\to\infty} \frac{1}{n_k} \sum_{j=0}^{n_k-1} \psi\bigl(f^{j+1}(x_k)\bigr)$$
$$= \lim_{k\to\infty} \int_X \psi\, d\mu_k + \lim_{k\to\infty} \frac{1}{n_k}\bigl[\psi\bigl(f^{n_k}(x_k)\bigr) - \psi(x_k)\bigr] = \int_X \psi\, d\nu$$

for every $\psi \in C^0(X)$, because $\|\psi \circ f^{n_k} - \psi\|_\infty \le 2\|\psi\|_\infty$, and hence ν is f-invariant. But

$$\left|\int_X \varphi\, d\nu - \int_X \varphi\, d\mu\right| = \lim_{k\to\infty} \left|\int_X \varphi\, d\mu_k - \int_X \varphi\, d\mu\right| = \lim_{k\to+\infty} \left|\frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi\bigl(f^j(x_k)\bigr) - \int_X \varphi\, d\mu\right| \ge \varepsilon;$$

so $\nu \ne \mu$, and f is not uniquely ergodic.

(iii) $\Rightarrow$ (ii) Obvious.

(ii) $\Rightarrow$ (i) Let $T\colon C^0(X) \to \mathbb{R}$ be the functional defined by

$$T(\varphi) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi\bigl(f^j(x)\bigr)$$

for any $x \in X$. Since T is clearly positive, bounded and linear, there is a measure $\nu \in \mathcal{M}(X)$ such that

$$\forall \varphi \in C^0(X) \qquad \int_X \varphi\, d\nu = T(\varphi).$$

Furthermore, ν is f-invariant, because

$$T(\varphi \circ f) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi\bigl(f^{j+1}(x)\bigr) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi\bigl(f^j(x)\bigr) + \lim_{n\to\infty} \frac{1}{n}\bigl[\varphi\bigl(f^n(x)\bigr) - \varphi(x)\bigr] = T(\varphi)$$

for all $\varphi \in C^0(X)$.

Now take any $\mu \in \mathcal{M}^f(X)$. Then the f-invariance of μ and the Dominated Convergence Theorem yield

$$\int_X \varphi\, d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \int_X (\varphi \circ f^j)\, d\mu = \int_X \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} (\varphi \circ f^j)\, d\mu = \int_X T(\varphi)\, d\mu = T(\varphi) = \int_X \varphi\, d\nu$$

for every $\varphi \in C^0(X)$. So $\mu = \nu$, and f is uniquely ergodic. □

4.5 Topological entropy

We now introduce a way to measure the complexity of the orbits of a topological dynamical system. The idea is to count the number of orbit segments of given length n we can distinguish up to a finite precision ε, and then let n go to infinity and ε go to zero.

Let us begin with some definitions.

Definition 4.5.1: Let $f\colon X \to X$ be a continuous self-map of a compact metric space. Given $n \in \mathbb{N}^*$, let $d^f_n\colon X \times X \to \mathbb{R}^+$ be the distance given by

$$\forall x, y \in X \qquad d^f_n(x,y) = \max_{0\le j\le n-1} d\bigl(f^j(x), f^j(y)\bigr).$$

We shall denote by $B_f(x,\varepsilon,n)$ the open ball of center x and radius ε for d^f_n.

Remark 4.5.1. We have $d^f_{n+1} \ge d^f_n$, and hence $B_f(x,\varepsilon,n+1) \subseteq B_f(x,\varepsilon,n)$.

Definition 4.5.2: Let $f: X \to X$ be a continuous self-map of a compact metric space. A set $E \subseteq X$ is (n, ε)-*spanning* if $X = \bigcup_{x \in E} B_f(x, \varepsilon, n)$. We shall denote by $S_d(f, \varepsilon, n)$ the minimal cardinality of an (n, ε)-spanning set.

Roughly speaking, $S_d(f, \varepsilon, n)$ is the minimal number of initial conditions needed to approximate *all* orbit segments of length n up to a precision ε.

Let us now put

$$h_d(f, \varepsilon) = \limsup_{n \to \infty} \frac{1}{n} \log S_d(f, \varepsilon, n) \in [0, +\infty].$$

Since $\varepsilon_1 \leq \varepsilon$ implies $S_d(f, n, \varepsilon_1) \geq S_d(f, n, \varepsilon)$ for all $n \in \mathbb{N}^*$ and hence $h_d(f, \varepsilon_1) \geq h_d(f, \varepsilon)$, the limit

$$\lim_{\varepsilon \to 0^+} h_d(f, \varepsilon) \in [0, +\infty] \tag{4.5.1}$$

exists.

Lemma 4.5.1: *The limit* (4.5.1) *does not depend on the distance chosen to define the topology of X.*

Definition 4.5.3: Let $f: X \to X$ be a continuous self-map of a compact metric space. Then the *topological entropy* $h(f) \in [0, +\infty]$ of f is

$$h(f) = \lim_{\varepsilon \to 0^+} h_d(f, \varepsilon) = \lim_{\varepsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} \log S_d(f, \varepsilon, n),$$

where d is any distance inducing the topology of X. We shall sometimes write $h_{\mathrm{top}}(f)$ instead of $h(f)$.

Lemma 4.5.2: *The topological entropy is invariant under topological conjugacy. More generally, if g is a factor of f, then $h(g) \leq h(f)$.*

Proof: Let $h: X \to Y$ be a semiconjugation between $f: X \to X$ and $g: Y \to Y$, so that $h \circ f = g \circ h$. Fix a distance d_X on X, and a distance d_Y on Y. Since h is uniformly continuous, for every $\varepsilon > 0$ there is a $\delta(\varepsilon) > 0$ so that $d_X(x_1, x_2) < \delta(\varepsilon)$ implies $d_Y\big(h(x_1), h(x_2)\big) < \varepsilon$. Then $h\big(B_f(x, \delta(\varepsilon), n)\big) \subseteq B_g\big(h(x), \varepsilon, n\big)$, and hence

$$S_{d_X}\big(f, \delta(\varepsilon), n\big) \geq S_{d_Y}(g, \varepsilon, n)$$

for every $\varepsilon > 0$ and $n \in \mathbb{N}^*$, because h is surjective. Taking logarithms and limits we get the assertion. □

There are two other ways of defining the topological entropy.

Definition 4.5.4: Let $f: X \to X$ be a continuous self-map of a compact metric space. Let $D_d(f, \varepsilon, n)$ be the minimal cardinality of a cover of X composed by sets of d_n^f-diameter at most ε.

We clearly have

$$S_d(f, \varepsilon, n) \leq D_d(f, \varepsilon, n) \leq S_d(f, \varepsilon/2, n). \tag{4.5.2}$$

Lemma 4.5.3: *For any $\varepsilon > 0$ the limit*

$$\lim_{n \to \infty} \frac{1}{n} \log D_d(f, \varepsilon, n)$$

exists. It follows that

$$\lim_{\varepsilon \to 0^+} \lim_{n \to \infty} \frac{1}{n} \log D_d(f, \varepsilon, n) = h(f).$$

Remark 4.5.2. Since (4.5.2) also implies

$$D_d(f, 2\varepsilon, n) \leq S_d(f, \varepsilon, n) \leq D_d(f, \varepsilon, n),$$

the previous lemma yields

$$\lim_{\varepsilon \to 0^+} \liminf_{n \to \infty} \frac{1}{n} \log S_d(f, \varepsilon, n) = h(f).$$

Definition 4.5.5: Let $f: X \to X$ be a continuous self-map of a compact metric space. We say that a set $E \subset X$ is (n, ε)-*separated* if $d_n^f(x, y) \geq \varepsilon$ for all x, $y \in E$. We denote by $N_d(f, \varepsilon, n)$ the maximal cardinality of an (n, ε)-separated set.

Since it is easy to see that

$$S_d(f,\varepsilon,n) \le N_d(f,\varepsilon,n) \le S_d(f,\varepsilon/2,n),$$

it follows that

$$\lim_{\varepsilon\to 0^+} \limsup_{n\to\infty} \frac{1}{n} \log N_d(f,\varepsilon,n) = \lim_{\varepsilon\to 0^+} \liminf_{n\to\infty} \frac{1}{n} \log N_d(f,\varepsilon,n) = h(f).$$

Definition 4.5.6: If $\mathfrak{U}$ is an open cover of a compact space X, let $N(\mathfrak{U})$ denote the minimal cardinality of a subcover of $\mathfrak{U}$. If $\mathfrak{V}$ is another open cover, let

$$\mathfrak{U} \vee \mathfrak{V} = \{U \cap V \mid U \in \mathfrak{U}, V \in \mathfrak{V}\}.$$

Finally, if $f\colon X \to X$ is continuous, set

$$f^{-1}(\mathfrak{U}) = \{f^{-1}(U) \mid U \in \mathfrak{U}\}.$$

Proposition 4.5.4: *Let $f\colon X \to X$ be a continuous self-map of a compact metric space X. Then:*

(i) *for every open cover $\mathfrak{U}$ of X the limit*

$$\lim_{n\to\infty} \frac{1}{n} \log N\big(\mathfrak{U} \vee f^{-1}(\mathfrak{U}) \vee \cdots \vee f^{-(n-1)}(\mathfrak{U})\big)$$

exists;

(ii) *we have*

$$h(f) = \sup_{\mathfrak{U}} \lim_{n\to\infty} \frac{1}{n} \log N\big(\mathfrak{U} \vee f^{-1}(\mathfrak{U}) \vee \cdots \vee f^{-(n-1)}(\mathfrak{U})\big),$$

where the supremum is taken over all open covers of X.

We now collect a few properties of the topological entropy:

Proposition 4.5.5: *Let $f\colon X \to X$ be a continuous self-map of a compact metric space. Then:*

(i) *if $\Lambda \subseteq X$ is a closed f-invariant set then $h(f|_\Lambda) \le h(f)$;*
(ii) *if $X = \Lambda_1 \cup \cdots \cup \Lambda_m$, where $\Lambda_1, \dots, \Lambda_m$ are closed f-invariant subsets, then $h(f) = \max_{1\le j\le m} h(f|_{\Lambda_j})$;*
(iii) *if $m \in \mathbb{N}$ then $h(f^m) = m h(f)$;*
(iv) *if f is a homeomorphism then $h(f^m) = |m| h(f)$ for all $m \in \mathbb{Z}$;*
(v) *if $g\colon Y \to Y$ is another continuous self-map of a compact metric space, then $h(f \times g) = h(f) + h(g)$.*

Example 4.5.1. If $f\colon X \to X$ is an isometry of a compact metric space, then $d_n^f = d$ for all $n \in \mathbb{N}^*$, and so $h(f) = 0$. In particular, the topological entropy of rotations of S^1 or of translations of the torus is zero.

Example 4.5.2. The topological entropy of $E_m\colon S^1 \to S^1$ is $\log|m|$.

Example 4.5.3. The topological entropy of the shift $\sigma_N\colon \Omega_N \to \Omega_N$ is $\log N$.

Example 4.5.4. If $F_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ is given by $F_L(x,y) = (2x+y, x+y) \pmod 1$ then $h(F_L) = (3+\sqrt{5})/2$.

Definition 4.5.7: Let (X,d) be a compact metric space. For $\varepsilon > 0$ let $b(\varepsilon)$ be the minimal cardinality of a cover of X by ε-balls. The *ball dimension* of X is

$$D(X) = \limsup_{\varepsilon\to 0} \frac{\log b(\varepsilon)}{|\log\varepsilon|} \in [0,+\infty].$$

Theorem 4.5.6: *Let $f\colon X \to X$ be a Lipschitz-continuous self-map of a compact metric space X. Then*

$$h(f) \le D(X) \max\{0, \log \mathrm{Lip}(f)\}.$$

4.6 Measure-theoretic entropy

We shall now describe a more quantitative concept of entropy, obtained using a probability measure.

Definition 4.6.1: Let $(X, \mathcal{A}, \mu)$ be a probability space. A *measurable partition* of X is a family of measurable subsets $\mathcal{P} = \{C_\alpha \mid \alpha \in I\}$, where I is a finite or countable set of indices, such that $\mu\bigl(X \setminus \bigcup_\alpha C_\alpha\bigr) = 0$ and $\mu(C_{\alpha_1} \cap C_{\alpha_2}) = 0$ when $\alpha_1 \neq \alpha_2$. Elements of $\mathcal{P}$ are called *atoms* of $\mathcal{P}$. Given two measurable partitions $\mathcal{P}$ and $\mathcal{Q}$, we say that $\mathcal{P} = \mathcal{Q}$ (mod 0) if for every $C \in \mathcal{P}$ there is $D \in \mathcal{Q}$ such that $\mu(C \triangle D) = 0$, and conversely. If $f\colon X \to X$ is a measurable map, we set $f^{-1}(\mathcal{P}) = \{f^{-1}(C_\alpha) \mid \alpha \in I\}$.

Definition 4.6.2: Let $\mathcal{P}$ be a measurable partition of a probability space (X, μ). For $x \in X$, let $\mathcal{P}(x)$ be the atom of $\mathcal{P}$ containing x (this is well-defined outside of a set of zero measure). The *information function* of the partition $\mathcal{P}$ is the measurable function $I_{\mathcal{P}}\colon X \to \mathbb{R}$ given by

$$I_{\mathcal{P}}(x) = -\log \mu\bigl(\mathcal{P}(x)\bigr).$$

We can think that a partition of X collects the elements of X we cannot tell apart using some measuring instrument. Then if the atom containing x is small then the information obtained measuring x is large; this is the meaning of the information function.

Definition 4.6.3: Let $(X, \mathcal{A}, \mu)$ be a probability space. The *entropy* of a measurable partition $\mathcal{P} = \{C_\alpha \mid \alpha \in I\}$ of X is given by

$$h_\mu(\mathcal{P}) = -\sum_{\substack{\alpha \in I \\ \mu(C_\alpha) > 0}} \mu(C_\alpha) \log \mu(C_\alpha) = \int_X I_{\mathcal{P}}\, d\mu \in [0, +\infty].$$

We shall also need a conditional notion of entropy.

Definition 4.6.4: Let $\mathcal{P} = \{C_\alpha \mid \alpha \in I\}$ and $\mathcal{Q} = \{D_\beta \mid \beta \in J\}$ two measurable partitions of a probability space (X, μ). The *conditional information function* $I_{\mathcal{P},\mathcal{Q}}\colon X \to \mathbb{R}$ of $\mathcal{P}$ with respect to $\mathcal{Q}$ is defined by

$$I_{\mathcal{P},\mathcal{Q}}(x) = -\log \mu\bigl(\mathcal{P}(x) \mid \mathcal{Q}(x)\bigr) = -\log \mu\bigl(\mathcal{P}(x) \cap \mathcal{Q}(x)\bigr) + \log \mu\bigl(\mathcal{Q}(x)\bigr).$$

The *conditional entropy* of $\mathcal{P}$ with respect to $\mathcal{Q}$ is given by

$$h_\mu(\mathcal{P} \mid \mathcal{Q}) = \int_X I_{\mathcal{P},\mathcal{Q}}\, d\mu = -\sum_{\beta \in J} \mu(D_\beta) \sum_{\alpha \in I} \mu(C_\alpha \mid D_\beta) \log \mu(C_\alpha \mid D_\beta),$$

where for simplicity we assume that $0 \log 0 = 0$.

Definition 4.6.5: Let $\mathcal{P} = \{C_\alpha \mid \alpha \in I\}$ and $\mathcal{Q} = \{D_\beta \mid \beta \in J\}$ two measurable partitions of a probability space (X, μ). We say that $\mathcal{P}$ is *subordinate* to $\mathcal{Q}$, or that $\mathcal{Q}$ is a *refinement* of $\mathcal{P}$, and we write $\mathcal{P} \le \mathcal{Q}$ (mod 0), if for every $D \in \mathcal{Q}$ there is $C \in \mathcal{P}$ so that $D \subseteq C$ (mod 0), that is $\mu(D \setminus C) = 0$. The *joint partition* of $\mathcal{P}$ and $\mathcal{Q}$ is

$$\mathcal{P} \vee \mathcal{Q} = \{C \cap D \mid C \in \mathcal{P}, D \in \mathcal{Q}, \mu(C \cap D) > 0\};$$

clearly $\mathcal{P} \vee \mathcal{Q} \le \mathcal{P}$, $\mathcal{Q}$. Finally, we say that $\mathcal{P}$ and $\mathcal{Q}$ are *independent* if

$$\forall C \in \mathcal{P}, D \in \mathcal{Q} \qquad \mu(C \cap D) = \mu(C)\mu(D).$$

The following proposition summarizes the technical properties of the entropy of partitions.

Proposition 4.6.1: *Let $(X, \mathcal{A}, \mu)$ be a probability space, and let $\mathcal{P} = \{C_\alpha \mid \alpha \in I\}$, $\mathcal{Q} = \{D_\beta \mid \beta \in J\}$ and $\mathcal{R} = \{E_\gamma \mid \gamma \in K\}$ be measurable partitions of X. Then:*

(i) *we have*

$$0 \le -\log\bigl(\sup_{\alpha \in I} \mu(C_\alpha)\bigr) \le h_\mu(\mathcal{P}) \le \log \operatorname{card}\mathcal{P};$$

furthermore, if $\mathcal{P}$ is finite then $h_\mu(\mathcal{P}) = \log \operatorname{card}\mathcal{P}$ if and only if all the atoms of $\mathcal{P}$ have equal measure;

(ii) *we have*
$$0 \le h_\mu(\mathcal{P} \mid \mathcal{Q}) \le h_\mu(\mathcal{P});$$
furthermore, we have $h_\mu(\mathcal{P} \mid \mathcal{Q}) = h_\mu(\mathcal{P})$ *if and only if* $\mathcal{P}$ *and* $\mathcal{Q}$ *are independent, and* $h_\mu(\mathcal{P} \mid \mathcal{Q}) = 0$ *if and only if* $\mathcal{P} \le \mathcal{Q}$ (mod 0)*;*

(iii) *if* $\mathcal{Q} \le \mathcal{R}$ *then* $h_\mu(\mathcal{P} \mid \mathcal{Q}) \ge h_\mu(\mathcal{P} \mid \mathcal{R})$*;*

(iv) *we have*
$$h_\mu(\mathcal{P} \vee \mathcal{Q}) = h_\mu(\mathcal{P}) + h_\mu(\mathcal{Q} \mid \mathcal{P}) = h_\mu(\mathcal{Q}) + h_\mu(\mathcal{P} \mid \mathcal{Q})$$
and, more generally,
$$h_\mu(\mathcal{P} \vee \mathcal{Q} \mid \mathcal{R}) = h_\mu(\mathcal{P} \mid \mathcal{R}) + h_\mu(\mathcal{Q} \mid \mathcal{P} \vee \mathcal{R});$$

(v) *we have*
$$h_\mu(\mathcal{P} \vee \mathcal{Q}) \le h_\mu(\mathcal{P}) + h_\mu(\mathcal{Q})$$
and, more generally,
$$h_\mu(\mathcal{P} \vee \mathcal{Q} \mid \mathcal{R}) \le h_\mu(\mathcal{P} \mid \mathcal{R}) + h_\mu(\mathcal{Q} \mid \mathcal{R});$$

(vi) *we have*
$$h_\mu(\mathcal{P} \mid \mathcal{R}) \le h_\mu(\mathcal{P} \mid \mathcal{Q}) + h_\mu(\mathcal{Q} \mid \mathcal{R});$$

(vii) *if* ν *is another probability measure on* X *and* $\mathcal{P}$ *is a partition measurable with respect both to* μ *and to* ν*, then*
$$\forall t \in [0,1] \qquad t h_\mu(\mathcal{P}) + (1-t) h_\nu(\mathcal{P}) \le h_{t\mu+(1-t)\nu}(\mathcal{P}).$$

Corollary 4.6.2: *If* $\mathcal{P}$ *and* $\mathcal{Q}$ *are two measurable partitions with finite entropy of a probability space* (X, μ)*, set*
$$d_R(\mathcal{P}, \mathcal{Q}) = h_\mu(\mathcal{P} \mid \mathcal{Q}) + h_\mu(\mathcal{Q} \mid \mathcal{P}).$$
Then d_R *is a distance on the set of (all equivalence classes mod 0 of) measurable partitions with finite entropy on* X*.*

Definition 4.6.6: The distance d_R is called the *Rokhlin distance.*

To define the metric entropy we need the following

Lemma 4.6.3: *Let* $\{a_n\} \subset \mathbb{R}^+$ *be a sequence of positive real numbers such that* $a_{n+m} \le a_n + a_m$ *for all* n, $m \in \mathbb{N}$*. Then*
$$\lim_{n\to+\infty} \frac{a_n}{n} = \inf_{n\in\mathbb{N}} \frac{a_n}{n} \ge 0.$$

Proof: Let $c = \inf_n a_n/n$. Given $\varepsilon > 0$, let $n_0 \in \mathbb{N}$ be such that $a_{n_0}/n_0 \le c + \varepsilon$. If $n > n_0$, write $n = n_0 p + q$ with $0 \le q < n_0$ and $p \ge 1$. Then
$$\frac{a_n}{n} = \frac{a_{n_0 p+q}}{n_0 p + q} \le \frac{p a_{n_0}}{n_0 p} + \frac{a_q}{n} = \frac{a_{n_0}}{n_0} + \frac{a_q}{n} \le c + \varepsilon + \frac{1}{n} \sup_{0\le j<n_0} a_j.$$
Hence for n large enough we have
$$c \le \frac{a_n}{n} \le c + 2\varepsilon,$$
and we are done. □

Let $f\colon X \to X$ be an endomorphism of the probability space $(X, \mathcal{A}, \mu)$, and $\mathcal{P}$ a measurable partition with finite entropy. Then we clearly have $h_\mu\big(f^{-1}(\mathcal{P})\big) = h_\mu(\mathcal{P})$, and so, setting
$$\mathcal{P}_n^f = \mathcal{P} \vee f^{-1}(\mathcal{P}) \vee \cdots \vee f^{-n-1}(\mathcal{P}),$$
Proposition 4.6.1.(v) yields
$$\forall m, n \in \mathbb{N} \qquad h_\mu\left(\mathcal{P}_{n+m}^f\right) \le h_\mu\left(\mathcal{P}_n^f\right) + h_\mu\left(\mathcal{P}_m^f\right).$$
Therefore the previous lemma allows us to make the following

Definition 4.6.7: Let $f\colon X \to X$ be an endomorphism of the probability space $(X, \mathcal{A}, \mu)$, and $\mathcal{P}$ a measurable partition with finite entropy. The *metric entropy* of f relative to $\mathcal{P}$ is

$$h_\mu(f, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} h_\mu(\mathcal{P}_n^f).$$

Lemma 4.6.4: *Let $f\colon X \to X$ be an endomorphism of the probability space $(X, \mathcal{A}, \mu)$, and $\mathcal{P}$ a measurable partition with finite entropy. Then $n \mapsto h_\mu\big(\mathcal{P} \mid f^{-1}\big(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\big)\big)$ is non-increasing, and*

$$h_\mu(f, \mathcal{P}) = \lim_{n\to\infty} h_\mu\big(\mathcal{P} \mid f^{-1}(\mathcal{P}_n^f)\big).$$

Definition 4.6.8: Let $f\colon X \to X$ be an endomorphism of the probability space $(X, \mathcal{A}, \mu)$. Then the *entropy* of f with respect to μ (or of μ with respect to f) is

$$h_\mu(f) = \sup\{h_\mu(f, \mathcal{P}) \mid \mathcal{P} \text{ is a measurable partition of } X \text{ with finite entropy}\} \in [0, +\infty].$$

Remark 4.6.1. It suffices to take the supremum with respect to the *finite* measurable partitions of X.

In some sense, $h_\mu(f, \mathcal{P})$ is the average amount of information given by knowing the present state (up to approximation $\mathcal{P}$) and an arbitrarily long past. Thus the metric entropy of f measure the maximum amount of average information we can extract from f if we disregard sets of μ-measure zero.

The definition we just gave is due to Kolmogorov. A slightly different approach is due to Shannon-McMillan-Breiman.

Let $f\colon X \to X$ be an endomorphism of the probability space $(X, \mathcal{A}, \mu)$. If $\mathcal{P}$ is a measurable partition of X, then it is easy to check that the atom of $x \in X$ in the partition $\mathcal{P}_n^f$ is given by

$$\mathcal{P}_n^f(x) = \{y \in X \mid f^j(y) \in \mathcal{P}\big(f^j(x)\big) \text{ for } 0 \le j \le n-1\} = \bigcap_{j=0}^{n-1} f^{-j}\Big(\mathcal{P}\big(f^j(x)\big)\Big).$$

In particular, $y \in \mathcal{P}_n^f(x)$ if and only if x and y have the same n-segment of orbit (up to the approximation $\mathcal{P}$). Then we have the following

Theorem 4.6.5: (Shannon-McMillan-Breiman) *Let $f\colon X \to X$ be an endomorphism of the probability space $(X, \mathcal{A}, \mu)$ and $\mathcal{P}$ is a measurable partition of X with finite entropy. Then the sequence of functions $I_{\mathcal{P}_n^f}/n$ converges μ-almost everywhere and in L^1 to a function $h_{\mathcal{P}}(f) \in L^1(X, \mu)$ which is f-invariant.*

Since the convergence is in L^1 we can apply the Dominated Convergence Theorem and hence we get

$$\int_X h_{\mathcal{P}}(f)\, d\mu = \int_X \lim_{n\to\infty} \frac{1}{n} I_{\mathcal{P}_n^f}\, d\mu = \lim_{n\to\infty} \frac{1}{n} \int_X I_{\mathcal{P}_n^f}\, d\mu = \lim_{n\to\infty} \frac{1}{n} h_\mu(\mathcal{P}_n^f) = h_\mu(f, \mathcal{P}),$$

which was the original definition of entropy. In particular, if μ is ergodic we have $h_{\mathcal{P}}(f) \equiv h_\mu(f, \mathcal{P})$ μ-almost everywhere.

The third definition is due to Brin and Katok:

Theorem 4.6.6: (Brin-Katok) *Let X be a compact metric space, $f\colon X \to X$ continuous, and $\mu \in \mathcal{M}^f(X)$ an f-invariant Borel probability measure, and set*

$$h_\mu^+(f, x, \varepsilon) = -\limsup_{n\to\infty} \frac{1}{n} \log \mu\big(B_f(x, \varepsilon, n)\big), \quad h_\mu^-(f, x, \varepsilon) = -\liminf_{n\to\infty} \frac{1}{n} \log \mu\big(B_f(x, \varepsilon, n)\big).$$

Then the limits

$$h_\mu(f, x) = \lim_{\varepsilon\to 0^+} h_\mu^+(f, x, \varepsilon) = \lim_{\varepsilon\to 0^+} h_\mu^-(f, x, \varepsilon)$$

exist for μ-almost every $x \in X$, are equal, are f-invariant and

$$h_\mu(f) = \int_X h_\mu(f, x)\, d\mu.$$

Now let us summarize the properties of the metric entropy.

Proposition 4.6.7: *Let $f: X \to X$ be an endomorphism of the probability space $(X, \mathcal{A}, \mu)$, and $\mathcal{P}$, $\mathcal{Q}$ measurable partitions with finite entropy. Then:*

(i) *we have*

$$0 \le -\limsup_{n\to\infty} \frac{1}{n} \log\left(\sup_{C\in\mathcal{P}_n^f} \mu(C)\right) \le h_\mu(f,\mathcal{P}) \le h_\mu(\mathcal{P});$$

(ii) $h_\mu(f, \mathcal{P}\vee\mathcal{Q}) \le h_\mu(f,\mathcal{P}) + h_\mu(f,\mathcal{Q})$;

(iii) $h_\mu(f,\mathcal{Q}) \le h_\mu(f,\mathcal{P}) + h_\mu(\mathcal{Q} \mid \mathcal{P})$; *in particular, if* $\mathcal{P} \le \mathcal{Q}$ *then* $h_\mu(f,\mathcal{P}) \le h_\mu(f,\mathcal{Q})$;

(iv) $|h_\mu(f,\mathcal{P}) - h_\mu(f,\mathcal{Q})| \le h_\mu(\mathcal{P} \mid \mathcal{Q}) + h_\mu(\mathcal{Q} \mid \mathcal{P})$;

(v) $h_\mu\big(f, f^{-1}(\mathcal{P})\big) = h_\mu(f,\mathcal{P})$ *and, if f is invertible,* $h_\mu(f,\mathcal{P}) = h_\mu\big(f, f(\mathcal{P})\big)$;

(vi) $h_\mu(f,\mathcal{P}) = h_\mu(f,\mathcal{P}_n^f)$ *for all* $n\in\mathbb{N}$;

(vii) *if ν is another f-invariant probability measure then*

$$\forall t\in[0,1] \qquad t h_\mu(f,\mathcal{P}) + (1-t)h_\nu(f,\mathcal{P}) \le h_{t\mu+(1-t)\nu}(f,\mathcal{P}).$$

Proposition 4.6.8: *Let $f: X \to X$ be an endomorphism of the probability space $(X, \mathcal{A}, \mu)$. Then:*

(i) *if the endomorphism $g: Y \to Y$ of the probability space (Y,ν) is a factor of f (that is there exists a measure-preserving $h: X \to Y$ such that $g\circ h = h \circ f$), then $h_\nu(g) \le h_\mu(f)$;*

(ii) *if A is completely f-invariant and $\mu(A) > 0$ then $h_\mu(f) = \mu(A)h_{\mu_A}(f) + \mu(X\setminus A)h_{\mu_{X\setminus A}}(f)$;*

(iii) *if ν is another f-invariant probability measure then*

$$\forall t\in[0,1] \qquad t h_\mu(f) + (1-t)h_\nu(f) \le h_{t\mu+(1-t)\nu}(f);$$

(iv) $h_\mu(f^k) = k h_\mu(f)$ *and, if f is invertible,* $h_\mu(f^{-k}) = |k| h_\mu(f)$;

(v) *if $g: Y\to Y$ is an endomorphism of a probability space (Y,ν) then $h_{\mu\times\nu}(f\times g) = h_\mu(f) + h_\nu(g)$.*

Example 4.6.1. The entropy of the rotations $R_\alpha: S^1 \to S^1$ and of translations $T_\gamma: \mathbb{T}^n \to \mathbb{T}^n$ of the torus with respect to the Lebesgue measure is zero.

Example 4.6.2. The entropy of $E_m: S^1 \to S^1$ with respect to the Lebesgue measure id $\log|m|$.

Example 4.6.3. The entropy of the shift $\sigma_N: \Omega_N \to \Omega_N$ with respect to the Bernoulli measure associated to $(p_0,\dots,p_{N-1})$ is $-p_0\log p_0 - \cdots - p_{N-1}\log p_{N-1}$.

Example 4.6.4. If $F_L: \mathbb{T}^2 \to \mathbb{T}^2$ is given by $F_L(x,y) = (2x+y, x+y) \pmod 1$ then its entropy with respect to the Lebesgue measure is $(3+\sqrt{5})/2$.

4.7 The variational principle

The aim of this section is to prove that the topological entropy is the supremum of the metric entropies. To do so we need two lemmas.

Lemma 4.7.1: *Let X be a compact metric space, and $\mu\in\mathcal{M}(X)$. Then:*

(i) *for every $x\in X$ and $\delta>0$ there is $\delta'\in(0,\delta)$ such that $\mu\big(\partial B(x,\delta')\big) = 0$;*

(ii) *given $\delta>0$ there is a finite measurable partition $\mathcal{P}$ of X such that $\operatorname{diam}(C) < \delta$ and $\mu(\partial C) = 0$ for all $C\in\mathcal{P}$.*

Proof: (i) $B(x,\delta) = \bigcup_{0<\delta'<\delta} \partial B(x,\delta')$ is an uncountable disjoint union with finite measure.

(ii) Let $\{B_1,\dots,B_k\}$ be a cover of X by balls of radius less than $\delta/2$ and with $\mu(\partial B_j) = 0$ for $j=1,\dots,k$. Put $C_1 = \overline{B_1}$ and $C_j = \overline{B_j} \setminus \bigcup_{l=1}^{j-1} \overline{B_l}$ for $j = 2,\dots,k$. Then $\mathcal{P} = \{C_1,\dots,C_k\}$ is as required. □

Lemma 4.7.2: *Let $f: X\to X$ be a continuous self-map of a compact metric space, and $\varepsilon > 0$ given. For every $n\in\mathbb{N}$ choose an (n,ε)-separated set $E_n\subset X$, and put*

$$\nu_n = \frac{1}{\operatorname{card}(E_n)} \sum_{x\in E_n} \delta_x \quad \text{and} \quad \mu_n = \frac{1}{n}\sum_{j=0}^{n-1} f_*^j \nu_n.$$

Then there is an accumulation point μ (in the weak- topology) of $\{\mu_n\}$ in $\mathcal{M}(X)$ which is f-invariant and satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \log \operatorname{card}(E_n) \le h_\mu(f).$$

Proof: Choose a sequence n_k so that

$$\lim_{k\to\infty} \frac{1}{n_k} \log \operatorname{card}(E_{n_k}) = \limsup_{n\to\infty} \frac{1}{n} \log \operatorname{card}(E_n).$$

Since $\mathcal{M}(X)$ is weak-* compact, we can also assume that $\mu_{n_k} \to \mu \in \mathcal{M}(X)$. But we have

$$f_*\mu_n - \mu_n = \frac{1}{n}(f_*^n \nu_n - \nu_n),$$

and hence μ is f-invariant, because the ν_n are probability measures.

Let now $\mathcal{P}$ be a finite measurable partition with atoms of diameter less than ε and satisfying the properties of Lemma 4.7.1.(ii). Since each $C \in \mathcal{P}_n^f$ contains at most one element of E_n, there are $\operatorname{card}(E_n)$ atoms of $\mathcal{P}_n^f$ with ν_n measure $1/\operatorname{card}(E_n)$, while the other atoms have vanishing ν_n-measure; in particular, $h_{\nu_n}(\mathcal{P}_n^f) = \log \operatorname{card}(E_n)$.

Now fix $0 < q < n$ and $0 \le k \le q-1$. If $a_k = \lfloor (n-k)/q \rfloor$, we have

$$\{0, 1, \ldots, n-1\} = \{k + rq + j \mid 0 \le r < a_k, 0 < j \le q\} \cup S,$$

where

$$S = \{0, 1, \ldots, k, k + a_k q + 1, \ldots, n-1\};$$

notice that $\operatorname{card}(S) < k + q < 2q$ by the definition of a_k. Now,

$$\mathcal{P}_n^f = \left(\bigvee_{r=0}^{a_k - 1} f^{-(k+rq)}(\mathcal{P}_q^f) \right) \vee \left(\bigvee_{j \in S} f^{-j}(\mathcal{P}) \right);$$

hence

$$\begin{aligned} \log \operatorname{card}(E_n) = h_{\nu_n}(\mathcal{P}_n^f) &\le \sum_{r=0}^{a_k-1} h_{\nu_n}\big(f^{-(k+rq)}(\mathcal{P}_q^f)\big) + \sum_{j\in S} h_{\nu_n}\big(f^{-j}(\mathcal{P})\big) \\ &\le \sum_{r=0}^{a_k-1} h_{f_*^{-(k+rq)}\nu_n}(\mathcal{P}_q^f) + 2q \log \operatorname{card}(\mathcal{P}), \end{aligned}$$

where we used Proposition 4.6.1.(i) and (v). Now recalling Proposition 4.6.1.(vii) we get

$$\begin{aligned} q \log \operatorname{card}(E_n) = \sum_{k=0}^{q-1} h_{\nu_n}(\mathcal{P}_n^f) &\le \sum_{k=0}^{q-1} \left(\sum_{r=0}^{a_k-1} h_{f_*^{-(k+rq)}\nu_n}(\mathcal{P}_q^f) + 2q \log \operatorname{card}(\mathcal{P}) \right) \\ &\le n h_{\mu_n}(\mathcal{P}_q^f) + 2q^2 \log \operatorname{card}(\mathcal{P}). \end{aligned}$$

Hence

$$\limsup_{n\to\infty} \frac{1}{n} \log \operatorname{card}(E_n) = \lim_{k\to\infty} \frac{1}{n_k} \log \operatorname{card}(E_{n_k}) \le \lim_{k\to\infty} \frac{h_{\mu_{n_k}}(\mathcal{P}_q^f)}{q} = \frac{h_\mu(\mathcal{P}_q^f)}{q}.$$

Since this is now true for all q we can pass to the limit in q and we get

$$\limsup_{n\to\infty} \frac{1}{n} \log \operatorname{card}(E_n) \le h_\mu(f, \mathcal{P}) \le h_\mu(f),$$

and we are done. □

Then we are able to prove the *variational principle:*

Theorem 4.7.3: *Let $f: X \to X$ be a continuous self-map of a compact metric space. Then*

$$h_{\text{top}}(f) = \sup\{h_\mu(f) \mid \mu \in \mathcal{M}^f(X)\}.$$

Proof: If $\mathcal{P} = \{C_1, \ldots, C_k\}$ is a finite measurable partition of X then, since μ is a Borel measure, we have

$$\forall C \in \mathcal{P} \qquad \mu(C) = \sup\{\mu(B) \mid B \subset C, C \text{ closed}\}.$$

So for $j = 1, \ldots, k$ we can choose a compact set $B_j \subseteq C_j$ so that if we take $\mathcal{Q} = \{B_0, B_1, \ldots, B_k\}$ where $B_0 = X \setminus (B_1 \cup \cdots \cup B_k)$ we have $h_\mu(\mathcal{P} \mid \mathcal{Q}) < 1$. Proposition 4.6.7.(iii) yields

$$h_\mu(f, \mathcal{P}) \le h_\mu(f, \mathcal{Q}) + h_\mu(\mathcal{P} \mid \mathcal{Q}) \le h_\mu(f, \mathcal{Q}) + 1.$$

Now $\mathfrak{U} = \{B_0 \cup B_1, \ldots, B_0 \cup B_k\}$ is an open cover of X. Proposition 4.6.1.(i) yields

$$h_\mu(\mathcal{Q}_n^f) \le \log \operatorname{card} \mathcal{Q}_n^f \le \log(2^n \operatorname{card} \mathfrak{U}_n^f).$$

If $\delta_0 > 0$ is the Lebesgue number of $\mathfrak{U}$, then it is also the Lebesgue number of $\mathfrak{U}_n^f$ with respect to the distance d_n^f. Now, $\mathfrak{U}$ is a minimal cover; hence also each $\mathfrak{U}_n^f$ is. This means that every $B \in \mathfrak{U}_n^f$ contains a point x_B that does not belong to any other element of $\mathfrak{U}_n^f$. In particular, the x_B form an (n, δ_0)-separated set. Consequently,

$$h_\mu(f, \mathcal{Q}) \le h_{\text{top}}(f) + \log 2$$

and

$$h_\mu(f, \mathcal{P}) \le h_{\text{top}}(f) + \log 2 + 1$$

for every finite measurable partition $\mathcal{P}$. Therefore using Propositions 4.6.8.(iv) and 4.5.5.(iii) we get

$$h_\mu(f) = \frac{h_\mu(f^n)}{n} \le \frac{h_{\text{top}}(f^n) + \log 2 + 1}{n} = h_{\text{top}}(f) + \frac{\log 2 + 1}{n}$$

for every $n \in \mathbb{N}^*$, and hence

$$h_\mu(f) \le h_{\text{top}}(f).$$

On the other hand, applying Lemma 4.7.2 to maximal (n, ε)-separated sets in X yields

$$\limsup_{n \to \infty} \frac{1}{n} N_d(f, \varepsilon, n) \le h_\mu(f),$$

where $\mu \in \mathcal{M}^f(X)$ is the accumulation point provided by the lemma. But then

$$\limsup_{n \to \infty} \frac{1}{n} N_d(f, \varepsilon, n) \le \sup_{\mu \in \mathcal{M}^f(X)} h_\mu(f),$$

and letting $\varepsilon \to 0$ we get the assertion. □

In general, the supremum in the variational principle is not achieved.

Definition 4.7.1: Let $f: X \to X$ be a continuous self-map of a compact metric space. A measure $\mu \in \mathcal{M}^f(X)$ satisfying $h_\mu(f) = h_{\text{top}}(f)$ is said *of maximal entropy.* If f has one and only one measure of maximal entropy, we say that f is *intrinsically ergodic.*

Exercise 4.7.1. Let $f: X \to X$ be an intrinsecally ergodic continuous self-map of a compact metric space. Prove that the unique measure of maximal entropy is ergodic.