Metric Entropy and Topological Entropy: The Variational Principle

João Pedro Casimiro Rijo

Dissertação para obtenção do Grau de Mestre em

Matemática e Aplicações

Orientadora: Professora Claudia Valls Anglés

Júri

Presidente: Professor Luís Manuel Gonçalves Barreira
Orientadora: Professora Claudia Valls Angles
Vogal: Professor João Luís Gonçalves Dias Ferreira Alves

Dezembro de 2014

"Jónas, o dia em que essa cabeça acordar tu vais ver."

Dedicado à memória do meu pai, Carlos Rijo.

Agradecimentos

Devo os meus primeiros agradecimentos à minha orientadora, Professora Claudia Valls e ao Professor Luís Barreira pela sugestão do tema e pela imensa paciência que demonstraram durante todo o tempo que demorou a compor esta dissertação. Agradeço também a todos os professores com quem tive a oportunidade de aprender durante a minha formação académica no IST, em particular ao Professor José Natário.

Quero agradecer a toda a minha família e amigos por todo o apoio e carinho, especialmente durante os últimos dois anos e meio. Poder contar com a vossa amizade tem sido uma grande fonte de alegria e orgulho na minha vida. Em particular, ao meu amigo de longa data, João Lázaro.

À Sandra Oliveira quero agradecer a companhia, encorajamento e cumplicidade ao longo de todo este percurso. Sem ti este texto provavelmente não veria a luz do dia.

Quero também deixar um especial agradecimento aos meus tios, José Virgílio Prata e Maria José Leitão que, juntamente com os meus pais, foram as pessoas que mais me motivaram e influenciaram academicamente.

À minha avó Fernanda, um agradecimento por vinte e seis anos de amor e dedicação. Irás para mim, ser sempre um exemplo de força e resiliência.

Finalmente devo o maior dos agradecimentos aos meus pais. À minha mãe, que nunca deixou de acreditar, mesmo quando nada o dava a entender, que eu conseguiria alcançar uma formação superior. E ao meu pai, que sempre insistiu para que eu estudasse e me esforçasse por ir mais longe do que ele pôde ir. Foi com eles sempre no pensamento que escrevi esta dissertação e a eles devo tudo o que sou.

Resumo

No contexto da teoria ergódica, damos uma prova auto-contida do princípio variacional para a entropia, que relaciona as noções de entropia métrica e de entropia topológica, de uma forma precisa.

Palavras-chave

Entropia métrica, entropia topológica, princípio variacional.

Abstract

In the context of ergodic theory, we give a self-contained proof of the variational principle for the entropy, which relates the notions of metric entropy and of topological entropy, in a precise manner.

Keywords

Metric entropy, topological entropy, variational principle.

Contents

Introduction **1**

1 Metric Entropy **5**

2 Conditional Entropy **13**

3 Topological Entropy **21**

4 Variational Principle **27**

A Space of Measures **37**

Bibliography **43**

Introduction

This text has the purpose of giving a self-contained proof of the variational principle for the entropy, which relates two important quantities in dynamical systems: metric entropy and topological entropy. It is written in such a way that is suitable for any student who had a first course in dynamical systems.

It aims to be self-contained, in the sense that all definitions and results that are necessary for the proof are included in the text. However, we assume that the reader is familiar with the basic concepts underlying the definitions of metric and topological entropies, including measure theory, metric spaces and basic topology.

The text also attempts to be "sufficient", in the sense that any subject that is not needed for the proof is avoided. It is worth mentioning that in particular we do not attempt to give an exhaustive introduction to metric entropy or topological entropy. In fact, and although we spare no effort in exposing any detail that is relevant for the proof, some of the results (and applications) that make these concepts important are not mentioned in the text.

The term *entropy* has its origin in 1864 with the work of the German physicist Clausius, and since then it became one of the most important concepts of physics. It roughly relates to the form of energy that can no longer produce work and it is the central concept of what became known as the *second Law of Thermodynamics*, which states that the entropy of an isolated system cannot decrease.

The notion of entropy first crossed the scope of thermodynamics and physics in 1948 with the publication of [11] by the American electronic engineer and mathematician Shannon. In his paper, Shannon defined the entropy of a probability vector $\mathbf{p}$, that is, a finite sequence of nonnegative numbers $\mathbf{p} = (p_1, \ldots, p_n)$ whose sum equals 1, to be

$$H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log p_i,$$

with the convention that $0 \log 0 = 0$.

Probability vectors are closely related to measurable partitions. Given a measure space (X, μ) with $\mu(X) = 1$, let $\xi = \{A_1, \ldots, A_n\}$ be a measurable partition, that is, a finite collection of pairwise disjoint measurable sets whose union is X. Then the sequence $\mathbf{p}_\xi = (\mu(A_1), \ldots, \mu(A_n))$ forms a probability vector, and we can define the entropy of the partition ξ with respect to the measure μ as the Shannon's entropy of $\mathbf{p}_\xi$, that is,

$$H_\mu(\xi) = H(\mathbf{p}_\xi) = -\sum_{i=1}^{n} \mu(A_i) \log \mu(A_i).$$

In 1958, Kolmogorov [9] brought the idea of Shannon's entropy to the theory of dynamical systems, by introducing one of the two main concepts considered in this text: the notion of *metric entropy* of a measurable transformation $T : X \to X$ with respect to a T-invariant probability measure μ (this means that $\mu(T^{-1}A) = \mu(A)$ for every measurable set $A \subset X$). His first definition was latter developed by Sinai [12] in 1959. The main idea behind the definition is, given a measurable partition ξ, to use the transformation T to construct a sequence of new partitions $(\xi_n)_{n\in\mathbb{N}}$, where ξ_n is formed by the nonempty sets of the form

$$A_1 \cap T^{-1}A_2 \cap \cdots \cap T^{-(n-1)}A_n$$

with $A_1, \ldots, A_n \in \xi$, and then by measuring the rate with which the metric entropies of these partitions ξ_n grow. More precisely, the *metric entropy* of

a measurable transformation $T\colon X \to X$ on a probability space (X, μ) with respect to a measurable partition ξ is defined by

$$h_\mu(T, \xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n)$$

(we will show in the text that the limit does exist). Taking the supremum of these entropies over all measurable partitions we obtain the *metric entropy* of T with respect to μ, that is,

$$h_\mu(T) = \sup \big\{ h_\mu(T, \xi) : \xi \,\text{is a measurable partition of}\, X \big\}.$$

The second main concept of this text, *topological entropy*, was introduced in 1965 by Adler, Konheim and McAndrew [1], a few years after Kolmogorov's definition of metric entropy. Their goal was to develop a similar concept for topological dynamical systems. Their original definition mimicked Kolmogorov's definition, using open covers instead of measurable partitions. The modern version in metric spaces is due to Dinaburg [5] in 1970 and (independently) to Bowen [4] in 1971. The notion of topological entropy is similar to the notion of metric entropy: instead of a probability measure space (X, μ), we have a metric space (X, d) and, instead of a measurable transformation preserving the measure μ, we have a continuous transformation $T\colon X \to X$. The topological entropy $h(T)$ of T measures the exponential growth rate of the maximum number of orbits that can be distinguished as time increases (a precise definition is given in the text).

The *Variational Principle* for the entropy gives a precise relation between the metric entropy and the topological entropy. In order to present it briefly, we note that to a metric space (X, d) one can associate the *Borel* σ-algebra of X, that is, the σ-algebra generated by all open balls $B_d(x, \varepsilon)$ with $x \in X$ and $\varepsilon > 0$, and we can consider the space $\mathcal{M}(X)$ of all probability measures on this σ-algebra. Since the transformation $T\colon X \to X$ is continuous, it is also measurable. We can now formulate the *Variational Principle*, which

states that the topological entropy is equal to the supremum of the metric entropies over all T-invariant measures in $\mathcal{M}(X)$, that is,

$$h(T) = \sup\left\{h_\mu(T) : \mu \,\text{is a}\, T\text{-invariant measure in } \mathcal{M}(X)\right\}.$$

Hints of this theorem began to arise right after the introduction of the topological entropy. Indeed, it was noticed that in various dynamical systems, topological entropy coincided with metric entropy with respect to some invariant measure. The first step towards the *Variational Principle* was made by Goodwyn [7] in 1969 when he showed that

$$h_\mu(T) \leq h(T)$$

for any μ. In 1970, Dinaburg [5] gave a proof of the complete theorem for a space with finite topological dimension and in 1971, Goodman [6] presented the proof for an arbitrary space. When we take a closer look at the result, it is natural to think that a great deal of work must be understanding the relation between the topology given by a distance and the Borel σ-algebra generated by the same distance. In particular, it will be important the fact that the space $\mathcal{M}(X)$ is metrizable and compact, as shown in the appendix.

The proof of the variational principle presented in this text is based on the simpler proof given by Misiurewicz in [10]. The main source for our work was [2]. We recommend [3] for a more comprehensive introduction to the entropy theory and [8] for a detailed historical account.

Chapter 1

Metric Entropy

In this first chapter, we introduce the notion of metric entropy. This was the first mathematical notion of entropy to be considered in dynamics. It was introduced by Andrei Kolmogorov in 1958 and developed by Yakov Sinai in 1959 who brought it to the modern version that we now present.

We first introduce the notion of entropy with respect to a measurable partition. Let (X, A, μ) be a measure space with $\mu(X) = 1$.

Definition 1.1. We say that a finite family $\xi \subset A$ is a *measurable partition* of (X, A, μ) or simply of X (with respect to μ) if:

1. $\mu\big(\bigcup_{C\in\xi} C\big) = 1$.
2. $\mu(C \cap D) = 0$ for any $C, D \in \xi$ with $C \neq D$.

Definition 1.2. The *entropy* of a measurable partition ξ of X (with respect to μ) is given by

$$H_\mu(\xi) = -\sum_{C\in\xi} \mu(C) \log \mu(C),$$

with the convention that $0 \log 0 = 0$.

Example 1.3. Consider the space $X = [0, 1]$ with the Lebesgue measure μ. Moreover, for each $k \in \mathbb{N}$, consider the measurable partition

$$\xi_k = \big\{C_j := (j/k, (j+1)/k) : j = 0, ..., k-1\big\}$$

of X, consisting of k intervals of equal measure $\mu(C_j) = 1/k$. We have

$$\begin{aligned} H_\mu(\xi_k) &= -\sum_{j=0}^{k-1} \mu(C_j) \log \mu(C_j) \\ &= -\sum_{j=0}^{k-1} \frac{1}{k} \log \frac{1}{k} = \log k. \end{aligned}$$

Now let $\psi\colon [0,1] \to \mathbb{R}$ be the continuous function defined by

$$\psi(x) = \begin{cases} -x \log x & \text{if } 0 < x \leq 1, \\ 0 & \text{if } x = 0. \end{cases} \tag{1.1}$$

We note that

$$H_\mu(\xi) = \sum_{C \in \xi} \psi(\mu(C)).$$

Since $\psi''(x) = -1/x < 0$ for $x > 0$, the function ψ is strictly concave and hence,

$$\sum_{i=1}^{p} a_i \psi(x_i) \leq \psi\left(\sum_{i=1}^{p} a_i x_i\right) \tag{1.2}$$

for any numbers

$$x_1, ..., x_p, a_1, ..., a_p \in [0,1] \quad \text{such that} \quad \sum_{i=1}^{p} a_i = 1.$$

This useful inequality will be used repeatedly throughout the chapter. In particular, it allows one to give an upper bound for the entropy in terms of the cardinality of ξ. Letting $k = \operatorname{card} \xi$, we obtain

$$\begin{aligned} H_\mu(\xi) &= k \sum_{C \in \xi} \frac{1}{k} \psi(\mu(C)) \\ &\leq k\psi\left(\sum_{C \in \xi} \frac{\mu(C)}{k}\right) \\ &= k\psi\left(\frac{1}{k}\right) = \log k, \end{aligned}$$

that is,

$$H_\mu(\xi) \leq \log \operatorname{card} \xi. \tag{1.3}$$

Definition 1.4. Given two measurable partitions ξ and η of X, we say that η *is a refinement of* ξ if for every $D \in \eta$ there exists $C \in \xi$ such that $\mu(D \backslash C) = 0$.

Proposition 1.5. *Let ξ and η be measurable partitions of X. If η is a refinement of ξ, then*

$$H_\mu(\xi) \leq H_\mu(\eta).$$

Proof. Since η is a refinement, one can write

$$\begin{aligned} H_\mu(\eta) &= \sum_{D \in \eta} \psi(\mu(D)) \\ &= \sum_{C \in \xi} \sum_{D \subset C} \psi(\mu(D)). \end{aligned} \tag{1.4}$$

For each $C \in \xi$, since $\sum_{D \subset C} \mu(D) = \mu(C)$, we have

$$\begin{aligned} \psi(\mu(C)) &= \psi\left(\sum_{D \subset C} \mu(D)\right) \\ &= -\sum_{D \subset C} \mu(D) \log \sum_{D \subset C} \mu(D) \\ &\leq -\sum_{D \subset C} \mu(D) \log \mu(D) = \sum_{D \subset C} \psi(\mu(D)), \end{aligned}$$

where the inequality follows from the fact that the logarithm is a nondecreasing function. Adding these quantities for all $C \in \xi$ yields that

$$\sum_{C \in \xi} \psi(\mu(C)) \leq \sum_{C \in \xi} \sum_{D \subset C} \psi(\mu(D))$$

and so it follows from (1.4) that $H_\mu(\xi) \leq H_\mu(\eta)$. □

In order to continue our discussion, we need to be able to construct new partitions from old ones. Given two measurable partitions ξ and η of X, we define the measurable partition

$$\xi \vee \eta = \big\{C \cap D : C \in \xi, D \in \eta\big\}.$$

We are only interested in partitions without sets of zero measure, and by discarding them one can always assume that all elements of the partition $\xi \vee \eta$ have positive measure.

We note that $\xi \vee \eta$ is the smallest measurable partition (in terms of cardinality) that is a refinement of both ξ and η.

Proposition 1.6. *If $T\colon X \to X$ is a measurable transformation preserving a probability measure μ on X and ξ is a measurable partition of X, then the limit*

$$\lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n) = \inf_{n\in\mathbb{N}} \frac{1}{n} H_\mu(\xi_n) \tag{1.5}$$

exists, where

$$\xi_n = \bigvee_{k=0}^{n-1} T^{-k}\xi \quad \textit{for} \quad n \in \mathbb{N}. \tag{1.6}$$

Proof. We first prove some lemmas.

Lemma 1.7. *If ξ and η are measurable partitions of X, then*

$$H_\mu(\xi \vee \eta) \le H_\mu(\xi) + H_\mu(\eta).$$

Proof of the lemma. Making some simple manipulations, we obtain

$$\begin{aligned} H_\mu(\xi \vee \eta) &= -\sum_{C\in\xi, D\in\eta} \mu(C\cap D)\log\mu(C\cap D) \\ &= -\sum_{C\in\xi, D\in\eta} \mu(C\cap D)\left(\log\frac{\mu(C\cap D)}{\mu(C)} + \log\mu(C)\right) \\ &= \sum_{D\in\eta}\sum_{C\in\xi} \mu(C)\psi\left(\frac{\mu(C\cap D)}{\mu(C)}\right) - \sum_{C\in\xi}\sum_{D\in\eta} \mu(C\cap D)\log\mu(C). \end{aligned}$$

Moreover, using inequality (1.2) for the first term, we obtain

$$\begin{aligned} H_\mu(\xi \vee \eta) &\le \sum_{D\in\eta} \psi\left(\sum_{C\in\xi} \mu(C)\frac{\mu(C\cap D)}{\mu(C)}\right) - \sum_{C\in\xi} \mu(C)\log\mu(C) \\ &= \sum_{D\in\eta} \psi(\mu(D)) + H_\mu(\xi) \\ &= H_\mu(\xi) + H_\mu(\eta). \end{aligned}$$

This completes the proof of the lemma. □

Now we consider the measurable partitions

$$T^{-n}\xi = \{T^{-n}C : C \in \xi\} \quad \text{for} \quad n \in \mathbb{N}.$$

Lemma 1.8. *If ξ is a measurable partition of X, then*

$$H_\mu(T^{-n}\xi) = H_\mu(\xi) \quad \textit{for all } n \in \mathbb{N}.$$

Proof of the lemma. We will use induction on n.

For $n = 1$, since μ is T-invariant, we obtain

$$\begin{aligned} H_\mu(T^{-1}\xi) &= -\sum_{C\in\xi} \mu(T^{-1}C)\log\mu(T^{-1}C) \\ &= -\sum_{C\in\xi} \mu(C)\log\mu(C) \\ &= H_\mu(\xi). \end{aligned}$$

Now assume that $H_\mu(T^{-n}\xi) = H_\mu(\xi)$ for some $n \in \mathbb{N}$. Using again the fact that μ is T-invariant, for $n+1$ we obtain

$$\begin{aligned} H_\mu(T^{-(n+1)}\xi) &= -\sum_{C\in\xi} \mu(T^{-1}(T^{-n}C))\log\mu(T^{-1}(T^{-n}C)) \\ &= -\sum_{C\in\xi} \mu(T^{-n}C)\log\mu(T^{-n}C) \\ &= H_\mu(T^{-n}\xi) \\ &= H_\mu(\xi), \end{aligned}$$

using the induction hypothesis on the last identity. □

Let us take a closer look at the measurable partitions defined in (1.6). Given $n, m \in \mathbb{N}$, we have

$$\xi_{n+m} = \xi_n \vee T^{-n}\xi_m.$$

Indeed, one can write

$$\begin{aligned} \xi_{n+m} &= \left\{ \bigcap_{j=0}^{n+m-1} T^{-j}C_j : C_j \in \xi \text{ for every } j \right\} \\ &= \left\{ \left(\bigcap_{j=0}^{n-1} T^{-j}C_j \right) \cap \left(\bigcap_{j=n}^{n+m-1} T^{-j}C_j \right) : C_j \in \xi \text{ for every } j \right\} \end{aligned}$$

$$= \left\{ \left(\bigcap_{j=0}^{n-1} T^{-j} C_j \right) \cap T^{-n} \left(\bigcap_{k=0}^{m-1} T^{-k} D_k \right) : C_j, D_k \in \xi \text{ for every } j, k \right\}$$
$$= \xi_n \vee T^{-n} \xi_m.$$

Using this observation and the two former lemmas, we obtain

$$\begin{aligned} H_\mu(\xi_{n+m}) &= H_\mu(\xi_n \vee T^{-n}\xi_m) \\ &\leq H_\mu(\xi_n) + H_\mu(T^{-n}\xi_m) \\ &= H_\mu(\xi_n) + H_\mu(\xi_m) \end{aligned} \tag{1.7}$$

for all $n, m \in \mathbb{N}$. This is the main property that guarantees the existence of the limit in (1.5). For this, we need to establish the following result.

Lemma 1.9. *If $(a_n)_{n\in\mathbb{N}}$ is a sequence of positive numbers such that*

$$a_{n+m} \leq a_n + a_m \quad \textit{for every } n, m \in \mathbb{N},$$

then the limit

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf \left\{ \frac{a_k}{k} : k \in \mathbb{N} \right\} \geq 0 \tag{1.8}$$

exists.

Proof of the lemma. Take $k \in \mathbb{N}$. Then $n = qk + r$ for some integers $q \in \mathbb{N}$ and $r \in \{0, \ldots, k-1\}$. We have

$$\frac{a_n}{n} \leq \frac{a_{qk} + a_r}{qk + r} \leq \frac{qa_k + a_r}{qk + r}.$$

Thus, letting $n \to \infty$ (and hence $q \to \infty$), we obtain

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \frac{a_k}{k}.$$

Therefore,

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \inf \left\{ \frac{a_k}{k} : k \in \mathbb{N} \right\} \leq \liminf_{n\to\infty} \frac{a_n}{n}.$$

This establishes the existence of the limit in (1.8). □

By (1.7), it follows from Lemma 1.9 that the sequence $H_\mu(\xi_n)/n$ converges and that identity (1.5) holds. $\square$

Definition 1.10. Let $T\colon X \to X$ be a measurable transformation preserving a probability measure μ on X. We define the *metric entropy* of T with respect to μ and a measurable partition ξ of X by

$$h_\mu(T,\xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n),$$

with ξ_n as in (1.6). Finally, we define the *metric entropy* of T with respect to μ by

$$h_\mu(T) = \sup\big\{h_\mu(T,\xi) : \xi \text{ is a measurable partition of } X\big\}.$$

Proposition 1.11. *We have $h_\mu(T^k) = kh_\mu(T)$ for each $k \in \mathbb{N}$.*

Proof. We first show that $h_\mu(T^k) \le kh_\mu(T)$. It is clear that the measurable partition $\bigvee_{i=0}^{nk-1} T^{-i}\xi$ is a refinement of $\bigvee_{i=0}^{n-1} T^{-ki}\xi$. So, by Proposition 1.5, we have

$$H_\mu\left(\bigvee_{i=0}^{n-1} T^{-ki}\xi\right) \le H_\mu\left(\bigvee_{i=0}^{nk-1} T^{-i}\xi\right).$$

Therefore,

$$h_\mu(T^k,\xi) \le \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{nk-1} T^{-i}\xi\right) = kh_\mu(T,\xi).$$

Taking the supremum in ξ, this implies that $h_\mu(T^k) \le kh_\mu(T)$.

Now we show that $h_\mu(T^k) \ge kh_\mu(T)$. Since

$$\bigvee_{\ell=0}^{n-1} T^{-k\ell}\xi_k = \bigvee_{i=0}^{nk-1} T^{-i}\xi,$$

we have

$$\frac{1}{n} H_\mu\left(\bigvee_{\ell=0}^{n-1} T^{-k\ell}\xi_k\right) = \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{nk-1} T^{-i}\xi\right).$$

Letting $n \to \infty$, we obtain

$$h_\mu(T^k, \xi_k) = k h_\mu(T, \xi)$$

and so,

$$\begin{aligned} h_\mu(T^k) &= \sup_\xi h_\mu(T^k, \xi) \\ &\geq \sup_\xi h_\mu(T^k, \xi_k) \\ &= \sup_\xi k h_\mu(T, \xi) = k h_\mu(T). \end{aligned}$$

This completes the proof of the proposition. $\square$

Chapter 2

Conditional Entropy

The goal of this chapter is to introduce the notion of conditional entropy. We use this notion to establish a few deeper results concerning metric entropy, which will be useful latter on. It also allows us to give an alternative characterization of the notion of metric entropy.

Definition 2.1. Given measurable partitions ξ and ζ of X, we define the *conditional entropy* of ξ with respect to ζ by

$$H_\mu(\xi|\zeta) = -\sum_{C\in\xi, D\in\zeta} \mu(C\cap D)\log\frac{\mu(C\cap D)}{\mu(D)}.$$

The notion of conditional entropy can be used to give the following characterization of metric entropy.

Theorem 2.2. *If $T\colon X\to X$ is a measurable transformation preserving a probability measure μ on X and ξ is a measurable partition of X, then*

$$h_\mu(T,\xi) = \lim_{n\to\infty} H_\mu\left(\xi|\bigvee_{i=1}^{n} T^{-i}\xi\right). \tag{2.1}$$

Proof. In order to establish the existence of the limit in (2.1) we will use the following auxiliary result showing that the sequence $H_\mu\left(\xi|\bigvee_{i=1}^{n} T^{-i}\xi\right)$ is nonincreasing.

Lemma 2.3. *If ζ is a refinement of η, then*

$$H_\mu(\xi|\zeta) \le H_\mu(\xi|\eta)$$

for every partition ξ.

Proof of the lemma. We have

$$\begin{aligned} H_\mu(\xi|\zeta) &= -\sum_{C\in\xi, D\in\zeta} \mu(C\cap D)\log\frac{\mu(C\cap D)}{\mu(D)} \\ &= -\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(D\cap E)\frac{\mu(C\cap D)}{\mu(D)}\log\frac{\mu(C\cap D)}{\mu(D)} \\ &= \sum_{C\in\xi, E\in\eta} \mu(E)\sum_{D\in\zeta}\frac{\mu(D\cap E)}{\mu(E)}\psi\left(\frac{\mu(C\cap D)}{\mu(D)}\right). \end{aligned}$$

Again using (1.2), we obtain

$$H_\mu(\xi|\zeta) \le \sum_{C\in\xi, E\in\eta} \mu(E)\psi\left(\sum_{D\in\zeta}\frac{\mu(D\cap E)}{\mu(E)}\frac{\mu(C\cap D)}{\mu(D)}\right).$$

But since ζ is a refinement of η, we have

$$\begin{aligned} \sum_{D\in\zeta}\frac{\mu(D\cap E)}{\mu(E)}\frac{\mu(C\cap D)}{\mu(D)} &= \sum_{D\in\zeta, D\subset E}\frac{\mu(C\cap D)}{\mu(E)} \\ &= \frac{\mu(C\cap E)}{\mu(E)} \end{aligned}$$

and hence,

$$\begin{aligned} H_\mu(\xi|\zeta) &\le \sum_{C\in\xi, E\in\eta} \mu(E)\psi\left(\frac{\mu(C\cap E)}{\mu(E)}\right) \\ &= -\sum_{C\in\xi, E\in\eta} \mu(C\cap E)\log\frac{\mu(C\cap E)}{\mu(E)} = H_\mu(\xi|\eta). \end{aligned}$$

This yields the desired inequality. □

Since $\bigvee_{i=1}^{n+1} T^{-i}\xi$ is a refinement of $\bigvee_{i=1}^{n} T^{-i}\xi$, it follows from Lemma 2.3 that the sequence $H_\mu(\xi|\bigvee_{i=1}^{n} T^{-i}\xi)$ is nonincreasing in n and so the limit in (2.1) exists.

In order to complete the proof, we also need the following result.

Lemma 2.4. *If ξ, ζ, and η are measurable partitions of X, then*

$$H_\mu(\xi \vee \zeta|\eta) = H_\mu(\xi|\zeta \vee \eta) + H_\mu(\zeta|\eta).$$

Proof of the lemma. We have

$$\begin{aligned} H_\mu(\xi \vee \zeta|\eta) &= -\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(C\cap D\cap E)\log\frac{\mu(C\cap D\cap E)}{\mu(D)} \\ &= -\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(C\cap D\cap E)\log\left(\frac{\mu(C\cap D\cap E)}{\mu(D)}\frac{\mu(D\cap E)}{\mu(D\cap E)}\right) \\ &= -\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(C\cap D\cap E)\log\frac{\mu(C\cap D\cap E)}{\mu(D\cap E)} \\ &\quad -\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(C\cap D\cap E)\log\frac{\mu(D\cap E)}{\mu(D)}. \end{aligned}$$

Therefore,

$$\begin{aligned} H_\mu(\xi \vee \zeta|\eta) &= H_\mu(\xi|\zeta \vee \eta) - \sum_{D\in\zeta, E\in\eta} \mu(D\cap E)\log\frac{\mu(D\cap E)}{\mu(D)} \\ &= H_\mu(\xi|\zeta \vee \eta) + H_\mu(\zeta|\eta), \end{aligned}$$

which yields the identity in the lemma. □

In particular, for the trivial partition $\eta = \{X\}$ it follows from Lemma 2.4 that

$$H_\mu(\xi \vee \zeta) = H_\mu(\xi|\zeta) + H_\mu(\zeta). \tag{2.2}$$

Now we use induction to show that

$$H_\mu(\xi_n) = H_\mu(\xi) + \sum_{j=1}^{n-1} H_\mu\left(\xi|\bigvee_{i=1}^{j} T^{-i}\xi\right) \tag{2.3}$$

for every $n \in \mathbb{N}$. The identity is clear when $n = 1$. Now we assume that it holds for some n. Then, by Lemma 1.8 and (2.2), we have

$$\begin{aligned} H_\mu(\xi_{n+1}) &= H_\mu\left(\xi \vee \bigvee_{i=1}^{n} T^{-i}\xi\right) \\ &= H_\mu\left(\bigvee_{i=1}^{n} T^{-i}\xi\right) + H_\mu\left(\xi|\bigvee_{i=1}^{n} T^{-i}\xi\right) \end{aligned}$$

$$= H_\mu(T^{-1}(\xi_n)) + H_\mu\left(\xi|\bigvee_{i=1}^{n} T^{-i}\xi\right)$$
$$= H_\mu(\xi) + \sum_{j=1}^{n} H_\mu\left(\xi|\bigvee_{i=1}^{j} T^{-i}\xi\right),$$

using the induction hypothesis for the last equality.

Since the sequence $H_\mu(\xi|\bigvee_{i=1}^{n} T^{-i}\xi)$ converges, it follows from (2.3) that

$$\begin{aligned} h_\mu(T,\xi) &= \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n-1} H_\mu\left(\xi|\bigvee_{i=1}^{j} T^{-i}\xi\right) \\ &= \lim_{n\to\infty} H_\mu\left(\xi|\bigvee_{i=1}^{n} T^{-i}\xi\right), \end{aligned}$$

where for the last identity we used the fact that if a_n is a sequence such that $a_n \to a$, then $1/n \sum_{j=1}^{n-1} a_j \to a$. □

Given a measurable partition ξ of X, we denote by $\mathcal{A}(\xi)$ the σ-algebra generated by ξ, that is, the smallest σ-algebra containing all the elements of ξ. Moreover, we denote by $\bigvee_{n\in I} \mathcal{A}(\xi_n)$ the σ-algebra generated by the measurable partitions ξ_n for $n \in I$.

Theorem 2.5. *Let $T\colon X \to X$ be a measurable transformation preserving a probability measure μ on X. If $(\alpha_n)_{n\in\mathbb{N}}$ is a sequence of measurable partitions of X with $\bigvee_{n=1}^{\infty} \mathcal{A}(\alpha_n) = \mathcal{A}$ such that α_{n+1} is a refinement of α_n for each $n \in \mathbb{N}$, then*

$$h_\mu(T) = \lim_{n\to\infty} h_\mu(T,\alpha_n) = \sup_{n\in\mathbb{N}} h_\mu(T,\alpha_n). \tag{2.4}$$

Proof. We first observe that the limit in (2.4) exists. Indeed, it follows from Proposition 1.5 that

$$H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha_m\right) \le H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha_{m+1}\right)$$

for any $m, n \in \mathbb{N}$ and hence,

$$h_\mu(T, \alpha_m) \leq h_\mu(T, \alpha_{m+1}) \quad \text{for each} \quad m \in \mathbb{N}.$$

So the sequence $h_\mu(T, \alpha_n)$ is nondecreasing and we have

$$\lim_{n\to\infty} h_\mu(T, \alpha_n) = \sup_{n\in\mathbb{N}} h_\mu(T, \alpha_n).$$

Now we establish two auxiliary results, which combined will yield the desired result.

Lemma 2.6. *If η and ζ are measurable partitions of X, then*

$$h_\mu(T, \eta) \leq h_\mu(T, \zeta) + H_\mu(\eta|\zeta). \tag{2.5}$$

Proof of the lemma. By Proposition 1.5 and (2.2), we have

$$H_\mu(\eta_n) \leq H_\mu(\eta_n \vee \zeta_n) = H_\mu(\zeta_n) + H_\mu(\eta_n|\zeta_n), \tag{2.6}$$

where

$$\eta_n = \bigvee_{i=0}^{n-1} T^{-i}\eta \quad \text{and} \quad \zeta_n = \bigvee_{i=0}^{n-1} T^{-i}\zeta. \tag{2.7}$$

On the other hand, it follows from Lemmas 2.3 and 2.4 that

$$\begin{aligned} H_\mu(\alpha \vee \beta|\gamma) &= H_\mu(\alpha|\beta \vee \gamma) + H_\mu(\beta|\gamma) \\ &\leq H_\mu(\alpha|\gamma) + H_\mu(\beta|\gamma) \end{aligned}$$

for any measurable partitions α, β, and γ of X. Therefore, by (2.6), (2.7), and Lemma 2.3, we obtain

$$\begin{aligned} H_\mu(\eta_n) &\leq H_\mu(\zeta_n) + H_\mu(\eta_n|\zeta_n) \\ &\leq H_\mu(\zeta_n) + \sum_{i=0}^{n-1} H_\mu(T^{-i}\eta|\zeta_n) \\ &\leq H_\mu(\zeta_n) + \sum_{i=0}^{n-1} H_\mu(T^{-i}\eta|T^{-i}\zeta) \\ &= H_\mu(\zeta_n) + nH_\mu(\eta|\zeta), \end{aligned}$$

using the invariance of the measure μ for the last identity. Finally, dividing by n and taking limits when $n \to \infty$ yields inequality (2.5). □

Lemma 2.7. *For any measurable partition η of X, we have $H_\mu(\eta|\alpha_n) \to 0$ when $n \to \infty$.*

Proof of the lemma. Write $\eta = \{C_1, \ldots, C_k\}$. For each sufficiently large $n \in \mathbb{N}$ and each $i = 1, \ldots, k$, let $D_i^n \subset C_i$ be a set in $\mathcal{A}(\alpha_n)$ such that $\mu(C_i \backslash D_i^n) \to 0$ when $n \to \infty$. Let $\beta_n = \{D_0^n, D_1^n, \ldots, D_k^n\}$, where $D_0^n = X \backslash \bigcup_{i=1}^k D_i^n$. Clearly, $\mu(C_i \cap D_0^n) \to 0$ when $n \to \infty$. Since

$$C_i \cap D_i^n = D_i^n \quad \text{and} \quad \mu(C_i \cap D_j^n) = 0$$

for $i = 1, \ldots, k$ and $j \neq i$, we obtain

$$\begin{aligned} H_\mu(\eta|\beta_n) &= -\sum_{i=1}^k \sum_{j=0}^k \mu(C_i \cap D_j^n) \log \frac{\mu(C_i \cap D_j^n)}{\mu(D_j^n)} \\ &= -\sum_{i=1}^k \mu(C_i \cap D_i^n) \log \frac{\mu(C_i \cap D_i^n)}{\mu(D_i^n)} \\ &\quad - \sum_{i=1}^k \mu(C_i \cap D_0^n) \log \frac{\mu(C_i \cap D_0^n)}{\mu(D_0^n)} \\ &\quad - \sum_{j=1}^k \sum_{i \neq j} \mu(C_i \cap D_j^n) \log \frac{\mu(C_i \cap D_j^n)}{\mu(D_j^n)} \\ &= -\sum_{i=1}^k \mu(C_i \cap D_0^n) \log \frac{\mu(C_i \cap D_0^n)}{\mu(D_0^n)}. \end{aligned} \tag{2.8}$$

Moreover, since

$$\log \frac{\mu(C_i \cap D_0^n)}{\mu(D_0^n)} = \log \mu(C_i \cap D_0^n) - \log \mu(D_0^n),$$

we have

$$H_\mu(\eta|\beta_n) = \sum_{i=1}^k \psi(\mu(C_i \cap D_0^n)) - \psi(\mu(D_0^n)) \to 0 \quad \text{when} \quad n \to \infty.$$

Finally, since α_n is a refinement of β_n, it follows from Lemma 2.3 that

$$H_\mu(\eta|\alpha_n) \leq H_\mu(\eta|\beta_n) \to 0 \quad \text{when} \quad n \to \infty,$$

which yields the desired statement. □

Taking $\zeta = \alpha_n$ in (2.5), it follows from Lemma 2.7 that

$$h_\mu(T, \eta) \leq \lim_{n\to\infty} h_\mu(T, \alpha_n) \quad \text{for any measurable partition } \eta.$$

This completes the proof of the theorem. $\square$

Chapter 3

Topological Entropy

In this chapter we introduce the notion of topological entropy and we establish some first results. The modern notion that we present is due to Efim Dinaburg in 1970 and (independently) to Rufus Bowen in 1971. We also give some alternative characterizations of topological entropy, which will be useful latter on.

Let $T\colon X \to X$ be a continuous transformation of a compact metric space (X, d). For each $n \in \mathbb{N}$, we introduce a new distance on X by

$$d_n(x, y) = d_{n,T}(x, y) = \max\big\{d(T^k(x), T^k(y)) : 0 \le k \le n-1\big\}.$$

Given $\varepsilon > 0$, we denote by $N(d, \varepsilon)$ the maximum number of points in X at a d-distance at least ε.

Definition 3.1. We define the *topological entropy* of T by

$$h(T) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log N(d_n, \varepsilon). \tag{3.1}$$

We note that since the function

$$\varepsilon \mapsto \limsup_{n \to \infty} \frac{1}{n} \log N(d_n, \varepsilon)$$

is nonincreasing, the limit in (3.1) when $\varepsilon \to 0$ is well defined.

We also give two equivalent descriptions of the topological entropy. These are not only helpful to compute the topological entropy, but in the context of this text they will be central pieces in later results and in particular in the proof of the variational principle.

We recall that the *d-diameter* of a set $A \subset X$ is the number

$$\sup\big\{d(x,y) : x, y \in A\big\}.$$

Given $\varepsilon > 0$, let $D(d,\varepsilon)$ be the minimum number of sets of d-diameter less than ε that are needed to cover X and let $M(d,\varepsilon)$ be the minimum number of points $p_1, \ldots, p_m$ of X such that the balls $B_d(p_1,\varepsilon), \ldots, B_d(p_m,\varepsilon)$ cover X.

We have the following relation between these numbers and $N(d,\varepsilon)$.

Proposition 3.2. *For any distance d on X and any $\varepsilon > 0$, we have*

$$D(d,2\varepsilon) \le M(d,\varepsilon) \le N(d,\varepsilon) \le M(d,\varepsilon/2) \le D(d,\varepsilon/2). \tag{3.2}$$

Proof. We establish the four inequalities in the same order:

1. Take points $p_1, \ldots, p_m \in X$ with $m = M(d,\varepsilon)$ such that the open balls $B_d(p_i,\varepsilon)$ cover X. Since the d-diameter of each of these open balls is less than 2ε, we have $D(d,2\varepsilon) \le m$.

2. Let $p_1, \ldots, p_m \in X$ be points such that $d(p_i,p_j) \ge \varepsilon$ for $i \ne j$, where $m = N(d,\varepsilon)$. Each point $x \in X$ satisfies $d(x,p_i) < \varepsilon$ for some i and hence $M(d,\varepsilon) \le m$.

3. Since it is impossible for any d-open ball of radius $\varepsilon/2$ to contain two points at a d-distance ε, we have $N(d,\varepsilon) \le M(d,\varepsilon/2)$.

4. Let $A_1, \ldots, A_m$ be a cover of X by sets of d-diameter less than $\varepsilon/2$, where $m = D(d,\varepsilon/2)$. For each $i = 1, \ldots, m$ take a point $p_i \in A_i$. It is easy to verify that $A_i \subset B_d(p_i,\varepsilon/2)$, and hence these d-balls form a cover of X. Therefore $M(d,\varepsilon/2) \le D(d,\varepsilon/2)$.

This completes the proof of the proposition. □

Now we can provide equivalent definitions of the topological entropy using these quantities.

Theorem 3.3. *If $T\colon X \to X$ is a continuous transformation of a compact metric space, then*

$$\begin{aligned} h(T) &= \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log N(d_n, \varepsilon) \\ &= \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log M(d_n, \varepsilon) \\ &= \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log M(d_n, \varepsilon) \\ &= \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log D(d_n, \varepsilon). \end{aligned} \tag{3.3}$$

Proof. First we show that the limit when $n \to \infty$ in the last formula in (3.3) exists.

Lemma 3.4. *If $m, n \in \mathbb{N}$ and $\varepsilon > 0$, then*

$$D(d_{m+n}, \varepsilon) \le D(d_m, \varepsilon) D(d_n, \varepsilon).$$

Proof of the lemma. Let $\{A_1, \dots, A_k\}$ be a cover of X by sets of d_n-diameter less than ε, where $k = D(d_n, \varepsilon)$. Let also $\{B_1, \dots, B_\ell\}$ be a cover of X by sets of d_m-diameter less than ε, where $\ell = D(d_m, \varepsilon)$. Then each set $A_i \cap T^{-n} B_j$ has d_{m+n}-diameter less than ε and

$$\mathcal{U} = \left\{ A_i \cap T^{-n} B_j : i = 1, \dots, k \text{ and } j = 1, \dots, \ell \right\}$$

is a cover of X. Therefore,

$$D(d_{m+n}, \varepsilon) \le \operatorname{card} \mathcal{U} \le \ell k = D(d_m, \varepsilon) D(d_n, \varepsilon),$$

which yields the desired inequality. □

It follows from Lemmas 1.9 and 3.4 that the limit

$$\lim_{n\to\infty} \frac{1}{n} \log D(d_n, \varepsilon)$$

exists. Hence, by (3.2), we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log D(d_n, 2\varepsilon) &\leq \liminf_{\to\infty} \frac{1}{n} \log M(d_n, \varepsilon) \\
&\leq \liminf_{n\to\infty} \frac{1}{n} \log N(d_n, \varepsilon) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log N(d_n, \varepsilon) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log M(d_n, \varepsilon/2) \\
&\leq \lim_{n\to\infty} \frac{1}{n} \log D(d_n, \varepsilon/2)
\end{aligned}$$

for all $\varepsilon > 0$. Taking limits when $\varepsilon \to 0$, we conclude that

$$\begin{aligned}
\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log D(d_n, 2\varepsilon) &\leq \lim_{\varepsilon\to 0} \liminf_{\to\infty} \frac{1}{n} \log M(d_n, \varepsilon) \\
&\leq \lim_{\varepsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log N(d_n, \varepsilon) \\
&\leq h(T) \qquad (3.4) \\
&\leq \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log M(d_n, \varepsilon/2) \\
&\leq \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log D(d_n, \varepsilon/2)
\end{aligned}$$

The fact that the first and the last terms in (3.4) are equal yields the desired identities in (3.3). □

Now we establish a similar result to Proposition 1.11, but for the topological entropy.

Theorem 3.5. *If $T\colon X \to X$ is a continuous transformation of a compact metric space, then $h(T^k) = kh(T)$ for each $k \in \mathbb{N}$.*

Proof. First we show that $h(T^k) \leq kh(T)$. We have

$$\begin{aligned}
d_{n,T^k}(x, y) &= \max\{d(T^{ik}(x), T^{ik}(y)) : 0 \leq i \leq n-1\} \\
&\leq \max\{d(T^i(x), T^i(y)) : 0 \leq i \leq nk-1\} \\
&= d_{nk,T}(x, y).
\end{aligned}$$

Therefore, $N(d_{n,T^k},\varepsilon) \le N(d_{nk,T},\varepsilon)$ and

$$\begin{aligned} h(T^k) &= \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log N(d_{n,T^k},\varepsilon) \\ &\le \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log N(d_{nk,T},\varepsilon) \\ &= k\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{nk}\log N(d_{nk,T},\varepsilon) \\ &\le k\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log N(d_{n,T},\varepsilon) = kh(T). \end{aligned}$$

Now we show that $kh(T) \le h(T^k)$. By the uniform continuity of T, given $\varepsilon > 0$, there exists $\delta(\varepsilon) \in (0,\varepsilon)$ such that $d_k(x,y) < \varepsilon$ whenever $d(x,y) < \delta(\varepsilon)$. Equivalently, $d(x,y) \ge \delta(\varepsilon)$ whenever $d_k(x,y) \ge \varepsilon$. Since

$$d_{kn,T}(x,y) = \max\left\{d_k(T^{ik}(x),T^{ik}(y)) : 0 \le i \le n-1\right\},$$

we conclude that if $d_{kn,T}(x,y) \ge \varepsilon$, then $d(T^{ik}(x),T^{ik}(y)) \ge \delta(\varepsilon)$ for some $i \in \{0,\dots,n-1\}$. Hence,

$$d_{n,T^k}(x,y) \ge \delta(\varepsilon) \quad \text{whenever} \quad d_{kn,T}(x,y) \ge \varepsilon.$$

Therefore, $N(d_{kn,T},\varepsilon) \le N(d_{n,T^k},\delta(\varepsilon))$ and so,

$$\begin{aligned} \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log N(d_{kn,T},\varepsilon) &\le \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log N(d_{n,T^k},\delta(\varepsilon)) \\ &= h(T^k), \end{aligned} \tag{3.5}$$

where the last equality is due to the fact that $\delta(\varepsilon) \to 0$ when $\varepsilon \to 0$. Finally, it follows from (3.2), (3.3) and (3.5) that

$$\begin{aligned} h(T^k) &\ge \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log N(d_{kn},\varepsilon) \\ &\ge \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log D(d_{kn},2\varepsilon) \\ &= k\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{kn}\log D(d_{kn},2\varepsilon) \\ &= k\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log D(d_n,2\varepsilon) = kh(T), \end{aligned}$$

which yields the desired identity. □

Chapter 4

Variational Principle

In this chapter we give a complete proof of the *Variational Principle* for the entropy, showing that the supremum of the metric entropies over all Borel probability invariant measures is equal to the topological entropy. The proof presented here is based on the simpler proof given by Misiurewicz in [10].

Let X be a topological space. The σ-algebra generated by all open subsets of X is called the *Borel* σ-algebra of X. A measure defined in this σ-algebra is called a Borel measure. In this chapter all measures are Borel measures. Finally we denote by $\mathcal{M}(X)$ the set of all Borel probability measures on X.

Theorem 4.1 (Variational principle for the entropy)**.** *If $T\colon X \to X$ is a continuous transformation of a compact metric space, then*

$$h(T) = \sup\big\{h_\mu(T) : \mu \textit{ is a } T\textit{-invariant measure in } \mathcal{M}(X)\big\}.$$

Proof. We first show that $h_\mu(T) \leq h(T)$ for any T-invariant probability measure μ on X. Let

$$\eta = \{C_1, \ldots, C_k\}$$

be a measurable partition of X. Given $\delta > 0$, for each $i = 1, \ldots, k$, let $D_i \subset C_i$ be a compact set such that

$$\mu(C_i \backslash D_i) < \delta \quad \text{and} \quad D_i \cap D_j = \emptyset \quad \text{for } i \neq j.$$

Now let

$$\beta = \{D_0, D_1, \ldots, D_k\},$$

where $D_0 = X \setminus \bigcup_{i=1}^k D_i$. Clearly, β is a measurable partition of X. Proceeding as in (2.8) with D_i^n replaced by D_i, we obtain

$$H_\mu(\eta|\beta) = \sum_{i=1}^k \psi(\mu(C_i \cap D_0)) - \psi(\mu(D_0)).$$

By construction, we have $\mu(C_i \cap D_0) < \delta$ and since ψ is nondecreasing near zero, we conclude that $H_\mu(\eta|\beta) < 1$ for any sufficiently small δ. Thus, it follows from (2.5) that

$$h_\mu(T, \eta) \leq h_\mu(T, \beta) + H_\mu(\eta|\beta) < h_\mu(T, \beta) + 1. \tag{4.1}$$

Now we construct a sequence of open covers that together with some delicate observations will allow us to relate the metric entropy and the topological entropy. Consider the open cover of X given by

$$\mathcal{U} = \{D_0 \cup D_1, \ldots, D_0 \cup D_k\},$$

as well as the open covers

$$\mathcal{U}_n = \left\{ \bigcap_{i=0}^{n-1} T^{-i} U_i : U_0, \ldots, U_{n-1} \in \mathcal{U} \right\}$$

for each $n \in \mathbb{N}$. Let $N(\mathcal{U}_n)$ denote the minimal cardinality of a subcover of $\mathcal{U}_n$. We wish to relate $N(\mathcal{U}_n)$ to the cardinality of the partitions $\beta_n = \bigvee_{i=0}^{n-1} T^{-i}\beta$. We first notice that since

$$T^{-i}(D_0 \cup D_j) = T^{-i}(D_0) \cup T^{-i}(D_j)$$

for $j = 1, \ldots, k$ and $i \in \mathbb{N}$, each element of $\mathcal{U}_n$ is the union of at most 2^n elements of β_n. Moreover, since the elements in β_n are disjoint, each element of β_n is contained in some element of any given subcover of $\mathcal{U}_n$. Therefore,

$$\operatorname{card} \beta_n \leq 2^n N(\mathcal{U}_n).$$

Moreover, by (1.3), we obtain

$$H_\mu(\beta_n) \leq \log \operatorname{card} \beta_n \leq n \log 2 + \log N(\mathcal{U}_n). \tag{4.2}$$

On the other hand, by the theory of metric spaces we know that since X is compact, there exists $\varepsilon > 0$ such that every ball of radius at most ε is contained in some element of $\mathcal{U}$ (any such ε is called a Lebesgue number of $\mathcal{U}$). Moreover, one can easily verify that if ε is a Lebesgue number of $\mathcal{U}$, then ε is also a Lebesgue number of the cover $\mathcal{U}_n$ with respect to the distance d_n. In fact, by the definition of d_n we have

$$B_{d_n}(x, \varepsilon) = \bigcap_{i=0}^{n-1} T^{-i}(B_d(T^i(x), \varepsilon)),$$

and since ε is a Lebesgue number of $\mathcal{U}$ each ball $B_d(T^i(x), \varepsilon)$ is contained in some element U_i of $\mathcal{U}$. Hence,

$$B_{d_n}(x, \varepsilon) \subset \bigcap_{i=0}^{n-1} T^{-i} U_i.$$

Now take points $x_1, \ldots, x_p \in X$ such that

$$\left\{B_{d_n}(x_1, \varepsilon), \ldots, B_{d_n}(x_p, \varepsilon)\right\}$$

is an open cover of X, with $p = M(d_n, \varepsilon)$, as defined in the previous chapter. Since ε is a Lebesgue number of $\mathcal{U}_n$ with respect to the distance d_n, for each $i = 1, \ldots, p$ one can choose an element $U_i \in \mathcal{U}_n$ such that $B_{d_n}(x_i, \varepsilon) \subset U_i$. Clearly, the collection of open sets $\{U_i : 1 \leq i \leq p\}$ forms a subcover of $\mathcal{U}_n$ and hence,

$$N(\mathcal{U}_n) \leq \operatorname{card}\{U_i : 1 \leq i \leq p\} \leq p = M(d_n, \varepsilon).$$

It follows from (4.2) that

$$\begin{aligned} h_\mu(T, \beta) &= \lim_{n\to\infty} \frac{1}{n} H_\mu(\beta_n) \\ &\leq \log 2 + \limsup_{n\to\infty} \frac{1}{n} \log M(d_n, \varepsilon). \end{aligned}$$

By (4.1) and (3.3), letting $\varepsilon \to 0$ yields the inequality

$$h_\mu(T, \eta) \leq h(T) + \log 2 + 1$$

and hence,

$$h_\mu(T) \leq h(T) + \log 2 + 1. \tag{4.3}$$

In order to obtain this inequality, we required nothing from the transformation T other than the fact that it is continuous and that the measure μ is T-invariant. Hence, inequality (4.3) still holds when T is replaced by T^m. Then

$$h_\mu(T^m) \leq h(T^m) + \log 2 + 1 \quad \text{for all} \quad m \in \mathbb{N}.$$

Therefore, by Proposition 1.11 and Theorem 3.5,

$$\begin{aligned} h_\mu(T) &= \frac{1}{m} h_\mu(T^m) \\ &\leq \frac{1}{m}(h(T^m) + \log 2 + 1) \\ &= h(T) + \frac{1}{m}(\log 2 + 1) \end{aligned}$$

for each $m \in \mathbb{N}$. Letting $m \to \infty$, we conclude that $h_\mu(T) \leq h(T)$.

Now we show that $h(T) \leq \sup_\nu h_\nu(T)$. Given $\varepsilon > 0$, for each $n \in \mathbb{N}$, we consider a set E_n of points at a d_n-distance at least ε such that $\operatorname{card} E_n = N(d_n, \varepsilon)$. We then define measures

$$\nu_n = \frac{1}{\operatorname{card} E_n} \sum_{x \in E_n} \delta_x \quad \text{and} \quad \mu_n = \frac{1}{n} \sum_{i=0}^{n-1} T_*^i \nu_n,$$

where the measures $T_*^i \nu_n$ are defined by

$$(T_*^i \nu_n)(A) = \nu_n(T^{-i} A).$$

One can show that the set of all Borel probability measures on a compact metric space is a compact metrizable space (see the appendix). Hence, each

sequence of measures has a convergent subsequence. Moreover, a sequence of measures μ_n converges to a measure μ if and only if

$$\int_X \varphi \, d\mu_n \to \int_X \varphi \, d\mu$$

when $n \to \infty$, for every continuous function φ (see also the appendix).

Given a sequence $(k_n)_{n\in\mathbb{N}} \subset \mathbb{N}$ such that

$$\lim_{n\to\infty} \frac{1}{k_n} \log \operatorname{card} E_{k_n} = \limsup_{n\to\infty} \frac{1}{n} \log \operatorname{card} E_n, \tag{4.4}$$

let $(\ell_n)_{n\in\mathbb{N}}$ be a subsequence of $(k_n)_{n\in\mathbb{N}}$ such that $(\mu_{\ell_n})_{n\in\mathbb{N}}$ converges to some probability measure μ. We will show that this measure μ is T-invariant. We first establish an auxiliary result.

Lemma 4.2. *The transformation $T_* : \mathcal{M}(X) \to \mathcal{M}(X)$ defined by*

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

is continuous.

Proof. We will show that if $\mu_n \to \mu$, then $T_*\mu_n \to T_*\mu$, thus establishing the continuity of T_*. We first observe that

$$\int_X \varphi \, d(T_*\mu) = \int_X (\varphi \circ T) \, d\mu$$

for every function $\varphi \in C(X)$. Since T is continuous, we conclude that if $\varphi \in C(X)$, then

$$\begin{aligned} \int_X \varphi \, d(T_*\mu_n) &= \int_X (\varphi \circ T) \, d\mu_n \\ &\to \int_X (\varphi \circ T) \, d\mu = \int_X \varphi \, d(T_*\mu) \end{aligned}$$

when $n \to \infty$, that is, the sequence $(T_*\mu_n)_{n\in\mathbb{N}}$ converges to $T_*\mu$. This shows that the transformation T_* is continuous. $\square$

Now we show that μ is a fixed point of T_*, and hence T-invariant. Indeed,

$$\begin{aligned} T_*\mu_{\ell_n} &= T_*\left(\frac{1}{\ell_n}\sum_{i=0}^{\ell_n-1} T_*^i\nu_{\ell_n}\right) \\ &= \frac{1}{\ell_n}\sum_{i=1}^{\ell_n} T_*^i\nu_{\ell_n} \\ &= \frac{1}{\ell_n}\sum_{i=0}^{\ell_n-1} T_*^i\nu_{\ell_n} + \frac{T_*^{\ell_n}\nu_{\ell_n}}{\ell_n} - \frac{\nu_{\ell_n}}{\ell_n} \\ &= \mu_{\ell_n} + \frac{T_*^{\ell_n}\nu_{\ell_n}}{\ell_n} - \frac{\nu_{\ell_n}}{\ell_n}, \end{aligned}$$

and letting $n \to \infty$ we obtain $T_*\mu_{\ell_n} \to \mu$. On the other hand, by the continuity of T_* we also have $T_*\mu_{\ell_n} \to T_*\mu$, and hence $T_*\mu = \mu$. We note that in general the measure μ may depend on ε.

Now we consider some particular measurable partitions. Let $\{B_1, \ldots, B_k\}$ be an open cover of X by balls of radius less than $\varepsilon/2$ such that $\mu(\partial B_i) = 0$ for $i = 1, \ldots, k$, where ∂B_i denotes the boundary of B_i. This is always possible since for each $x \in X$ there are at most countably many values of $r > 0$ such that $\mu(\partial B(x, r)) > 0$. We define a partition $\xi = \{C_1, \ldots, C_k\}$ of X by

$$C_1 = \overline{B_1} \quad \text{and} \quad C_i = \overline{B_i} \setminus \bigcup_{j=1}^{i-1} \overline{B_j} \quad \text{for} \quad i = 2, \ldots, k.$$

Then diam $C_i < \varepsilon$ and $\mu(\partial C_i) = 0$ for each i since $\partial C_i \subset \bigcup_{j=1}^k \partial B_j$. Moreover, since diam $C_i < \varepsilon$, each element of $\xi_n = \bigvee_{i=0}^{n-1} T^{-i}\xi$ has d_n-diameter less than ε, and hence each element of ξ_n contains at most one point in E_n. Therefore, there is exactly a number $\operatorname{card} E_n$ of elements of ξ_n with ν_n-measure equal to $1/\operatorname{card} E_n$ and so,

$$\begin{aligned} H_{\nu_n}(\xi_n) &= -\sum_{C\in\xi_n} \nu_n(C)\log(\nu_n(C)) \\ &= \log \operatorname{card} E_n. \end{aligned} \tag{4.5}$$

Given $m, n \in \mathbb{N}$, we write $n = qm + r$, where $q \geq 0$ and $0 \leq r < m$. We have

$$\xi_n = \xi_{qm+r} = \bigvee_{j=0}^{q-1} T^{-jm}\xi_m \vee \bigvee_{j=qm}^{qm+r-1} T^{-j}\xi,$$

and thus, for each $i = 0, \ldots, m-1$, the measurable partition

$$\eta = \bigvee_{j=0}^{q-1} T^{-jm-i}\xi_m \vee \left(\bigvee_{j=qm}^{qm+r-1} T^{-j}\xi \vee \xi_i \right)$$

is a refinement of ξ_n. Since

$$\operatorname{card}\left(\bigvee_{j=qm}^{qm+r-1} T^{-j}\xi \vee \xi_i \right) \leq (\operatorname{card}\xi)^{2m},$$

it follows from Proposition 1.5, Lemma 1.7 and (1.3) that

$$\begin{aligned} H_{\nu_n}(\xi_n) &\leq H_{\nu_n}(\eta) \\ &\leq H_{\nu_n}\left(\bigvee_{j=0}^{q-1} T^{-jm-i}\xi_m \right) + 2m \log \operatorname{card}\xi \\ &\leq \sum_{j=0}^{q-1} H_{\nu_n}(T^{-jm-i}\xi_m) + 2m \log \operatorname{card}\xi. \end{aligned} \tag{4.6}$$

Now we observe that since the function ψ in (1.1) is concave, we have

$$\begin{aligned} \frac{1}{n}\sum_{i=0}^{m-1}\sum_{j=0}^{q-1} H_{\nu_n}(T^{-jm-i}\xi_m) &\leq \frac{1}{n}\sum_{\ell=0}^{n-1} H_{\nu_n}(T^{-\ell}\xi_m) \\ &= \sum_{A\in\xi_m}\sum_{\ell=0}^{n-1} \frac{1}{n}\psi(\nu_n(T^{-\ell}A)) \\ &\leq \sum_{A\in\xi_m} \psi\left(\sum_{\ell=0}^{n-1} \frac{1}{n}\nu_n(T^{-\ell}A) \right) \\ &= \sum_{A\in\xi_m} \psi(\mu_n(A)) = H_{\mu_n}(\xi_m), \end{aligned}$$

and since (4.6) holds for $i = 0, \ldots, m-1$, we obtain

$$\frac{m}{n} H_{\nu_n}(\xi_n) \leq H_{\mu_n}(\xi_m) + \frac{2m^2}{n} \log \operatorname{card}\xi.$$

It follows from (4.5) with n replaced by ℓ_n that

$$\frac{1}{\ell_n}\log\operatorname{card}E_{\ell_n}=\frac{1}{\ell_n}H_{\nu_{\ell_n}}(\xi_{\ell_n})\le\frac{1}{m}H_{\mu_{\ell_n}}(\xi_m)+\frac{2m}{\ell_n}\log\operatorname{card}\xi.$$

On the other hand, by (4.4), letting $n\to\infty$ we obtain

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log N(d_n,\varepsilon)&=\limsup_{n\to\infty}\frac{1}{n}\log\operatorname{card}E_n\\&=\limsup_{n\to\infty}\frac{1}{\ell_n}\log\operatorname{card}E_{\ell_n}\\&\le\frac{1}{m}\limsup_{n\to\infty}H_{\mu_{\ell_n}}(\xi_m).\end{aligned}\tag{4.7}$$

We will use the next lemma to show that in fact the limit in the last line of (4.7) converges to $H_\mu(\xi_m)$. It is here that we use the requirement that $\mu(\partial B_i)=0$ for $i=1,\dots,k$.

Lemma 4.3. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of measures converging to μ and let ξ be a measurable partition such that $\mu(\partial A)=0$ for $A\in\xi$. Then*

$$\limsup_{n\to\infty}H_{\mu_n}(\xi)=\lim_{n\to\infty}H_{\mu_n}(\xi)=H_\mu(\xi).$$

Proof of the lemma. Let A be a measurable set in ξ. Since $(\mu_n)_{n\in\mathbb{N}}$ converges to μ, if $\varphi_k:X\to\mathbb{R}_0^+$ is a sequence of continuous functions decreasing to $\chi_{\overline{A}}$ when $k\to\infty$ (such a sequence exists since $\overline{A}$ is closed), then

$$\begin{aligned}\limsup_{n\to\infty}\mu_n(\overline{A})&\le\limsup_{n\to\infty}\int_X\varphi_k\,d\mu_n\\&=\int_X\varphi_k\,d\mu\to\mu(\overline{A})\end{aligned}$$

when $k\to\infty$, by the Monotone convergence theorem[1]. Therefore, since $\mu(\partial A)=0$, we obtain

$$\limsup_{n\to\infty}\mu_n(A)\le\limsup_{n\to\infty}\mu_n(\overline{A})\le\mu(\overline{A})=\mu(A).\tag{4.8}$$

[1]The Monotone convergence theorem says that if $\varphi_n:X\to\mathbb{R}_0^+$ is a nondecreasing sequence of measurable functions, then $\int_X\lim_{n\to\infty}\varphi_n\,d\mu=\lim_{n\to\infty}\int_X\varphi_n\,d\mu$.

Similarly, since $\partial(X\backslash A) = \partial A$, we also have

$$\limsup_{n\to\infty} \mu_n(X\backslash A) \leq \mu(X\backslash A),$$

that is,

$$\liminf_{n\to\infty} \mu_n(A) \geq \mu(A).$$

Together with (4.8), this implies that

$$\lim_{n\to\infty} \mu_n(A) = \mu(A)$$

and hence, using the continuity of the function ψ in (1.1), we obtain

$$\begin{aligned}\lim_{n\to\infty} H_{\mu_n}(\xi) &= \lim_{n\to\infty} \sum_{A\in\xi} \psi(\mu_n(A)) \\ &= \sum_{A\in\xi} \psi\left(\lim_{n\to\infty} \mu_n(A)\right) \\ &= \sum_{A\in\xi} \psi(\mu(A)) = H_\mu(\xi).\end{aligned}$$

This establishes the desired equality. □

In order to complete the proof, we observe that since $\mu(\partial C) = 0$ for $C \in \xi$, we have

$$\mu(\partial C) = 0 \quad \text{for } C \in \xi_n.$$

Therefore, using Lemma 4.3 in (4.7), we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log N(d_n, \varepsilon) \leq \frac{1}{m} H_\mu(\xi_m).$$

Letting $m \to \infty$, we finally obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log N(d_n, \varepsilon) \leq h_\mu(T, \xi) \leq h_\mu(T).$$

Therefore,

$$\limsup_{n\to\infty} \frac{1}{n} \log N(d_n, \varepsilon) \leq \sup_\nu h_\nu(T),$$

with the supremum taken over all T-invariant probability measures ν on X. This inequality holds for all $\varepsilon > 0$. Letting $\varepsilon \to 0$ yields the inequality

$$h(T) \leq \sup_{\nu} h_{\nu}(T).$$

This completes the proof of the theorem. $\square$

Appendix A

Space of Measures

In this appendix we describe the construction of a distance in $\mathcal{M}(X)$, the space of all Borel probability measures on a compact metric space X, and we show that with this distance the space $\mathcal{M}(X)$ is compact.

Let X be a compact metric space. We denote by $C(X)$ the space of all continuous functions $\varphi\colon X \to \mathbb{R}$. We consider the usual supremum norm

$$\|\varphi\|_\infty = \max\{|\varphi(x)| : x \in X\} \tag{A.1}$$

in $C(X)$. In order to define a distance in $\mathcal{M}(X)$, we first recall (without proof) an important result concerning the space $C(X)$.

Proposition A.1. *If X is a compact metric space, then equipped with the norm in (A.1), the space $C(X)$ is separable, that is, it contains a countable dense set.*

Now we define our candidate for the distance in $\mathcal{M}(X)$. Proposition A.1 guarantees the existence of a sequence of continuous functions $(\varphi_n)_{n\in\mathbb{N}}$ with closure equal to the ball

$$\{\varphi \in C(X) : \|\varphi\|_\infty \leq 1\}.$$

We define a function $d\colon \mathcal{M}(X) \times \mathcal{M}(X) \to \mathbb{R}_0^+$ by

$$d(\mu,\nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left| \int_X \varphi_n \, d\mu - \int_X \varphi_n \, d\nu \right|. \tag{A.2}$$

One can easily verify that the series in (A.2) converges. Indeed,

$$\begin{aligned} \left| \int_X \varphi_n \, d\mu - \int_X \varphi_n \, d\nu \right| &\leq \left| \int_X \varphi_n \, d\mu \right| + \left| \int_X \varphi_n \, d\nu \right| \\ &\leq \int_X |\varphi_n| \, d\mu + \int_X |\varphi_n| \, d\nu \\ &\leq 2\|\varphi_n\|_\infty \leq 2 \end{aligned}$$

for $\mu, \nu \in \mathcal{M}(X)$ and $n \in \mathbb{N}$. Therefore,

$$d(\mu,\nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left| \int_X \varphi_n \, d\mu - \int_X \varphi_n \, d\nu \right| \leq \sum_{n=1}^{\infty} \frac{2}{2^n} = 2$$

for $\mu, \nu \in \mathcal{M}(X)$ and so d is well defined.

Proposition A.2. *If X is a compact metric space, then d is a distance in $\mathcal{M}(X)$. Moreover, given a sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}(X)$ and $\mu \in \mathcal{M}(X)$, we have $d(\mu_n, \mu) \to 0$ when $n \to \infty$ if and only if*

$$\lim_{n\to\infty} \int_X \varphi \, d\mu_n = \int_X \varphi \, d\mu$$

for every function $\varphi \in C(X)$.

Proof. We first show that d is a distance in $\mathcal{M}(X)$. Note that $d(\mu,\nu) = d(\nu,\mu)$ and $d(\mu,\nu) \geq 0$ for any measures μ and ν. The triangular inequality is also easily verified. Hence, it remains to show that $\mu = \nu$ whenever $d(\mu,\nu) = 0$. If $d(\mu,\nu) = 0$, then

$$\int_X \varphi_m \, d\mu = \int_X \varphi_m \, d\nu$$

for $m \in \mathbb{N}$. Since the sequence $(\varphi_m)_{m\in\mathbb{N}}$ is dense in the closed unit ball $\{\varphi \in C(X) : \|\varphi\|_\infty \leq 1\}$, given a nonzero function $\varphi \in C(X)$ and $\varepsilon > 0$, there exists $m \in \mathbb{N}$ such that

$$\left\| \frac{\varphi}{\|\varphi\|_\infty} - \varphi_m \right\|_\infty < \varepsilon. \tag{A.3}$$

Hence,

$$\int_X \varphi_m \, d\mu - \varepsilon \leq \frac{1}{\|\varphi\|_\infty} \int_X \varphi \, d\mu \leq \int_X \varphi_m \, d\mu + \varepsilon$$

and

$$-\int_X \varphi_m \, d\nu - \varepsilon \leq -\frac{1}{\|\varphi\|_\infty} \int_X \varphi \, d\nu \leq -\int_X \varphi_m \, d\nu + \varepsilon.$$

Adding these inequalities, we obtain

$$\frac{1}{\|\varphi\|_\infty} \left| \int_X \varphi \, d\mu - \int_X \varphi \, d\nu \right| \leq \left| \int_X \varphi_m \, d\mu - \int_X \varphi_m \, d\nu \right| + 2\varepsilon = 2\varepsilon.$$

Since ε is arbitrary, we finally conclude that

$$\int_X \varphi \, d\mu = \int_X \varphi \, d\nu$$

for every $\varphi \in C(X)$. Now let F be any closed subset of X and consider a sequence of continuous functions $\varphi_n \colon X \to \mathbb{R}_0^+$ decreasing to χ_F when $n \to \infty$. Using the Monotone convergence theorem we obtain

$$\begin{aligned} \mu(F) = \int_X \chi_F \, d\mu &= \lim_{n\to\infty} \int_X \varphi_n \, d\mu \\ &= \lim_{n\to\infty} \int_X \varphi_n \, d\nu = \lim_{n\to\infty} \int_X \varphi_n \, d\nu = \nu(F). \end{aligned}$$

Hence, μ and ν coincide when evaluated on closed subsets. Given any measurable set A and $\varepsilon > 0$, there exists a closed set $F \subset A$ such that

$$\mu(A \backslash F) < \varepsilon \quad \text{and} \quad \nu(A \backslash F) < \varepsilon.$$

Therefore,

$$\begin{aligned} |\mu(A) - \nu(A)| &= |\mu(F) + \mu(A \backslash F) - \nu(F) - \nu(A \backslash F)| \\ &= |\mu(A \backslash F) - \nu(A \backslash F)| \leq 2\varepsilon, \end{aligned}$$

Since ε is arbitrary, we have $\mu(A) = \nu(A)$ for every measurable set A and so $\mu = \nu$.

Now we prove the second property. First we assume that $d(\mu_n, \mu) \to 0$ when $n \to \infty$. Then

$$\lim_{n\to\infty} \int_X \varphi_m \, d\mu_n = \int_X \varphi_m \, d\mu$$

for $m \in \mathbb{N}$. Again, given a nonzero function $\varphi \in C(X)$ and $\varepsilon > 0$, there exists $m \in \mathbb{N}$ such that inequality (A.3) holds. Therefore, following the same steps as before, for each $n \in \mathbb{N}$, we have

$$\frac{1}{\|\varphi\|_\infty}\left|\int_X \varphi\, d\mu_n - \int_X \varphi\, d\mu\right| \leq \left|\int_X \varphi_m\, d\mu_n - \int_X \varphi_m\, d\mu\right| + 2\varepsilon.$$

Letting $n \to \infty$, we obtain

$$\frac{1}{\|\varphi\|_\infty}\limsup_{n\to\infty}\left|\int_X \varphi\, d\mu_n - \int_X \varphi\, d\mu\right| \leq 2\varepsilon$$

and since ε is arbitrary, we conclude that

$$\lim_{n\to\infty}\left|\int_X \varphi\, d\mu_n - \int_X \varphi\, d\mu\right| = 0,$$

which yields the identity

$$\lim_{n\to\infty}\int_X \varphi\, d\mu_n = \int_X \varphi\, d\mu. \tag{A.4}$$

Finally we show that if (A.4) holds for any $\varphi \in C(X)$ then $d(\mu_n, \mu) \to 0$ when $n \to \infty$. Given $\varepsilon > 0$, let $m \in \mathbb{N}$ such that $1/2^{m-1} < \varepsilon$. There exists $k_m \in \mathbb{N}$ such that if $n > k_m$, then

$$\left|\int_X \varphi_i\, d\mu_n - \int_X \varphi_i\, d\mu\right| < \varepsilon$$

for $i = 1, \ldots, m$. Therefore,

$$\begin{aligned} d(\mu_n, \mu) &< \varepsilon + \sum_{i=m+1}^{\infty} \frac{1}{2^i}\left|\int_X \varphi_i\, d\mu_n - \int_X \varphi_i\, d\mu\right| \\ &\leq \varepsilon + \sum_{i=m+1}^{\infty} \frac{2}{2^i} \\ &= \varepsilon + \frac{1}{2^{m-1}} \leq 2\varepsilon \end{aligned}$$

for every $n > k_m$. Letting $\varepsilon \to 0$, we conclude that $d(\mu_n, \mu) \to 0$ when $n \to \infty$. This completes the proof of the proposition. □

Now we show that with the distance d in (A.2), the space $\mathcal{M}(X)$ is compact.

Theorem A.3. *If X is a compact metric space, then $\mathcal{M}(X)$ is compact.*

Proof. Since $\mathcal{M}(X)$ is a metric space, it is sufficient to show that each sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}(X)$ has a convergent subsequence. Let $(\varphi_n)_{n\in\mathbb{N}}$ be again the sequence used to define the distance d in (A.2). For each $n \in \mathbb{N}$, we consider the bounded sequence $\beta_n \colon \mathbb{N} \to [-1, 1]$ defined by

$$\beta_n(k) = \int_X \varphi_k \, d\mu_n.$$

By Tychonoff's theorem[1], the space of sequences $[-1, 1]^{\mathbb{N}}$ with the product topology is compact, and there exists a subsequence $(\beta_{m_n})_{n\in\mathbb{N}}$ converging in this topology. This implies that the sequence $(\int_X \varphi_k \, d\mu_{m_n})_{n\in\mathbb{N}}$ converges for each $k \in \mathbb{N}$. On the other hand, given a nonzero function $\varphi \in C(X)$ and $\varepsilon > 0$, there exists $k \in \mathbb{N}$ such that

$$\left\| \frac{\varphi}{\|\varphi\|_\infty} - \varphi_k \right\|_\infty < \varepsilon.$$

Therefore,

$$\frac{1}{\|\varphi\|_\infty} \left| \int_X \varphi \, d\mu_{m_n} - \int_X \varphi \, d\mu_{m_\ell} \right| \leq \left| \int_X \varphi_k \, d\mu_{m_n} - \int_X \varphi_k \, d\mu_{m_\ell} \right| + 2\varepsilon$$

for $n, \ell \in \mathbb{N}$. From these observations we conclude that if n and ℓ are sufficiently large, then

$$\left| \int_X \varphi \, d\mu_{m_n} - \int_X \varphi \, d\mu_{m_\ell} \right| \leq 3\varepsilon \|\varphi\|_\infty,$$

and so $(\int_X \varphi \, d\mu_{m_n})_{n\in\mathbb{N}}$ is a Cauchy sequence. Hence, one can define a linear functional $J \colon C(X) \to \mathbb{R}$ by

$$J(\varphi) = \lim_{n\to\infty} \int_X \varphi \, d\mu_{m_n}. \tag{A.5}$$

Clearly,

$$|J(\varphi)| \leq \|\varphi\|_\infty$$

[1]Tychonoff's theorem say that any product of compact topological spaces is compact with the product topology.

and so J is continuous. Moreover, J is positive, that is, $J(\varphi) \geq 0$ whenever $\varphi \geq 0$. By Riesz's representation theorem[2], there exists a unique Borel measure μ such that

$$J(\varphi) = \int_X \varphi \, d\mu$$

for every function $\varphi \in C(X)$. Moreover, since $J(1) = 1$ we have

$$\int_X 1 \, d\mu = \mu(X) = 1$$

and so $\mu \in \mathcal{M}(X)$. By Proposition A.2 and (A.5), we find that $d(\mu_{m_n}, \mu) \to 0$ when $n \to \infty$. This shows that the space $\mathcal{M}(X)$ is compact. $\square$

[2]Riesz's representation theorem says that if X is a compact metric space and $J\colon C(X) \to \mathbb{R}$ is a positive continuous linear functional, then there exists exactly one measure μ in the Borel σ-algebra of X such that $J(\varphi) = \int_X \varphi \, d\mu$ for every $\varphi \in C(X)$.

Bibliography

[1] R. Adler, A. Konheim and M. McAndrew, *Topological entropy*, Transactions of the American Mathematical Society **114** (1965), 309–319.

[2] L. Barreira, *Ergodic Theory, Hyperbolic Dynamics and Dimension Theory*, Springer, 2012.

[3] L. Barreira and C. Valls, *Dynamical Systems: An Introduction*, Springer, 2012.

[4] R. Bowen, *Entropy for group endomorphisms and homogeneous spaces*, Transactions of the American Mathematical Society **153** (1971), 401–414.

[5] E. Dinaburg, *A correlation between topological entropy and metric entropy*, Doklady Akademii Nauk SSSR **190** (1970), 19–22.

[6] T. Goodman, *Relating topological entropy and measure entropy*, Bulletin of the London Mathematical Society **3** (1971), 176–180.

[7] L. Goodwyn, *Topological entropy bounds measure-theoretic entropy*, Proceedings of the American Mathematical Society **23** (1969), 679–688.

[8] A. Katok, *Fifty years of entropy in dynamics: 1958-2007*, Journal of Modern Dynamics **1** (2007), 545–596.

[9] A. Kolmogorov, *A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces*, Doklady Akademii Nauk SSSR **119** (1958), 861–864.

[10] M. Misiurewicz, *A short proof of the variational principle for a $\mathbb{Z}_+^n$ action on a compact space*, Astérisque **40** (1976), 147–157.

[11] C. Shannon, *A mathematical theory of communication*, Bell Systems Technical Journal **27** (1948), 379–423, 623–656.

[12] Ya. Sinai, *On the concept of entropy for a dynamic system*, Doklady Akademii Nauk SSSR **124** (1959), 768–771.