

## Variational principle

In 1965, the IBM researchers R. Adler, A. Konheim and M. McAndrew proposed [AKM65] a notion of *topological entropy*, inspired by the Kolmogorov–Sinai entropy that we studied in the previous chapter, but whose definition does not involve any invariant measure. This notion applies to any continuous transformation in a compact topological space.

Subsequently, Efim Dinaburg [Din70] and Rufus Bowen [Bow71, Bow75a] gave a different, yet equivalent, definition for continuous transformations in compact metric spaces. Despite being a bit more restrictive, the Bowen–Dinaburg definition has the advantage of making more transparent the meaning of this concept: the topological entropy is the rate of exponential growth of the number of orbits that can be distinguished within a certain precision, arbitrarily small. Moreover, Bowen extended the definition to non-compact spaces, which is also very useful for applications.

These definitions of topological entropy and their properties are studied in Section 10.1 where, in particular, we observe that the topological entropy is an invariant of topological equivalence (topological conjugacy). In Section 10.2 we analyze several concrete examples.

The main result is the following remarkable relation between the topological entropy and the entropies of the transformation with respect to its invariant measures:

**Theorem 10.1** (Variational principle). *If  $f : M \rightarrow M$  is a continuous transformation in a compact metric space then its topological entropy  $h(f)$  coincides with the supremum of the entropies  $h_\mu(f)$  of  $f$  with respect to all the invariant probability measures.*

This theorem was proved by Dinaburg [Din70, Din71], Goodman [Goo71a] and Goodwin [Goo71b]. Here, it arises as a special case of a more general statement, the variational principle for the pressure, which is due to Walters [Wal75].

The *pressure*  $P(f, \phi)$  is a weighted version of the topological entropy  $h(f)$ , where the “weights” are determined by a continuous function  $\phi : M \rightarrow \mathbb{R}$ ,

which we call a *potential*. We study these notions and their properties in Section 10.3. The topological entropy corresponds to the special case when the potential is identically zero. The notion of pressure was brought from statistical mechanics to ergodic theory by the Belgium mathematician and theoretical physicist David Ruelle, one of the founders of differentiable ergodic theory, and was then extended by the British mathematician Peter Walters.

The variational principle (Theorem 10.1) extends to the setting of the pressure, as we are going to see in Section 10.4:

$$P(f, \phi) = \sup \left\{ h_\mu(f) + \int \phi d\mu : \mu \text{ is invariant under } f \right\} \quad (10.0.1)$$

for every continuous function  $\phi : M \rightarrow \mathbb{R}$ . An invariant probability measure  $\mu$  is called an *equilibrium state* for the potential  $\phi$  if it realizes the supremum in (10.0.1), that is, if  $h_\mu(f) + \int \phi d\mu = P(f, \phi)$ . The set of all equilibrium states is studied in Section 10.5.

## 10.1 Topological entropy

Initially, we present the definitions of Adler–Konheim–McAndrew and Bowen–Dinaburg and we prove that they are equivalent when the ambient is a compact metric space.

### 10.1.1 Definition via open covers

The original definition of the topological entropy is very similar to that of the Kolmogorov–Sinai entropy, with open covers in the place of partitions into measurable sets.

Let  $M$  be a compact topological space. An *open cover* of  $M$  is any family  $\alpha$  of open sets whose union is the whole of  $M$ . By compactness, every open cover admits a *subcover* (that is, a subfamily that is still an open cover) with finitely many elements. We call the *entropy* of the open cover  $\alpha$  the number

$$H(\alpha) = \log N(\alpha), \quad (10.1.1)$$

where  $N(\alpha)$  is the smallest number such that  $\alpha$  admits some finite subcover with that number of elements.

Given two open covers  $\alpha$  and  $\beta$ , we say that  $\alpha$  is *coarser* than  $\beta$  (or  $\beta$  is finer than  $\alpha$ ), and we write  $\alpha < \beta$ , if every element of  $\beta$  is contained in some element of  $\alpha$ . For example, if  $\beta$  is a subcover of  $\alpha$  then  $\alpha < \beta$ . By Exercise 10.1.1,

$$\alpha < \beta \Rightarrow H(\alpha) \leq H(\beta). \quad (10.1.2)$$

Given open covers  $\alpha_1, \dots, \alpha_n$ , we denote by  $\alpha_1 \vee \dots \vee \alpha_n$  their *sum*, that is, the open cover whose elements are the intersections  $A_1 \cap \dots \cap A_n$  with  $A_j \in \alpha_j$  for each  $j$ . Note that  $\alpha_j < \alpha_1 \vee \dots \vee \alpha_n$  for every  $j$ .

Let  $f : M \rightarrow M$  be a continuous transformation. If  $\alpha$  is an open cover of  $M$  then so is  $f^{-j}(\alpha) = \{f^{-j}(A) : A \in \alpha\}$ . For each  $n \geq 1$ , let us denote

$$\alpha^n = \alpha \vee f^{-1}(\alpha) \vee \dots \vee f^{-n+1}(\alpha).$$

Using Exercise 10.1.2, we see that

$$H(\alpha^{m+n}) = H(\alpha^m \vee f^{-m}(\alpha^n)) \leq H(\alpha^m) + H(f^{-m}(\alpha^n)) \leq H(\alpha^m) + H(\alpha^n)$$

for every  $m, n \geq 1$ . In other words, the sequence  $H(\alpha^n)$  is subadditive. Consequently (Lemma 3.3.4),

$$h(f, \alpha) = \lim_n \frac{1}{n} H(\alpha^n) = \inf_n \frac{1}{n} H(\alpha^n) \tag{10.1.3}$$

always exists and is finite. It is called the entropy of  $f$  with respect to the open cover  $\alpha$ . The relation (10.1.2) implies that

$$\alpha < \beta \implies h(f, \alpha) \leq h(f, \beta). \tag{10.1.4}$$

Finally, we define the *topological entropy* of  $f$  to be

$$h(f) = \sup\{h(f, \alpha) : \alpha \text{ is an open cover of } M\}. \tag{10.1.5}$$

In particular, if  $\beta$  is a subcover of  $\alpha$  then  $h(f, \alpha) \leq h(f, \beta)$ . Therefore, the definition (10.1.5) does not change when one restricts the supremum to the finite open covers.

Observe that the entropy  $h(f)$  is a non-negative number, possibly infinite (see Exercise 10.1.6).

**Example 10.1.1.** Let  $f : S^1 \rightarrow S^1$  be any homeomorphism (for example, a rotation  $R_\theta$ ) and let  $\alpha$  be an open cover of the circle formed by a finite number of open intervals. Let  $\partial\alpha$  be the set consisting of the endpoints of those intervals. For each  $n \geq 1$ , the open cover  $\alpha^n$  is formed by intervals whose endpoints are in

$$\partial\alpha^n = \partial\alpha \cup f^{-1}(\partial\alpha) \cup \dots \cup f^{-n+1}(\partial\alpha).$$

Note that  $\#\alpha^n \leq \#\partial\alpha^n \leq n\#\partial\alpha$ . Therefore,

$$h(f, \alpha) = \lim_n \frac{1}{n} H(\alpha^n) \leq \liminf_n \frac{1}{n} \log \#\alpha^n \leq \liminf_n \frac{1}{n} \log n = 0.$$

Proposition 10.1.12 below gives that  $h(f) = \lim_k h(f, \alpha_k)$  for any sequence of open covers  $\alpha_k$  with  $\text{diam } \alpha_k \rightarrow 0$ . Then, considering open covers  $\alpha_k$  by intervals of length less than  $1/k$ , we conclude from the previous calculation that  $h(f) = 0$  for every homeomorphism of the circle.

**Example 10.1.2.** Let  $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$  and  $\alpha$  be the cover of  $\Sigma$  by the cylinders  $[0; a]$ ,  $a = 1, \dots, d$ . Consider the shift map  $\sigma : \Sigma \rightarrow \Sigma$ . For each  $n$ , the open cover  $\alpha^n$  consists of the cylinders of length  $n$ :

$$\alpha^n = \{[0; a_0, \dots, a_{n-1}] : a_j = 1, \dots, d\}.$$

Therefore,  $H(\alpha^n) = \log \# \alpha^n = \log d^n$  and, consequently,  $h(f, \alpha) = \log d$ . Observe also that  $\text{diam} \alpha^n$  converges to zero when  $n \rightarrow \infty$ , relative to the distance defined by (A.2.7). Then, it follows from Corollary 10.1.13 below that  $h(f) = h(f, \alpha) = \log d$ . The same holds for the two-sided shift  $\sigma : \Sigma \rightarrow \Sigma$  in  $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$ .

Now we show that the topological entropy is an invariant of topological equivalence. Let  $f : M \rightarrow M$  and  $g : N \rightarrow N$  be continuous transformations in compact topological spaces  $M$  and  $N$ . We say that  $g$  is a *topological factor* of  $f$  if there exists a surjective continuous map  $\theta : M \rightarrow N$  such that  $\theta \circ f = g \circ \theta$ . When  $\theta$  may be chosen to be invertible (a homeomorphism), we say that the two transformations are *topologically equivalent*, or *topologically conjugate*, and we call  $\theta$  a *topological conjugacy* between  $f$  and  $g$ .

**Proposition 10.1.3.** *If  $g$  is a topological factor of  $f$  then  $h(g) \leq h(f)$ . In particular, if  $f$  and  $g$  are topologically equivalent then  $h(f) = h(g)$ .*

*Proof.* Let  $\theta : M \rightarrow N$  be a surjective continuous map such that  $\theta \circ f = g \circ \theta$ . Given any open cover  $\alpha$  of  $N$ , the family

$$\theta^{-1}(\alpha) = \{\theta^{-1}(A) : A \in \alpha\}$$

is an open cover of  $M$ . Recall that, by definition, the iterated sum  $\alpha^n$  is the open cover formed by the sets  $\bigcap_{j=0}^{n-1} g^{-j}(A_j)$  with  $A_0, A_1, \dots, A_{n-1} \in \alpha$ . Analogously, the iterated sum  $\theta^{-1}(\alpha)^n$  consists of the sets  $\bigcap_{j=0}^{n-1} f^{-j}(\theta^{-1}(A_j))$ . Clearly,

$$\bigcap_{j=0}^{n-1} f^{-j}(\theta^{-1}(A_j)) = \bigcap_{j=0}^{n-1} \theta^{-1}(g^{-j}(A_j)) = \theta^{-1}\left(\bigcap_{j=0}^{n-1} g^{-j}(A_j)\right).$$

Noting that the sets of the form on the right-hand side of this identity constitute the pre-image  $\theta^{-1}(\alpha^n)$  of  $\alpha^n$ , we conclude that  $\theta^{-1}(\alpha^n) = \theta^{-1}(\alpha)^n$ . Since  $\theta$  is surjective, a family  $\gamma \subset \alpha^n$  covers  $N$  if and only if  $\theta^{-1}(\gamma)$  covers  $M$ . Therefore,

$$H(\theta^{-1}(\alpha)^n) = H(\theta^{-1}(\alpha^n)) = H(\alpha^n).$$

Since  $n$  is arbitrary, it follows that  $h(f, \theta^{-1}(\alpha)) = h(g, \alpha)$ . Then, taking the supremum over all the open covers  $\alpha$  of  $N$ :

$$h(g) = \sup_{\alpha} h(g, \alpha) = \sup_{\alpha} h(f, \theta^{-1}(\alpha)) \leq h(f).$$

This proves the first part of the proposition. The second part is an immediate consequence, since in that case  $f$  is also a factor of  $g$ .

The converse to Proposition 10.1.3 is false, in general. For example, all the homeomorphisms of the circle have topological entropy equal to zero (recall Example 10.1.1) but they are not necessarily topologically equivalent (for example, the identity is not topologically equivalent to any other homeomorphism).

10.1.2 Generating sets and separated sets

Next, we present the definition of topological entropy of Bowen–Dinaburg. Let  $f : M \rightarrow M$  be a continuous transformation in a metric space  $M$ , not necessarily compact, and let  $K \subset M$  be any compact subset. When  $M$  is compact it suffices to consider  $K = M$ , as observed in (10.1.12) below.

Given  $\varepsilon > 0$  and  $n \in \mathbb{N}$ , we say that a set  $E \subset M$   $(n, \varepsilon)$ -generates  $K$  if for every  $x \in K$  there exists  $a \in E$  such that  $d(f^i(x), f^i(a)) < \varepsilon$  for every  $i \in \{0, \dots, n - 1\}$ . In other words,

$$K \subset \bigcup_{a \in E} B(a, n, \varepsilon),$$

where  $B(a, n, \varepsilon) = \{x \in M : d(f^i(x), f^i(a)) < \varepsilon \text{ for } i = 0, \dots, n - 1\}$  is the dynamical ball of center  $a$ , length  $n$  and radius  $\varepsilon$ . Note that  $\{B(x, n, \varepsilon) : x \in K\}$  is an open cover of  $K$ . Hence, by compactness, there always exist finite  $(n, \varepsilon)$ -generating sets.

Let us denote by  $g_n(f, \varepsilon, K)$  the smallest cardinality of an  $(n, \varepsilon)$ -generating set of  $K$ . We define

$$g(f, \varepsilon, K) = \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, K). \tag{10.1.6}$$

Observe that the function  $\varepsilon \mapsto g(f, \varepsilon, K)$  is monotone non-increasing. Indeed, it is clear from the definition that if  $\varepsilon_1 < \varepsilon_2$  then every  $(n, \varepsilon_1)$ -generating set is also  $(n, \varepsilon_2)$ -generating. Therefore,  $g_n(f, \varepsilon_1, K) \geq g_n(f, \varepsilon_2, K)$  for every  $n \geq 1$  and, taking the limit,  $g(f, \varepsilon_1, K) \geq g(f, \varepsilon_2, K)$ . This ensures, in particular, that

$$g(f, K) = \lim_{\varepsilon \rightarrow 0} g(f, \varepsilon, K) \tag{10.1.7}$$

exists. Finally, we define

$$g(f) = \sup\{g(f, K) : K \subset M \text{ compact}\}. \tag{10.1.8}$$

We also introduce the following dual notion. Given  $\varepsilon > 0$  and  $n \in \mathbb{N}$ , we say that a set  $E \subset K$  is  $(n, \varepsilon)$ -separated if, given  $x, y \in E$ , there exists  $j \in \{0, \dots, n - 1\}$  such that  $d(f^j(x), f^j(y)) \geq \varepsilon$ . In other words, if  $x \in E$  then  $B(x, n, \varepsilon)$  contains no other point of  $E$ . We denote by  $s_n(f, \varepsilon, K)$  the largest cardinality of an  $(n, \varepsilon)$ -separated set. We define

$$s(f, \varepsilon, K) = \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, K). \tag{10.1.9}$$

It is clear that if  $0 < \varepsilon_1 < \varepsilon_2$ , then every  $(n, \varepsilon_2)$ -separated set is also  $(n, \varepsilon_1)$ -separated. Therefore,  $s_n(f, \varepsilon_1, K) \geq s_n(f, \varepsilon_2, K)$  for every  $n \geq 1$  and, taking the limit,  $s(f, \varepsilon_1, K) \geq s(f, \varepsilon_2, K)$ . In particular,

$$s(f, K) = \lim_{\varepsilon \rightarrow 0} s(f, \varepsilon, K) \tag{10.1.10}$$

always exists. Finally, we define

$$s(f) = \sup\{s(f, K) : K \subset M \text{ compact}\}. \tag{10.1.11}$$

It is clear that  $g(f, K_1) \leq g(f, K_2)$  and  $s(f, K_1) \leq s(f, K_2)$  if  $K_1 \subset K_2$ . In particular,

$$g(f) = g(f, M) \quad \text{and} \quad s(f) = s(f, M) \quad \text{if } M \text{ is compact.} \tag{10.1.12}$$

Another interesting observation (Exercise 10.1.7) is that the definitions (10.1.8) and (10.1.11) are not affected when we restrict the supremum to compact sets with small diameter.

**Proposition 10.1.4.** *We have  $g(f, K) = s(f, K)$  for every compact  $K \subset M$ . Consequently,  $g(f) = s(f)$ .*

*Proof.* For the proof we need the following lemma:

**Lemma 10.1.5.**  *$g_n(f, \varepsilon, K) \leq s_n(f, \varepsilon, K) \leq g_n(f, \varepsilon/2, K)$  for every  $n \geq 1$ , every  $\varepsilon > 0$  and every compact  $K \subset M$ .*

*Proof.* Let  $E \subset K$  be an  $(n, \varepsilon)$ -separated set with maximal cardinality. Given any  $y \in K \setminus E$ , the set  $E \cup \{y\}$  is not  $(n, \varepsilon)$ -separated, and so there exists  $x \in E$  such that  $d(f^i(x), f^i(y)) < \varepsilon$  for every  $i \in \{0, \dots, n-1\}$ . This shows that  $E$  is an  $(n, \varepsilon)$ -generating set of  $K$ . Consequently,  $g_n(f, \varepsilon, K) \leq \#E = s_n(f, \varepsilon, K)$ .

To prove the other inequality, let  $E \subset K$  be an  $(n, \varepsilon)$ -separated set and  $F \subset M$  be an  $(n, \varepsilon/2)$ -generating set of  $K$ . The hypothesis ensures that, given any  $x \in E$  there exists some  $y \in F$  such that  $d(f^i(x), f^i(y)) < \varepsilon/2$  for every  $i \in \{0, \dots, n-1\}$ . Let  $\phi : E \rightarrow F$  be a map such that each  $\phi(x)$  is a point  $y$  satisfying this condition. We claim that the map  $\phi$  is injective. Indeed, suppose that  $x, z \in E$  are such that  $\phi(x) = y = \phi(z)$ . Then

$$d(f^i(x), f^i(z)) \leq d(f^i(x), f^i(y)) + d(f^i(y), f^i(z)) < \varepsilon/2 + \varepsilon/2$$

for every  $i \in \{0, \dots, n-1\}$ . Since  $E$  is  $(n, \varepsilon)$ -separated, this implies that  $x = z$ . Therefore,  $\phi$  is injective, as we claimed. It follows that  $\#E \leq \#F$  and, since  $E$  and  $F$  are arbitrary, this proves that  $s_n(f, \varepsilon, K) \leq g_n(f, \varepsilon/2, K)$ .

Then, given any  $\varepsilon > 0$  and any compact  $K \subset M$ ,

$$\begin{aligned} g(f, \varepsilon, K) &= \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, K) \\ &\leq \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, K) = s(f, \varepsilon, K) \\ &\leq \limsup_n \frac{1}{n} \log g_n(f, \frac{\varepsilon}{2}, K) = g(f, \frac{\varepsilon}{2}, K). \end{aligned} \tag{10.1.13}$$

Taking the limit when  $\varepsilon \rightarrow 0$ , we get that

$$\begin{aligned} g(f, K) &= \lim_{\varepsilon \rightarrow 0} g(f, \varepsilon, K) \leq \lim_{\varepsilon \rightarrow 0} s(f, \varepsilon, K) = s(f, K) \\ &\leq \lim_{\varepsilon \rightarrow 0} g(f, \frac{\varepsilon}{2}, K) = g(f, K). \end{aligned}$$

This proves the first part of the proposition. The second part is an immediate consequence.

By definition, the *diameter* of an open cover  $\alpha$  of a metric space  $M$  is the supremum of the diameters of all the sets  $A \in \alpha$ .

**Proposition 10.1.6.** *If  $M$  is a compact metric space then  $h(f) = g(f) = s(f)$ .*

*Proof.* By Proposition 10.1.4, it suffices to show that  $s(f) \leq h(f) \leq g(f)$ .

Start by fixing  $\varepsilon > 0$  and  $n \geq 1$ . Let  $E \subset M$  be an  $(n, \varepsilon)$ -separated set and  $\alpha$  be any open cover of  $M$  with diameter less than  $\varepsilon$ . If  $x$  and  $y$  are in the same element of  $\alpha^n$  then

$$d(f^i(x), f^i(y)) \leq \text{diam } \alpha < \varepsilon \quad \text{for every } i = 0, \dots, n - 1.$$

In particular, each element of  $\alpha^n$  contains at most one element of  $E$ . Consequently,  $\#E \leq N(\alpha^n)$ . Taking  $E$  with maximal cardinality, we conclude that  $s_n(f, \varepsilon, M) \leq N(\alpha^n)$  for every  $n \geq 1$ . So,

$$\begin{aligned} s(f, \varepsilon, M) &= \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, M) \\ &\leq \lim_n \frac{1}{n} \log N(\alpha^n) = h(f, \alpha) \leq h(f). \end{aligned} \tag{10.1.14}$$

Making  $\varepsilon \rightarrow 0$ , we find that  $s(f) = s(f, M) \leq h(f)$ .

Next, given any open cover  $\alpha$  of  $M$ , let  $\varepsilon > 0$  be a Lebesgue number for  $\alpha$ , that is, a positive number such that every ball of radius  $\varepsilon$  is contained in some element of  $\alpha$ . Let  $E \subset M$  be an  $(n, \varepsilon)$ -generating set of  $M$  with minimal cardinality. For each  $x \in E$  and  $i = 0, \dots, n - 1$ , there exists  $A_{x,i} \in \alpha$  such that  $B(f^i(x), \varepsilon)$  is contained in  $A_{x,i}$ . Then,

$$B(x, n, \varepsilon) \subset \bigcap_{i=0}^{n-1} f^{-i}(A_{x,i}).$$

Therefore, the hypothesis that  $E$  is a generating set implies that the family  $\gamma = \{\bigcap_{i=0}^{n-1} f^{-i}(A_{x,i}) : x \in E\}$  is an open cover of  $M$ . Since  $\gamma \subset \alpha^n$ , it follows that  $N(\alpha^n) \leq \#E = g_n(f, \varepsilon, M)$  for every  $n$ . Therefore,

$$\begin{aligned} h(f, \alpha) &= \lim_n \frac{1}{n} \log N(\alpha^n) \leq \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M) \\ &\leq \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, M) = g(f, \varepsilon, M). \end{aligned} \tag{10.1.15}$$

Making  $\varepsilon \rightarrow 0$ , we get that  $h(f, \alpha) \leq g(f, M) = g(f)$ . Since the open cover  $\alpha$  is arbitrary, it follows that  $h(f) \leq g(f)$ .

We define the *topological entropy* of a continuous transformation  $f : M \rightarrow M$  in a metric space  $M$  to be  $g(f) = s(f)$ . Proposition 10.1.6 shows that this definition is compatible with the one we gave in Section 10.1.1

for transformations in compact topological spaces. A relevant difference is that, while for compact spaces the topological entropy depends only on the topology (because  $h(f)$  is defined solely in terms of the open sets), in the non-compact case the topological entropy may also depend on the distance function in  $M$ . In this regard, see Exercises 10.1.4 and 10.1.5. They also show that in the non-compact case the topological entropy is no longer an invariant of topological conjugacy, although it remains an invariant of *uniformly* continuous conjugacy.

**Example 10.1.7.** Assume that  $f : M \rightarrow M$  does not expand distances, that is, that  $d(f(x), f(y)) \leq d(x, y)$  for every  $x, y \in M$ . Then the topological entropy of  $f$  is equal to zero. Indeed, the hypothesis implies that  $B(x, n, \varepsilon) = B(x, \varepsilon)$  for every  $n \geq 1$ . Hence, a set  $E$  is  $(n, \varepsilon)$ -generating if and only if it is  $(1, \varepsilon)$ -generating. In particular, the sequence  $g_n(f, \varepsilon, K)$  does not depend on  $n$  and, hence,  $g(f, \varepsilon, K) = 0$  for every  $\varepsilon > 0$  and every compact set  $K$ . Making  $\varepsilon \rightarrow 0$  and taking the supremum over  $K$  we get that  $g(f) = 0$  (analogously,  $s(f) = 0$ ).

There are two important special cases: *contractions*, such that there exists  $\lambda < 1$  satisfying  $d(f(x), f(y)) \leq \lambda d(x, y)$  for every  $x, y \in M$ ; and *isometries*, such that  $d(f(x), f(y)) = d(x, y)$  for every  $x, y \in M$ . We saw in Lemma 6.3.6 that every compact metrizable group admits a distance relative to which every translation is an isometry. Therefore, it also follows from the previous observations that the topological entropy of every translation in a compact metrizable group is zero.

Recalling that  $g(f) = g(f, M)$  and  $s(f) = s(f, M)$  when  $M$  is compact, we see that the conclusion of Proposition 10.1.6 may be rewritten as follows:

$$h(f) = \lim_{\varepsilon \rightarrow 0} \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, M) = \lim_{\varepsilon \rightarrow 0} \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, M).$$

From the proof of the proposition we may also obtain the following related identity:

**Corollary 10.1.8.** *If  $f : M \rightarrow M$  is a continuous transformation in a compact metric space then*

$$h(f) = \lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M) = \lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M).$$

*Proof.* The relation (10.1.15) gives that

$$h(f, \alpha) \leq \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M)$$

whenever  $\varepsilon > 0$  is a Lebesgue number for the open cover  $\alpha$ . Making  $\varepsilon \rightarrow 0$ , we conclude that

$$h(f) \leq \lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M). \tag{10.1.16}$$



The first inequality in Lemma 10.1.5 implies that

$$\lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M) \leq \lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M). \tag{10.1.17}$$

Also, it is clear that

$$\lim_{\varepsilon \rightarrow 0} \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M) \leq \lim_{\varepsilon \rightarrow 0} \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, M). \tag{10.1.18}$$

As we have just observed, the expression on the right-hand side is equal to  $h(f)$ . Therefore, the inequalities (10.1.16)–(10.1.18) imply the conclusion.

### 10.1.3 Calculation and properties

We start by proving a version of Lemma 9.1.13 for the topological entropy. The proof is a bit more elaborate because, unlike what happens for partitions, given an open cover  $\alpha$  the covers  $(\alpha^k)^n$  and  $\alpha^{n+k-1}$  need not coincide if the elements of  $\alpha$  are not pairwise disjoint.

**Example 10.1.9.** Let  $f : M \rightarrow M$  be the shift map in  $M = \{1, 2, 3\}^{\mathbb{N}}$  (or  $M = \{1, 2, 3\}^{\mathbb{Z}}$ ) and  $\alpha$  be the open cover of  $M$  consisting of the cylinders  $[0; \{1, 2\}]$  and  $[0; \{1, 3\}]$ . For each  $n \geq 1$ , the cover  $\alpha^n$  consists of the  $2^n$  cylinders of the form  $[0; A_0, \dots, A_{n-1}]$  with  $A_j = [0; \{1, 2\}]$  or  $A_j = [0; \{1, 3\}]$ . In particular,  $\#\alpha^3 = 8$ . On the other hand,  $(\alpha^2)^2$  contains 12 elements: the 8 elements of  $\alpha^3$  together with the 4 cylinders of the form  $[0; A_0, \{1\}, A_2]$  with  $A_j = [0; \{1, 2\}]$  or  $A_j = [0; \{1, 3\}]$  for  $j = 0$  and  $j = 2$ . Hence,  $\alpha^{n+k-1} \neq (\alpha^k)^n$  for  $n = k = 2$ .

**Proposition 10.1.10.** *Let  $M$  be a compact topological space,  $f : M \rightarrow M$  be a continuous transformation and  $\alpha$  be an open cover of  $M$ . Then  $h(f, \alpha) = h(f, \alpha^k)$  for every  $k \geq 1$ . Moreover, if  $f : M \rightarrow M$  is a homeomorphism then  $h(f, \alpha) = h(f, \alpha^{\pm k})$  for every  $k \geq 1$ , where  $\alpha^{\pm k} = \bigvee_{j=-k}^{k-1} f^{-j}(\alpha)$ .*

*Proof.* The main point is to show that the open covers  $(\alpha^k)^n$  and  $\alpha^{n+k-1}$  have the same entropy, for every  $n \geq 1$ . We use the following simple fact, which will be useful again later:

**Lemma 10.1.11.** *Given any open cover  $\alpha$  and any  $n, k \geq 1$ ,*

1.  $\alpha^{n+k-1}$  is a subcover of  $(\alpha^k)^n$  and, in particular,  $(\alpha^k)^n \prec \alpha^{n+k-1}$ ;
2. for any subcover  $\beta$  of  $(\alpha^k)^n$  there exists a subcover  $\gamma$  of  $\alpha^{n+k-1}$  such that  $\#\gamma \leq \#\beta$  and  $\gamma \prec \beta$ .

*Proof.* By definition, every element  $\alpha^{n+k-1}$  has the form  $B = \bigcap_{l=0}^{n+k-2} f^{-l}(B_l)$  with  $B_l \in \alpha$  for every  $l$ . It is clear that this may be written in the form  $B = \bigcap_{i=0}^{n-1} f^{-i}(\bigcap_{j=0}^{k-1} f^{-j}(B_{i+j}))$  and, thus,  $B \in (\alpha^k)^n$ . This proves the first claim. Next, let  $\beta$  be a subcover of  $(\alpha^k)^n$ . Every element of  $\beta$  has the form

$$A = \bigcap_{i=0}^{n-1} f^{-i} \left( \bigcap_{j=0}^{k-1} f^{-j}(A_{i,j}) \right) = \bigcap_{l=0}^{n+k-2} f^{-l} \left( \bigcap_{i+j=l} A_{i,j} \right),$$

with  $A_{i,j} \in \alpha$ . Consider  $B = \bigcap_{l=0}^{n+k-2} f^{-l}(B_l)$ , where  $B_l = A_{i,j}$  for some pair  $(i,j)$  such that  $i+j=l$ . Observe that  $A \subset B$  and  $B \in \alpha^{n+k-1}$ . Therefore, the family  $\gamma$  formed by all the sets  $B$  obtained in this way satisfies all the conditions in the second claim.

According to the relation (10.1.2), the first part of Lemma 10.1.11 implies that  $H((\alpha^k)^n) \leq H(\alpha^{n+k-1})$ . Clearly, the second part of the lemma implies the opposite inequality. Hence,

$$H(\alpha^{n+k-1}) = H((\alpha^k)^n) \quad \text{for any } n, k \geq 1, \tag{10.1.19}$$

as we claimed. Therefore,

$$h(f, \alpha^k) = \lim_n \frac{1}{n} H((\alpha^k)^n) = \lim_n \frac{1}{n} H(\alpha^{n+k-1}) = h(f, \alpha) \quad \text{for every } k.$$

When  $f$  is invertible, it follows from the definitions that  $\alpha^{\pm k} = f^k(\alpha^{2k})$ . Using Exercise 10.1.3, we get that  $h(f, \alpha^{\pm k}) = h(f, f^k(\alpha^{2k})) = h(f, \alpha^{2k}) = h(f, \alpha)$ .

The next proposition and its corollary simplify the calculation of the topological entropy significantly in concrete examples. Recall that, when  $M$  is a metric space, the *diameter* of an open cover is defined to be the supremum of the diameters of its elements.

**Proposition 10.1.12.** *Assume that  $M$  is a compact metric space. Let  $(\beta_k)_k$  be any sequence of open covers of  $M$  such that  $\text{diam } \beta_k$  converges to zero. Then*

$$h(f) = \sup_k h(f, \beta_k) = \lim_k h(f, \beta_k).$$

*Proof.* Given any open cover  $\alpha$ , let  $\varepsilon > 0$  be a Lebesgue number of  $\alpha$ . Take  $n \geq 1$  such that  $\text{diam } \beta_k < \varepsilon$  for every  $k \geq n$ . By the definition of Lebesgue number, it follows that every element of  $\beta_k$  is contained in some element of  $\alpha$ . In other words,  $\alpha \prec \beta_k$  and, hence,  $h(f, \beta_k) \geq h(f, \alpha)$ . In view of the definition (10.1.5), this proves that

$$\liminf_k h(f, \beta_k) \geq h(f).$$

It is also clear from the definitions that  $h(f) \geq \sup_k h(f, \beta_k) \geq \limsup_k h(f, \beta_k)$ . Combining these observations, we obtain the conclusion of the proposition.

**Corollary 10.1.13.** *Assume that  $M$  is a compact metric space. If  $\beta$  is an open cover such that*

- (1) *the diameter of the one-sided iterated sum  $\beta^k = \bigvee_{j=0}^{k-1} f^{-j}(\beta)$  converges to zero when  $k \rightarrow \infty$ , or*
- (2)  *$f : M \rightarrow M$  is a homeomorphism and the diameter of the two-sided iterated sum  $\beta^{\pm k} = \bigvee_{j=-k}^{k-1} f^{-j}(\beta)$  converges to zero when  $k \rightarrow \infty$ ,*

*then  $h(f) = h(f, \beta)$ .*

*Proof.* In case (1), Propositions 10.1.10 and 10.1.12 yield

$$h(f) = \lim_k h(f, \beta^k) = h(f, \beta).$$

The proof in case (2) is analogous.

Next, we check that the topological entropy behaves as one could expect with respect to positive iterates, at least when the transformation is uniformly continuous:

**Proposition 10.1.14.** *If  $f : M \rightarrow M$  is a uniformly continuous transformation in a metric space then  $h(f^k) = kh(f)$  for every  $k \in \mathbb{N}$ .*

*Proof.* Fix  $k \geq 1$  and let  $K \subset M$  be any compact set. Consider any  $n \geq 1$  and  $\varepsilon > 0$ . It is clear that if  $E \subset M$  is an  $(nk, \varepsilon)$ -generating set of  $K$  for the transformation  $f$  then it is also an  $(n, \varepsilon)$ -generating set of  $K$  for the iterate  $f^k$ . Therefore,  $g_n(f^k, \varepsilon, K) \leq g_{nk}(f, \varepsilon, K)$ . Hence,

$$g(f^k, \varepsilon, K) = \lim_n \frac{1}{n} g_n(f^k, \varepsilon, K) \leq \lim_n \frac{1}{n} g_{nk}(f, \varepsilon, K) = kg(f, \varepsilon, K).$$

Making  $\varepsilon \rightarrow 0$  and taking the supremum over  $K$ , we see that  $h(f^k) \leq kh(f)$ .

The proof of the other inequality uses the assumption that  $f$  is uniformly continuous. Take  $\delta > 0$  such that  $d(x, y) < \delta$  implies  $d(f^j(x), f^j(y)) < \varepsilon$  for every  $j \in \{0, \dots, k-1\}$ . If  $E \subset M$  is an  $(n, \delta)$ -generating set of  $K$  for  $f^k$  then  $E$  is an  $(nk, \varepsilon)$ -generating set of  $K$  for  $f$ . Therefore,  $g_{nk}(f, \varepsilon, K) \leq g_n(f^k, \delta, K)$ . This shows that  $kg(f, \varepsilon, K) \leq g(f^k, \delta, K)$ . Making  $\varepsilon$  and  $\delta$  go to zero, we get that  $kg(f, K) \leq g(f^k, K)$  for every compact set  $K$ . Hence,  $kh(f) \leq h(f^k)$ .

In particular, Proposition 10.1.14 holds for every continuous transformation in a compact metric space. On the other hand, in the case of homeomorphisms in compact spaces the conclusion extends to negative iterates:

**Proposition 10.1.15.** *If  $f : M \rightarrow M$  is a homeomorphism of a compact metric space then  $h(f^{-1}) = h(f)$ . Consequently,  $h(f^n) = |n|h(f)$  for every  $n \in \mathbb{Z}$ .*

*Proof.* Let  $\alpha$  be an open cover of  $M$ . For every  $n \geq 1$ , denote

$$\alpha_+^n = \alpha \vee f^{-1}(\alpha) \vee \dots \vee f^{-n+1}(\alpha) \quad \text{and} \quad \alpha_-^n = \alpha \vee f(\alpha) \vee \dots \vee f^{n-1}(\alpha).$$

Observe that  $\alpha_-^n = f^{n-1}(\alpha_+^n)$ . Moreover,  $\gamma$  is a finite subcover of  $\alpha_+^n$  if and only if  $f^{n-1}(\gamma)$  is a finite subcover of  $\alpha_-^n$ . Since the two subcovers have the same number of elements, it follows that  $H(\alpha_+^n) = H(\alpha_-^n)$ . Therefore,

$$h(f, \alpha) = \lim_n \frac{1}{n} H(\alpha_+^n) = \lim_n \frac{1}{n} H(\alpha_-^n) = h(f^{-1}, \alpha).$$

Since  $\alpha$  is arbitrary, this proves that  $h(f) = h(f^{-1})$ . The second part of the statement follows from combining the first part with Proposition 10.1.14.

The claim in Proposition 10.1.15 is generally false when the space  $M$  is not compact:

**Example 10.1.16.** Let  $M = \mathbb{R}$  with the distance  $d(x, y) = |x - y|$  and take  $f : \mathbb{R} \rightarrow \mathbb{R}$  to be given by  $f(x) = 2x$ . We are going to check that  $h(f) \neq h(f^{-1})$ . Let  $K = [0, 1]$  and, given  $n \geq 1$  and  $\varepsilon > 0$ , take  $E \subset \mathbb{R}$  to be any  $(n, \varepsilon)$ -generating set of  $K$ . In particular, every point of  $f^{n-1}(K) = [0, 2^{n-1}]$  is within less than  $\varepsilon$  from some point of  $f^{n-1}(E)$ . Hence,

$$2\varepsilon \#E = 2\varepsilon \#f^{n-1}(E) \geq 2^{n-1}.$$

This proves that  $g_n(f, \varepsilon, K) \geq 2^{n-2}/\varepsilon$  for every  $n$  and, thus,  $g(f, \varepsilon, K) \geq \log 2$ . It follows that  $h(f) \geq g(f, K) \geq \log 2$ . On the other hand,  $f^{-1}$  is a contraction and so it follows from Example 10.1.7 that its topological entropy  $h(f^{-1})$  is zero.

### 10.1.4 Exercises

- 10.1.1. Let  $M$  be a compact topological space. Show that if  $\alpha$  and  $\beta$  are open covers of  $M$  such that  $\alpha < \beta$  then  $H(\alpha) \leq H(\beta)$ .
- 10.1.2. Let  $f : M \rightarrow M$  be a continuous transformation and  $\alpha, \beta$  be open covers of a compact topological space  $M$ . Show that  $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$  and  $H(f^{-1}(\beta)) \leq H(\beta)$ . Check that if  $f$  is surjective then  $H(f^{-1}(\beta)) = H(\beta)$ .
- 10.1.3. Let  $M$  be a compact topological space. Show that if  $f : M \rightarrow M$  is a surjective continuous transformation and  $\beta$  is an open cover of  $M$  then  $h(f, \beta) = h(f, f^{-1}(\beta))$ . Moreover, if  $f$  is a homeomorphism then  $h(f, \beta) = h(f, f(\beta))$ .
- 10.1.4. Let  $M = (0, \infty)$  and  $f : M \rightarrow M$  be given by  $f(x) = 2x$ . Calculate the topological entropy of  $f$  when one considers in  $M$ :
- the usual distance  $d(x, y) = |x - y|$ ;
  - the distance  $d(x, y) = |\log x - \log y|$ .
- [Observation: Hence, in non-compact spaces the topological entropy may depend on the distance function, not just the topology.]
- 10.1.5. Consider in  $M$  two distances  $d_1$  and  $d_2$  that are uniformly equivalent: for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that

$$d_1(x, y) < \delta \Rightarrow d_2(x, y) < \varepsilon \quad \text{and} \quad d_2(x, y) < \delta \Rightarrow d_1(x, y) < \varepsilon.$$

Show that if  $f : M \rightarrow M$  is continuous with respect to either of the two distances then the value of the topological entropy is the same relative to both distances.

- 10.1.6. Let  $f : M \rightarrow M$  and  $g : N \rightarrow N$  be continuous transformations in compact metric spaces. Show that if there exists a continuous injective map  $\psi : M \rightarrow N$  such that  $\psi \circ f = g \circ \psi$  then  $h(f) \leq h(g)$ . Use this fact to show that the topological entropy of the shift map  $\sigma : [0, 1]^{\mathbb{Z}} \rightarrow [0, 1]^{\mathbb{Z}}$  is infinite (thus, the topological entropy of a homeomorphism of a compact space need not be finite). [Observation: The first claim remains valid for non-compact spaces, as long as we require the inverse  $\psi^{-1} : \psi(M) \rightarrow M$  to be uniformly continuous.]

10.1.7. Show that if  $K, K_1, \dots, K_l$  are compact sets such that  $K$  is contained in  $K_1 \cup \dots \cup K_l$  then  $g(f, K) \leq \max_i g(f, K_i)$ . Conclude that, given any  $\delta > 0$ ,

$$g(f) = \sup\{g(f, K) : K \text{ compact with } \text{diam} K < \delta\}$$

and analogously for  $s(f)$ .

10.1.8. Prove that the *logistic* map  $f : [0, 1] \rightarrow [0, 1], f(x) = 4x(1 - x)$  is topologically conjugate to the map  $g : [0, 1] \rightarrow [0, 1]$  defined by  $g(x) = 1 - |2x - 1|$ . Use this fact to calculate  $h(f)$ .

10.1.9. Let  $\mathcal{A}$  be a finite alphabet and  $\sigma : \Sigma \rightarrow \Sigma$  be the shift map in  $\Sigma = \mathcal{A}^{\mathbb{N}}$ . The *complexity* of a sequence  $x \in \Sigma$  is defined by  $c(x) = \lim_n n^{-1} \log c_n(x)$ , where  $c_n(x)$  is the number of distinct words of length  $n$  that appear in  $x$ . Show that this limit exists and coincides with the topological entropy of the restriction  $\sigma : \mathcal{X} \rightarrow \mathcal{X}$  of the shift map to the closure  $\mathcal{X}$  of the orbit of  $x$ . [Observation: One interesting application we have in mind is in the context of Example 6.3.10, where  $x$  is the fixed point of a substitution.]

10.1.10. Check that if  $\theta$  is the fixed point of the Fibonacci substitution in  $\mathcal{A} = \{0, 1\}$  (see Example 6.3.10) then  $c_n(\theta) = n + 1$  for every  $n$  and so the complexity  $c(\theta)$  is equal to zero. Hence, the topological entropy of the shift map  $\sigma : \mathcal{X} \rightarrow \mathcal{X}$  associated with the Fibonacci substitution is equal to zero.

## 10.2 Examples

Let us use a few concrete situations to illustrate the ideas introduced in the previous section.

### 10.2.1 Expansive maps

Recall (Section 9.2.3) that a continuous transformation  $f : M \rightarrow M$  in a compact metric space is said to be *expansive* if there exists  $\varepsilon_0 > 0$  such that  $d(f^j(x), f^j(y)) < \varepsilon_0$  for every  $j \in \mathbb{N}$  implies that  $x = y$ . When  $f : M \rightarrow M$  is invertible, we say that it is *two-sided expansive* if there exists  $\varepsilon_0 > 0$  such that  $d(f^j(x), f^j(y)) < \varepsilon_0$  for every  $j \in \mathbb{Z}$  implies that  $x = y$ . In both cases,  $\varepsilon_0$  is called a *constant of expansivity* for  $f$ .

**Proposition 10.2.1.** *If  $\varepsilon_0 > 0$  is a constant of expansivity for  $f$  then*

- (i)  $h(f) = h(f, \alpha)$  for every open cover  $\alpha$  with diameter less than  $\varepsilon_0$ ;
- (ii)  $h(f) = g(f, \varepsilon, M) = s(f, \varepsilon, M)$  for every  $\varepsilon < \varepsilon_0/2$ .

*In particular,  $h(f) < \infty$ .*

*Proof.* Let  $\alpha$  be any open cover of  $M$  with diameter less than  $\varepsilon_0$ . We claim that  $\lim_k \text{diam } \alpha^k = 0$ . Indeed, suppose that this is not so. It is clear that the sequence of diameters is non-increasing. Then, there exists  $\delta > 0$  and for each  $k \geq 1$  there exist points  $x_k$  and  $y_k$  in the same element of  $\alpha^k$  such that  $d(x_k, y_k) \geq \delta$ .

By compactness, we may find a subsequence  $(k_j)_j$  such that both  $x = \lim_j x_{k_j}$  and  $y = \lim_j y_{k_j}$  exist. On the one hand,  $d(x, y) \geq \delta$  and so  $x \neq y$ . On the other hand, the fact that  $x_k$  and  $y_k$  are in the same element of  $\alpha^k$  implies that

$$d(f^i(x_k), f^i(y_k)) \leq \text{diam } \alpha \quad \text{for every } 0 \leq i < k.$$

Passing to the limit, we get that  $d(f^i(x), f^i(y)) \leq \text{diam } \alpha < \varepsilon_0$  for every  $i \geq 0$ . This contradicts the hypothesis that  $\varepsilon_0$  is a constant of expansivity for  $f$ . This contradiction proves our claim. Using Corollary 10.1.13, it follows that  $h(f) = h(f, \alpha)$ , as claimed in part (i).

To prove part (ii), let  $\alpha$  be the open cover of  $M$  formed by the balls of radius  $\varepsilon$ . Note that  $\alpha^n$  contains every dynamical ball  $B(x, n, \varepsilon)$ :

$$B(x, n, \varepsilon) = \bigcap_{j=0}^{n-1} f^{-j}(B(f^j(x), \varepsilon)) \quad \text{and each } B(f^j(x), \varepsilon) \in \alpha.$$

If  $E$  is an  $(n, \varepsilon)$ -generating set of  $M$  then  $\{B(a, n, \varepsilon) : a \in E\}$  is an open cover of  $M$ ; in view of what we have just said, it is a subcover of  $\alpha^n$ . Therefore (recall also Lemma 10.1.5),

$$N(\alpha^n) \leq g_n(f, \varepsilon, M) \leq s_n(f, \varepsilon, M) \quad \text{for every } n.$$

Passing to the limit, we get that  $h(f, \alpha) \leq g(f, \varepsilon, M) \leq s(f, \varepsilon, M)$ . Recall that  $s(f, \varepsilon, M) \leq s(f, M) = h(f)$ . Since  $\text{diam } \alpha < \varepsilon_0$ , the first part of the proposition yields that  $h(f) = h(f, \alpha)$ . These relations imply part (ii).

The last claim in the proposition is a direct consequence, since  $g(f, \varepsilon, M)$ ,  $s(f, \varepsilon, M)$  and  $h(f, \alpha)$  are always finite. Indeed, that  $h(f, \alpha) < \infty$  for every open cover was observed right after the definition (10.1.3). Then (10.1.14) implies that  $s(f, \varepsilon, M) < \infty$  and (10.1.13) implies that  $g(f, \varepsilon, M) < \infty$  for every  $\varepsilon > 0$ .

Exercise 10.2.8 contains an extension of Proposition 10.2.1 to *h-expansive* transformations, due to Rufus Bowen [Bow72]. Exercise 10.1.6 shows that the topological entropy of a continuous transformation, or even a homeomorphism, in a compact metric space may be infinite, if one omits the expansivity assumption.

Next, we prove that for expansive maps the topological entropy is an upper bound on the rate of growth of the number of periodic points. Let  $\text{Fix}(f^n)$  denote the set of all points  $x \in M$  such that  $f^n(x) = x$ .

**Proposition 10.2.2.** *If  $M$  is a compact metric space and  $f : M \rightarrow M$  is expansive then*

$$\limsup_n \frac{1}{n} \log \# \text{Fix}(f^n) \leq h(f).$$

*Proof.* Let  $\varepsilon_0$  be a constant of expansivity for  $f$  and  $\alpha$  be any open cover of  $M$  with  $\text{diam } \alpha < \varepsilon_0$ . We claim that every element of  $\alpha^n$  contains at most one point of  $\text{Fix}(f^n)$ . Indeed, if  $x, y \in \text{Fix}(f^n)$  are in the same element of  $\alpha^n$  then

$d(f^i(x), f^i(y)) < \text{diam } \alpha < \varepsilon_0$  for every  $i = 0, \dots, n - 1$ . Since  $f^n(x) = x$  and  $f^n(y) = y$ , it follows that  $d(f^i(x), f^i(y)) < \varepsilon_0$  for every  $i \geq 0$ . By expansivity, this implies that  $x = y$ , which proves our claim. It follows that

$$\limsup_n \frac{1}{n} \log \# \text{Fix}(f^n) \leq \limsup_n \frac{1}{n} \log N(\alpha^n) = h(f, \alpha).$$

Taking the limit when the diameter of  $\alpha$  goes to zero, we get the conclusion of the proposition.

In some interesting situations, one can show that the topological entropy actually coincides with the rate of growth of the number of periodic points:

$$\lim_n \frac{1}{n} \log \# \text{Fix}(f^n) = h(f). \tag{10.2.1}$$

That is the case, for example, for the shifts of finite type, which we are going to study in Section 10.2.2 (check Proposition 10.2.5 below). More generally, (10.2.1) holds whenever  $f : M \rightarrow M$  is an expanding transformation in a compact metric space, as we are going to see in Section 11.3.

### 10.2.2 Shifts of finite type

Let  $X = \{1, \dots, d\}$  be a finite set and  $A = (A_{i,j})_{i,j}$  be a *transition matrix*, that is, a square matrix of dimension  $d \geq 2$  with coefficients in the set  $\{0, 1\}$  and such that no row is identically zero: for every  $i$  there exists  $j$  such that  $A_{i,j} = 1$ . Consider the subset  $\Sigma_A$  of  $\Sigma = X^{\mathbb{N}}$  consisting of all the sequences  $(x_n)_n \in \Sigma$  that are *A-admissible*, meaning that

$$A_{x_n, x_{n+1}} = 1 \quad \text{for every } n \in \mathbb{N}. \tag{10.2.2}$$

It is clear that  $\Sigma_A$  is invariant under the shift map  $\sigma : \Sigma \rightarrow \Sigma$ , in the sense that  $\sigma(\Sigma_A) \subset \Sigma_A$ . Note also that  $\Sigma_A$  is closed in  $\Sigma$  and, hence, it is a compact metric space (this is similar to Lemma 7.2.5).

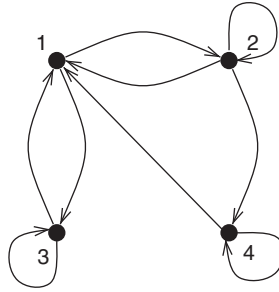
The restriction  $\sigma_A : \Sigma_A \rightarrow \Sigma_A$  of the shift map  $\sigma : \Sigma \rightarrow \Sigma$  to this invariant compact set is called the *one-sided shift of finite type* associated with  $A$ . The two-sided shift of finite type associated with a transition matrix  $A$  is defined analogously, considering  $\Sigma = X^{\mathbb{Z}}$  and requiring (10.2.2) for every  $n \in \mathbb{Z}$ . In this case, as part of the definition of a transition matrix, we also require the columns (not just the rows) of  $A$  to be non-zero.

The restriction of the shift map  $\sigma : \Sigma \rightarrow \Sigma$  to the support of any Markov measure is a shift of finite type:

**Example 10.2.3.** Given a stochastic matrix  $P = (P_{i,j})_{i,j}$ , define  $A = (A_{i,j})_{i,j}$  by

$$A_{i,j} = \begin{cases} 1 & \text{if } P_{i,j} > 0 \\ 0 & \text{if } P_{i,j} = 0. \end{cases}$$

Note that  $A$  is a transition matrix: the definition of a stochastic matrix implies that no row  $P$  is identically zero (in the two-sided situation we must assume



**Figure 10.1.** Graph associated with a transition matrix

that the columns of  $P$  are also not zero; this is automatic, for example, if the matrix  $P$  is aperiodic). Comparing (7.2.7) and (10.2.2), we see that a sequence is  $A$ -admissible if and only if it is  $P$ -admissible. Let  $\mu$  be the Markov measure determined by a probability vector  $p = (p_j)_j$  with positive coefficients and such that  $P^*p = p$  (recall Example 7.2.2). By Lemma 7.2.5, the support of  $\mu$  coincides with the set  $\Sigma_A = \Sigma_P$  of all admissible sequences.

It is useful to associate with any transition matrix  $A$  the oriented graph whose vertices are the points of  $X = \{1, \dots, d\}$  and such that there exists an edge from vertex  $a$  to vertex  $b$  if and only if  $A_{a,b} = 1$ . In other words,

$$G_A = \{(a, b) \in X \times X : A_{a,b} = 1\}.$$

For example, Figure 10.1 describes the graph associated with the matrix

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

A path of length  $l \geq 1$  in the graph  $G_A$  is a sequence  $a_0, \dots, a_l$  in  $X$  such that  $A_{a_{i-1}, a_i} = 1$  for every  $i$ , that is, such that there always exists an edge connecting  $a_{i-1}$  to  $a_i$ . Given  $a, b \in X$  and  $l \geq 1$ , denote by  $A_{a,b}^l$  the number of paths of length  $l$  starting at  $a$  and ending at  $b$ , that is, with  $a_0 = a$  and  $a_l = b$ . Observe that:

1.  $A_{a,b}^1 = 1$  if there exists an edge connecting  $a$  to  $b$  and  $A_{a,b}^1 = 0$  otherwise. In other words,  $A_{a,b}^1 = A_{a,b}$  for every  $a, b$ .
2. The paths of length  $l+m$  starting at  $a$  and ending at  $b$  are the concatenations of the paths of length  $l$  starting at  $a$  and ending at some point  $z \in X$  with the paths of length  $m$  starting at that point  $z$  and ending at  $b$ . Therefore,

$$A_{a,b}^{l+m} = \sum_{z=1}^d A_{a,z}^l A_{z,b}^m \quad \text{for every } a, b \in X \text{ and every } l, m \geq 1.$$

It follows, by induction on  $l$ , that  $A_{a,b}^l$  coincides with the coefficient in row  $a$  and column  $b$  of the matrix  $A^l$ .



The basic topological properties of shifts of finite type are analyzed in Exercise 10.2.2. In the proposition that follows we calculate the topological entropy of these transformations. We need a few prior observations about transition matrices.

Recall that the *spectral radius*  $\rho(B)$  of a linear map  $B : \mathbb{R}^d \rightarrow \mathbb{R}^d$  (that is, the largest absolute value of an eigenvalue of  $B$ ) is given by

$$\rho(B) = \lim_n \|B^n\|^{1/n} = \lim_n |\text{trc } B^n|^{1/n}, \tag{10.2.3}$$

where  $\text{trc}$  denotes the *trace* of the matrix and  $\|\cdot\|$  denotes any norm in the vector space of linear maps (all norms are equivalent, as we are in finite dimension). Most of the time, one uses the *operator norm*  $\|B\| = \sup\{\|Bv\|/\|v\| : v \neq 0\}$ , but it will also be useful to consider the norm  $\|\cdot\|_s$  defined by

$$\|B\|_s = \sum_{i,j=1}^d |B_{i,j}|.$$

Now take  $A$  to be a transition matrix. Since the coefficients of  $A$  are non-negative, we may use the Perron–Frobenius theorem (Theorem 7.2.3) to conclude that  $A$  admits a non-negative eigenvalue  $\lambda_A$  that is equal to the spectral radius. By our definition of the transition matrix, we also have that all the rows of  $A$  are non-zero. Then the same is true about  $A^n$ , for any  $n \geq 1$  (Exercise 10.2.5). This implies that all the coefficients of the vector  $A^n(1, \dots, 1)$  are positive (and integer) and, thus,

$$\|A^n\| \geq \frac{\|A^n(1, \dots, 1)\|}{\|(1, \dots, 1)\|} \geq 1 \quad \text{for every } n \geq 1.$$

Using (10.2.3), we get that  $\lambda_A = \rho(A) \geq 1$  for every transition matrix  $A$ .

**Proposition 10.2.4.** *The topological entropy  $h(\sigma_A)$  of a shift of finite type  $\sigma_A : \Sigma_A \rightarrow \Sigma_A$  is given by  $h(\sigma_A) = \log \lambda_A$ , where  $\lambda_A$  is the largest eigenvalue of the transition matrix  $A$ .*

*Proof.* We treat the case of one-sided shifts; the two-sided case is analogous, as the reader may readily check. Consider the open cover  $\alpha$  of  $\Sigma_A$  formed by the restrictions

$$[0; a]_A = \{(x_j)_j \in \Sigma_A : x_0 = a\}$$

of the cylinders  $[0; a]$  of  $\Sigma$ . For each  $n \geq 1$ , the open cover  $\alpha^n$  is formed by the restrictions

$$[0; a_0, \dots, a_{n-1}]_A = \{(x_j)_j \in \Sigma_A : x_j = a_j \text{ for } j = 0, \dots, n-1\}$$

of the cylinders of length  $n$ . Observe that  $[0; a_0, \dots, a_{n-1}]_A$  is non-empty if and only if  $a_0, \dots, a_{n-1}$  is a path (of length  $n-1$ ) in the graph  $G_A$ : it is evident that this condition is necessary; to see that it is also sufficient, use the assumption that for every  $i$  there exists  $j$  such that  $A_{i,j} = 1$ . Since the cylinders are pairwise

disjoint, this observation shows that  $N(\alpha^n)$  is equal to the total number of paths of length  $n - 1$  in the graph  $G_A$ . In other words,

$$N(\alpha^n) = \sum_{i,j=1}^d A_{i,j}^{n-1} = \|A^{n-1}\|_s.$$

By the spectral radius formula (10.2.3), it follows that

$$h(\sigma_A, \alpha) = \lim_n \frac{1}{n} \log N(\alpha^n) = \lim_n \frac{1}{n} \log \|A^{n-1}\|_s = \log \rho(A) = \log \lambda_A.$$

Finally, since  $\text{diam } \alpha^n \rightarrow 0$ , Corollary 10.1.13 yields that  $h(\sigma_A) = h(\sigma_A, \alpha)$ .

**Proposition 10.2.5.** *If  $\sigma_A : \Sigma_A \rightarrow \Sigma_A$  is a shift of finite type then*

$$h(\sigma_A) = \lim_n \frac{1}{n} \log \# \text{Fix}(\sigma_A^n).$$

*Proof.* We treat the case of one-sided shifts, leaving the two-sided case for the reader. Note that  $(x_k)_k \in \Sigma_A$  is a fixed point of  $\sigma_A^n$  if and only if  $x_k = x_{k-n}$  for every  $k \geq n$ . In particular, every cylinder  $[0; a_0, \dots, a_{n-1}]_A$  contains at most one element of  $\text{Fix}(\sigma_A^n)$ . Moreover, the cylinder does contain a fixed point if and only if  $a_0, \dots, a_{n-1}, a_0$  is a path (of length  $n$ ) in the graph  $G_A$ . This proves that

$$\# \text{Fix}(\sigma_A^n) = \sum_{i=1}^d A_{i,i}^n = \text{trc } A^n$$

for every  $n$ . Consequently,

$$\lim_n \frac{1}{n} \log \# \text{Fix}(\sigma_A^n) = \lim_n \frac{1}{n} \log \text{trc } A^n = \log \rho(A).$$

Now the conclusion is a direct consequence of the previous proposition.

### 10.2.3 Topological entropy of flows

The definition of topological entropy extends easily to the context of continuous flows  $\phi = \{\phi^t : M \rightarrow M : t \in \mathbb{R}\}$  in a metric space  $M$ , as we now explain.

Given  $x \in M$  and  $T > 0$  and  $\varepsilon > 0$ , the *dynamical ball* of center  $x$ , length  $T$  and radius  $\varepsilon > 0$  is the set

$$B(x, T, \varepsilon) = \{y \in M : d(\phi^t(x), \phi^t(y)) < \varepsilon \text{ for every } 0 \leq t \leq T\}.$$

Let  $K$  be any compact subset of  $M$ . We say that  $E \subset M$  is a  $(T, \varepsilon)$ -generating set for  $K$  if

$$K \subset \bigcup_{x \in E} B(x, T, \varepsilon),$$

and we say that  $E \subset K$  is a  $(T, \varepsilon)$ -separated set if the dynamical ball  $B(x, T, \varepsilon)$  of each  $x \in E$  contains no other element of  $E$ .

Denote by  $g_T(\phi, \varepsilon, K)$  the smallest cardinality of a  $(T, \varepsilon)$ -generating set of  $K$  and by  $s_T(\phi, \varepsilon, K)$  the largest cardinality of a  $(T, \varepsilon)$ -separated set  $E \subset K$ . Then, take

$$g(\phi, K) = \lim_{\varepsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \frac{1}{T} \log g_T(\phi, \varepsilon, K) \quad \text{and}$$

$$s(\phi, K) = \lim_{\varepsilon \rightarrow 0} \limsup_{T \rightarrow \infty} \frac{1}{T} \log s_T(\phi, \varepsilon, K)$$

and define

$$g(\phi) = \sup_K g(\phi, K) \quad \text{and} \quad s(\phi) = \sup_K s(\phi, K),$$

where both suprema are taken over all the compact sets  $K \subset M$ .

The next result, a continuous-time analogue of Proposition 10.1.4, ensures that these two last numbers coincide. We leave the proof up to the reader (Exercise 10.2.3). By definition, the *topological entropy* of the flow  $\phi$  is the number  $h(\phi) = g(\phi) = s(\phi)$ .

**Proposition 10.2.6.** *We have  $g(\phi, K) = s(\phi, K)$  for every compact  $K \subset M$ . Consequently,  $g(\phi) = s(\phi)$ .*

In the statement that follows we take the flow to be *uniformly continuous*, that is, such that for every  $T > 0$  and  $\varepsilon > 0$  there exists  $\delta > 0$  such that

$$d(x, y) < \delta \quad \Rightarrow \quad d(\phi^t(x), \phi^t(y)) < \varepsilon \quad \text{for every } t \in [-T, T].$$

Observe that this is automatic for continuous flows when  $M$  is compact.

**Proposition 10.2.7.** *If the flow  $\phi$  is uniformly continuous then its topological entropy  $h(\phi)$  coincides with the topological entropy  $h(\phi^1)$  of its time-1 map.*

*Proof.* It suffices to prove that  $g(\phi, K) = g(\phi^1, K)$  for every compact  $K \subset M$ .

It is clear that if  $E \subset M$  is  $(T, \varepsilon)$ -generating for  $K$  relative to the flow  $\phi$  then  $E$  is also  $(n, \varepsilon)$ -generating for  $K$  relative to the time-1 map, for any  $n \leq T + 1$ . In particular,  $g_n(\phi^1, \varepsilon, K) \leq g_T(\phi, \varepsilon, K)$ . It follows that

$$\limsup_n \frac{1}{n} \log g_n(\phi^1, \varepsilon, K) \leq \limsup_{T \rightarrow \infty} \frac{1}{T} \log g_T(\phi, \varepsilon, K),$$

and so  $g(\phi^1, K) \leq g(\phi, K)$ .

The hypothesis of uniform continuity is used for the opposite inequality. Given  $\varepsilon > 0$ , fix  $\delta \in (0, \varepsilon)$  such that if  $d(x, y) < \delta$  then  $d(\phi^t(x), \phi^t(y)) < \varepsilon$  for every  $t \in [0, 1]$ . If  $E \subset M$  is an  $(n, \delta)$ -generating set of  $K$  relative to  $\phi^1$  then  $E$  is a  $(T, \varepsilon)$ -generating set of  $K$  relative to the flow  $\phi$ , for any  $T \leq n$ . In particular,  $g_T(\phi, \varepsilon, K) \leq g_n(\phi^1, \delta, K)$ . It follows that

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \log g_T(\phi, \varepsilon, K) \leq \limsup_n \frac{1}{n} \log g_n(\phi^1, \delta, K)$$

(given a sequence  $(T_j)_j$  that realizes the supremum on the left-hand side, consider the sequence  $(n_j)_j$  given by  $n_j = [T_j] + 1$ ). Making  $\varepsilon \rightarrow 0$  (then  $\delta \rightarrow 0$ ), we get that  $g(\phi, K) \leq g(\phi^1, K)$ .

We have seen previously that for transformations the topological entropy is an invariant of topological (uniformly continuous) conjugacy. The same is true for flows: this follows from Proposition 10.2.7 and the obvious observation that any flow conjugacy also conjugates the corresponding time-1 maps. However, in the continuous-time context, one more often uses the concept of *topological equivalence*, which allows for rescaling of time. Clearly, topological equivalence need not preserve the topological entropy.

### 10.2.4 Differentiable maps

In this section we take  $M$  to be a Riemannian manifold (Appendix A.4.5). Let  $f : M \rightarrow M$  be a differentiable map and  $Df(x) : T_x M \rightarrow T_{f(x)} M$  denote the derivative of  $f$  at each point  $x \in M$ . Our goal is to prove that the norm of the derivative, defined by

$$\|Df(x)\| = \sup \left\{ \frac{\|Df(x)v\|}{\|v\|} : v \in T_x M \text{ and } v \neq 0 \right\},$$

determines an upper bound for the topological entropy  $h(f)$  of  $f$ . For  $x > 0$ , we denote  $\log^+ x = \max\{\log x, 0\}$ .

**Proposition 10.2.8.** *Let  $f : M \rightarrow M$  be a differentiable map in a Riemannian manifold of dimension  $d$  such that  $\|Df\|$  is bounded. Then*

$$h(f) \leq d \log^+ \sup \|Df\| < \infty.$$

*Proof.* Let  $L = \sup\{\|Df(x)\| : x \in M\}$ . By the mean value theorem,

$$d(f(x), f(y)) \leq Ld(x, y) \quad \text{for every } x, y \in M.$$

If  $L \leq 1$  then, as we have seen in Example 10.1.7, the entropy of  $f$  is zero. Thus, from now on we may suppose that  $L > 1$ .

Let  $\mathcal{A}$  be an atlas of the manifold  $M$  consisting of charts  $\varphi_\alpha : U_\alpha \rightarrow X_\alpha$  with  $X_\alpha = (-2, 2)^d$ . Given any compact set  $K \subset M$ , we may find a finite family  $\mathcal{A}_K \subset \mathcal{A}$  such that

$$\{\varphi_\alpha^{-1}((-1, 1)^d) : \varphi_\alpha \in \mathcal{A}_K\}$$

covers  $K$ . Fix  $B > 0$  such that  $d(u, v) \leq Bd(\varphi_\alpha(u), \varphi_\alpha(v))$  for all  $u, v \in [-1, 1]^d$  and  $\varphi_\alpha \in \mathcal{A}_K$ . Given  $n \geq 1$  and  $\varepsilon > 0$ , fix  $\delta = (\varepsilon/B\sqrt{d})L^{-n}$ . Denote by  $\delta\mathbb{Z}^d$  the set of all points of the form  $(\delta k_1, \dots, \delta k_d)$  with  $k_j \in \mathbb{Z}$  for every  $j = 1, \dots, d$ . Let  $E \subset M$  be the union of the pre-images  $\varphi_\alpha^{-1}(\delta\mathbb{Z}^d \cap (-1, 1)^d)$ , with  $\varphi_\alpha \in \mathcal{A}_K$ .

Note that every point of  $(-1, 1)^d$  is at a distance less than  $\delta\sqrt{d}$  from some point of  $\delta\mathbb{Z}^d \cap (-1, 1)^d$ . Therefore, for any  $\varphi_\alpha \in \mathcal{A}_K$ , every  $x \in \varphi_\alpha^{-1}((-1, 1)^d)$

is at a distance less than  $B\delta\sqrt{d}$  from some point  $a \in \varphi(\delta\mathbb{Z}^d \cap (-1, 1)^d)$ . Then, by the choice of  $\delta$ ,

$$d(f^j(x), f^j(a)) \leq L^j B \delta \sqrt{d} < L^n B \delta \sqrt{d} = \varepsilon$$

for every  $j = 0, \dots, n - 1$ . This proves that  $E$  is an  $(n, \varepsilon)$ -generating set for  $K$ . On the other hand, by construction,

$$\#E \leq \#\mathcal{A}_K \#(\delta\mathbb{Z}^d \cap (-1, 1)^d) \leq \#\mathcal{A}_K (2/\delta)^d \leq \#\mathcal{A}_K (2B\sqrt{d}L^n/\varepsilon)^d,$$

so the expression on the right-hand side is an upper bound for  $g_n(f, \varepsilon, K)$ . Consequently,

$$g(f, \varepsilon, K) \leq \limsup_n \frac{1}{n} \log(2B\sqrt{d}L^n/\varepsilon)^d = d \log L.$$

Making  $\varepsilon \rightarrow 0$  and taking the supremum over  $K$ , we get that  $h(f) \leq d \log L$ .

Combining Propositions 10.1.14 and 10.2.8, we find that

$$h(f) \leq \frac{1}{n} \log^+ \sup \|Df^n\| \quad \text{for every } n \geq 1.$$

When  $f$  is a homeomorphism, using Proposition 10.1.15 we also get that

$$h(f) \leq \frac{1}{n} \log^+ \sup \|Df^{-n}\| \quad \text{for every } n \geq 1.$$

The following conjecture of Michael Shub [Shu74] is central to the theory of topological entropy:

**Conjecture 10.2.9** (Entropy conjecture). *If  $f : M \rightarrow M$  is a diffeomorphism of class  $C^1$  in a Riemannian manifold of dimension  $d$ , then*

$$h(f) \geq \max_{1 \leq k \leq d} \log \rho(f_k), \tag{10.2.4}$$

where each  $\rho(f_k)$  denotes the spectral radius of the action  $f_k : H_k(M) \rightarrow H_k(M)$  induced by  $f$  in the real homology of dimension  $k$ .

The full statement of the conjecture remains open to date, but several partial answers and related results have been obtained, both positive and negative. Let us summarize what is known in this regard.

It follows from a result of Yano [Yan80] that the inequality (10.2.4) is true for an open and dense subset of the space of homeomorphisms in any manifold of dimension  $d \geq 2$ . Moreover, it is true for every homeomorphism in certain classes of manifolds, such as the spheres or the infranilmanifolds [MP77b, MP77a, MP08]. On the other hand, Shub [Shu74] exhibited a Lipschitz homeomorphism, with zero topological entropy, for which (10.2.4) is false. See Exercise 10.2.7.

A useful way to approach (10.2.4) is by comparing the topological entropy with each one of the spectral radii  $\rho(f_k)$ . The case  $k = d$  is relatively easy. Indeed, for any continuous map  $f$  in a manifold of dimension  $d$ , the spectral

radius  $\rho(f_d)$  is equal to the absolute value  $|\deg f|$  of the degree of the map. In particular, the inequality  $h(f) \geq \log \rho(f_d)$  is trivial for any homeomorphism. For non-invertible continuous maps, the topological entropy may be less than the logarithm of the absolute value of the degree. However, it was shown in [MP77b] that for differentiable maps one always has  $h(f) \geq \log |\deg f|$ .

Anthony Manning [Man75] proved that the inequality  $h(f) \geq \log \rho(f_1)$  is true for every homeomorphism in a manifold of any dimension  $d$ . It follows that  $h(f) \geq \log \rho(f_{d-1})$ , since the duality theorem of Poincaré implies that

$$\rho(f_k) = \rho(f_{d-k}) \quad \text{for every } 0 < k < d.$$

In particular, the theorem of Manning together with the observations in the previous paragraph prove that entropy conjecture is true for every homeomorphisms in any manifold of dimension  $d \leq 3$ .

Rufus Bowen [Bow78] proved that for any homeomorphism in a manifold the topological entropy  $h(f)$  is greater than or equal to the logarithm of the rate of growth of the fundamental group. One can show that this rate of growth is greater than or equal to  $\rho(f_1)$ . Thus, this result of Bowen implies the theorem of Manning that we have just mentioned.

The main result concerning the entropy conjecture is the theorem of Yosef Yomdin [Yom87], according to which the conjecture is true for every diffeomorphism of class  $C^\infty$ . The crucial ingredient in the proof is a relation between the topological entropy  $h(f)$  and the diffeomorphism's *rate of growth of volume*, which is defined as follows. For each  $1 \leq k < d$ , let  $B^k$  be the unit ball in  $\mathbb{R}^k$ . Denote by  $v(\sigma)$  the  $k$ -dimensional volume of the image of any differentiable embedding  $\sigma : B^k \rightarrow M$ . Then, define

$$v_k(f) = \sup_{\sigma} \limsup_n \frac{1}{n} \log v(f^n \circ \sigma),$$

where the supremum is taken over all the embeddings  $\sigma : B^k \rightarrow M$  of class  $C^\infty$ . Define also  $v(f) = \max\{v_k(f) : 1 \leq k < d\}$ . It is not difficult to check that

$$\log \rho(f_k) \leq v_k(f) \quad \text{for every } 1 \leq k < d. \quad (10.2.5)$$

On the one hand, Sheldon Newhouse [New88] proved that  $h(f) \leq v(f)$  for every diffeomorphism of class  $C^r$  with  $r > 1$ . On the other hand, Yomdin [Yom87] proved the opposite inequality:

$$v(f) \leq h(f), \quad (10.2.6)$$

for every diffeomorphism of class  $C^\infty$  (this inequality is false, in general, in the  $C^r$  case with  $r < \infty$ ). Combining (10.2.5) with (10.2.6), one gets the entropy conjecture (10.2.4) for every diffeomorphism of class  $C^\infty$ .

Concerning systems of class  $C^1$ , it is also known that the inequality (10.2.4) is true for every Axiom A diffeomorphism with no cycles [SW75], for certain partially hyperbolic diffeomorphisms [SX10] and, more generally, for any  $C^1$  diffeomorphism far from homoclinic tangencies [LVY13].

### 10.2.5 Linear endomorphisms of the torus

In this section we calculate the topological entropy of the linear endomorphisms of the torus:

**Proposition 10.2.10.** *Let  $f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$  be the endomorphism induced on the torus  $\mathbb{T}^d$  by some square matrix  $A$  of dimension  $d$  with integer coefficients and non-zero determinant. Then*

$$h(f_A) = \sum_{j=1}^d \log^+ |\lambda_j|, \tag{10.2.7}$$

where  $\lambda_1, \dots, \lambda_d$  are the eigenvalues of  $A$ , counted with multiplicity.

We have seen in Proposition 9.4.3 that the entropy of  $f_A$  with respect to the Haar measure  $\mu$  is equal to the expression on the right-hand side of (10.2.7). By the variational principle (Theorem 10.1), whose proof is contained in Section 10.4 below, the topological entropy is greater than or equal to the entropy of the transformation with respect to any invariant probability measure. Thus,

$$h(f_A) \geq h_\mu(f) = \sum_{j=1}^d \log^+ |\lambda_j|.$$

In what follows, we focus on proving the opposite inequality:

$$h(f_A) \leq \sum_{j=1}^d \log^+ |\lambda_j|. \tag{10.2.8}$$

Initially, assume that  $A$  is diagonalizable, that is, that there exists a basis  $v_1, \dots, v_d$  of  $\mathbb{R}^d$  with  $Av_i = \lambda_i v_i$  for each  $i$ . Then, clearly, we may take the elements of such a basis to be unit vectors. Moreover, up to renumbering the eigenvalues, we may assume that there exists  $u \in \{0, \dots, d\}$  such that  $|\lambda_i| > 1$  for  $1 \leq i \leq u$  and  $|\lambda_i| \leq 1$  for every  $i > u$ . Let  $e_1, \dots, e_d$  be the canonical basis of  $\mathbb{R}^d$  and  $P : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be the linear isomorphism defined by  $P(e_i) = v_i$  for each  $i$ . Then  $P^{-1}AP$  is a diagonal matrix. Fix  $L > 0$  large enough so that  $P((0, L)^d)$  contains some unit cube  $\prod_{i=1}^d [b_i, b_i + 1]^d$ . See Figure 10.2. Let  $\pi : \mathbb{R}^d \rightarrow \mathbb{T}^d$  be the canonical projection. Then  $\pi P((0, L)^d)$  contains the whole torus  $\mathbb{T}^d$ .

Given  $n \geq 1$  and  $\varepsilon > 0$ , fix  $\delta > 0$  such that  $\|P\| \delta \sqrt{d} < \varepsilon$ . Moreover, for each  $i = 1, \dots, d$ , take

$$\delta_i = \begin{cases} \delta |\lambda_i|^{-n} & \text{if } i \leq u \\ \delta & \text{if } i > u. \end{cases}$$

Consider the set

$$E = \pi P(\{(k_1 \delta_1, \dots, k_d \delta_d) \in (0, L)^d : k_1, \dots, k_d \in \mathbb{Z}\}).$$

Observe also that, given any  $j \geq 0$ ,

$$f_A^j(E) \subset \pi P(\{(k_1 \lambda_1^j \delta_1, \dots, k_d \lambda_d^j \delta_d) : k_1, \dots, k_d \in \mathbb{Z}\}).$$

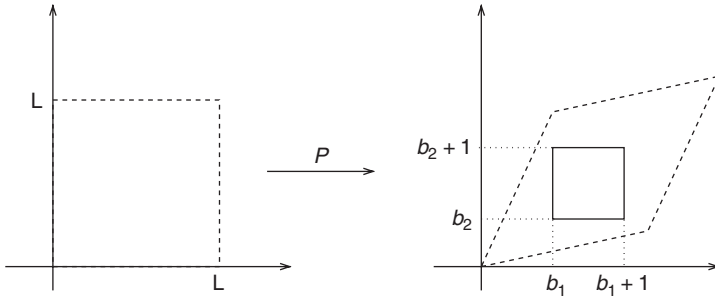


Figure 10.2. Building an  $(n, \varepsilon)$ -generating set in  $\mathbb{T}^d$

Consider  $0 \leq j < n$ . By construction,  $|\lambda_i^j \delta_i| \leq \delta$  for every  $i = 1, \dots, d$ . Therefore, every point of  $\mathbb{R}^d$  is at a distance less than or equal to  $\delta\sqrt{d}$  from some point of the form  $(k_1 \lambda_1^j \delta_1, \dots, k_d \lambda_d^j \delta_d)$ . Then (see Figure 10.2), for each  $x \in \mathbb{T}^d$  we may find  $a \in E$  such that  $d(f^j(x), f^j(a)) \leq \|P\| \delta\sqrt{d} < \varepsilon$  for every  $0 \leq j < n$ . This shows that  $E$  is an  $(n, \varepsilon)$ -generating set for  $\mathbb{T}^d$ . On the other hand,

$$\#E \leq \prod_{i=1}^d \frac{L}{\delta_i} = \left(\frac{L}{\delta}\right)^d \prod_{i=1}^u |\lambda_i|^n.$$

These observations show that  $g_n(f_A, \varepsilon, \mathbb{T}^d) \leq (L/\delta)^d \prod_{i=1}^u |\lambda_i|^n$  for every  $n \geq 1$  and  $\varepsilon > 0$ . Hence,

$$h(f) = \lim_{\varepsilon \rightarrow 0} \limsup_n \frac{1}{n} g_n(f_A, \varepsilon, \mathbb{T}^d) \leq \sum_{i=1}^u \log |\lambda_i| = \sum_{i=1}^d \log^+ |\lambda_i|.$$

This proves Proposition 10.2.10 in the case when  $A$  is diagonalizable.

The general case may be treated in a similar fashion, writing the matrix  $A$  in its Jordan canonical form. The reader is invited to carry out the details.

### 10.2.6 Exercises

10.2.1. Let  $(M_i, d_i)$ ,  $i = 1, 2$  be metric spaces and  $f_i : M_i \rightarrow M_i$ ,  $i = 1, 2$  be continuous transformations. Let  $M = M_1 \times M_2$ ,  $d$  be the distance defined in  $M$  by

$$d((x_1, x_2), (y_1, y_2)) = \max\{d_1(x_1, y_1), d_2(x_2, y_2)\}$$

and  $f : M \rightarrow M$  be the transformation defined by  $f(x_1, x_2) = (f_1(x_1), f_2(x_2))$ . Show that  $h(f) \leq h(f_1) + h(f_2)$  and the identity holds if at least one of the spaces is compact.

10.2.2. Let  $\sigma_A : \Sigma_A \rightarrow \Sigma_A$  be a shift of finite type, either one-sided or two-sided. We say that a transition matrix  $A$  is *irreducible* if for any  $i, j \in X$  there exists  $n \geq 1$  such that  $A_{ij}^n > 0$  and that  $A$  is *aperiodic* if there exists  $n \geq 1$  such that  $A_{ij}^n > 0$  for every  $i, j \in X$ . Show that:

(a) If  $A$  is irreducible then the set of periodic points of  $\sigma_A$  is dense in  $\Sigma_A$ .



- (b)  $\sigma_A$  is transitive if and only if  $A$  is irreducible.  
 (c)  $\sigma_A$  is topologically mixing if and only if  $A$  is aperiodic.  
 [Observation: Condition (b) means that the oriented graph  $G_A$  is connected: given any  $a, b \in X$  there exists some path in  $G_A$  starting at  $a$  and ending at  $b$ .]
- 10.2.3. Prove Proposition 10.2.6.  
 10.2.4. Let  $M$  be a compact metric space. Show that, given any  $\varepsilon > 0$ , the restriction of the topological entropy function  $f \mapsto h(f)$  to the set of continuous transformations  $f : M \rightarrow M$  that are  $\varepsilon$ -expansive is upper semi-continuous (with respect to the topology of uniform convergence).  
 10.2.5. Show that if  $A$  is a transition matrix then, for every  $k \geq 1$ , no row of  $A^k$  is identically zero. The same is true for the columns of  $A^k$ ,  $k \geq 1$ , if we assume that  $A$  is a transition matrix in the two-sided sense.  
 10.2.6. (a) Let  $f : M \rightarrow M$  be a surjective local homeomorphism in a compact metric space and let  $d = \inf_y \#f^{-1}(y)$ . Prove that  $h(f) \geq \log d$ .  
 (b) Let  $f : S^1 \rightarrow S^1$  be a continuous map in the circle. Show that  $h(f)$  is greater than or equal to the logarithm of the absolute value of the degree of  $f$ , that is,  $h(f) \geq \log |\deg f|$ .  
 [Observation: Misiurewicz and Przytycki [MP77b] proved that  $h(f) \geq \log |\deg f|$  for every map  $f : M \rightarrow M$  of class  $C^1$  in a compact manifold.]  
 10.2.7. Consider the map  $f : \overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$  defined by  $f(z) = z^d / (2|z|^{d-1})$ , with  $d \geq 2$ . Prove that the topological entropy of  $f$  is zero, but the degree of  $f$  is  $d$ . Why is this not in contradiction with Exercise 10.2.6?  
 10.2.8. Let  $f : M \rightarrow M$  be a continuous map in a compact metric space  $M$ . Given  $\varepsilon > 0$ , define

$$g_*(f, \varepsilon) = \sup\{g(f, B(x, \infty, \varepsilon)) : x \in M\},$$

where  $B(x, \infty, \varepsilon)$  denotes the set of all  $y \in M$  such that  $d(f^i(x), f^i(y)) \leq \varepsilon$  for every  $n \geq 0$ . Bowen [Bow72] has shown that, given  $b > 0$  and  $\delta > 0$ , there exists  $c > 0$  such that

$$\log g_n(f, \delta, B(x, n, \varepsilon)) < c + (g_*(f, \varepsilon) + b)n \quad \text{for every } x \in M \text{ and } n \geq 1.$$

Using this fact, prove that  $h(f) \leq g(f, \varepsilon, M) + g_*(f, \varepsilon)$ . One says that  $f$  is *h-expansive* if  $g_*(f, \varepsilon) = 0$  for some  $\varepsilon > 0$ . Conclude that in that case  $h(f) = g(f, \varepsilon, M)$ . [Observation: This generalizes Proposition 10.2.1, since every expansive transformation is also *h-expansive*.]

## 10.3 Pressure

In this section we introduce an important extension of the concept of topological entropy, called (*topological*) *pressure*, and we study its main properties. Throughout, we consider only continuous transformations in compact metric spaces. Related to this, check Exercises 10.3.4 and 10.3.5.

### 10.3.1 Definition via open covers

Let  $f : M \rightarrow M$  be a continuous transformation in a compact metric space. We call a *potential* in  $M$  any continuous function  $\phi : M \rightarrow \mathbb{R}$ . For each  $n \in \mathbb{N}$ , define  $\phi_n : M \rightarrow \mathbb{R}$  by  $\phi_n = \sum_{i=0}^{n-1} \phi \circ f^i$ . Given an open cover  $\alpha$  of  $M$ , let

$$P_n(f, \phi, \alpha) = \inf \left\{ \sum_{U \in \gamma} \sup_{x \in U} e^{\phi_n(x)} : \gamma \text{ is a finite subcover of } \alpha^n \right\}. \tag{10.3.1}$$

This sequence  $\log P_n(f, \phi, \alpha)$  is subadditive (Exercise 10.3.1) and so the limit

$$P(f, \phi, \alpha) = \lim_n \frac{1}{n} \log P_n(f, \phi, \alpha) \tag{10.3.2}$$

exists. Define the *pressure* of the potential  $\phi$  with respect to  $f$  to be the limit  $P(f, \phi)$  of  $P(f, \phi, \alpha)$  when the diameter of  $\alpha$  goes to zero. The existence of this limit is guaranteed by the following lemma:

**Lemma 10.3.1.** *There exists  $\lim_{\text{diam } \alpha \rightarrow 0} P(f, \phi, \alpha)$ , that is, there exists some  $P(f, \phi) \in \mathbb{R}$  such that*

$$\lim_k P(f, \phi, \alpha_k) = P(f, \phi)$$

for every sequence  $(\alpha_k)_k$  of open covers with  $\text{diam } \alpha_k \rightarrow 0$ .

*Proof.* Let  $(\alpha_k)_k$  and  $(\beta_k)_k$  be any sequences of open covers with diameters converging to zero. Given any  $\varepsilon > 0$ , fix  $\delta > 0$  such that  $|\phi(x) - \phi(y)| \leq \varepsilon$  whenever  $d(x, y) \leq \delta$ . By assumption,  $\text{diam } \alpha_k < \delta$  for every  $k$  sufficiently large. For fixed  $k$ , let  $\rho > 0$  be a Lebesgue number for  $\alpha_k$ . By assumption,  $\text{diam } \beta_l < \rho$  for every  $l$  sufficiently large. By the definition of Lebesgue number, it follows that every  $B \in \beta_l$  is contained in some  $A \in \alpha_k$ . Observe also that

$$\sup_{x \in A} \phi_n(x) \leq n\varepsilon + \sup_{y \in B} \phi_n(y)$$

for every  $n \geq 1$ , since  $\text{diam } \alpha_k < \delta$ . This implies that

$$P_n(f, \phi, \alpha_k) \leq e^{n\varepsilon} P_n(f, \phi, \beta_l) \quad \text{for every } n \geq 1$$

and, hence,  $P(f, \phi, \alpha_k) \leq \varepsilon + P(f, \phi, \beta_l)$ . Making  $l \rightarrow \infty$  and then  $k \rightarrow \infty$ , we get that

$$\limsup_k P(f, \phi, \alpha_k) \leq \varepsilon + \liminf_l P(f, \phi, \beta_l).$$

Since  $\varepsilon > 0$  is arbitrary, it follows that  $\limsup_k P(f, \phi, \alpha_k) \leq \liminf_l P(f, \phi, \beta_l)$ . Exchanging the roles of the two sequences of covers, we conclude that the limits  $\lim_k P(f, \phi, \alpha_k)$  and  $\lim_l P(f, \phi, \beta_l)$  exist and are equal.

Before we proceed, let us mention a few simple consequences of the definitions. The first is that the pressure of the zero potential coincides with the topological entropy. Indeed, it is immediate from (10.3.1) that  $P_n(f, 0, \alpha) = N(\alpha^n)$  for every  $n \geq 1$  and, thus,  $P(f, 0, \alpha) = h(f, \alpha)$  for every open cover  $\alpha$ .

Let  $(\alpha_k)_k$  be any sequence of open covers with diameters going to zero. Then, by Proposition 10.1.12 and the definition of the pressure,

$$h(f) = \lim_k h(f, \alpha_k) = \lim_k P(f, 0, \alpha_k) = P(f, 0). \tag{10.3.3}$$

Observe, however, that for general potentials  $P(f, \phi)$  need not coincide with the supremum of  $P(f, \phi, \alpha)$  over all open covers  $\alpha$  (see Exercise 10.3.5).

Given any constant  $c \in \mathbb{R}$ , we have that  $P_n(f, \phi + c, \alpha) = e^{cn} P_n(f, \phi, \alpha)$  for every  $n \geq 1$  and, consequently,  $P(f, \phi + c, \alpha) = P(f, \phi, \alpha) + c$  for any open cover  $\alpha$ . Hence,

$$P(f, \phi + c) = P(f, \phi) + c. \tag{10.3.4}$$

Analogously, if  $\phi \leq \psi$  then  $P_n(f, \phi, \alpha) \leq P_n(f, \psi, \alpha)$  for every  $n \geq 1$ , which implies that  $P(f, \phi, \alpha) = P(f, \psi, \alpha)$  for every open cover  $\alpha$ . That is,

$$\phi \leq \psi \Rightarrow P(f, \phi) \leq P(f, \psi). \tag{10.3.5}$$

In particular, since  $\inf \phi \leq \phi \leq \sup \phi$ , we have that

$$h(f) + \inf \phi \leq P(f, \phi) \leq h(f) + \sup \phi \tag{10.3.6}$$

for every potential  $\phi$ . An interesting corollary is that if  $h(f)$  is finite then  $P(f, \phi) < \infty$  for every potential  $\phi$  and, otherwise,  $P(f, \phi) = \infty$  for every potential  $\phi$ . An example of this last situation is given in Exercise 10.1.6.

Another simple consequence of the definition is that the pressure is an invariant of topological equivalence:

**Proposition 10.3.2.** *Let  $f : M \rightarrow M$  and  $g : N \rightarrow N$  be continuous transformations in compact metric spaces. If there exists a homeomorphism  $h : M \rightarrow N$  such that  $h \circ f = g \circ h$  then  $P(g, \phi) = P(f, \phi \circ h)$  for every potential  $\phi$  in  $N$ .*

*Proof.* The correspondence  $\alpha \mapsto h(\alpha)$  is a bijection between the spaces of open covers of  $M$  and  $N$ , respectively. Moreover, since  $h$  and its inverse are (uniformly) continuous,  $\text{diam } \alpha_k \rightarrow 0$  if and only if  $\text{diam } h(\alpha_k) \rightarrow 0$ . Consider the potential  $\psi = \phi \circ h$  in  $M$ . Note that  $\psi_n = \phi_n \circ h$  and so

$$\sup_{x \in U} \psi_n(x) = \sup_{y \in h(U)} \phi_n(y)$$

for every  $U \subset M$  and every  $n \geq 1$ . Hence,  $P_n(f, \psi, \alpha) = P_n(g, \phi, h(\alpha))$  for every  $n$  and every open cover  $\alpha$  of  $M$ . Thus,  $P(f, \psi, \alpha) = P(g, \phi, h(\alpha))$  and, taking the limit when the diameter of  $\alpha$  goes to zero,  $P(f, \psi) = P(g, \phi)$ .

One may replace the supremum by the infimum in (10.3.1), that is, replace  $P_n(f, \phi, \alpha)$  with

$$Q_n(f, \phi, \alpha) = \inf \left\{ \sum_{U \in \gamma} \inf_{x \in U} e^{\phi_n(x)} : \gamma \text{ is a finite subcover of } \alpha^n \right\},$$

although this makes the definition a bit more complicated. In contrast with  $\log P_n(f, \phi, \alpha)$ , the sequence  $\log Q_n(f, \phi, \alpha)$  need not be subadditive. Denote

$$Q^-(f, \phi, \alpha) = \liminf_n \frac{1}{n} \log Q_n(f, \phi, \alpha) \quad \text{and}$$

$$Q^+(f, \phi, \alpha) = \limsup_n \frac{1}{n} \log Q_n(f, \phi, \alpha).$$

Clearly,  $Q^-(f, \phi, \alpha) \leq Q^+(f, \phi, \alpha)$  for every open cover  $\alpha$  of  $M$ . Furthermore,  $Q_n(f, 0, \alpha) = P_n(f, 0, \alpha) = N(\alpha^n)$  for every  $n$  and so  $Q^-(f, 0, \alpha) = Q^+(f, 0, \alpha) = P(f, 0, \alpha) = h(f, \alpha)$ .

**Corollary 10.3.3.** *For any potential  $\phi : M \rightarrow \mathbb{R}$ ,*

$$P(f, \phi) = \lim_{\text{diam } \alpha \rightarrow 0} Q^+(f, \phi, \alpha) = \lim_{\text{diam } \alpha \rightarrow 0} Q^-(f, \phi, \alpha).$$

*Proof.* Since  $\phi$  is (uniformly) continuous, given any  $\varepsilon > 0$  there exists  $\delta > 0$  such that

$$\inf_{x \in C} \phi_n(x) \leq \sup_{x \in C} \phi_n(x) \leq n\varepsilon + \inf_{x \in C} \phi_n(x)$$

whenever  $\text{diam } C \leq \delta$ . So,

$$Q_n(f, \phi, \alpha) \leq P_n(f, \phi, \alpha) \leq e^{n\varepsilon} Q_n(f, \phi, \alpha)$$

for every open cover  $\alpha$  with  $\text{diam } \alpha \leq \delta$ . It follows that

$$\limsup_n \frac{1}{n} \log Q_n(f, \phi, \alpha) \leq P(f, \phi, \alpha) \leq \varepsilon + \liminf_n \frac{1}{n} \log Q_n(f, \phi, \alpha).$$

As the diameter of  $\alpha$  goes to zero, we may take  $\varepsilon \rightarrow 0$ . Thus,

$$\lim_{\text{diam } \alpha \rightarrow 0} Q^-(f, \phi, \alpha) = \lim_{\text{diam } \alpha \rightarrow 0} Q^+(f, \phi, \alpha) = \lim_{\text{diam } \alpha \rightarrow 0} P(f, \phi, \alpha) = P(f, \phi),$$

as claimed.

### 10.3.2 Generating sets and separated sets

Now we present two alternative definitions of pressure, in terms of generating sets and separated sets. As before,  $f : M \rightarrow M$  is a continuous transformation in a compact metric space and  $\phi : M \rightarrow \mathbb{R}$  is a continuous function.

Given  $n \geq 1$  and  $\varepsilon > 0$ , define

$$G_n(f, \phi, \varepsilon) = \inf \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ is an } (n, \varepsilon)\text{-generating set for } M \right\} \quad \text{and}$$

$$S_n(f, \phi, \varepsilon) = \sup \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ is an } (n, \varepsilon)\text{-separated set in } M \right\}.$$

(10.3.7)

Next, define

$$G(f, \phi, \varepsilon) = \limsup_n \frac{1}{n} \log G_n(f, \phi, \varepsilon) \quad \text{and} \tag{10.3.8}$$

$$S(f, \phi, \varepsilon) = \limsup_n \frac{1}{n} \log S_n(f, \phi, \varepsilon),$$

and also

$$G(f, \phi) = \lim_{\varepsilon \rightarrow 0} G(f, \phi, \varepsilon) \quad \text{and} \quad S(f, \phi) = \lim_{\varepsilon \rightarrow 0} S(f, \phi, \varepsilon) \tag{10.3.9}$$

(these limits exist because the functions are monotonic in  $\varepsilon$ ).

Note that  $G_n(f, 0, \varepsilon) = g_n(f, \varepsilon)$  and  $S_n(f, 0, \varepsilon) = s_n(f, \varepsilon)$  for every  $n \geq 1$  and every  $\varepsilon > 0$ . Therefore (Proposition 10.1.6),  $G(f, 0) = g(f)$  and  $S(f, 0) = s(f)$  coincide with the topological entropy  $h(f)$ . In fact,

**Proposition 10.3.4.**  $P(f, \phi) = G(f, \phi) = S(f, \phi)$  for every potential  $\phi$  in  $M$ .

*Proof.* Consider  $n \geq 1$  and  $\varepsilon > 0$ . It is clear from the definitions that every maximal  $(n, \varepsilon)$ -separated set is  $(n, \varepsilon)$ -generating. Then,

$$\begin{aligned} S_n(f, \phi, \varepsilon) &= \sup \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ is } (n, \varepsilon)\text{-separated} \right\} \\ &= \sup \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ is } (n, \varepsilon)\text{-separated maximal} \right\} \tag{10.3.10} \\ &\geq \inf \left\{ \sum_{x \in E} e^{\phi_n(x)} : E \text{ is } (n, \varepsilon)\text{-generating} \right\} = G_n(f, \phi, \varepsilon) \end{aligned}$$

for every  $n$  and every  $\varepsilon$ . This implies that  $G(f, \phi, \varepsilon) \leq S(f, \phi, \varepsilon)$  for every  $\varepsilon$  and, thus,  $G(f, \phi) \leq S(f, \phi)$ .

Next, we prove that  $S(f, \phi) \leq P(f, \phi)$ . Let  $\varepsilon$  and  $\delta$  be positive numbers such that  $d(x, y) \leq \delta$  implies  $|\phi(x) - \phi(y)| \leq \varepsilon$ . Let  $\alpha$  be any open cover of  $M$  with  $\text{diam } \alpha < \delta$  and  $E \subset M$  be any  $(n, \delta)$ -separated set. Given any subcover  $\gamma$  of  $\alpha^n$ , it is obvious that every point of  $E$  is contained in some element of  $\gamma$ . On the other hand, the hypothesis that  $E$  is  $(n, \delta)$ -separated implies that each element of  $\gamma$  contains at most one element of  $E$ . Therefore,

$$\sum_{x \in E} e^{\phi_n(x)} \leq \sum_{U \in \gamma} \sup_{y \in U} e^{\phi_n(y)}.$$

Taking the supremum in  $E$  and the infimum in  $\gamma$ , we get that

$$S_n(f, \phi, \delta) \leq P_n(f, \phi, \alpha). \tag{10.3.11}$$

It follows that  $S(f, \phi, \delta) \leq P(f, \phi, \alpha)$ . Making  $\delta \rightarrow 0$  (hence  $\text{diam } \alpha \rightarrow 0$ ), we conclude that  $S(f, \phi) \leq P(f, \phi)$ , as stated.

Finally, we prove that  $P(f, \phi) \leq G(f, \phi)$ . Let  $\varepsilon$  and  $\delta$  be positive numbers such that  $d(x, y) \leq \delta$  implies  $|\phi(x) - \phi(y)| \leq \varepsilon$ . Let  $\alpha$  be any open cover of  $M$  with  $\text{diam } \alpha < \delta$  and  $\rho > 0$  be a Lebesgue number of  $\alpha$ . Let  $E \subset M$  be any

$(n, \rho)$ -generating set for  $M$ . For each  $x \in E$  and  $i = 0, \dots, n - 1$ , there exists  $A_{x,i} \in \alpha$  such that  $B(f^i(x), \rho)$  is contained in  $A_{x,i}$ . Denote

$$\gamma(x) = \bigcap_{i=0}^{n-1} f^{-i}(A_{x,i}).$$

Observe that  $\gamma(x) \in \alpha^n$  and  $B(x, n, \rho) \subset \gamma(x)$ . Hence, the hypothesis that  $E$  is  $(n, \rho)$ -generating implies that  $\gamma = \{\gamma(x) : x \in E\}$  is a subcover of  $\alpha$ . Observe also that

$$\sup_{y \in \gamma(x)} \phi_n(y) \leq n\varepsilon + \phi_n(x) \quad \text{for every } x \in E,$$

since  $\text{diam} A_{x,i} < \delta$  for every  $i$ . It follows that

$$\sum_{U \in \gamma} \sup_{y \in U} e^{\phi_n(y)} \leq e^{n\varepsilon} \sum_{x \in E} e^{\phi_n(x)}.$$

This proves that  $P_n(f, \phi, \alpha) \leq e^{n\varepsilon} G_n(f, \phi, \rho)$  for every  $n \geq 1$  and, consequently,

$$P(f, \phi, \alpha) \leq \varepsilon + \liminf_n \frac{1}{n} G_n(f, \phi, \rho) \leq \varepsilon + G(f, \phi, \rho). \tag{10.3.12}$$

Making  $\rho \rightarrow 0$  we find that  $P(f, \phi, \alpha) \leq \varepsilon + G(f, \phi)$ . Hence, making  $\varepsilon, \delta$  and  $\text{diam} \alpha$  go to zero,  $P(f, \phi) \leq G(f, \phi)$ .

The conclusion of Proposition 10.3.4 may be rewritten as follows:

$$\begin{aligned} P(f, \phi) &= \lim_{s \rightarrow 0} \limsup_n \frac{1}{n} \log G_n(f, \phi, s) \\ &= \lim_{s \rightarrow 0} \limsup_n \frac{1}{n} \log S_n(f, \phi, s). \end{aligned} \tag{10.3.13}$$

The relations (10.3.12) and (10.3.10) in the proof also give that

$$P(f, \phi) \leq \lim_{s \rightarrow 0} \liminf_n \frac{1}{n} \log G_n(f, \phi, s) \leq \lim_{s \rightarrow 0} \liminf_n \frac{1}{n} \log S_n(f, \phi, s).$$

Combining these observations, we get:

$$\begin{aligned} P(f, \phi) &= \lim_{s \rightarrow 0} \liminf_n \frac{1}{n} \log G_n(f, \phi, s) \\ &= \lim_{s \rightarrow 0} \liminf_n \frac{1}{n} \log S_n(f, \phi, s). \end{aligned} \tag{10.3.14}$$

### 10.3.3 Properties

Properties of the pressure function in the spirit of Proposition 10.1.10 and Corollary 10.1.13 are stated in Exercise 10.3.3. Let us also extend Propositions 10.1.14 and 10.1.15 to the present context:

**Proposition 10.3.5.** *Let  $f : M \rightarrow M$  be a continuous transformation in a compact metric space and  $\phi$  be a potential in  $M$ . Then:*

- (1)  $P(f^k, \phi_k) = kP(f, \phi)$  for every  $k \geq 1$ .
- (2) If  $f$  is a homeomorphism then  $P(f^{-1}, \phi) = P(f, \phi)$ .

*Proof.* Given a potential  $\phi : M \rightarrow \mathbb{R}$  and an open cover  $\alpha$ , denote  $\psi = \sum_{i=0}^{k-1} \phi \circ f^i$  and  $\beta = \bigvee_{i=0}^{k-1} f^{-i}(\alpha)$ . Let  $g = f^k$ . It is clear that

$$\sum_{j=0}^{n-1} \psi \circ g^j = \sum_{l=0}^{kn-1} \phi \circ f^l \quad \text{and} \quad \bigvee_{j=0}^{n-1} g^{-j}(\beta) = \bigvee_{l=0}^{nk-1} f^{-l}(\alpha).$$

Then,

$$\begin{aligned} P_n(g, \psi, \beta) &= \inf \left\{ \sum_{U \in \gamma} \sup_{x \in U} e^{\sum_{j=0}^{n-1} \psi(g^j(x))} : \gamma \subset \bigvee_{j=0}^{n-1} g^{-j}(\beta) \right\} \\ &= \inf \left\{ \sum_{U \in \gamma} \sup_{x \in U} e^{\sum_{l=0}^{kn-1} \phi(f^l(x))} : \gamma \subset \bigvee_{l=0}^{nk-1} f^{-l}(\alpha) \right\} = P_{kn}(f, \phi, \alpha). \end{aligned}$$

Consequently,  $P(f^k, \psi, \beta) = kP(f, \phi, \alpha)$  for any  $\alpha$ . Making  $\text{diam } \alpha \rightarrow 0$  (note that  $\text{diam } \beta \rightarrow 0$ ), we deduce that  $P(f^k, \psi) = kP(f, \phi)$ . This proves part (1).

Suppose that  $f$  is a homeomorphism. Given an open cover  $\alpha$  and an integer number  $n \geq 1$ , denote

$$\phi_n^- = \sum_{j=0}^{n-1} \phi \circ f^{-j} \quad \text{and} \quad \alpha_n^- = \alpha \vee f(\alpha) \vee \dots \vee f^{n-1}(\alpha).$$

It is clear that  $\phi_n^- = \phi_n \circ f^{n-1}$  and  $\alpha_n^- = f^{n-1}(\alpha_n)$ . Moreover,  $\gamma$  is a subcover of  $\alpha_n^-$  if and only if  $\delta = f^{n-1}(\gamma)$  is a subcover of  $\alpha_n$ . Combining these facts, we find that

$$\begin{aligned} P_n(f^{-1}, \phi, \alpha) &= \inf \left\{ \sum_{U \in \delta} \sup_{x \in U} e^{\phi_n^-(x)} : \gamma \subset \alpha_n^- \right\} \\ &= \inf \left\{ \sum_{V \in \gamma} \sup_{y \in V} e^{\phi_n(y)} : \delta \subset \alpha_n \right\} = P_n(f, \phi, \alpha) \end{aligned}$$

for every  $n \geq 1$ . Hence,  $P(f^{-1}, \phi, \alpha) = P(f, \phi, \alpha)$  for every open cover  $\alpha$ . Making  $\text{diam } \alpha \rightarrow 0$ , we reach the conclusion in part (2).

Next, we fix the transformation  $f : M \rightarrow M$  and we consider  $P(f, \cdot)$  as a function in the space  $C^0(M)$  of all continuous functions, with the norm defined by

$$\|\varphi\| = \sup\{|\varphi(x)| : x \in M\}.$$

We have seen in (10.3.6) that if the topological entropy  $h(f)$  is infinite then the pressure function is constant and equal to  $\infty$ . In what follows we assume that  $h(f)$  is finite. Then,  $P(f, \phi)$  is finite for every potential  $\phi$ .

**Proposition 10.3.6.** *The pressure function is Lipschitz, with Lipschitz constant equal to 1:  $|P(f, \phi) - P(f, \psi)| \leq \|\phi - \psi\|$  for any potentials  $\phi$  and  $\psi$ .*

*Proof.* Clearly,  $\phi \leq \psi + \|\phi - \psi\|$ . Hence, by (10.3.4) and (10.3.5), we have that  $P(f, \phi) \leq P(f, \psi) + \|\phi - \psi\|$ . Exchanging the roles of  $\phi$  and  $\psi$ , one gets the other inequality.

**Proposition 10.3.7.** *The pressure function is convex:*

$$P(f, (1 - t)\phi + t\psi) \leq (1 - t)P(f, \phi) + tP(f, \psi)$$

for any potentials  $\phi$  and  $\psi$  in  $M$  and any  $0 \leq t \leq 1$ .

*Proof.* Write  $\zeta = (1 - t)\phi + t\psi$ . Then  $\zeta_n = (1 - t)\phi_n + t\psi_n$  for every  $n \geq 1$  and, thus,  $\sup(\zeta_n | U) \leq (1 - t)\sup(\phi_n | U) + t\sup(\psi_n | U)$  for every  $U \subset M$ . Then, by the Hölder inequality (Theorem A.5.5),

$$\sum_{U \in \gamma} \sup_{x \in U} e^{\zeta_n(x)} \leq \left( \sum_{U \in \gamma} \sup_{x \in U} e^{\phi_n(x)} \right)^{1-t} \left( \sum_{U \in \gamma} \sup_{x \in U} e^{\psi_n(x)} \right)^t$$

for any finite family  $\gamma$  of subsets of  $M$ . This implies that, given any open cover  $\alpha$ ,

$$P_n(f, \zeta, \alpha) \leq P_n(f, \phi, \alpha)^{1-t} P_n(f, \psi, \alpha)^t$$

for every  $n \geq 1$  and, hence,  $P(f, \zeta, \alpha) \leq (1 - t)P(f, \phi, \alpha) + tP(f, \psi, \alpha)$ . Passing to the limit when  $\text{diam } \alpha \rightarrow 0$ , we get the conclusion of the proposition.

We say that two potentials  $\phi, \psi : M \rightarrow \mathbb{R}$  are *cohomologous* if there exists a continuous function  $u : M \rightarrow \mathbb{R}$  such that  $\phi = \psi + u \circ f - u$ . Note that this is an equivalence relation in the space of potentials (Exercise 10.3.6).

**Proposition 10.3.8.** *Let  $f : M \rightarrow M$  be a continuous transformation in a compact topological space. If  $\phi, \psi : M \rightarrow \mathbb{R}$  are cohomologous potentials then  $P(f, \phi) = P(f, \psi)$ .*

*Proof.* If  $\psi = \phi + u \circ f - u$  then  $\psi_n(x) = \phi_n(x) + u(f^n(x)) - u(x)$  for every  $n \in \mathbb{N}$ . Let  $K = \sup |u|$ . Then  $|\sup_{x \in C} \psi_n(x) - \sup_{x \in C} \phi_n(x)| \leq 2K$  for every set  $C \subset M$ . Hence, for any open cover  $\gamma$ ,

$$e^{-2K} \sum_{U \in \gamma} \sup_{x \in U} e^{\phi_n(x)} \leq \sum_{U \in \gamma} \sup_{x \in U} e^{\psi_n(x)} \leq e^{2K} \sum_{U \in \gamma} \sup_{x \in U} e^{\phi_n(x)}.$$

This implies that, given any open cover  $\alpha$  of  $M$ ,

$$e^{-2K} P_n(f, \phi, \alpha) \leq P_n(f, \psi, \alpha) \leq e^{2K} P_n(f, \phi, \alpha)$$

for every  $n$ . Therefore,  $P(f, \phi, \alpha) = P(f, \psi, \alpha)$  for every  $\alpha$  and, consequently,  $P(f, \phi) = P(f, \psi)$ .



### 10.3.4 Comments in statistical mechanics

Let us take a pause to explain the relation between the mathematical concept of pressure and the issues in physics that originated it. This also serves as a preview to Chapter 12, where this theory will be developed in the context of expanding maps in metric spaces. The discussion that follows is a combination of mathematical results and physical considerations, not necessarily rigorous, and is quite brief: we refer the reader to the classical works of David Ruelle [Rue04] and Oscar Lanford [Lan73] for actual presentations of the subject.

The goal of statistical mechanics is to describe the properties of physical systems consisting of a large number of units that interact with each other. For example, these units may be particles, such as molecules of a gas, or sites in a crystal grid, which may or may not be occupied by particles. The *constant of Avogadro*  $6.022 \times 10^{23}$  illustrates what one means by “large” in specific situations in this context.

The main challenge in this area of mathematical physics is to understand the phenomena of *phase transitions*, that is, sudden changes from one physical state to another: for example, what happens when liquid water turns into ice? Why does this occur suddenly, at a given freezing temperature? Mathematical methods developed for tackling this kind of question turn out to be very useful in other areas of science, such as quantum field theory and, closer to the scope of this book, the ergodic theory of hyperbolic dynamical systems (Bowen [Bow75a]).

In order to formulate these problems in mathematical terms, it is convenient to assume that the set  $L$  of units in the system is actually infinite, because finite systems do not have genuine phase transitions. The best-studied examples are the *lattice systems*, for which  $L = \mathbb{Z}^d$  with  $d \geq 1$ . It is assumed that each unit has a finite set  $F$  of possible values (or “states”). For example,  $F = \{-1, +1\}$  in the case of *spin systems*, with  $\pm 1$  representing the two possible orientations of the particle’s “spin”, and  $F = \{0, 1\}$  in the case of *lattice gases*: 1 means that the site  $k \in L$  is occupied by a gas molecule, whereas 0 means that the site is empty.

Then, the system’s *configuration space* is a subset  $\Omega$  of the product space  $F^L$ . We assume  $\Omega$  to be closed in  $F^L$  and invariant under the shift map

$$\sigma^n : F^L \rightarrow F^L, \quad (\zeta_k)_{k \in L} \mapsto (\zeta_{k+n})_{k \in L}$$

for every  $n \in L$ . A *state* of the system is a probability measure  $\mu$  on  $\Omega$ : intuitively, one presumes that at the microscopic level the system oscillates randomly between different configurations  $\zeta \in \Omega$  (for example, different positions or velocities of the molecules), all corresponding to the same macroscopic parameters (same temperature, etc.); then, the measure  $\mu$  describes the probability distribution of these microscopic configurations.

States corresponding to macroscopic configurations that can be physically observed, that is, that actually occur in Nature, are called *equilibrium states*. This notion has a central role in the theory, in particular, because *phase transitions are associated with the coexistence of more than one equilibrium state*. Under our hypotheses on  $\Omega$ , one can show that every equilibrium state  $\mu$  is invariant under the shift maps  $\sigma^n$ ,  $n \in L$ . Thus, the study of lattice systems is naturally inserted in the scope of ergodic theory.

According to the *variational principle* of statistical mechanics, which goes back to the *principle of least action* of Maupertuis, the equilibrium states are characterized by the fact that they minimize a certain fundamental quantity, called *Gibbs free energy*, whose definition involves the energy  $E$ , the temperature  $T$  and the entropy  $S$  of the system's state. The *pressure* of that state is, simply, the product of the Gibbs free energy and a negative factor  $-\beta$  whose nature will be explained shortly.<sup>1</sup> Therefore, the equilibrium states are also characterized by the fact that they maximize the pressure among all probability measures invariant under the shift maps  $\sigma^n$ ,  $n \in L$ .

From these facts, one can obtain a rather explicit description of the equilibrium states for lattice systems: under suitable hypotheses, *the equilibrium states are precisely the Gibbs states invariant under the shift maps*. In the remainder of this section we are going to motivate and define this concept of *Gibbs state*, which will also allow us to illustrate the ideas outlined in the previous paragraphs. By the end of the section we briefly comment on the case of one-dimensional lattice systems, that is, the case  $d = 1$ , whose theory is much simpler and which is more closely related to the topics treated in this book.

Let us start by considering the particularly simple case of finite systems, that is, such that the configuration space  $\Omega$  is finite. The *entropy* of a state  $\mu$  in  $\Omega$  is the number

$$S(\mu) = \sum_{\xi \in \Omega} -\mu(\{\xi\}) \log \mu(\{\xi\}).$$

To each configuration  $\xi \in \Omega$  corresponds a value  $E(\xi)$  for the energy of the system. Denote by  $E(\mu)$  the energy of the state  $\mu$ , that is, the mean

$$E(\mu) = \sum_{\xi \in \Omega} \mu(\{\xi\}) E(\{\xi\}).$$

Take the system's absolute temperature  $T$  to be constant in time. Then, the *Gibbs free energy* is defined by

$$G(\mu) = E(\mu) - \kappa TS(\mu),$$

<sup>1</sup> From the mathematical point of view, the two quantities are equivalent. Preference for one denomination or the other has mostly to do with the physical interpretation of the set  $F$ : for spin systems one usually refers to the Gibbs free energy, whereas for lattice gases, where the elements of  $F$  describe the rate of occupation of each site, it is more natural to refer to the pressure.

where  $\kappa = 1.380 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}$  is the *Boltzmann constant*. In other words, denoting  $\beta = 1/(\kappa T)$ ,

$$-\beta G(\mu) = \sum_{\zeta \in \Omega} \mu(\{\zeta\}) [ -\beta E(\zeta) - \log \mu(\{\zeta\}) ]. \tag{10.3.15}$$

This expression is denoted by  $P(\mu)$  and is called *pressure* of the state  $\mu$ .

It is easy to check that the pressure (10.3.15) is maximum (hence, the Gibbs free energy  $G(\mu)$  is minimum) if and only if

$$\mu(\{\zeta\}) = \frac{e^{-\beta E(\zeta)}}{\sum_{\eta \in \Omega} e^{-\beta E(\eta)}} \quad \text{for every } \zeta \in \Omega \tag{10.3.16}$$

(see Lemma 10.4.4 below). Therefore, the *Gibbs distribution*  $\mu$  given by (10.3.16) is the unique equilibrium state of the system. In particular, in this simple context there are no phase transitions.

Now we sketch how this analysis can be extended to infinite lattice systems, assuming that the interaction between sites that are far apart is sufficiently weak. It is part of the hypotheses that the energy associated with each configuration  $\zeta \in \Omega$  comes from the pairwise interactions between the different sites in the lattice (including self-interactions) and that this interaction is invariant under the shift maps  $\sigma^n$ ,  $n \in L$ . Then, the energy  $E_{k,l}$  resulting from the action of any site  $k \in L$  on any other site  $l \in L$  depends only on their relative position and on the values of  $\zeta_k$  and  $\zeta_l$ . In other words, there exists a function  $\Psi : L \times F \times F \rightarrow \bar{\mathbb{R}}$  such that

$$E_{k,l} = \Psi(k - l, \zeta_k, \zeta_l) \quad \text{for any } k, l \in L.$$

It is also assumed that the strength of the interaction decays exponentially with the distance between the sites, in the following sense: there exist constants  $K > 0$  and  $\theta > 0$  such that

$$|\Psi(m, a, b)| \leq Ke^{-\theta|m|} \quad \text{for any } m \in L \text{ and } a, b \in F, \tag{10.3.17}$$

where  $|m| = \max\{|m_1|, \dots, |m_d|\}$ . In particular (Exercise 10.3.9), the energy

$$\varphi(\zeta) = \sum_{k \in L} \Psi(k, \zeta_k, \zeta_0)$$

resulting from the action of all the sites on the site 0 at the origin is uniformly bounded.

Initially, given any finite set  $\Lambda \subset L$ , let us consider the system one obtains by observing only the sites  $k \in \Lambda$  and “switching off” their interactions with the sites in the complement of  $\Lambda$ . This is a finite system, as the configuration space is contained in  $F^\Lambda$ , with energy function given by

$$E_\Lambda(x) = \sum_{l \in \Lambda} \sum_{k \in \Lambda} \Psi(k - l, x_k, x_l) \quad \text{for every } x \in F^\Lambda.$$

Hence, according to (10.3.16), its Gibbs distribution  $\mu_\Lambda$  is given by

$$\mu_\Lambda(\{x\}) = \frac{e^{-\beta E_\Lambda(x)}}{\sum_{y \in F^\Lambda} e^{-\beta E_\Lambda(y)}} \quad \text{for each } x \in F^\Lambda. \tag{10.3.18}$$

The notion of Gibbs state is obtained from this one by “switching back on” the interaction with the sites outside  $\Lambda$ , in the way we are going to explain. Denote by  $r_\Lambda(x)$  the expression on the right-hand side of (10.3.18). Observe that

$$r_\Lambda(x) = \left[ \sum_{y \in F^\Lambda} e^{\beta E_\Lambda(x) - \beta E_\Lambda(y)} \right]^{-1}$$

and recall that

$$E_\Lambda(x) - E_\Lambda(y) = \sum_{l \in \Lambda} \sum_{k \in \Lambda} \Psi(k-l, x_k, x_l) - \Psi(k-l, y_k, y_l).$$

For  $\zeta, \eta \in F^L$ , define

$$\begin{aligned} \mathcal{E}(\zeta, \eta) &= \sum_{l \in L} \sum_{k \in L} \Psi(k-l, \zeta_k, \zeta_l) - \Psi(k-l, \eta_k, \eta_l) \\ &= \sum_{l \in L} \sum_{j \in L} \Psi(j, \zeta_{j+l}, \zeta_l) - \Psi(j, \eta_{j+l}, \eta_l) = \sum_{l \in L} \varphi(\sigma^l(\zeta)) - \varphi(\sigma^l(\eta)). \end{aligned}$$

It follows from the condition (10.3.17) that this sum converges whenever the two configurations are such that  $\zeta_k = \eta_k$  for every  $k$  in the complement of some finite set (Exercise 10.3.9). Then,

$$\rho_\Lambda(\zeta) = \left[ \beta \sum_{\eta | \Lambda^c = \zeta | \Lambda^c} e^{\beta \mathcal{E}(\zeta, \eta)} \right]^{-1}$$

is well defined for every  $\zeta \in F^\Lambda$ .

A probability measure  $\mu$  supported in  $\Omega \subset F^L$  is called a *Gibbs state* if, for every finite set  $\Lambda \subset L$ , the disintegration  $\{\mu_{\Lambda, \theta} : \theta \in F^{\Lambda^c}\}$  of  $\mu$  relative to the partition  $\{F^\Lambda \times \{\theta\} : \theta \in F^{\Lambda^c}\}$  of the space  $F^L = F^\Lambda \times F^{\Lambda^c}$  is given by

$$\mu_{\Lambda, \theta}(\{x\} \times \{\theta\}) = \begin{cases} \rho_\Lambda(x, \theta) & \text{if } (x, \theta) \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

To conclude this section, we state one of the main results of this formalism that we have been describing: *one-dimensional lattice systems exhibit no phase transitions*. More precisely:

**Theorem 10.3.9** (Ruelle). *If  $d = 1$  and the interactions decay exponentially with the distance, then there exists a unique Gibbs state and it is also the unique equilibrium state.*

The arguments in the proof of this theorem (Ruelle [Rue04]) are at the basis of the thermodynamic formalism of expanding maps, which we are going to present in Chapter 12. Let us point out that the theorem is false in

dimension  $d \geq 2$ . Exercise 10.3.10 highlights one of the specificities of the one-dimensional case that are behind this result.

### 10.3.5 Exercises

- 10.3.1. Check that the sequence  $\log P_n(f, \phi, \alpha)$  is subadditive.
- 10.3.2. Show that if  $f$  is a homeomorphism then  $P(f, \phi, f(\alpha)) = P(f, \phi, \alpha)$ ,  $Q^+(f, \phi, f(\alpha)) = Q^+(f, \phi, \alpha)$  and  $Q^-(f, \phi, f(\alpha)) = Q^-(f, \phi, \alpha)$  for every open cover  $\alpha$ .
- 10.3.3. Show that, for any potential  $\phi : M \rightarrow \mathbb{R}$ :
  - (a) If  $\alpha, \beta$  are open covers with  $\alpha < \beta$  then  $Q^+(f, \phi, \alpha) \leq Q^+(f, \phi, \beta)$  and  $Q^-(f, \phi, \alpha) \leq Q^-(f, \phi, \beta)$ .
  - (b)  $Q^+(f, \phi, \alpha) = Q^+(f, \phi, \alpha^k)$  and  $Q^-(f, \phi, \alpha) = Q^-(f, \phi, \alpha^k)$  for every  $k \geq 1$  and every open cover  $\alpha$ .
  - (c)  $Q^+(f, \phi, \alpha) = P(f, \phi) = Q^-(f, \phi, \alpha)$  for any open cover  $\alpha$  such that  $\text{diam } \alpha^k \rightarrow 0$ .
  - (d)  $P(f, \phi, \alpha) = P(f, \phi, \alpha^k)$  for every  $k \geq 1$  and any open cover  $\alpha$  whose elements are pairwise disjoint.
  - (e)  $P(f, \phi, \alpha) = P(f, \phi)$  for any open cover  $\alpha$  such that  $\text{diam } \alpha^k \rightarrow 0$  and whose elements are pairwise disjoint.
  - (f) If  $f$  is a homeomorphism, one may replace  $\alpha^k$  by  $\alpha^{\pm k}$  in statements (b), (c), (d) and (e).
- 10.3.4. (Walters). Prove that

$$\begin{aligned}
 P(f, \phi) &= \sup\{Q^-(f, \phi, \alpha) : \alpha \text{ is an open cover of } M\} \\
 &= \sup\{Q^+(f, \phi, \alpha) : \alpha \text{ is an open cover of } M\}.
 \end{aligned}$$

[Observation: In particular, the pressure depends only on the topology of  $M$ , not the distance. This also provides a way to extend the definition to continuous transformations in compact topological spaces.]

- 10.3.5. Exhibit a homeomorphism  $f : M \rightarrow M$ , a potential  $\phi : M \rightarrow \mathbb{R}$  and open covers  $\alpha$  and  $\beta$  of a compact metric space  $M$  such that  $\alpha < \beta$  and  $P(f, \phi, \alpha) > P(f, \phi, \beta) = P(f, \phi)$ . [Observation: Thus, the conclusions of Exercise 10.3.3(a) and Exercise 10.3.4 are no longer valid if one replaces  $Q^\pm(f, \phi, \alpha)$  by  $P(f, \phi, \alpha)$ .]
- 10.3.6. Check that the cohomology relation

$$\phi \sim \psi \Leftrightarrow \psi = \phi + u \circ f - u \text{ for some continuous function } u : M \rightarrow \mathbb{R}$$

is an equivalence relation.

- 10.3.7. Let  $f_i : M_i \rightarrow M_i$ ,  $i = 1, 2$  be continuous transformations in compact metric spaces and, for each  $i$ , let  $\phi_i$  be a potential in  $M_i$ . Define

$$\begin{aligned}
 f_1 \times f_2 : M_1 \times M_2 &\rightarrow M_1 \times M_2, & f_1 \times f_2(x_1, x_2) &= (f_1(x_1), f_2(x_2)) \\
 \phi_1 \times \phi_2 : M_1 \times M_2 &\rightarrow \mathbb{R}, & \phi_1 \times \phi_2(x_1, x_2) &= \phi_1(x_1) + \phi_2(x_2).
 \end{aligned}$$

Show that  $P(f_1 \times f_2, \phi_1 \times \phi_2) = P(f_1, \phi_1) + P(f_2, \phi_2)$ .

- 10.3.8. Consider the transformation  $f : S^1 \rightarrow S^1$  defined by  $f(x) = 2x \pmod{\mathbb{Z}}$ . Prove that if  $\phi : S^1 \rightarrow \mathbb{R}$  is a Hölder function then

$$P(f, \phi) = \lim_n \frac{1}{n} \log \sum_{p \in \text{Fix}(f^n)} e^{\phi_n(p)}.$$

[Observation: We will get a more general result in Exercise 11.3.4.]

- 10.3.9. Assuming the conditions in Section 10.3.4, prove that:
- (a) There exists  $C > 0$  such that  $|\varphi(\xi)| \leq C$  and  $|\varphi(\xi) - \varphi(\eta)| \leq Ce^{-\theta N/2}$  for any  $\xi, \eta \in F^L$  such that  $\xi_k = \eta_k$  whenever  $|k| < N$ .
  - (b) For any finite set  $\Lambda \subset L$ , there exists  $C_\Lambda > 0$  such that  $|\mathcal{E}(\xi, \eta)| \leq C_\Gamma$  for any  $\xi, \eta \in F^L$  such that  $\xi_k = \eta_k$  for every  $k \in \Lambda^c$ .
- 10.3.10. Assuming the conditions in Section 10.3.4, prove that if  $d = 1$  then there exists  $C > 0$  such that  $|\mathcal{E}(\xi, \eta) - E_\Lambda(\xi) + E_\Lambda(\eta)| \leq C$  for every finite interval  $\Lambda \subset \mathbb{Z}$  and any  $\xi, \eta \in F^{\mathbb{Z}}$  with  $\xi|_{\Lambda^c} = \eta|_{\Lambda^c}$ . Consequently,

$$e^{-\beta C} \leq \frac{r_\Lambda(\xi | \Lambda)}{\rho_\Lambda(\xi)} \leq e^{\beta C}$$

for every  $\xi \in F^{\mathbb{Z}}$  and every finite interval  $\Lambda$ . [Observation: Therefore, the probability distribution of the configurations restricted to each finite interval is not affected in a significant way by the interactions with the sites outside that interval.]

### 10.4 Variational principle

The variational principle for the pressure, which we state below, was proved originally by Ruelle [Rue73], under more restrictive assumptions, and was then extended by Walters [Wal75] to the context we consider here:

**Theorem 10.4.1** (Variational principle). *Let  $f : M \rightarrow M$  be a continuous transformation in a compact metric space and  $\mathcal{M}_1(f)$  denote the set of probability measures invariant under  $f$ . Then, for every continuous function  $\phi : M \rightarrow \mathbb{R}$ ,*

$$P(\phi, f) = \sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}_1(f)\}.$$

Theorem 10.1 corresponds to the special case  $\phi \equiv 0$ . In particular, it follows that the topological entropy of  $f$  is zero if and only if  $h_\nu(f) = 0$  for every invariant probability measure  $\nu$ . That is the case, for example, for every circle homeomorphism (Example 10.1.1) and every translation in a compact metrizable group (Example 10.1.7). The compactness hypothesis is crucial: as observed in Exercise 10.4.4, there exist transformations (in non-compact spaces) without invariant measures and whose topological entropy is positive.

In Sections 10.4.1 and 10.4.2 we present a proof of Theorem 10.4.1 that is due to Misiurewicz [Mis76]. Before that, let us mention a couple of consequences.

**Corollary 10.4.2.** *Let  $f : M \rightarrow M$  be a continuous transformation in a compact metric space and  $\mathcal{M}_e(f)$  denote the set of probability measures invariant and ergodic. Then*

$$P(\phi, f) = \sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}_e(f)\}.$$

*Proof.* Given any  $\nu \in \mathcal{M}_1(f)$ , let  $\{\nu_P : P \in \mathcal{P}\}$  be its ergodic decomposition. By Theorems 5.1.3 and 9.6.2,

$$h_\nu(f) + \int \phi d\nu = \int \left( h_{\nu_P}(f) + \int \phi d\nu_P \right) d\hat{\mu}(P).$$

This implies that

$$\sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}_1(f)\} \leq \sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}_e(f)\}.$$

The converse inequality is trivial, since  $\mathcal{M}_e(f) \subset \mathcal{M}_1(f)$ . Now it suffices to apply Theorem 10.4.1.

Another interesting consequence is that for transformations with finite topological entropy the pressure function determines the set of all invariant probability measures:

**Corollary 10.4.3** (Walters). *Let  $f : M \rightarrow M$  be a continuous transformation in a compact metric space with topological entropy  $h(f) < \infty$ . Let  $\eta$  be any finite signed measure on  $M$ . Then,  $\eta$  is a probability measure invariant under  $f$  if and only if  $\int \phi d\eta \leq P(f, \phi)$  for every continuous function  $\phi : M \rightarrow \mathbb{R}$ .*

*Proof.* The “only if” claim is an immediate consequence of Theorem 10.4.1: if  $\eta$  is an invariant probability measure then

$$P(f, \phi) \geq h_\eta(f) + \int \phi d\eta \geq \int \phi d\eta$$

for every continuous function  $\phi$ . In what follows we prove the converse.

Let  $\eta$  be a finite signed measure such that  $\int \phi d\eta \leq P(f, \phi)$  for every  $\phi$ . Consider any  $\phi \geq 0$ . For any  $c > 0$  and  $\varepsilon > 0$ ,

$$c \int (\phi + \varepsilon) d\eta = - \int -c(\phi + \varepsilon) d\eta \geq -P(f, -c(\phi + \varepsilon)).$$

By the relation (10.3.6), we also have that

$$P(f, -c(\phi + \varepsilon)) \leq h(f) + \sup(-c(\phi + \varepsilon)) = h(f) - c \inf(\phi + \varepsilon).$$

Therefore,  $c \int (\phi + \varepsilon) d\eta \geq -h(f) + c \inf(\phi + \varepsilon)$ . When  $c > 0$  is sufficiently large, the right-hand side of this inequality is positive. Hence,  $\int (\phi + \varepsilon) d\eta > 0$ . Since  $\varepsilon > 0$  is arbitrary, this implies that  $\int \phi d\eta \geq 0$  for every  $\phi \geq 0$ . So,  $\eta$  is a positive measure.

The next step is to show that  $\eta$  is a probability measure. By assumption,

$$\int c d\eta \leq P(f, c) = h(f) + c$$

for every  $c \in \mathbb{R}$ . For  $c > 0$ , this implies that  $\eta(M) \leq 1 + h(f)/c$ . Passing to the limit when  $c \rightarrow +\infty$ , we get that  $\eta(M) \leq 1$ . Analogously, considering  $c < 0$  and taking the limit when  $c \rightarrow -\infty$ , we get  $\eta(M) \geq 1$ . Therefore,  $\eta$  is a probability measure, as stated.

We are left to prove that  $\eta$  is invariant under  $f$ . By assumption, given any  $c \in \mathbb{R}$  and any potential  $\phi$ ,

$$c \int (\phi \circ f - \phi) d\eta \leq P(f, c(\phi \circ f - \phi)).$$

By Proposition 10.3.8, the expression on the right-hand side is equal to  $P(f, 0) = h(f)$ . For  $c > 0$ , this implies that

$$\int (\phi \circ f - \phi) d\eta \leq \frac{h(f)}{c}$$

and, taking the limit when  $c \rightarrow +\infty$ , it follows that  $\int (\phi \circ f - \phi) d\eta \leq 0$ . The same argument, applied to the function  $-\phi$ , shows that  $\int (\phi \circ f - \phi) d\eta \geq 0$ . Hence,  $\int \phi \circ f d\eta = \int \phi d\eta$  for every  $\phi$ . By Proposition A.3.3, this implies that  $f_*\eta = \eta$ .

### 10.4.1 Proof of the upper bound

In this section we prove that, given any invariant probability measure  $\nu$ ,

$$h_\nu(f) + \int \phi d\nu \leq P(f, \phi). \tag{10.4.1}$$

To do this, let  $\mathcal{P} = \{P_1, \dots, P_s\}$  be any finite partition. We are going to show that if  $\alpha$  is an open cover of  $M$  with sufficiently small diameter, depending only on  $\mathcal{P}$ , then

$$h_\nu(f, \mathcal{P}) + \int \phi d\nu \leq \log 4 + P(f, \phi, \alpha). \tag{10.4.2}$$

Making  $\text{diam } \alpha \rightarrow 0$ , it follows that  $h_\nu(f, \mathcal{P}) + \int \phi d\nu \leq \log 4 + P(f, \phi)$  for every finite partition  $\mathcal{P}$ . Hence,  $h_\nu(f) + \int \phi d\nu \leq \log 4 + P(f, \phi)$ . Now replace the transformation  $f$  by  $f^k$  and the potential  $\phi$  by  $\phi_k$ . Note that  $\int \phi_k d\nu = k \int \phi d\nu$ , since  $\nu$  is invariant under  $f$ . Using Propositions 9.1.14 and 10.3.5, we get that

$$kh_\nu(f, \mathcal{P}) + k \int \phi d\nu \leq \log 4 + kP(f, \phi)$$

for every  $k \geq 1$ . Dividing by  $k$  and taking the limit when  $k \rightarrow \infty$ , we get the inequality (10.4.1).

For proving (10.4.2) we need the following elementary fact:

**Lemma 10.4.4.** *Let  $a_1, \dots, a_k$  be real numbers and  $p_1, \dots, p_k$  be non-negative numbers such that  $p_1 + \dots + p_k = 1$ . Let  $A = \sum_{i=1}^k e^{a_i}$ . Then*

$$\sum_{i=1}^k p_i(a_i - \log p_i) \leq \log A.$$

Moreover, the identity holds if and only if  $p_j = e^{a_j} / A$  for every  $j$ .



*Proof.* Write  $t_i = e^{a_i}/A$  and  $x_i = p_i/e^{a_i}$ . Note that  $\sum_{i=1}^k t_i = 1$ . By the concavity property (9.1.8) of the function  $\phi(x) = -x \log x$ ,

$$\sum_{i=1}^k t_i \phi(x_i) \leq \phi\left(\sum_{i=1}^k t_i x_i\right).$$

Note that  $t_i \phi(x_i) = (p_i/A)(a_i - \log p_i)$  and  $\sum_{i=1}^k t_i x_i = 1/A$ . So, the previous inequality may be rewritten as follows:

$$\sum_{i=1}^k \frac{p_i}{A} (a_i - \log p_i) \leq \frac{1}{A} \log A.$$

Multiplying by  $A$  we get the inequality in the statement of the lemma. Moreover, the identity holds if and only if the  $x_i$  are all equal, that is, if and only if there exists  $c$  such that  $p_i = ce^{a_i}$  for every  $i$ . Summing over  $i = 1, \dots, k$  we see that in that case  $c = 1/A$ , as stated.

Since the measure  $\nu$  is regular (Proposition A.3.2), given  $\varepsilon > 0$  we may find compact sets  $Q_i \subset P_i$  such that  $\nu(P_i \setminus Q_i) < \varepsilon$  for every  $i = 1, \dots, s$ . Let  $Q_0$  be the complement of  $\bigcup_{i=1}^s Q_i$  and let  $P_0 = \emptyset$ . Then  $\mathcal{Q} = \{Q_0, Q_1, \dots, Q_s\}$  is a finite partition of  $M$  such that  $\nu(P_i \Delta Q_i) < s\varepsilon$  for every  $i = 0, 1, \dots, s$ . Hence, by Lemma 9.1.6,

$$H_\nu(\mathcal{P}/\mathcal{Q}) \leq \log 2$$

as long as  $\varepsilon > 0$  is sufficiently small (depending only on  $s$ ). Let  $\varepsilon$  and  $\mathcal{Q}$  be fixed from now on and assume that the open cover  $\alpha$  satisfies

$$\text{diam } \alpha < \min\{d(Q_i, Q_j) : 1 \leq i < j \leq s\}. \tag{10.4.3}$$

By Lemma 9.1.11, we have that  $h_\nu(f, \mathcal{P}) \leq h_\nu(f, \mathcal{Q}) + H_\nu(\mathcal{P}/\mathcal{Q}) \leq h_\nu(f, \mathcal{Q}) + \log 2$ . Hence, to prove (10.4.2) it suffices to show that

$$h_\nu(f, \mathcal{Q}) + \int \phi \, d\nu \leq \log 2 + P(f, \phi, \alpha). \tag{10.4.4}$$

To that end, observe that

$$H_\nu(\mathcal{Q}^n) + \int \phi_n \, d\nu \leq \sum_{Q \in \mathcal{Q}^n} \nu(Q) \left( -\log \nu(Q) + \sup_{x \in Q} \phi_n(x) \right)$$

for every  $n \geq 1$ . Then, by Lemma 10.4.4,

$$H_\nu(\mathcal{Q}^n) + \int \phi_n \, d\nu \leq \log \left( \sum_{Q \in \mathcal{Q}^n} \sup_{x \in Q} e^{\phi_n(x)} \right). \tag{10.4.5}$$

Let  $\gamma$  be any finite subcover of  $\alpha^n$ . For each  $Q \in \mathcal{Q}^n$ , consider any point  $x_Q$  in the closure of  $Q$  such that  $\phi_n(x_Q) = \sup_{x \in Q} \phi_n(x)$ . Pick  $U_Q \in \gamma$  such that  $x_Q \in U_Q$ . Then  $\sup_{x \in Q} \phi_n(x) \leq \sup_{y \in U_Q} \phi_n(y)$  for every  $Q \in \mathcal{Q}^n$ . The condition (10.4.3) implies that each element of  $\alpha$  intersects the closure of not more than two elements of  $\mathcal{Q}$ . Therefore, each element of  $\alpha^n$  intersects the closure of not

more than  $2^n$  elements of  $\mathcal{Q}^n$ . In particular, for each  $U \in \gamma$  there exist not more than  $2^n$  sets  $Q \in \mathcal{Q}^n$  such that  $U_Q = U$ . Therefore,

$$\sum_{Q \in \mathcal{Q}^n} \sup_{x \in Q} e^{\phi_n(x)} \leq 2^n \sum_{U \in \gamma} \sup_{y \in U} e^{\phi_n(y)}, \tag{10.4.6}$$

for any finite subcover  $\gamma$  of  $\alpha^n$ . Combining (10.4.5) and (10.4.6),

$$H_\nu(\mathcal{Q}^n) + \int \phi_n d\nu \leq n \log 2 + \log P_n(f, \phi, \alpha).$$

Dividing by  $n$  and taking the limit when  $n \rightarrow \infty$ , we get (10.4.4). This completes the proof of the upper bound (10.4.1).

### 10.4.2 Approximating the pressure

To finish the proof of Theorem 10.4.1, we now show that for every  $\varepsilon > 0$  there exists a probability measure  $\mu$  invariant under  $f$  and such that

$$h_\mu(f) + \int \phi d\mu \geq S(f, \phi, \varepsilon). \tag{10.4.7}$$

Clearly, this implies that the supremum of the values of  $h_\nu(f) + \int \phi d\nu$  when  $\nu$  varies in  $\mathcal{M}_1(f)$  is greater than or equal to  $S(f, \phi) = P(f, \phi)$ .

For each  $n \geq 1$ , let  $E$  be an  $(n, \varepsilon)$ -separated set such that

$$\sum_{y \in E} e^{\phi_n(y)} \geq \frac{1}{2} S_n(f, \phi, \varepsilon). \tag{10.4.8}$$

Denote by  $A$  the expression on the left-hand side of this inequality. Consider the probability measures  $\nu_n$  and  $\mu_n$  defined on  $M$  by

$$\nu_n = \frac{1}{A} \sum_{x \in E} e^{\phi_n(x)} \delta_x \quad \text{and} \quad \mu_n = \sum_{j=0}^{n-1} f_*^j \nu_n.$$

By the definition (10.3.8), recalling also that the space of probability measures is compact (Theorem 2.1.5), we may choose a subsequence  $(n_j)_j \rightarrow \infty$  such that

1.  $\frac{1}{n_j} \log S_{n_j}(f, \phi, \varepsilon)$  converges to  $S(f, \phi, \varepsilon)$ , and
2.  $\mu_{n_j}$  converges, in the weak\* topology, to some probability measure  $\mu$ .

We are going to check that such a measure  $\mu$  is invariant under  $f$  and satisfies (10.4.7). For the reader's convenience, we split the argument into four steps.

**Step 1:** First, we prove that  $\mu$  is invariant. Let  $\varphi : M \rightarrow \mathbb{R}$  be any continuous function. For each  $n \geq 1$ ,

$$\int \varphi d(f_* \mu_n) = \frac{1}{n} \sum_{j=1}^n \int \varphi \circ f^j d\nu_n = \int \varphi d\mu_n + \frac{1}{n} \left( \int \varphi \circ f^n d\nu_n - \int \varphi d\nu_n \right)$$

and, consequently,

$$\left| \int \varphi d(f_*\mu_n) - \int \varphi d\mu_n \right| \leq \frac{2}{n} \sup |\varphi|.$$

Restricting to  $n = n_j$  and taking the limit when  $j \rightarrow \infty$ , we see that  $\int \varphi df_*\mu = \int \varphi d\mu$  for every continuous function  $\varphi : M \rightarrow \mathbb{R}$ . This proves (recall Proposition A.3.3) that  $f_*\mu = \mu$ , as claimed.

**Step 2:** Next, we estimate the entropy with respect to  $\nu_n$ . Let  $\mathcal{P}$  be any finite partition of  $M$  such that  $\text{diam } \mathcal{P} < \varepsilon$  and  $\mu(\partial\mathcal{P}) = 0$ , where  $\partial\mathcal{P}$  denotes the union of the boundaries  $\partial P$  of all sets  $P \in \mathcal{P}$ . The first condition implies that each element of  $\mathcal{P}^n$  contains at most one element of  $E$ . On the other hand, it is clear that every element of  $E$  is contained in some element of  $\mathcal{P}^n$ . Hence,

$$\begin{aligned} H_{\nu_n}(\mathcal{P}^n) &= \sum_{x \in E} -\nu_n(\{x\}) \log \nu_n(\{x\}) = \sum_{x \in E} -\frac{1}{A} e^{\phi_n(x)} \log \left( \frac{1}{A} e^{\phi_n(x)} \right) \\ &= \log A - \frac{1}{A} \sum_{x \in E} e^{\phi_n(x)} \phi_n(x) = \log A - \int \phi_n d\nu_n \end{aligned} \tag{10.4.9}$$

(the last equality follows directly from the definition of  $\nu_n$ ).

**Step 3:** Now we calculate the entropy with respect to  $\mu_n$ . Consider  $1 \leq k < n$ . For each  $r \in \{0, \dots, k-1\}$ , let  $q_r \geq 0$  be the largest integer number such that  $r + kq_r \leq n$ . In other words,  $q_r = \lfloor (n-r)/k \rfloor$ . Then,

$$\mathcal{P}^n = \mathcal{P}^r \vee \left[ \bigvee_{j=0}^{q_r-1} f^{-(kj+r)}(\mathcal{P}^k) \right] \vee f^{-(kq_r+r)}(\mathcal{P}^{n-(kq_r+r)})$$

(the first term is void if  $r = 0$  and the third one is void if  $n = kq_r + r$ ). Therefore,

$$H_{\nu_n}(\mathcal{P}^n) \leq \sum_{j=0}^{q_r-1} H_{\nu_n}(f^{-(kj+r)}(\mathcal{P}^k)) + H_{\nu_n}(\mathcal{P}^r) + H_{\nu_n}(f^{-(kq_r+r)}(\mathcal{P}^{n-(kq_r+r)})).$$

Clearly,  $\#\mathcal{P}^r \leq (\#\mathcal{P})^k$ . Using Lemma 9.1.3, we find that  $H_{\nu_n}(\mathcal{P}^r) \leq k \log \#\mathcal{P}$ . For the same reason, the last term in the previous inequality is also bounded by  $k \log \#\mathcal{P}$ . Then, using the property (9.1.12),

$$H_{\nu_n}(\mathcal{P}_n) \leq \sum_{j=0}^{q_r-1} H_{f_*^{(kj+r)}\nu_n}(\mathcal{P}^k) + 2k \log \#\mathcal{P} \tag{10.4.10}$$

for every  $r \in \{0, \dots, k-1\}$ . Now, it is clear that every number  $i \in \{0, \dots, n-1\}$  may be written in a unique way as  $i = kj + r$  with  $0 \leq j \leq q_r - 1$ . Then, summing (10.4.10) over all the values of  $r$ ,

$$kH_{\nu_n}(\mathcal{P}_n) \leq \sum_{i=0}^{n-1} H_{f_*^i\nu_n}(\mathcal{P}^k) + 2k^2 \log \#\mathcal{P}. \tag{10.4.11}$$

The concavity property (9.1.8) of the function  $\phi(x) = -x \log x$  implies that

$$\frac{1}{n} \sum_{i=0}^{n-1} H_{f_*^i v_n}(\mathcal{P}^k) \leq H_{\mu_n}(\mathcal{P}^k).$$

Combining this inequality with (10.4.11), we see that

$$\frac{1}{n} H_{v_n}(\mathcal{P}_n) \leq \frac{1}{k} H_{\mu_n}(\mathcal{P}^k) + \frac{2k}{n} \log \# \mathcal{P}.$$

On the other hand, by the definition of  $\mu_n$ ,

$$\frac{1}{n} \int \phi_n dv_n = \frac{1}{n} \sum_{j=0}^{n-1} \int \phi \circ f^j dv_n = \int \phi d\mu_n.$$

Thus, the previous inequality yields

$$\frac{1}{n} H_{v_n}(\mathcal{P}_n) + \frac{1}{n} \int \phi_n dv_n \leq \frac{1}{k} H_{\mu_n}(\mathcal{P}^k) + \int \phi d\mu_n + \frac{2k}{n} \log \# \mathcal{P}. \tag{10.4.12}$$

**Step 4:** Finally, we translate the previous estimates to the limit measure  $\mu$ . From (10.4.9) and (10.4.12), we get

$$\frac{1}{k} H_{\mu_n}(\mathcal{P}^k) + \int \phi d\mu_n \geq \frac{1}{n} \log A - \frac{2k}{n} \log \# \mathcal{P}.$$

By the choice of  $E$  in (10.4.8), it follows that

$$\frac{1}{k} H_{\mu_n}(\mathcal{P}^k) + \int \phi d\mu_n \geq \frac{1}{n} \log S_n(f, \phi, \varepsilon) - \frac{1}{n} \log 2 - \frac{2k}{n} \log \# \mathcal{P}. \tag{10.4.13}$$

The choice of the partition  $\mathcal{P}$  with  $\mu(\partial \mathcal{P}) = 0$  implies that  $\mu(\partial \mathcal{P}^k) = 0$  for every  $k \geq 1$ , since

$$\partial \mathcal{P}^k \subset \partial \mathcal{P} \cup f^{-1}(\partial \mathcal{P}) \cup \dots \cup f^{-k+1}(\partial \mathcal{P}).$$

In other words, every element of  $\mathcal{P}^k$  is a continuity set for the measure  $\mu$ . According to Exercise 2.1.1, it follows that  $\mu(P) = \lim_j \mu_{n_j}(P)$  for every  $P \in \mathcal{P}^k$  and, hence,

$$H_{\mu}(\mathcal{P}^k) = \lim_j H_{\mu_{n_j}}(\mathcal{P}^k) \quad \text{for every } k \geq 1.$$

Since the function  $\phi$  is continuous, we also have that  $\int \phi d\mu = \lim_j \int \phi d\mu_{n_j}$ . Therefore, restricting (10.4.13) to the subsequence  $(n_j)_j$  and taking the limit when  $j \rightarrow \infty$ ,

$$\frac{1}{k} H_{\mu}(\mathcal{P}^k) + \int \phi d\mu \geq S(f, \phi, \varepsilon).$$

Taking the limit when  $k \rightarrow \infty$ , we find that

$$h_{\mu}(f, \mathcal{P}) + \int \phi d\mu \geq S(f, \phi, \varepsilon).$$

Then, making  $\varepsilon \rightarrow 0$  (and, consequently,  $\text{diam } \mathcal{P} \rightarrow 0$ ), we get (10.4.7).

This completes the proof of the variational principle (Theorem 10.4.1).

10.4.3 Exercises

- 10.4.1. Let  $f : M \rightarrow M$  be a continuous transformation in a compact metric space  $M$ . Check that  $P(f, \varphi) \leq h(f) + \sup \varphi$  for every continuous function  $\varphi : M \rightarrow \mathbb{R}$ .
- 10.4.2. Show that if  $f : M \rightarrow M$  is a continuous transformation in a compact metric space and  $X \subset M$  is a forward invariant set, meaning that  $f(X) \subset X$ , then  $P(f | X, \varphi | X) \leq P(f, \varphi)$ .
- 10.4.3. Give an alternative proof of Proposition 10.3.8, using the variational principle.
- 10.4.4. Exhibit a continuous transformation  $f : M \rightarrow M$  in a non-compact metric space  $M$  such that  $f$  has no invariant probability measure and yet the topological entropy  $h(f)$  is positive. [Observation: Thus, the variational principle need not hold when the ambient space is not compact.]
- 10.4.5. Given numbers  $\alpha, \beta > 0$  such that  $\alpha + \beta < 1$ , define

$$g : [0, \alpha] \cup [1 - \beta, 1] \rightarrow [0, 1] \quad g(x) = \begin{cases} x/\alpha & \text{if } x \in [0, \alpha] \\ (x - 1)/\beta + 1 & \text{if } x \in [1 - \beta, 1]. \end{cases}$$

Let  $K \subset [0, 1]$  be the Cantor set formed by the points  $x$  such that  $g^n(x)$  is defined for every  $n \geq 0$  and  $f : K \rightarrow K$  be the restriction of  $g$ . Calculate the function  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  defined by  $\psi(t) = P(f, -t \log g')$ . Check that  $\psi$  is convex and decreasing and admits a (unique) zero in  $(0, 1)$ . Show that  $h_\mu(f) < \int \log g' d\mu$  for every probability measure  $\mu$  invariant under  $f$ .

- 10.4.6. Let  $f : M \rightarrow M$  be a continuous transformation in a compact metric space, such that the set of ergodic invariant probability measures is finite. Show that for every potential  $\varphi : M \rightarrow \mathbb{R}$  there exists some invariant probability measure that realizes the supremum in (10.0.1).

10.5 Equilibrium states

Let  $f : M \rightarrow M$  be a continuous transformation and  $\phi : M \rightarrow \mathbb{R}$  be a potential in a compact metric space. In this section we study the fundamental properties of the set  $\mathcal{E}(f, \phi)$  formed by the *equilibrium states*, that is, the invariant probability measures  $\mu$  such that

$$h_\mu(f) + \int \phi d\mu = P(f, \phi) = \sup\{h_\nu(f) + \int \phi d\nu : \nu \in \mathcal{M}_1(f)\}.$$

In the special case  $\phi \equiv 0$  the elements of  $\mathcal{E}(f, \phi)$  are also called *measures of maximal entropy*. Let us start with a few simple examples.

**Example 10.5.1.** If  $f : M \rightarrow M$  has zero topological entropy then every invariant probability measure  $\mu$  is a measure of maximal entropy:  $h_\mu(f) = 0 = h(f)$ . For any potential  $\phi : M \rightarrow \mathbb{R}$ ,

$$P(f, \phi) = \sup\{\int \phi d\nu : \nu \in \mathcal{M}_1(f)\}.$$

Hence,  $\nu$  is an equilibrium state if and only if  $\nu$  maximizes the integral of  $\phi$ . Since the function  $\nu \mapsto \int \phi d\nu$  is continuous and  $\mathcal{M}_1(f)$  is compact, with respect to the weak\* topology, maxima do exist for every potential  $\phi$ .

**Example 10.5.2.** Let  $f_A : M \rightarrow M$  be the linear endomorphism induced in  $\mathbb{T}^d$  by some matrix  $A$  with integer coefficients and non-zero determinant. Let  $\mu$  be the Haar measure on  $\mathbb{T}^d$ . By Propositions 9.4.3 and 10.2.10,

$$h_\mu(f_A) = \sum_{i=1}^d \log^+ |\lambda_i| = h(f),$$

where  $\lambda_1, \dots, \lambda_d$  are the eigenvalues of  $A$ . In particular, the Haar measure is a measure of maximal entropy for  $f$ .

**Example 10.5.3.** Let  $\sigma : \Sigma \rightarrow \Sigma$  be the shift map in  $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$  and  $\mu$  be the Bernoulli measure associated with a probability vector  $p = (p_1, \dots, p_d)$ . As observed in Example 9.1.10,

$$h_\mu(\sigma, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \sum_{i=1}^d -p_i \log p_i.$$

We leave it to the reader (Exercise 10.5.1) to check that this function attains its maximum precisely when the coefficients  $p_i$  are all equal to  $1/d$ . Moreover, in that case  $h_\mu(\sigma) = \log d$ . Recall also (Example 10.1.2) that  $h(\sigma) = \log d$ . Therefore, the Bernoulli measure associated with the vector  $p = (1/d, \dots, 1/d)$  is the only measure of maximal entropy among the Bernoulli measures. In fact, it follows from the theory that we develop in Chapter 12 that  $\mu$  is the unique measure of maximal entropy among *all* invariant measures.

Let us start with the following extension of the variational principle:

**Proposition 10.5.4.** *For every potential  $\phi : M \rightarrow \mathbb{R}$ ,*

$$P(f, \phi) = \sup\{h_\mu(f) + \int \phi d\mu : \mu \text{ invariant and ergodic for } f\}.$$

*Proof.* Consider the function  $\Psi : \mathcal{M}_1(f) \rightarrow \mathbb{R}$  given by  $\Psi(\mu) = h_\mu(f) + \int \phi d\mu$ . For each invariant probability measure  $\mu$ , let  $\{\mu_P : P \in \mathcal{P}\}$  be the corresponding ergodic decomposition. It follows from Theorem 9.6.2 that

$$\Psi(\mu) = \int \Psi(\mu_P) d\hat{\mu}(P). \tag{10.5.1}$$

This implies that the supremum of  $\Psi$  over all the invariant probability measures is less than or equal to the supremum of  $\Psi$  over the ergodic invariant probability measures. Since the opposite inequality is obvious, it follows that the two suprema coincide. By the variational principle (Theorem 10.4.1), the supremum of  $\Psi$  over all the invariant probability measures is equal to  $P(f, \phi)$ .

**Proposition 10.5.5.** *Assume that  $h(f) < \infty$ . Then the set of equilibrium states for any potential  $\phi : M \rightarrow \mathbb{R}$  is a convex subset of  $\mathcal{M}_1(f)$ : more precisely, given  $t \in (0, 1)$  and  $\mu_1, \mu_2 \in \mathcal{M}_1(f)$ ,*

$$(1 - t)\mu_1 + t\mu_2 \in \mathcal{E}(f, \phi) \iff \{\mu_1, \mu_2\} \subset \mathcal{E}(f, \phi).$$

*Moreover, an invariant probability measure  $\mu$  is in  $\mathcal{E}(f, \phi)$  if and only if almost every ergodic component of  $\mu$  is in  $\mathcal{E}(f, \phi)$ .*

*Proof.* As we have seen in (10.3.6), the hypothesis that the topological entropy is finite ensures that  $P(f, \phi) < \infty$  for every potential  $\phi$ . Let us consider the functional  $\Psi(\mu) = h_\mu(f) + \int \phi d\mu$  introduced in the proof of the previous result. By Proposition 9.6.1, this functional is convex:

$$\Psi((1 - t)\mu_1 + t\mu_2) = (1 - t)\Psi(\mu_1) + t\Psi(\mu_2)$$

for every  $t \in (0, 1)$  and any  $\mu_1, \mu_2 \in \mathcal{M}_1(f)$ . Then,  $\Psi((1 - t)\mu_1 + t\mu_2)$  is equal to the supremum of  $\Psi$  if and only if both  $\Psi(\mu_1)$  and  $\Psi(\mu_2)$  are. This proves the first part of the proposition. The proof of the second part is analogous: the relation (10.5.1) implies that  $\Psi(\mu) = \sup \Psi$  if and only if  $\Psi(\mu_P) = \sup \Psi$  for  $\hat{\mu}$ -almost every  $P$ .

**Corollary 10.5.6.** *If  $\mathcal{E}(f, \phi)$  is non-empty then it contains ergodic invariant probability measures. Moreover, the extremal elements of the convex set  $\mathcal{E}(f, \phi)$  are precisely the ergodic measures contained in it.*

*Proof.* To get the first claim it suffices to consider the ergodic components of any element of  $\mathcal{E}(f, \phi)$ . Let us move on to proving the second claim. If  $\mu \in \mathcal{E}(f, \phi)$  is ergodic then (Proposition 4.3.2)  $\mu$  is an extremal element of  $\mathcal{M}_1(f)$  and so it must be an extremal element of  $\mathcal{E}(f, \phi)$ . Conversely, if  $\mu \in \mathcal{E}(f, \phi)$  is not ergodic then we may write

$$\mu = (1 - t)\mu_1 + t\mu_2, \quad \text{with } 0 < t < 1 \text{ and } \mu_1, \mu_2 \in \mathcal{M}_1(f).$$

By Proposition 10.5.5 we have that  $\mu_1, \mu_2 \in \mathcal{E}(f, \phi)$ , which implies that  $\mu$  is not an extremal element of the set  $\mathcal{E}(f, \phi)$ .

In general, the set of equilibrium states *may* be empty. The first example of this kind was given by Gurevič. The following construction is taken from Walters [Wal82]:

**Example 10.5.7.** Let  $f_n : M_n \rightarrow M_n$  be a sequence of homeomorphisms in compact metric spaces such that the sequence  $(h(f_n))_n$  is increasing and bounded. We are going to build a metric space  $M$  and a homeomorphism  $f : M \rightarrow M$  with the following features:

- $M$  is the union of all  $M_n$  with an additional point, denoted as  $\infty$ , endowed with a distance function relative to which  $(M_n)_n$  converges to  $\infty$ .

- $f$  fixes the point  $\infty$  and its restriction to each  $M_n$  coincides with  $f_n$ .

Then we are going to check that  $f : M \rightarrow M$  has no measure of maximal entropy. Let us explain how this is done.

Denote by  $d_n$  the distance in each metric space  $M_n$ . It is no restriction to assume that  $d_n \leq 1$  for every  $n$ . Define  $M = \bigcup_n M_n \cup \{\infty\}$  and consider the distance  $d$  defined in  $M$  by:

$$d(x,y) = \begin{cases} n^{-2}d_n(x,y) & \text{if } x \in X_n \text{ and } y \in X_n \text{ with } n \geq 1 \\ \sum_{i=n}^m i^{-2} & \text{if } x \in X_n \text{ and } y \in X_m \text{ with } n < m \\ \sum_{i=n}^{\infty} i^{-2} & \text{if } x \in X_n \text{ and } y = \infty. \end{cases}$$

We leave it to the reader to check that  $d$  is indeed a distance in  $M$  and that  $(M, d)$  is a compact space. Let  $\beta = \sup_n h(f_n)$ . Since the sets  $\{\infty\}$  and  $M_n, n \geq 1$  are invariant and cover the whole of  $M$ , any ergodic probability measure  $\mu$  of  $f$  satisfies either  $\mu(\{\infty\}) = 1$  or  $\mu(M_n) = 1$  for some  $n \geq 1$ . In the first case,  $h_\mu(f) = 0$ . In the second,  $\mu$  may be viewed as a probability measure invariant under  $f_n$  and, consequently,  $h_\mu(f) \leq h(f_n)$ . In particular,  $h_\mu(f) < \beta$  for every ergodic probability measure  $\mu$  of  $f$ . The previous observation also shows that

$$\begin{aligned} \sup\{h_\mu(f) : \mu \text{ invariant and ergodic for } f\} \\ = \sup_n \sup\{h_\mu(f) : \mu \text{ invariant and ergodic for } f_n\}. \end{aligned}$$

According to Proposition 10.5.4, this means that  $h(f) = \sup_n h(f_n) = \beta$ . Thus, no ergodic invariant measure of  $f$  realizes the topological entropy. By Proposition 10.5.5, it follows that  $f$  has no measure of maximal entropy.

Nevertheless, there is a broad class of transformations for which equilibrium states do exist for every potential:

**Lemma 10.5.8.** *If the entropy function  $\nu \mapsto h_\nu(f)$  is upper semi-continuous then  $\mathcal{E}(f, \phi)$  is compact, relative to the weak\* topology, and non-empty, for any potential  $\phi : M \rightarrow \mathbb{R}$ .*

*Proof.* Let  $(\mu_n)_n$  be a sequence in  $\mathcal{M}_1(f)$  such that

$$h_{\mu_n}(f) + \int \phi d\mu_n \text{ converges to } P(f, \phi).$$

Since  $\mathcal{M}_1(f)$  is compact (Theorem 2.1.5), there exists some accumulation point  $\mu$ . The assumption implies that  $\nu \mapsto h_\nu(f) + \int \phi d\nu$  is upper semi-continuous. Consequently,

$$h_\mu(f) + \int \phi d\mu \geq \liminf_n h_{\mu_n}(f) + \int \phi d\mu_n = P(f, \phi)$$

and so  $\mu$  is an equilibrium state, as stated. Analogously, taking any sequence  $(\nu_n)_n$  in  $\mathcal{E}(f, \phi)$  we see that every accumulation point  $\nu$  is an equilibrium state. This shows that  $\mathcal{E}(f, \phi)$  is closed and, thus, compact.



**Corollary 10.5.9.** *Assume that  $f : M \rightarrow M$  is an expansive continuous transformation in a compact metric space  $M$ . Then every potential  $\phi : M \rightarrow \mathbb{R}$  admits some equilibrium state.*

*Proof.* Just combine Corollary 9.2.17 with Lemma 10.5.8.

The conclusions of Corollaries 9.2.17 and 10.5.9 remain valid when  $f$  is just  $h$ -expansive, in the sense of Exercise 10.2.8. See Bowen [Bow72]. Misiurewicz [Mis73] noted that the same is still true when  $f$  is just *asymptotically  $h$ -expansive*, meaning that  $g_*(f, \varepsilon) \rightarrow 0$  when  $\varepsilon \rightarrow 0$ . Buzzi [Buz97] proved that every  $C^\infty$  diffeomorphism is asymptotically  $h$ -expansive. The corresponding statement for  $h$ -expansivity is false: Burguet, Liao and Yang [BLY] found open sets, in the  $C^2$  topology, formed by diffeomorphisms that are not  $h$ -expansive;  $C^\infty$  diffeomorphisms are dense in such sets.

Combining these results of Misiurewicz and Buzzi, one gets the following theorem of Newhouse [New90]: for every  $C^\infty$  diffeomorphism  $f$ , the entropy function  $\nu \mapsto h_\nu(f)$  is upper semi-continuous and so equilibrium states always exist. Yomdin [Yom87] also proved that the topological entropy function  $f \mapsto h(f)$  is upper semi-continuous in the realm of  $C^\infty$  diffeomorphisms. Both conclusions, Newhouse's and Yomdin's, are usually false for  $C^r$  diffeomorphisms with  $r < \infty$ , according to Misiurewicz [Mis73]. But Liao, Viana and Yang [LVY13] proved that they both extend to  $C^1$  diffeomorphisms away from homoclinic tangencies and any such diffeomorphism is  $h$ -expansive. In particular, equilibrium states always exist in that generality.

Uniqueness of equilibrium states is a much more delicate problem. It is very easy to exhibit transformations with infinitely many ergodic equilibrium states. For example, let  $f : S^1 \rightarrow S^1$  be a circle homeomorphism with infinitely many fixed points. The Dirac measures on those points are ergodic invariant probability measures. Since the topological entropy  $h(f)$  is equal to zero (Example 10.5.1), each of those measures is an equilibrium state for any potential that attains its maximum at the corresponding point.

This type of example is trivial, of course, because the transformation is not transitive. A more interesting question is whether an indivisibility property, such as transitivity or topological mixing, ensures uniqueness of the equilibrium state. It turns out that this is also not true. The first counter-example (called Dyck shift) was exhibited by Krieger [Kri75]. Next, we present a particularly transparent and flexible construction, due to Haydn [Hay]. Other interesting examples were studied by Hofbauer [Hof77].

**Example 10.5.10.** Let  $X = \{*, 1, 2, 3, 4\}$  and consider the subsets  $E = \{2, 4\}$  (the *even* symbols) and  $O = \{1, 3\}$  (the *odd* symbols). We are going to exhibit a compact  $\mathcal{H} \subset X^{\mathbb{Z}}$  invariant under the shift map in  $X^{\mathbb{Z}}$  such that the restriction  $\sigma : \mathcal{H} \rightarrow \mathcal{H}$  is topologically mixing and yet admits two mutually singular invariant

measures,  $\mu_v$  and  $\mu_a$ , such that

$$h_{\mu_v}(\sigma) = h_{\mu_a}(\sigma) = \log 2 = h(\sigma).$$

Let us describe this example. By definition,  $\mathcal{H} = E^{\mathbb{Z}} \cup O^{\mathbb{Z}} \cup \mathcal{H}_*$ , where  $\mathcal{H}_*$  consists of the sequences  $x \in X^{\mathbb{Z}}$  that satisfy the following rule: Whenever one block with  $m$  symbols of one type, even or odd, is followed by another block with  $n$  symbols of the other type, odd or even, the two of them are separated by a block of no less than  $m + n$  symbols  $*$ . In other words, the following configurations are admissible in sequences  $x \in \mathcal{H}_*$ :

$$x = (\dots, *, e_1, \dots, e_m, \underbrace{*, \dots, *}_k, o_1, \dots, o_n, * \dots) \quad \text{or}$$

$$x = (\dots, *, o_1, \dots, o_m, \underbrace{*, \dots, *}_k, e_1, \dots, e_n, * \dots),$$

with  $e_i \in E$ ,  $o_j \in O$  and  $k \geq m + n$ . Observe that a sequence  $x \in \mathcal{H}_*$  may start and/or end with an infinite block of  $*$  but it can neither start nor end with an infinite block of either even or odd type. It is clear that  $\mathcal{H}$  is invariant under the shift map. Haydn [Hay] proved that (see Exercise 10.5.6):

- (i) the shift map  $\sigma : \mathcal{H} \rightarrow \mathcal{H}$  is topologically mixing;
- (ii)  $h(\sigma) = \log 2$ .

We know that  $E^{\mathbb{Z}}$  and  $O^{\mathbb{Z}}$  support Bernoulli measures  $\mu_v$  and  $\mu_a$  with entropy equal to  $\log 2$ . Then,  $\mu_v$  and  $\mu_a$  are measures of maximal entropy for  $\sigma : \mathcal{H} \rightarrow \mathcal{H}$  and they are mutually singular.

Clearly, this construction may be modified to yield transformations with any given number of ergodic measures of maximal entropy. Haydn [Hay] has also shown how to adapt it to construct examples with multiple equilibrium states for other potentials as well.

In Chapters 11 and 12 we study a class of transformations, called *expanding*, for which every Hölder potential admits exactly one equilibrium state. In particular, these transformations are *intrinsically ergodic*, that is, they have a unique measure of maximal entropy.

### 10.5.1 Exercises

- 10.5.1. Show that, among the Bernoulli measures of the shift map  $\sigma : \Sigma \rightarrow \Sigma$  in the space  $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$ , the one with the largest entropy is given by the probability vector  $(1/d, \dots, 1/d)$ .
- 10.5.2. Let  $\sigma : \Sigma \rightarrow \Sigma$  be the shift map in  $\Sigma = \{1, \dots, d\}^{\mathbb{Z}}$  and  $\phi : \Sigma \rightarrow \mathbb{R}$  be a locally constant potential, that is, such that  $\phi$  is constant on each cylinder  $[0; i]$ . Calculate  $P(f, \phi)$  and show that there exists some equilibrium state that is a Bernoulli measure.

- 10.5.3. Let  $\sigma : \Sigma \rightarrow \Sigma$  be the shift map in  $\Sigma = \{1, \dots, d\}^{\mathbb{N}}$ . An invariant probability measure  $\mu$  is called a *Gibbs state* for a potential  $\varphi : \Sigma \rightarrow \mathbb{R}$  if there exist  $P \in \mathbb{R}$  and  $K > 0$  such that

$$K^{-1} \leq \frac{\mu(C)}{\exp(\varphi_n(x) - nP)} \leq K \quad (10.5.2)$$

for every cylinder  $C = [0; i_0, \dots, i_{n-1}]$  and any  $x \in C$ . Prove that if  $\mu$  is a Gibbs state then  $h_\mu(\sigma) + \int \varphi d\mu$  coincides with the constant  $P$  in (10.5.2). Therefore,  $\mu$  is an equilibrium state if and only if  $P = P(\sigma, \varphi)$ . Prove that for each choice of the constant  $P$  there exists at most one ergodic Gibbs state.

- 10.5.4. Let  $f : M \rightarrow M$  be a continuous transformation in a compact metric space and  $\phi : M \rightarrow \mathbb{R}$  be a continuous function. If  $\mu$  is an equilibrium state for  $\phi$ , then the functional  $F_\mu : C^0(M) \rightarrow \mathbb{R}$  defined by  $F_\mu(\psi) = \int \psi d\mu$  is such that  $F_\mu(\psi) \leq P(f, \phi + \psi) - P(f, \phi)$  for every  $\psi \in C^0(M)$ . Conclude that if the pressure function  $P(f, \cdot) : C^0(M) \rightarrow \mathbb{R}$  is differentiable in every direction at a point  $\phi$  then  $\phi$  admits at most one equilibrium state.
- 10.5.5. Let  $f : M \rightarrow M$  be a continuous transformation in a compact metric space. Show that the subset of functions  $\phi : M \rightarrow \mathbb{R}$  for which there exists a unique equilibrium state is residual in  $C^0(M)$ .
- 10.5.6. Check the claims (i) and (ii) in Example 10.5.10.