

## 2

# Existence of invariant measures

In this chapter we prove the following result, which guarantees the existence of invariant measures for a broad class of transformations:

**Theorem 2.1** (Existence of invariant measures). *Let  $f : M \rightarrow M$  be a continuous transformation on a compact metric space. Then there exists some probability measure on  $M$  invariant under  $f$ .*

The main point in the proof is to introduce a certain topology in the set  $\mathcal{M}_1(M)$  of probability measures on  $M$ , that we call weak\* topology. The idea is that two measures are close, with respect to this topology, if the integrals they assign to (many) bounded continuous functions are close. The precise definition and some of the properties of the weak\* topology are presented in Section 2.1. The crucial property, that makes this topology so useful for proving the existence theorem, is that it turns  $\mathcal{M}_1(M)$  into a compact space (Theorem 2.1.5).

The proof of Theorem 2.1 is given in Section 2.2. We will also see, through examples, that the hypotheses of continuity and compactness cannot be omitted.

In Section 2.3 we insert the construction of the weak\* topology into a broader framework from functional analysis and we also take the opportunity to introduce the notion of the Koopman operator of a transformation, which will be very useful in the sequel. In particular, as we are going to see, it allows us to give an alternative proof of Theorem 2.1, based on tools from functional analysis.

In Section 2.4 we describe certain explicit constructions of invariant measures for two important classes of systems: skew-products and natural extensions (or inverse limits) of non-invertible transformations.

Finally, in Section 2.5 we discuss some important applications of the idea of multiple recurrence (Section 1.5) in the context of combinatorial arithmetics. Theorem 2.1.5 has an important role in the arguments, which is the reason why this discussion was postponed to the present chapter.

### 2.1 Weak\* topology

In this section  $M$  will always be a metric space. Our goal is to define the so-called weak\* topology in the set  $\mathcal{M}_1(M)$  of Borel probability measures on  $M$  and to discuss its main properties.

Let  $d(\cdot, \cdot)$  be the distance function on  $M$  and  $B(x, \delta)$  denote the ball of center  $x \in M$  and radius  $\delta > 0$ . Given  $B \subset M$ , we define

$$d(x, B) = \inf\{d(x, y) : y \in B\}$$

and we call the  $\delta$ -neighborhood of  $B$  the set  $B^\delta$  of points  $x \in M$  with  $d(x, B) < \delta$ .

#### 2.1.1 Definition and properties of the weak\* topology

Given a measure  $\mu \in \mathcal{M}_1(M)$ , a finite set  $\Phi = \{\phi_1, \dots, \phi_N\}$  of bounded continuous functions  $\phi_i : M \rightarrow \mathbb{R}$  and a number  $\varepsilon > 0$ , we define

$$V(\mu, \Phi, \varepsilon) = \{v \in \mathcal{M}_1(M) : \left| \int \phi_i dv - \int \phi_i d\mu \right| < \varepsilon \text{ for every } i\}. \quad (2.1.1)$$

Note that the intersection of any two such sets contains some set of this form. Thus, the family  $\{V(\mu, \Phi, \varepsilon) : \Phi, \varepsilon\}$  may be taken as a basis of neighborhoods of each  $\mu \in \mathcal{M}_1(M)$ .

The *weak\* topology* is the topology defined by these bases of neighborhoods. In other words, the open sets in the weak\* topology are the sets  $\mathcal{A} \subset \mathcal{M}_1(M)$  such that for every  $\mu \in \mathcal{A}$  there exists some  $V(\mu, \Phi, \varepsilon)$  contained in  $\mathcal{A}$ . Observe that the definition depends only on the topology of  $M$ , not on its distance. Furthermore, this topology is Hausdorff: Proposition A.3.3 implies that if  $\mu$  and  $\nu$  are distinct probabilities then there exist  $\varepsilon > 0$  and some bounded continuous function  $\phi : M \rightarrow \mathbb{R}$  such that  $V(\mu, \{\phi\}, \varepsilon) \cap V(\nu, \{\phi\}, \varepsilon) = \emptyset$ .

**Lemma 2.1.1.** *A sequence  $(\mu_n)_{n \in \mathbb{N}}$  converges to a measure  $\mu \in \mathcal{M}_1(M)$  in the weak\* topology if and only if*

$$\int \phi d\mu_n \rightarrow \int \phi d\mu \quad \text{for every bounded continuous function } \phi : M \rightarrow \mathbb{R}.$$

*Proof.* To prove the “only if” claim, consider any set  $\Phi = \{\phi\}$  consisting of a single bounded continuous function  $\phi$ . Since  $(\mu_n)_n \rightarrow \mu$ , for any  $\varepsilon > 0$  there exists  $\bar{n} \geq 1$  such that  $\mu_n \in V(\mu, \Phi, \varepsilon)$  for every  $n \geq \bar{n}$ . This means, precisely, that

$$\left| \int \phi d\mu_n - \int \phi d\mu \right| < \varepsilon \quad \text{for every } n \geq \bar{n}.$$

In other words, the sequence  $(\int \phi d\mu_n)_n$  converges to  $\int \phi d\mu$ .

The converse asserts that if  $(\int \phi d\mu_n)_n$  converges to  $\int \phi d\mu$  for every bounded continuous function  $\phi$  then, given any  $\Phi = \{\phi_1, \dots, \phi_N\}$  and  $\varepsilon > 0$ , there exists  $\bar{n} \geq 1$  such that  $\mu_n \in V(\mu, \Phi, \varepsilon)$  for  $n \geq \bar{n}$ . To check that this is

so, let  $\Phi = \{\phi_1, \dots, \phi_N\}$ . The hypothesis ensures that for every  $i$  there exists  $\bar{n}_i$  such that

$$\left| \int \phi_i d\mu_n - \int \phi_i d\mu \right| < \varepsilon \quad \text{for every } n \geq \bar{n}_i.$$

Taking  $\bar{n} = \max\{\bar{n}_1, \dots, \bar{n}_N\}$  we get that  $\mu_n \in V(\mu, \Phi, \varepsilon)$  for every  $n \geq \bar{n}$ .

### 2.1.2 Portmanteau theorem

Now let us discuss other useful ways of defining the weak\* topology. Indeed, the relations (2.1.2), (2.1.3), (2.1.4) and (2.1.5) below introduce other natural choices for neighborhoods of a probability measure  $\mu \in \mathcal{M}_1$ . In Theorem 2.1.2 we prove that all these choices give rise to the same topology in  $\mathcal{M}_1(M)$ , which coincides with the weak\* topology.

A direct variation of the definition of weak\* topology is obtained by taking as the basis of neighborhoods the family of sets

$$V(\mu, \Psi, \varepsilon) = \{\eta \in \mathcal{M}_1(M) : \left| \int \psi_i d\eta - \int \psi_i d\mu \right| < \varepsilon \text{ for every } i\}, \quad (2.1.2)$$

where  $\varepsilon > 0$  and  $\Psi = \{\psi_1, \dots, \psi_N\}$  is a family of Lipschitz functions. The next definition is formulated in terms of closed subsets. Given any finite family  $\mathcal{F} = \{F_1, \dots, F_N\}$  of closed subsets of  $M$  and given any  $\varepsilon > 0$ , consider

$$V_f(\mu, \mathcal{F}, \varepsilon) = \{v \in \mathcal{M}_1 : v(F_i) < \mu(F_i) + \varepsilon \text{ for every } i\}. \quad (2.1.3)$$

The next construction is analogous, just with open subsets instead of closed subsets. Given any finite family  $\mathcal{A} = \{A_1, \dots, A_N\}$  of open subsets of  $M$  and given any  $\varepsilon > 0$ , consider

$$V_a(\mu, \mathcal{A}, \varepsilon) = \{v \in \mathcal{M}_1 : v(A_i) > \mu(A_i) - \varepsilon \text{ for every } i\}. \quad (2.1.4)$$

We call a *continuity set* of a measure  $\mu$  any Borel subset  $B$  of  $M$  whose boundary  $\partial B$  has zero measure for  $\mu$ . Given any finite family  $\mathcal{B} = \{B_1, \dots, B_N\}$  of continuity sets of  $\mu$  and given any  $\varepsilon > 0$ , consider

$$V_c(\mu, \mathcal{B}, \varepsilon) = \{v \in \mathcal{M}_1 : |\mu(B_i) - v(B_i)| < \varepsilon \text{ for every } i\}. \quad (2.1.5)$$

Given any two topologies  $\mathcal{T}_1$  and  $\mathcal{T}_2$  in the same set, we say that  $\mathcal{T}_1$  is *weaker* than  $\mathcal{T}_2$  (or, equivalently, that  $\mathcal{T}_2$  is *stronger* than  $\mathcal{T}_1$ ) if every subset that is open for  $\mathcal{T}_1$  is also open for  $\mathcal{T}_2$ . We say that the two topologies are equivalent if they have exactly the same open sets.

**Theorem 2.1.2.** *The topologies defined by the bases of neighborhoods (2.1.1), (2.1.2), (2.1.3), (2.1.4) and (2.1.5) are all equivalent.*

*Proof.* Since every Lipschitz function is continuous, it is clear that the topology (2.1.2) is weaker than the topology (2.1.1).

To show that the topology (2.1.3) is weaker than the topology (2.1.2), consider any finite family  $\mathcal{F} = \{F_1, \dots, F_N\}$  of closed subsets of  $M$ . According

to Lemma A.3.4, for each  $\delta > 0$  and each  $i$  there exists a Lipschitz function  $\psi_i : M \rightarrow [0, 1]$  such that  $\mathcal{X}_{F_i} \leq \psi_i \leq \mathcal{X}_{F_i^\delta}$ . Observe that  $\bigcap_{\delta} F_i^\delta = F_i$ , because  $F_i$  is closed, and so  $\mu(F_i^\delta) \rightarrow \mu(F_i)$  when  $\delta \rightarrow 0$ . Fix  $\delta > 0$  small enough so that  $\mu(F_i^\delta) - \mu(F_i) < \varepsilon/2$  for every  $i$ . Let  $\Psi$  be the set of functions  $\psi_1, \dots, \psi_N$  obtained in this way. Observe that

$$\left| \int \psi_i dv - \int \psi_i d\mu \right| < \varepsilon/2 \Rightarrow v(F_i) - \mu(F_i^\delta) < \varepsilon/2 \Rightarrow v(F_i) \leq \mu(F_i) + \varepsilon$$

for every  $i$ . In other words,  $V(\mu, \Psi, \varepsilon/2)$  is contained in  $V_f(\mu, \mathcal{F}, \varepsilon)$ .

It is easy to see that the topologies (2.1.3) and (2.1.4) are equivalent. Indeed, let  $\mathcal{F} = \{F_1, \dots, F_n\}$  be any finite family of closed subsets and let  $\mathcal{A} = \{A_1, \dots, A_N\}$ , where each  $A_i$  is the complement of  $F_i$ . Clearly,

$$\begin{aligned} V_f(\mu, \mathcal{F}, \varepsilon) &= \{v \in \mathcal{M}_1 : v(F_i) < \mu(F_i) + \varepsilon \text{ for every } i\} \\ &= \{v \in \mathcal{M}_1 : v(A_i) > \mu(A_i) - \varepsilon \text{ for every } i\} = V_a(\mu, \mathcal{A}, \varepsilon). \end{aligned}$$

Next, let us show that the topology (2.1.5) is weaker than these equivalent topologies (2.1.3) and (2.1.4). Given any finite family  $\mathcal{B} = \{B_1, \dots, B_N\}$  of continuity sets of  $\mu$ , let  $F_i$  be the closure and  $A_i$  be the interior of each  $B_i$ . Denote  $\mathcal{F} = \{F_1, \dots, F_N\}$  and  $\mathcal{A} = \{A_1, \dots, A_N\}$ . Since  $\mu(F_i) = \mu(B_i) = \mu(A_i)$ ,

$$\begin{aligned} v(F_i) < \mu(F_i) + \varepsilon &\Rightarrow v(B_i) < \mu(B_i) + \varepsilon \\ v(A_i) > \mu(A_i) - \varepsilon &\Rightarrow v(B_i) > \mu(B_i) - \varepsilon \end{aligned}$$

for every  $i$ . This means that  $V_f(\mu, \mathcal{F}, \varepsilon) \cap V_a(\mu, \mathcal{A}, \varepsilon)$  is contained in  $V_c(\mu, \mathcal{B}, \varepsilon)$ .

Finally, let us prove that the topology (2.1.1) is weaker than the topology (2.1.5). Let  $\Phi = \{\phi_1, \dots, \phi_N\}$  be a finite family of bounded continuous functions. Fix an integer number  $\ell$  such that  $\sup|\phi_i(x)| < \ell$  for every  $i$ . For each  $i$ , the pre-images  $\phi_i^{-1}(s)$ ,  $s \in [-\ell, \ell]$  are pairwise disjoint. Hence,  $\mu(\phi_i^{-1}(s)) = 0$  except for a countable set of values of  $s$ . In particular, we may choose  $k \in \mathbb{N}$  and points  $-\ell = t_0 < t_1 < \dots < t_{k-1} < t_k = \ell$  such that  $t_j - t_{j-1} < \varepsilon/2$  and  $\mu(\{\phi_i^{-1}(t_j)\}) = 0$  for every  $j$ . Then, each

$$B_{i,j} = \phi_i^{-1}((t_{j-1}, t_j])$$

is a continuity set of  $\mu$ . Moreover,

$$\sum_{j=1}^k t_j \mu(B_{i,j}) \geq \int \phi_i d\mu \geq \sum_{j=1}^k t_{j-1} \mu(B_{i,j}) > \sum_{j=1}^k t_j \mu(B_{i,j}) - \varepsilon/2,$$

and we also have similar inequalities for the integrals relative to  $v$ . It follows that

$$\left| \int \phi_i d\mu - \int \phi_i dv \right| \leq \sum_{j=1}^k \ell |\mu(B_{i,j}) - v(B_{i,j})| + \varepsilon/2 \tag{2.1.6}$$

for every  $i$ . Denote  $\mathcal{B} = \{B_{i,j} : i = 1, \dots, N \text{ and } j = 1, \dots, k\}$ . Then the relation (2.1.6) implies that  $V_c(\mu, \mathcal{B}, \varepsilon/(2k\ell))$  is contained in  $V(\mu, \Phi, \varepsilon)$ .

**2.1.3 The weak\* topology is metrizable**

Now assume that the metric space  $M$  is separable. We will see in Exercise 2.1.3 that the weak\* topology on  $\mathcal{M}_1(M)$  is separable. Here we show that it is also metrizable: we exhibit a distance function on  $\mathcal{M}_1(M)$  that induces the weak\* topology.

Given  $\mu, \nu \in \mathcal{M}_1(M)$ , define  $D(\mu, \nu)$  to be the infimum of all numbers  $\delta > 0$  such that

$$\mu(B) < \nu(B^\delta) + \delta \quad \text{and} \quad \nu(B) < \mu(B^\delta) + \delta \tag{2.1.7}$$

for every Borel set  $B \subset M$ .

**Lemma 2.1.3.** *The function  $D$  is a distance on  $\mathcal{M}_1(M)$ .*

*Proof.* Let us start by showing that  $D(\mu, \nu) = 0$  implies  $\mu = \nu$ . Indeed, the hypothesis implies

$$\mu(B) \leq \nu(\bar{B}) \quad \text{and} \quad \nu(B) \leq \mu(\bar{B})$$

for every Borel set  $B \subset M$ , where  $\bar{B}$  denotes the closure of  $B$ . When  $B$  is closed, these inequalities mean that  $\mu(B) = \nu(B)$ . As we have seen previously, any two measures that coincide on the closed subsets are necessarily the same.

We leave it to the reader to check all the other conditions in the definition of a distance (Exercise 2.1.5).

This distance  $D$  is called the *Levy–Prohorov metric* on  $\mathcal{M}_1(M)$ . In what follows we denote by  $B_D(\mu, r)$  the ball of radius  $r > 0$  around any  $\mu \in \mathcal{M}_1(M)$ .

**Proposition 2.1.4.** *If  $M$  is a separable metric space then the topology induced by the Levy–Prohorov distance  $D$  coincides with the weak\* topology on  $\mathcal{M}_1(M)$ .*

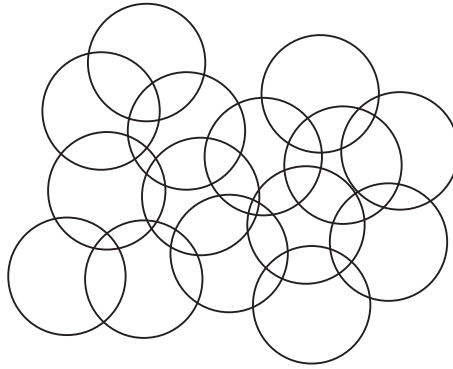
*Proof.* Let  $\varepsilon > 0$  and  $\mathcal{F} = \{F_1, \dots, F_N\}$  be a finite family of closed subsets of  $M$ . Fix  $\delta \in (0, \varepsilon/2)$  such that  $\mu(F_i^\delta) < \mu(F_i) + \varepsilon/2$  for every  $i$ . If  $\nu \in B_D(\mu, \delta)$  then

$$\nu(F_i) < \mu(F_i^\delta) + \delta < \mu(F_i) + \varepsilon \quad \text{for every } i,$$

which means that  $\nu \in V_{\mathcal{F}}(\mu, \mathcal{F}, \varepsilon)$ . This shows that the topology induced by the distance  $D$  is stronger than the topology (2.1.3) which, as we have seen, is equivalent to the weak\* topology.

We are left to prove that if  $M$  is separable then the weak\* topology is stronger than the topology induced by  $D$ . For that, let  $\{p_1, p_2, \dots\}$  be any countable dense subset of  $M$ . Given  $\varepsilon > 0$ , let us fix  $\delta \in (0, \varepsilon/3)$ . For each  $j$ , the spheres  $\partial B(p_j, r) = \{x : d(x, p_j) = r\}$ ,  $r > 0$  are pairwise disjoint. So, we may find  $r > 0$  arbitrarily small such that  $\mu(\partial B(p_j, r)) = 0$  for every  $j$ . Fix any such  $r$ , with  $r \in (0, \delta/3)$ . The family  $\{B(p_j, r) : j = 1, 2, \dots\}$  is a countable cover of  $M$  by continuity sets of  $\mu$ . Fix  $k \geq 1$  such that the set  $U = \bigcup_{j=1}^k B(p_j, r)$  satisfies

$$\mu(U) > 1 - \delta. \tag{2.1.8}$$



**Figure 2.1.** Partition defined by a finite cover

Next, let us consider the (finite) partition  $\mathcal{P}$  of  $U$  defined by the family of balls

$$\{B(p_j, r) : j = 1, \dots, k\}.$$

That is, the elements of  $\mathcal{P}$  are the maximal sets  $P \subset U$  such that, for each  $j$ , either  $P$  is contained in  $B(p_j, r)$  or  $P$  is disjoint from  $B(p_j, r)$ . See Figure 2.1. Now let  $\mathcal{E}$  be the family of all finite unions of elements of  $\mathcal{P}$ . Note that the boundary of every element of  $\mathcal{E}$  has measure zero, since it is contained in the union of the boundaries of the balls  $B(p_j, r)$ ,  $1 \leq j \leq k$ . That is, every element of  $\mathcal{E}$  is a continuity set of  $\mu$ .

If  $\nu \in V_c(\mu, \mathcal{E}, \delta)$  then

$$|\mu(E) - \nu(E)| < \delta \quad \text{for every } E \in \mathcal{E}. \tag{2.1.9}$$

In particular, (2.1.8) together with (2.1.9) imply that

$$\nu(U) > 1 - 2\delta. \tag{2.1.10}$$

Now, given any Borel subset  $B$ , denote by  $E_B$  the union of all the elements of  $\mathcal{P}$  that intersect  $B$ . Then  $E_B \in \mathcal{E}$  and so the relation (2.1.9) yields

$$|\mu(E_B) - \nu(E_B)| < \delta.$$

Observe that  $B$  is contained in  $E_B \cup U^c$ . Moreover,  $E_B \subset B^\delta$  because every element of  $\mathcal{P}$  has diameter less than  $2r < \delta$ . These facts, together with (2.1.8) and (2.1.10), imply that

$$\begin{aligned} \mu(B) &\leq \mu(E_B) + \delta < \nu(E_B) + 2\delta \leq \nu(B^\delta) + 2\delta \\ \nu(B) &\leq \nu(E_B) + 2\delta < \mu(E_B) + 3\delta \leq \mu(B^\delta) + 3\delta. \end{aligned}$$

Since  $3\delta < \varepsilon$ , these relations imply that  $\nu \in B_D(\mu, \varepsilon)$ .

One can show that if  $M$  is a complete separable metric space then the Levy–Prohorov metric on  $\mathcal{M}_1(M)$  is complete (and separable, by Exercise 2.1.3). See, for example, Theorem 6.8 in Billingsley [Bil68].

**2.1.4 The weak\* topology is compact**

In this section we take the metric space  $M$  to be compact. We are going to prove the following fundamental result:

**Theorem 2.1.5.** *The space  $\mathcal{M}_1(M)$  is compact for the weak\* topology.*

Since we already know that  $\mathcal{M}_1(M)$  is metrizable, it suffices to prove:

**Proposition 2.1.6.** *Every sequence  $(\mu_k)_{k \in \mathbb{N}}$  in  $\mathcal{M}_1(M)$  has some subsequence that converges in the weak\* topology.*

*Proof.* Let  $\{\phi_n : n \in \mathbb{N}\}$  be a countable dense subset of the unit ball of  $C^0(M)$  (recall Theorem A.3.13). For each  $n \in \mathbb{N}$ , the sequence of real numbers  $\int \phi_n d\mu_k, k \in \mathbb{N}$  is bounded by 1. Hence, for each  $n \in \mathbb{N}$  there exists a sequence  $(k_j^n)_{j \in \mathbb{N}}$  such that

$$\int \phi_n d\mu_{k_j^n} \text{ converges to some number } \Phi_n \in \mathbb{R} \text{ when } j \rightarrow \infty.$$

Moreover, each sequence  $(k_j^{n+1})_{j \in \mathbb{N}}$  may be chosen to be a subsequence of the previous  $(k_j^n)_{j \in \mathbb{N}}$ . Define  $\ell_j = k_j^j$  for each  $j \in \mathbb{N}$ . By construction,  $(\ell_j)_{j \in \mathbb{N}}$  is a subsequence of every  $(k_j^n)_{j \in \mathbb{N}}$ , up to finitely many terms. Hence,

$$\left( \int \phi_n d\mu_{\ell_j} \right)_j \rightarrow \Phi_n \quad \text{for every } n \in \mathbb{N}.$$

One can easily deduce that

$$\Phi(\varphi) = \lim_j \int \varphi d\mu_{\ell_j} \tag{2.1.11}$$

exists, for every function  $\varphi \in C^0(M)$ . Indeed, suppose first that  $\varphi$  is in the unit ball of  $C^0(M)$ . Given any  $\varepsilon > 0$ , we may find  $n \in \mathbb{N}$  such that  $\|\varphi - \phi_n\| \leq \varepsilon$ . Then,

$$\left| \int \varphi d\mu_{\ell_j} - \int \phi_n d\mu_{\ell_j} \right| \leq \varepsilon$$

for every  $j$ . Since  $\int \phi_n d\mu_{\ell_j}$  converges (to  $\Phi_n$ ), it follows that

$$\limsup_j \int \varphi d\mu_{\ell_j} - \liminf_j \int \varphi d\mu_{\ell_j} \leq 2\varepsilon.$$

Since  $\varepsilon$  is arbitrary, we find that  $\lim_j \int \varphi d\mu_{\ell_j}$  exists. This proves (2.1.11) when the function is in the unit ball. The general case reduces immediately to this one, just replacing  $\varphi$  with  $\varphi/\|\varphi\|$ . In this way, we have completed the proof of (2.1.11).

Finally, it is clear that the operator  $\Phi : C^0(M) \rightarrow \mathbb{R}$  defined by (2.1.11) is linear and positive:  $\Phi(\varphi) \geq \min \varphi \geq 0$  whenever  $\varphi \geq 0$  at all points. Moreover,  $\Phi(1) = 1$ . Thus, by Theorem A.3.11, there exists some Borel probability

measure  $\mu$  on  $M$  such that  $\Phi(\varphi) = \int \varphi d\mu$  for every continuous function  $\varphi$ . Now, the equality in (2.1.11) may be rewritten as

$$\int \varphi d\mu = \lim_j \int \varphi d\mu_{\ell_j} \quad \text{for every } \varphi \in C^0(M).$$

According to Lemma 2.1.1, this means that the subsequence  $(\mu_{\ell_j})_{j \in \mathbb{N}}$  converges to  $\mu$  in the weak\* topology.

As we observed previously, Theorem 2.1.5 is an immediate consequence of the proposition we have just proved.

### 2.1.5 Theorem of Prohorov

The theorem that we are going to state in this section provides a very general criterion for a family of probability measures to be compact. Indeed, the class of metric spaces  $M$  to which it applies includes virtually all interesting examples.

**Definition 2.1.7.** A set  $\mathcal{M}$  of Borel measures in a topological space is *tight* if for every  $\varepsilon > 0$  there exists a compact set  $K \subset M$  such that  $\mu(K^c) < \varepsilon$  for every measure  $\mu \in \mathcal{M}$ .

Note that when  $\mathcal{M}$  consists of a single measure this definition corresponds exactly to Definition A.3.6. Clearly, tightness is a hereditary property: if a set is tight then all its subsets are also tight. Note also that if  $M$  is a compact metric space then the space  $\mathcal{M}_1(M)$  of all probability measures is a tight set. So, the next result is an extension of Theorem 2.1.5:

**Theorem 2.1.8 (Prohorov).** *Let  $M$  be a complete separable metric space. A set  $\mathcal{K} \subset \mathcal{M}_1(M)$  is tight if and only if every sequence in  $\mathcal{K}$  admits some subsequence that is convergent in the weak\* topology of  $\mathcal{M}_1(M)$ .*

*Proof.* We only prove the necessary condition, which is the most useful part of the statement. Then, in Exercise 2.1.8, we invite the reader to prove the converse.

Suppose that  $\mathcal{K}$  is tight. Consider an increasing sequence  $(K_l)_l$  of compact subsets of  $M$  such that  $\eta(K_l^c) \leq 1/l$  for every  $l$  and every  $\eta \in \mathcal{K}$ . Fix any sequence  $(\mu_n)_n$  in  $\mathcal{K}$ . To begin with, we claim that for every  $l$  there exists a subsequence  $(n_j)_j$  and there exists a measure  $\nu_l$  on  $M$  such that  $\nu_l(K_l^c) = 0$  and  $(\mu_{n_j} |_{K_l})_j$  converges to  $\nu_l$ , in the sense that

$$\int_{K_l} \psi d\mu_{n_j} \rightarrow \int_{K_l} \psi d\nu_l \quad \text{for every continuous function } \psi : K_l \rightarrow \mathbb{R}. \quad (2.1.12)$$

Indeed, that is a simple consequence of Theorem 2.1.5: up to restricting to a subsequence, we may suppose that the limit  $b_l = \lim_n \mu_n(K_l)$  exists (note that



$1 \geq b_l \geq 1 - 1/l$ ; it follows from the theorem that the sequence of normalized restrictions

$$((\mu_n | K_l) / \mu_n(K_l))_n$$

admits a subsequence converging to some probability measure  $\eta_l \in \mathcal{M}_1(K_l)$ ; to conclude the proof of the claim it suffices to take  $\eta_l$  to be a probability measure on  $M$  with  $\eta_l(K_l^c) = 0$  and to choose  $\nu_l = b_l \eta_l$ .

Next, using a diagonal argument analogous to the one in Proposition 2.1.6, we may choose a subsequence  $(n_j)_j$  in such a way that (2.1.12) holds, simultaneously, for every  $l \geq 1$ . Observe that the sequence  $(\nu_l)_l$  is non-decreasing: given  $k > l$  and any continuous function  $\phi : M \rightarrow [0, 1]$ ,

$$\int \phi d\nu_l = \lim_j \int_{K_l} \phi d\mu_{n_j} \leq \lim_j \int_{K_k} \phi d\mu_{n_j} = \int \phi d\nu_k.$$

Analogously, for any  $k > l$  and any continuous function  $\phi : M \rightarrow [0, 1]$ ,

$$\int \phi d\nu_k - \int \phi d\nu_l = \lim_j \int_{K_k \setminus K_l} \phi d\mu_{n_j} \leq \limsup_j \mu_{n_j}(K_l^c) \leq 1/l.$$

Using Exercise A.3.5, we may translate this in terms of measures of sets (rather than integrals of functions): for every  $k > l$  and every Borel set  $E \subset M$ ,

$$\nu_l(E) \leq \nu_k(E) \leq \nu_l(E) + 1/l. \tag{2.1.13}$$

Define  $\nu(E) = \lim_l \nu_l(E)$  for each Borel set  $E$ . We claim that  $\nu$  is a probability measure on  $M$ . It is immediate from the definition that  $\nu(\emptyset) = 0$  and that  $\nu$  is additive. Furthermore,  $\nu(M) = \lim_l \nu(K_l) = \lim_l b_l = 1$ . To show that  $\nu$  is countably additive ( $\sigma$ -additive), we use the criterion of continuity at the empty set (Theorem A.1.14). Consider any decreasing sequence  $(B_n)_n$  of Borel subsets of  $M$  with  $\bigcap_n B_n = \emptyset$ . Given  $\varepsilon > 0$ , choose  $l$  such that  $1/l < \varepsilon$ . Since  $\nu_l$  is countably additive, Theorem A.1.14 shows that  $\nu_l(B_n) < \varepsilon$  for every  $n$  sufficiently large. Hence,  $\nu(B_n) \leq \nu_l(B_n) + 1/l < 2\varepsilon$  for every  $n$  sufficiently large. This proves that  $(\nu(B_n))_n$  converges to zero and, by Theorem A.1.14, it follows that  $\nu$  is indeed countably additive.

The definition of  $\nu$  implies (see Exercise 2.1.1 or Exercise 2.1.4) that  $(\nu_l)_l$  converges to  $\nu$  in the weak\* topology. So, given  $\varepsilon > 0$  and any bounded continuous function  $\phi : M \rightarrow \mathbb{R}$ , we have that  $|\int \phi d\nu_l - \int \phi d\nu| < \varepsilon$  for every  $l$  sufficiently large. Fix  $l$  such that, in addition,  $\sup |\phi|/l < \varepsilon$ . Then,

$$\left| \int \phi d\mu_{n_j} - \int \phi d\nu_l \right| \leq \left| \int_{K_l^c} \phi d\mu_{n_j} \right| + \left| \int_{K_l} \phi d\mu_{n_j} - \int_{K_l} \phi d\nu_l \right| \leq 2\varepsilon$$

for every  $j$  sufficiently large. This shows that  $|\int \phi d\mu_{n_j} - \int \phi d\nu| < 3\varepsilon$  whenever  $j$  is large enough. Thus,  $(\mu_{n_j})_j$  converges to  $\nu$  in the weak\* topology.

### 2.1.6 Exercises

2.1.1. Let  $M$  be a metric space and  $(\mu_n)_n$  be a sequence in  $\mathcal{M}_1(M)$ . Show that the following conditions are all equivalent:

1.  $(\mu_n)_n$  converges to a probability measure  $\mu$  in the weak\* topology.
2.  $\limsup_n \mu_n(F) \leq \mu(F)$  for every closed set  $F \subset M$ .
3.  $\liminf_n \mu_n(A) \geq \mu(A)$  for every open set  $A \subset M$ .
4.  $\lim_n \mu_n(B) = \mu(B)$  for every continuity set  $B$  of  $\mu$ .
5.  $\lim_n \int \psi d\mu_n = \int \psi d\mu$  for every Lipschitz function  $\psi : M \rightarrow \mathbb{R}$ .

2.1.2. Fix any dense subset  $\mathcal{F}$  of the unit ball of  $C^0(M)$ . Show that a sequence  $(\mu_n)_{n \in \mathbb{N}}$  of probability measures on  $M$  converges to some  $\mu \in \mathcal{M}_1(M)$  in the weak\* topology if and only if

$$\int \phi d\mu_n \text{ converges to } \int \phi d\mu, \quad \text{for every } \phi \in \mathcal{F}.$$

2.1.3. Show that the subset formed by the measures with finite support is dense in  $\mathcal{M}_1(M)$ , relative to the weak\* topology. Assuming that the metric space  $M$  is separable, conclude that  $\mathcal{M}_1(M)$  is also separable.

2.1.4. The *uniform topology* in  $\mathcal{M}_1(M)$  is defined by the basis of neighborhoods

$$V_u(\mu, \varepsilon) = \{v \in \mathcal{M}_1(M) : |\mu(B) - v(B)| < \varepsilon \text{ for every } B \in \mathcal{B}\},$$

and the *pointwise topology* is defined by the basis of neighborhoods

$$V_p(\mu, \mathcal{B}, \varepsilon) = \{v \in \mathcal{M}_1(M) : |\mu(B_i) - v(B_i)| < \varepsilon \text{ for every } i\},$$

where  $\varepsilon > 0$ ,  $n \geq 1$  and  $\mathcal{B} = \{B_1, \dots, B_N\}$  is a finite family of measurable sets. Check that the uniform topology is stronger than the pointwise topology and the latter is stronger than the weak\* topology. Show, by means of examples, that these relations may be strict.

2.1.5. Complete the proof of Lemma 2.1.3.

2.1.6. Let  $V_k$ ,  $k = 1, 2, \dots$  be random variables with real values, that is, measurable functions  $V_k : (X, \mathcal{B}, \mu) \rightarrow \mathbb{R}$  defined in some probability space  $(X, \mathcal{B}, \mu)$ . The *distribution function* of  $V_k$  is the monotone function  $F_k : \mathbb{R} \rightarrow [0, 1]$  defined by  $F_k(a) = \mu(\{x \in X : V_k(x) \leq a\})$ . We say that  $(V_k)_k$  *converges in distribution* to some random variable  $V$  if  $\lim_k F_k(a) = F(a)$  for every continuity point  $a$  of the distribution function  $F$  of the random variable  $V$ . What does this have to do with the weak\* topology?

2.1.7. Let  $(\mu_n)_{n \in \mathbb{N}}$  be a sequence of probability measures converging to some  $\mu$  in the weak\* topology. Let  $B$  be a continuity set of  $\mu$  with  $\mu(B) > 0$ . Prove that the normalized restrictions  $(\mu_n | B) / \mu_n(B)$  converge to the normalized restriction  $(\mu | B) / \mu(B)$  when  $n \rightarrow \infty$ . What can be said if we replace continuity sets by closed sets or by open sets?

2.1.8. (Converse to the theorem of Prohorov) Prove that if  $\mathcal{K} \subset \mathcal{M}_1(M)$  is such that every sequence in  $\mathcal{K}$  admits some convergent subsequence then  $\mathcal{K}$  is tight.

### 2.2 Proof of the existence theorem

Given any  $f : M \rightarrow M$  and any measure  $\eta$  on  $M$ , we denote by  $f_*\eta$  and call the *iterate (or image) of  $\eta$  under  $f$*  the measure defined by

$$f_*\eta(B) = \eta(f^{-1}(B))$$

for each measurable set  $B \subset M$ . Note that the measure  $\eta$  is invariant under  $f$  if and only if  $f_*\eta = \eta$ .

**Lemma 2.2.1.** *Let  $\eta$  be a measure and  $\phi$  be a bounded measurable function. Then*

$$\int \phi df_*\eta = \int \phi \circ f d\eta. \tag{2.2.1}$$

*Proof.* If  $\phi$  is the characteristic function of a measurable set  $B$  then the relation (2.2.1) means that  $f_*\eta(B) = \eta(f^{-1}(B))$ , which holds by definition. By linearity of the integral, it follows that (2.2.1) holds whenever  $\phi$  is a simple function. Finally, since every bounded measurable function can be approximated by simple functions (see Proposition A.1.33), it follows that the claim in the lemma is true in general.

**Proposition 2.2.2.** *If  $f : M \rightarrow M$  is continuous then  $f_* : \mathcal{M}_1(M) \rightarrow \mathcal{M}_1(M)$  is continuous relative to the weak\* topology.*

*Proof.* Let  $\varepsilon > 0$  and  $\Phi = \{\phi_1, \dots, \phi_n\}$  be any family of bounded continuous functions. Since  $f$  is continuous, the family  $\Psi = \{\phi_1 \circ f, \dots, \phi_n \circ f\}$  also consists of bounded continuous functions. By the previous lemma,

$$\left| \int \phi_i d(f_*\mu) - \int \phi_i d(f_*\nu) \right| = \left| \int (\phi_i \circ f) d\mu - \int (\phi_i \circ f) d\nu \right|$$

and so the left-hand side is smaller than  $\varepsilon$  if the right-hand side is smaller than  $\varepsilon$ . That means that

$$f_*(V(\mu, \Psi, \varepsilon) \subset V(f_*\mu, \Phi, \varepsilon)) \quad \text{for every } \mu, \Phi \text{ and } \varepsilon,$$

and this last fact shows that  $f_*$  is continuous.

At this point, Theorem 2.1 could be deduced from the classical Schauder–Tychonoff fixed point theorem for continuous operators in topological vector spaces. A *topological vector space* is a vector space  $V$  endowed with a topology relative to which both operations of  $V$  (addition and multiplication by a scalar) are continuous. A set  $K \subset V$  is said to be *convex* if  $(1 - t)x + ty \in K$  for every  $x, y \in K$  and every  $t \in [0, 1]$ .

**Theorem 2.2.3** (Schauder–Tychonoff). *Let  $F : V \rightarrow V$  be a continuous transformation on a topological vector space  $V$ . Suppose that there exists a compact convex set  $K \subset V$  such that  $F(K) \subset K$ . Then  $F(v) = v$  for some  $v \in K$ .*

Theorem 2.1 corresponds to the special case when  $V = \mathcal{M}(M)$  is the space of complex measures,  $K = \mathcal{M}_1(M)$  is the space of probability measures on  $M$  and  $F = f_*$  is the action of  $f$  on  $\mathcal{M}(M)$ .

However, the situation of Theorem 2.1 is a lot simpler than the general case of the Schauder–Tychonoff theorem because the operator  $f_*$  besides being continuous is also *linear*. This allows for a direct and elementary proof of Theorem 2.1 that also provides some additional information about the invariant measure.

To that end, let  $\nu$  be any probability measure on  $M$ : for example,  $\nu$  could be the Dirac mass at any point. Form the sequence of probability measures

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} f_*^j \nu, \tag{2.2.2}$$

where  $f_*^j \nu$  is the image of  $\nu$  under the iterate  $f^j$ . By Theorem 2.1.5, this sequence has some accumulation point, that is, there exists some subsequence  $(n_k)_{k \in \mathbb{N}}$  and some probability measure  $\mu \in \mathcal{M}_1(M)$  such that

$$\frac{1}{n_k} \sum_{j=0}^{n_k-1} f_*^j \nu \rightarrow \mu \tag{2.2.3}$$

in the weak\* topology. Now we only need to prove:

**Lemma 2.2.4.** *Every accumulation point of a sequence  $(\mu_n)_{n \in \mathbb{N}}$  of the form (2.2.2) is a probability measure invariant under  $f$ .*

*Proof.* The relation (2.2.3) asserts that, given any family  $\Phi = \{\phi_1, \dots, \phi_N\}$  of bounded continuous functions and given any  $\varepsilon > 0$ , we have

$$\left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \int (\phi_i \circ f^j) d\nu - \int \phi_i d\mu \right| < \varepsilon/2 \tag{2.2.4}$$

for every  $i$  and every  $k$  sufficiently large. By Proposition 2.2.2,

$$f_* \mu = f_* \left( \lim_k \frac{1}{n_k} \sum_{j=0}^{n_k-1} f_*^j \nu \right) = \lim_k \frac{1}{n_k} \sum_{j=1}^{n_k} f_*^j \nu. \tag{2.2.5}$$

Now observe that

$$\begin{aligned} & \left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \int (\phi_i \circ f^j) d\nu - \frac{1}{n_k} \sum_{j=1}^{n_k} \int (\phi_i \circ f^j) d\nu \right| \\ &= \frac{1}{n_k} \left| \int \phi_i d\nu - \int (\phi_i \circ f^{n_k}) d\nu \right| \leq \frac{2}{n_k} \sup |\phi_i|, \end{aligned}$$

and the latter expression is smaller than  $\varepsilon/2$  for every  $i$  and every  $k$  sufficiently large. Combining this fact with (2.2.4), we conclude that

$$\left| \frac{1}{n_k} \sum_{j=1}^{n_k} \int (\phi_i \circ f^j) dv - \int \phi_i d\mu \right| < \varepsilon \tag{2.2.6}$$

for every  $i$  and every  $k$  sufficiently large. This means that

$$\frac{1}{n_k} \sum_{j=1}^{n_k} f_*^j \nu \rightarrow \mu$$

when  $k \rightarrow \infty$ . However, (2.2.5) means that this sequence converges to  $f_* \mu$ . By uniqueness of the limit, it follows that  $f_* \mu = \mu$ .

Now the proof of Theorem 2.1 is complete. The examples that follow show that neither of the two hypotheses in the theorem, continuity and compactness, may be omitted.

**Example 2.2.5.** Consider  $f : (0, 1] \rightarrow (0, 1]$  given by  $f(x) = x/2$ . Suppose that  $f$  admits some invariant probability measure: we are going to show that this is actually not true. By the recurrence theorem (Theorem 1.2.4), relative to that probability measure almost every point of  $(0, 1]$  is recurrent. However, it is clear that there are no recurrent points: the orbit of every  $x \in (0, 1]$  converges to zero and, in particular, does not accumulate on the initial point  $x$ . Hence,  $f$  is an example of a continuous transformation (on a non-compact space) that does not have any invariant probability measure.

**Example 2.2.6.** Modifying a little the previous construction, we see that the same phenomenon may occur in compact spaces, if the transformation is not continuous. Consider  $f : [0, 1] \rightarrow [0, 1]$  given by  $f(x) = x/2$  if  $x \neq 0$  and  $f(0) = 1$ . For the same reason as before, no point  $x \in (0, 1]$  is recurrent. So, if there exists some invariant probability measure  $\mu$  then it must give full weight to the sole recurrent point  $x = 0$ . In other words,  $\mu$  must be the Dirac mass supported at zero, that is, the measure  $\delta_0$  defined by

$$\delta_0(E) = 1 \text{ if } 0 \in E \quad \text{and} \quad \delta_0(E) = 0 \text{ if } 0 \notin E.$$

However, the measure  $\delta_0$  is *not* invariant under  $f$ : for example, the measurable set  $E = \{0\}$  has measure 1 and yet its pre-image  $f^{-1}(E)$  is the empty set, which has measure zero. Thus, this transformation  $f$  has no invariant probability measures.

Our third example is of a different nature. We include it to stress the limitations of Theorem 2.1 (which are inherent to its great generality): the measures whose existence is ensured by the theorem may be completely trivial; for example, in the situation that we are going to describe “almost every point” just means the point  $x = 0$ . For this reason, an important objective in ergodic

theory is to construct more sophisticated invariant measures, with additional interesting properties such as, for instance, being equivalent to the Lebesgue measure.

**Example 2.2.7.** Consider  $f : [0, 1] \rightarrow [0, 1]$  given by  $f(x) = x/2$ . This is a continuous transformation on a compact space. So, by Theorem 2.1,  $f$  admits some invariant probability measure. Using the same arguments as in the previous example, we find that there exists a unique invariant probability measure, namely, the Dirac mass  $\delta_0$  at the origin. Note that in this case the measure  $\delta_0$  is indeed invariant.

As an immediate application of Theorem 2.1, we have the following alternative proof of the Birkhoff recurrence theorem (Theorem 1.2.6). Suppose that  $f : M \rightarrow M$  is a continuous transformation on a compact metric space. By Theorem 2.1, there exists some  $f$ -invariant probability measure  $\mu$ . Every compact metric space admits a countable basis of open sets. So, we may apply Theorem 1.2.4 to conclude that  $\mu$ -almost every point is recurrent. In particular, the set of recurrent points is non-empty, as stated by Theorem 1.2.6.

## 2.2.1 Exercises

2.2.1. Prove the following generalization of Lemma 2.2.4. Let  $f : M \rightarrow M$  be a continuous transformation on a compact metric space,  $\nu$  be a probability measure on  $M$  and  $(I_n)_n$  be a sequence of intervals of natural numbers such that  $\#I_n$  converges to infinity when  $n$  goes to infinity. Then every accumulation point of the sequence

$$\mu_n = \frac{1}{\#I_n} \sum_{j \in I_n} f_*^j \nu$$

is an  $f$ -invariant probability measure.

2.2.2. Let  $f_1, \dots, f_q : M \rightarrow M$  be any finite family of *commuting* continuous transformations on a compact metric space. Prove that there exists some probability measure  $\mu$  that is invariant under  $f_i$  for every  $i \in \{1, \dots, q\}$ . In fact, the conclusion remains true for any countable family  $\{f_j : j \in \mathbb{N}\}$  of commuting continuous transformations on a compact metric space.

2.2.3. Let  $f : [0, 1] \rightarrow [0, 1]$  be the decimal expansion transformation. Show that for every  $k \geq 1$  there exists some invariant probability measure whose support is formed by exactly  $k$  points (in particular,  $f$  admits infinitely many invariant probability measures). Determine whether there are invariant probability measures  $\mu$  such that

- the support of  $\mu$  is infinite countable;
- the support of  $\mu$  is non-countable but has empty interior;
- the support of  $\mu$  has non-empty interior but  $\mu$  is singular with respect to the Lebesgue measure  $m$ .

- 2.2.4. Prove the theorem of existence of invariant measures for continuous flows: every continuous flow  $(f^t)_{t \in \mathbb{R}}$  on a compact metric space admits some invariant probability measure.
- 2.2.5. Show that the transformation  $f : [-1, 1] \rightarrow [-1, 1]$ ,  $f(x) = 1 - 2x^2$  has some invariant probability measure equivalent to the Lebesgue on the interval.
- 2.2.6. Let  $f : M \rightarrow M$  be an invertible measurable transformation and  $m$  be a probability measure on  $M$  such that  $m(A) = 0$  if and only if  $m(f(A)) = 0$ . We say that the pair  $(f, m)$  is *totally dissipative* if there exists a measurable set  $W \subset M$  whose iterates  $f^j(W)$ ,  $j \in \mathbb{Z}$  are pairwise disjoint and such that their union has full measure. Prove that if  $(f, m)$  is totally dissipative then  $f$  admits some  $\sigma$ -finite invariant measure equivalent to Lebesgue measure  $m$ . This measure is necessarily infinite.
- 2.2.7. Let  $f : M \rightarrow M$  be an invertible measurable transformation and  $m$  be a probability measure on  $M$  such that  $m(A) = 0$  if and only if  $m(f(A)) = 0$ . We say that the pair  $(f, m)$  is *conservative* if there is no measurable set  $W \subset M$  with positive measure whose iterates  $f^j(W)$ ,  $j \in \mathbb{Z}$  are pairwise disjoint. Show that if  $(f, m)$  is conservative then, for every measurable set  $X \subset M$ ,  $m$ -almost every point of  $X$  returns to  $X$  infinitely times.
- 2.2.8. Suppose that  $(f, m)$  is conservative. Show that  $f$  admits a  $\sigma$ -finite invariant measure  $\mu$  equivalent to  $m$  if and only if there exist sets  $X_1 \subset \dots \subset X_n \subset \dots$  with  $M = \bigcup_n X_n$  and  $m(X_n) < \infty$  for every  $n$ , such that the first-return map  $f_n$  to each  $X_n$  admits a finite invariant measure  $\mu_n$  absolutely continuous with respect to the restriction of  $m$  to  $X_n$ .
- 2.2.9. Find conservative pairs  $(f, m)$  such that  $f$  has no finite invariant measures equivalent to  $m$ . [Observation: Ornstein [Orn60] gave examples such that  $f$  does not even have  $\sigma$ -finite invariant measures equivalent to  $m$ .]

## 2.3 Comments in functional analysis

The definition of weak\* topology in the space of probability measures is a special case of a construction from functional analysis that is worthwhile recalling here. It leads us to introducing a certain linear isometry  $U_f$  in the space  $L^1(\mu)$ , called the *Koopman operator* of the system  $(f, \mu)$ . These operators have an important role in ergodic theory because they allow for powerful tools from Analysis to be used in the study of invariant measures. To illustrate this fact, we present an alternative proof of Theorem 2.1 based on the spectral properties of the Koopman operator.

### 2.3.1 Duality and weak topologies

Let  $E$  be a Banach space, that is, a vector space endowed with a complete norm. The *dual* of  $E$  is the space  $E^*$  of all continuous linear functionals defined on  $E$ . This is also a Banach space, with the norm

$$\|g\| = \sup \left\{ \frac{|g(v)|}{\|v\|} : v \in E \setminus \{0\} \right\}. \quad (2.3.1)$$

The *weak topology* in the space  $E$  is the topology defined by the following basis of neighborhoods:

$$V(v, \{g_1, \dots, g_N\}, \varepsilon) = \{w \in E : |g_i(v) - g_i(w)| < \varepsilon \text{ for every } i\}, \quad (2.3.2)$$

where  $g_1, \dots, g_N \in E^*$ . In terms of sequences, it satisfies

$$(v_n)_n \rightarrow v \quad \Rightarrow \quad (g(v_n))_n \rightarrow g(v) \quad \text{for every } g \in E^*.$$

The *weak\* topology* in the dual space  $E^*$  is the topology defined by the following basis of neighborhoods:

$$V^*(g, \{v_1, \dots, v_N\}, \varepsilon) = \{h \in E^* : |g(v_i) - h(v_i)| < \varepsilon \text{ for every } i\}, \quad (2.3.3)$$

where  $v_1, \dots, v_N \in E$ . It satisfies

$$(g_n)_n \rightarrow g \quad \Rightarrow \quad (g_n(v))_n \rightarrow g(v) \quad \text{for every } v \in E.$$

The weak\* topology has the following remarkable property:

**Theorem 2.3.1** (Banach–Alaoglu). *The closed unit ball of  $E^*$  is compact for the weak\* topology.*

The construction carried out in the previous sections corresponds to the situation where  $E$  is the space  $C^0(M)$  of (complex) continuous functions and  $E^*$  is the space  $\mathcal{M}(M)$  of complex measures on a compact metric space  $M$ : according to the theorem of Riesz–Markov (Theorem A.3.12),  $\mathcal{M}(M)$  corresponds to the dual of  $C^0(M)$  when we identify each measure  $\mu \in \mathcal{M}(M)$  with the linear functional  $I_\mu(\phi) = \int \phi d\mu$ . Note that the definition of the norm (2.3.1) implies that

$$\|\mu\| = \sup \left\{ \frac{|\int \phi d\mu|}{\sup |\phi|} : \phi \in C^0(M) \setminus \{0\} \right\}.$$

In particular, the set  $\mathcal{M}_1(M)$  of probability measures is contained in the unit ball of  $\mathcal{M}(M)$ . Since this set is closed for the weak\* topology, we conclude that Theorem 2.1.5 is also a direct consequence of the theorem of Banach–Alaoglu.

Now consider any continuous transformation  $f : M \rightarrow M$  and the corresponding action  $f_* : \mathcal{M}(M) \rightarrow \mathcal{M}(M)$ ,  $\mu \mapsto f_*\mu$  in the space of complex measures.

Then  $f_*$  is a linear operator on  $\mathcal{M}(M)$  and it is continuous with respect to the weak\* topology. There exists another continuous linear operator naturally associated with  $f$ , namely  $U_f : C^0(M) \rightarrow C^0(M)$ ,  $\phi \mapsto \phi \circ f$ . Observe that these two operators are dual, in the following sense (remember Lemma 2.2.1):

$$\int U_f(\phi) d\mu = \int (\phi \circ f) d\mu = \int \phi d(f_*\mu). \quad (2.3.4)$$

These observations motivate the important notion that we are going to introduce in the next section.



### 2.3.2 Koopman operator

Let  $(M, \mathcal{B})$  be a measurable space,  $f : M \rightarrow M$  be a measurable transformation and  $\mu$  be an  $f$ -invariant measure. The *Koopman operator* of  $(f, \mu)$  is the linear operator

$$U_f : L^1(\mu) \rightarrow L^1(\mu), \quad U_f(\phi) = \phi \circ f.$$

Note that  $U_f$  is well defined and is an *isometry*, that is, it preserves the norm in the Banach space  $L^1(\mu)$ : since  $\mu$  is invariant under  $f$ ,

$$\|U_f(\phi)\|_1 = \int |U_f(\phi)| d\mu = \int |\phi| \circ f d\mu = \int |\phi| d\mu = \|\phi\|_1. \quad (2.3.5)$$

Moreover,  $U_f$  is a *positive* linear operator:  $U_f(\phi) \geq 0$  at  $\mu$ -almost every point whenever  $\phi \geq 0$  at  $\mu$ -almost every point. For future reference, we summarize these facts in the following proposition:

**Proposition 2.3.2.** *The Koopman operator  $U_f : L^1(M) \rightarrow L^1(M)$  of any system  $(f, \mu)$  is a positive linear isometry.*

The property (2.3.5) implies that the operator  $U_f$  is injective. In general,  $U_f$  is not surjective (see Exercise 2.3.5). It is clear that if  $f$  is invertible then  $U_f$  is an isomorphism: the inverse is just the Koopman operator  $U_{f^{-1}}$  of the inverse of  $f$ .

We may also consider versions of the Koopman operator defined on the spaces  $L^p(\mu)$ ,

$$U_f : L^p(\mu) \rightarrow L^p(\mu), \quad U_f(\phi) = \phi \circ f$$

for each  $p \in [1, \infty]$ . Proposition 2.3.2 remains valid in all these cases: all these operators are positive linear isometries.

When  $M$  is a compact metric space and  $f$  is continuous, it is particularly interesting to investigate the action of  $U_f$  restricted to the space  $C^0(M)$  of continuous functions:

$$U_f : C^0(M) \rightarrow C^0(M).$$

It is clear that this operator is continuous relative to the norm of uniform convergence. As we have seen previously, the dual space of  $C^0(M)$  is naturally identified with the space  $\mathcal{M}(M)$  of complex measures on  $M$ . Moreover, the relation (2.3.4) shows that, under that identification, the dual operator

$$U_f^* : C^0(M)^* \rightarrow C^0(M)^*$$

corresponds precisely to the action  $f_* : \mathcal{M}(M) \rightarrow \mathcal{M}(M)$  of the transformation  $f$  on  $\mathcal{M}(M)$ . This fact allows us to give an alternative proof of Theorem 2.1, based on certain facts from spectral theory.

For that, we need to recall some notions from the theory of positive linear operators. The reader can find a lot more details in Deimling [Dei85], including the proofs of the results quoted in the following.

Let  $E$  be a Banach space. A closed convex subset  $C$  is called a *cone* of  $E$  if it satisfies:

$$\lambda C \subset C \text{ for every } \lambda \geq 0 \quad \text{and} \quad C \cap (-C) = \{0\}. \tag{2.3.6}$$

We call the cone  $C$  *normal* if

$$\inf\{\|x + y\| : x, y \in C \text{ such that } \|x\| = \|y\| = 1\} > 0.$$

Let us fix a cone  $C$  of  $E$ . Given any continuous linear operator  $T : E \rightarrow E$ , we say that  $T$  is *positive over  $C$*  if the image  $T(C) \subset C$ . Given a continuous linear functional  $\phi : E \rightarrow \mathbb{R}$ , we say that  $\phi$  is *positive over  $C$*  if  $\phi(v) \geq 0$  for every  $v \in C$ . By definition, the *dual cone*  $C^*$  is the cone of  $E^*$  formed by all the linear functionals positive over  $C$ .

**Example 2.3.3.** The cone  $C_+^0(M) = \{\varphi \in C^0(M) : \varphi \geq 0\}$  is a normal cone of  $C^0(M)$  (Exercise 2.3.3). By the Riesz–Markov theorem (Theorem A.3.11), the dual cone is naturally identified with the space of finite positive measures on  $M$ .

Denote by  $r(T)$  the *spectral radius* of the continuous linear operator  $T$ :

$$r(T) = \lim_n \sqrt[n]{\|T^n\|}.$$

Then  $r(T) = r(T^*)$ , where  $T^* : E^* \rightarrow E^*$  represents the linear operator dual to  $T$ . The next result is a consequence of the theorem of Banach–Mazur; see Proposition 7.2 in Deimling [Dei85]:

**Theorem 2.3.4.** *Let  $C$  be a normal cone of a Banach space  $E$  and  $T : E \rightarrow E$  be a linear operator positive over  $C$ . Then  $r(T^*)$  is an eigenvalue of the dual operator  $T^* : E^* \rightarrow E^*$  and it admits some eigenvector  $v^* \in C^*$ .*

As an application, let us give an alternative proof of the existence of invariant measures for continuous transformations on compact spaces. Consider the cone  $C = C_+^0(M)$  of  $E = C^0(M)$ . As we observed before, the dual cone  $C^*$  is the space of finite positive measures on  $M$ . It is clear from the definition that the operator  $T = U_f$  is positive over  $C$ . Also, its spectral radius is equal to 1, since  $\sup|T(\varphi)| \leq \sup|\varphi|$  for every  $\varphi \in C^0(M)$  and  $T(1) = 1$ . So, by Theorem 2.3.4, there exists some finite positive measure  $\mu$  on  $M$  that is an eigenvalue of the dual operator  $T^* = f_*$  associated with the eigenvalue 1. In other words, the measure  $\mu$  is invariant. Multiplying by a suitable constant, we may suppose that  $\mu$  is a probability measure.

### 2.3.3 Exercises

- 2.3.1. Let  $\ell^1$  be the space of summable sequences of complex numbers, endowed with the norm  $\|(a_n)_n\|_1 = \sum_{n=0}^\infty |a_n|$ . Let  $\ell^\infty$  be the space of bounded sequences and  $c_0$  be the space of sequences converging to zero, both endowed with the norm  $\|(a_n)_n\|_\infty = \sup_{n \geq 0} |a_n|$ .

- (a) Check that  $\ell^\infty$ ,  $\ell^1$  and  $c_0$  are Banach spaces.
  - (b) Show that the map  $(a_n)_n \mapsto [(b_n)_n \mapsto \sum_n a_n b_n]$  defines norm-preserving isomorphisms from  $\ell^\infty$  to the dual space  $(\ell^1)^*$  and from  $\ell^1$  to the dual space  $(c_0)^*$ .
- 2.3.2. Show that a sequence  $(x^k)_k$  in  $\ell^1$  (write  $x^k = (x_n^k)_n$  for each  $k$ ) converges in the topology defined by the norm if and only if it converges in the weak topology, that is, if and only if  $(\sum_n a_n x_n^k)_k$  converges for every  $(a_n)_n \in \ell^\infty$ . [Observation: This does not imply that the two topologies are the same. Why not?] Show that this is no longer true if we replace the weak topology by the weak\* topology.
- 2.3.3. Check that  $C_+^0(M)$  is a normal cone.
- 2.3.4. Let  $R_\theta : S^1 \rightarrow S^1$  be an irrational rotation and  $m$  be the Lebesgue measure on the circle. Calculate the eigenvalues and the eigenvectors of the Koopman operator  $U_\theta : L^2(m) \rightarrow L^2(m)$ . Show that the spectrum of  $U_\theta$  coincides with the unit circle  $\{z \in \mathbb{C} : |z| = 1\}$ .
- 2.3.5. Show, through examples, that the Koopman operator  $U_f$  need not be surjective.
- 2.3.6. Let  $U : H \rightarrow H$  be an isometry of a Hilbert space  $H$ . By Exercise A.6.8, the image of  $U$  is a closed subspace of  $H$ . Deduce that there exist closed subspaces  $V$  and  $W$  such that  $U(V) = V$ , the iterates of  $W$  are pairwise orthogonal and orthogonal to  $V$ , and

$$H = V \oplus \bigoplus_{n=0}^{\infty} U^n(W).$$

Furthermore,  $U$  is an isomorphism if and only if  $W = \{0\}$ .

- 2.3.7. Let  $\phi : E \rightarrow \mathbb{R}$  be a continuous convex functional on a separable Banach space  $E$ . Assume that  $\phi$  is differentiable in all directions at some point  $u \in E$ . Prove that there exists at most one bounded linear functional  $T : E \rightarrow \mathbb{R}$  tangent to  $\phi$  at  $u$ , that is, such that  $T(v) \leq \phi(u+v) - \phi(u)$  for every  $v \in E$ . If  $\phi$  is differentiable at  $u$  then the derivative  $D\phi(u)$  is a linear functional tangent to  $\phi$  at  $u$ . [Observation: The smoothness theorem of Mazur (Theorem 1.20 in Phelps [Phe93]) states that the set of points where  $\phi$  is differentiable and, consequently, there exists a unique linear functional tangent to  $\phi$  is a residual subset of  $E$ .]

## 2.4 Skew-products and natural extensions

In this section we describe two general constructions that are quite useful in ergodic theory. The first one is a basic model for the situation where two dynamical systems are coupled in the following way: the first system is autonomous but the second one is not, because its evolution depends on the evolution of the former. The second construction associates an invertible system with any given system, in such a way that their invariant measures are in one-to-one correspondence. This permits reduction to the invertible case for many statements about general, not necessarily invertible systems.

### 2.4.1 Measures on skew-products

Let  $(X, \mathcal{A})$  and  $(Y, \mathcal{B})$  be measurable spaces. We call a *skew-product* any measurable transformation  $F : X \times Y \rightarrow X \times Y$  of the form  $F(x, y) =$

$(f(x), g(x, y))$ . Represent by  $\pi : X \times Y \rightarrow X$  the canonical projection to the first coordinate. By definition,

$$\pi \circ F = f \circ \pi. \tag{2.4.1}$$

Let  $m$  be a probability measure on  $X \times Y$  invariant under  $F$  and let  $\mu = \pi_* m$  be its projection to  $X$ . Using that  $m$  is invariant under  $F$ , we get that

$$f_* \mu = f_* \pi_* m = \pi_* F_* m = \pi_* m = \mu,$$

that is,  $\mu$  is invariant under  $f$ . The next proposition provides a partial converse to this conclusion: under suitable hypotheses, every  $f$ -invariant measure is the projection of some  $F$ -invariant measure.

**Proposition 2.4.1.** *Let  $X$  be a complete separable metric space,  $Y$  be a compact metric space and  $F$  be continuous. Then, for every probability measure  $\mu$  on  $X$  invariant under  $f$  there exists some probability measure  $m$  on  $X \times Y$  invariant under  $F$  and such that  $\pi_* m = \mu$ .*

*Proof.* Given any  $f$ -invariant probability measure  $\mu$  on  $X$ , let  $\mathcal{K} \subset \mathcal{M}_1(X \times Y)$  be the set of measures  $\eta$  on  $X \times Y$  such that  $\pi_* \eta = \mu$ . Consider any  $\eta \in \mathcal{K}$ . Then,  $\pi_* F_* \eta = f_* \pi_* \eta = f_* \mu = \mu$ . This proves that  $\mathcal{K}$  is invariant under  $F_*$ . Next, note that the projection  $\pi : X \times Y \rightarrow X$  is continuous and, thus, the operator  $\pi_*$  is continuous relative to the weak\* topology. So,  $\mathcal{K}$  is closed in  $\mathcal{M}_1(X \times Y)$ . By Proposition A.3.7, given any  $\varepsilon > 0$  there exists a compact set  $K \subset X$  such that  $\mu(K^c) < \varepsilon$ . Then  $K \times Y$  is compact and  $\eta((K \times Y)^c) = \mu(K^c) < \varepsilon$  for every  $\eta \in \mathcal{K}$ . This proves that the set  $\mathcal{K}$  is tight. Consider any  $\eta \in \mathcal{K}$ . By the theorem of Prohorov (Theorem 2.1.8), the sequence

$$\frac{1}{n} \sum_{j=0}^{n-1} F_*^j \eta$$

has some accumulation point  $m \in \mathcal{K}$ . Arguing as in the proof of Lemma 2.2.4, we conclude that  $m$  is invariant under  $F$ .

### 2.4.2 Natural extensions

We are going to see that, given any surjective transformation  $f : M \rightarrow M$ , one can always find an extension  $\hat{f} : \hat{M} \rightarrow \hat{M}$  that is invertible. By *extension* we mean that there exists a surjective map  $\pi : \hat{M} \rightarrow M$  such that  $\pi \circ \hat{f} = f \circ \pi$ . This fact is very useful, for it makes it possible to reduce to the invertible case the proofs of many statements about general systems. We comment on the surjective hypothesis in Example 2.4.2: we will see that this hypothesis can be omitted in many interesting cases.

To begin with, take  $\hat{M}$  to be the set of all *pre-orbits* of  $f$ , that is, all sequences  $(x_n)_{n \leq 0}$  indexed by the non-positive integers and satisfying  $f(x_n) = x_{n+1}$  for

every  $n < 0$ . Consider the map  $\pi : \hat{M} \rightarrow M$  sending each sequence  $(x_n)_{n \leq 0}$  to its term  $x_0$  of order zero. Observe that  $\pi(\hat{M}) = M$ . Finally, define  $\hat{f} : \hat{M} \rightarrow \hat{M}$  to be the shift by one unit to the left:

$$\hat{f}(\dots, x_n, \dots, x_0) = (\dots, x_n, \dots, x_0, f(x_0)). \tag{2.4.2}$$

It is clear that  $\hat{f}$  is well defined and satisfies  $\pi \circ \hat{f} = f \circ \pi$ . Moreover,  $\hat{f}$  is invertible: the inverse is the shift to the right:

$$(\dots, y_n, \dots, y_{-1}, y_0) \mapsto (\dots, y_n, \dots, y_{-2}, y_{-1}).$$

If  $M$  is a measurable space then we may turn  $\hat{M}$  into a measurable space by endowing it with the  $\sigma$ -algebra generated by the *measurable cylinders*

$$[A_k, \dots, A_0] = \{(x_n)_{n \leq 0} \in \hat{M} : x_i \in A_i \text{ for } i = k, \dots, 0\}, \tag{2.4.3}$$

where  $k \leq 0$  and  $A_k, \dots, A_0$  are measurable subsets of  $M$ . Then  $\pi$  is a measurable map, since

$$\pi^{-1}(A) = [A]. \tag{2.4.4}$$

Moreover,  $\hat{f}$  is measurable if  $f$  is measurable:

$$\hat{f}^{-1}([A_k, \dots, A_0]) = [A_k, \dots, A_{-2}, A_{-1} \cap f^{-1}(A_0)]. \tag{2.4.5}$$

The inverse of  $\hat{f}$  is also measurable, since

$$\hat{f}([A_k, \dots, A_0]) = [A_k, \dots, A_0, M]. \tag{2.4.6}$$

Analogously, if  $M$  is a topological space then we may turn  $\hat{M}$  into a topological space by endowing it with the topology generated by the *open cylinders*  $[A_k, \dots, A_0]$ , where  $k \leq 0$  and  $A_k, \dots, A_0$  are open subsets of  $M$ . The relations (2.4.4) and (2.4.6) show that  $\pi$  and  $\hat{f}^{-1}$  are continuous, whereas (2.4.5) shows that  $\hat{f}$  is continuous if  $f$  is continuous. Observe that if  $M$  admits a countable basis  $\mathcal{U}$  of open sets then the cylinders  $[A_k, \dots, A_0]$  with  $k \geq 0$  and  $A_0, \dots, A_k \in \mathcal{U}$  constitute a countable basis of open sets for  $\hat{M}$ .

If  $M$  is a metric space, with distance  $d$ , then the following function is a distance on  $\hat{M}$ :

$$\hat{d}(\hat{x}, \hat{y}) = \sum_{n=-\infty}^0 2^n \min\{d(x_n, y_n), 1\}, \tag{2.4.7}$$

where  $\hat{x} = (x_n)_{n \leq 0}$  and  $\hat{y} = (y_n)_{n \leq 0}$ . It follows immediately from the definition that if  $\hat{x}$  and  $\hat{y}$  belong to the same pre-image  $\pi^{-1}(x)$  then

$$\hat{d}(\hat{f}^j(\hat{x}), \hat{f}^j(\hat{y})) \leq 2^{-j} \hat{d}(\hat{x}, \hat{y}) \quad \text{for every } j \geq 0.$$

So, every pre-image  $\pi^{-1}(x)$  is a *stable set*, that is, a subset restricted to which the transformation  $\hat{f}$  is uniformly contracting.

**Example 2.4.2.** Given any transformation  $g : M \rightarrow M$ , consider its maximal invariant set  $M_g = \bigcap_{n=1}^{\infty} g^n(M)$ . Clearly,  $g(M_g) \subset M_g$ . Suppose that

- (i)  $M$  is compact and  $g$  is continuous
- or
- (ii)  $\#g^{-1}(y) < \infty$  for every  $y$ .

Then (Exercise 2.4.3), the restriction  $f = (g | M_g) : M_g \rightarrow M_g$  is surjective. This restriction contains all the interesting dynamics of  $g$ . For example, assuming that  $f^n(M)$  is a measurable set for every  $n$ , every probability measure invariant under  $g$  is also invariant under  $f$ . Analogously, every point that is recurrent for  $g$  is also recurrent for  $f$ , at least in case (i). For this reason, we also refer to the natural extension of  $f = (g | M_g)$  as the natural extension of  $g$ .

A set  $\Lambda \subset M$  such that  $f^{-1}(\Lambda) = \Lambda$  is called an *invariant set* of  $f$ . There is a corresponding notion for the transformation  $\hat{f}$ . The next proposition shows that every closed invariant set of  $f$  admits a unique lift to a closed invariant set of the transformation  $\hat{f}$ :

**Proposition 2.4.3.** *Assume that  $M$  is a topological space. If  $\Lambda \subset M$  is a closed set invariant under  $f$  then  $\hat{\Lambda} = \pi^{-1}(\Lambda)$  is the only closed set invariant under  $\hat{f}$  and satisfying  $\pi(\hat{\Lambda}) = \Lambda$ .*

*Proof.* Since  $\pi$  is continuous, if  $\Lambda$  is closed then  $\hat{\Lambda} = \pi^{-1}(\Lambda)$  is also closed. Moreover, if  $\Lambda$  is invariant under  $f$  then  $\hat{\Lambda}$  is invariant under  $\hat{f}$ :

$$\hat{f}^{-1}(\hat{\Lambda}) = (\pi \circ \hat{f})^{-1}(\Lambda) = (f \circ \pi)^{-1}(\Lambda) = \pi^{-1}(\Lambda) = \hat{\Lambda}.$$

In the converse direction, let  $\hat{\Lambda} \subset \hat{M}$  be a closed set invariant under  $\hat{f}$  and such that  $\pi(\hat{\Lambda}) = \Lambda$ . It is clear that  $\hat{\Lambda} \subset \pi^{-1}(\Lambda)$ . To prove the other inclusion, we must show that, given any  $x_0 \in \Lambda$ , if  $\hat{x} \in \pi^{-1}(x_0)$  then  $\hat{x} \in \hat{\Lambda}$ . Let us write  $\hat{x} = (x_n)_{n \leq 0}$ . Consider  $n \leq 0$  and any neighborhood of  $\hat{x}$  of the form

$$V = [A_n, \dots, A_0], \quad A_n, \dots, A_0 \text{ open subsets of } M.$$

By the definition of natural extension,  $x_0 = f^{-n}(x_n)$  and, hence,  $x_n \in f^n(\Lambda) = \Lambda$ . Then, the hypothesis  $\pi(\hat{\Lambda}) = \Lambda$  implies that  $\pi(\hat{y}_n) = x_n$  for some  $\hat{y}_n \in \hat{\Lambda}$ . Since  $\hat{\Lambda}$  is invariant under  $\hat{f}$ , we have that  $\hat{f}^{-n}(\hat{y}_n) \in \hat{\Lambda}$ . Moreover, the property  $\pi(\hat{y}_n) = x_n$  implies that

$$f^{-n}(\hat{y}_n) = (\dots, y_{n,k}, \dots, y_{n,-1}, y_{n,0} = x_n, x_{n-1}, \dots, x_{-1}, x_0).$$

It follows that  $f^{-n}(\hat{y}_n) \in V$ , since  $V$  contains  $\hat{x}$  and its definition only depends on the coordinates indexed by  $j \in \{n, \dots, 0\}$ . This proves that  $\hat{x}$  is accumulated by elements of  $\hat{\Lambda}$ . Since  $\hat{\Lambda}$  is closed, it follows that  $\hat{x} \in \hat{\Lambda}$ .

Now let  $\hat{\mu}$  be an invariant measure of  $\hat{f}$  and let  $\mu = \pi_*\hat{\mu}$ . The property  $\pi \circ \hat{f} = f \circ \pi$  implies that  $\mu$  is invariant under  $f$ :

$$f_*\mu = f_*\pi_*\hat{\mu} = \pi_*\hat{f}_*\hat{\mu} = \pi_*\hat{\mu} = \mu.$$

We say that  $\hat{\mu}$  is a *lift* of  $\mu$ . The next result, which is a kind of version of Proposition 2.4.3 for measures, is due to Rokhlin [Rok61]:

**Proposition 2.4.4.** *Assume that  $M$  is a complete separable metric space and  $f : M \rightarrow M$  is continuous. Then every probability measure  $\mu$  invariant under  $f$*

admits a unique lift, that is, there is a unique measure  $\hat{\mu}$  on  $\hat{M}$  invariant under  $\hat{f}$  and such that  $\pi_*\hat{\mu} = \mu$ .

Uniqueness is easy to establish and is independent of the hypotheses on the space  $M$  and the transformation  $f$ . Indeed, if  $\hat{\mu}$  is a lift of  $\mu$  then (2.4.4) and (2.4.5) imply that the measure of every cylinder is uniquely determined:

$$\hat{\mu}([A_k, \dots, A_0]) = \hat{\mu}([A_k \cap \dots \cap f^{-k}(A_0)]) = \mu(A_k \cap \dots \cap f^{-k}(A_0)). \quad (2.4.8)$$

The proof of existence will be proposed to the reader in Exercise 5.2.4, using ideas to be developed in Chapter 5. We will also see in Exercise 8.5.7 that those arguments remain valid in the somewhat more general setting of Lebesgue spaces. But existence of the lift is *not* true in general, for arbitrary probability spaces, as shown by the example in Exercise 1.15 in the book of Przytycki and Urbański [PU10]).

### 2.4.3 Exercises

- 2.4.1. Let  $M$  be a compact metric space and  $X$  be a set of continuous maps  $f : M \rightarrow M$ , endowed with a probability measure  $\nu$ . Consider the skew-product  $F : X^{\mathbb{N}} \times M \rightarrow X^{\mathbb{N}} \times M$  defined by  $F((f_n)_n, x) = ((f_{n+1})_n, f_0(x))$ . Show that  $F$  admits some invariant probability measure  $m$  of the form  $m = \nu^{\mathbb{N}} \times \mu$ . Moreover, a measure  $m$  of this form is invariant under  $F$  if and only if the measure  $\mu$  is *stationary* for  $\nu$ , that is, if and only if  $\mu(E) = \int f_*\mu(E) d\nu(f)$  for every measurable set  $E \subset M$ .
- 2.4.2. Let  $f : M \rightarrow M$  be a surjective transformation,  $\hat{f} : \hat{M} \rightarrow \hat{M}$  be its natural extension and  $\pi : \hat{M} \rightarrow M$  be the canonical projection. Show that if  $g : N \rightarrow N$  is an invertible transformation such that  $f \circ p = p \circ g$  for some map  $p : N \rightarrow M$  then there exists a unique map  $\hat{p} : N \rightarrow \hat{M}$  such that  $\pi \circ \hat{p} = p$  and  $\hat{p} \circ g = \hat{f} \circ \hat{p}$ . Suppose that  $M$  and  $N$  are compact spaces and the maps  $p$  and  $g$  are continuous. Show that if  $p$  is surjective then  $\hat{p}$  is surjective (and so  $g : N \rightarrow N$  is an extension of  $\hat{f} : \hat{M} \rightarrow \hat{M}$ ).
- 2.4.3. Check the claims in Example 2.4.2.
- 2.4.4. Show that if  $(M, d)$  is a complete separable metric space then the same holds for the space  $(\hat{M}, \hat{d})$  of the pre-orbits of any continuous surjective transformation  $f : M \rightarrow M$ .
- 2.4.5. The purpose of this exercise and the next is to generalize the notion of natural extension to finite families of commuting transformations. Let  $M$  be a compact space and  $f_1, \dots, f_q : M \rightarrow M$  be commuting surjective continuous transformations. Let  $\hat{M}$  be the set of all sequences  $(x_{n_1, \dots, n_q})_{n_1, \dots, n_q \leq 0}$ , indexed by the  $q$ -tuples of non-positive integer numbers, such that

$$f_i(x_{n_1, \dots, n_i+1, \dots, n_q}) = x_{n_1, \dots, n_i+1, \dots, n_q} \quad \text{for every } i \text{ and every } (n_1, \dots, n_q).$$

Let  $\pi : \hat{M} \rightarrow M$  be the map sending  $(x_{n_1, \dots, n_q})_{n_1, \dots, n_q \leq 0}$  to the point  $x_{0, \dots, 0}$ . For each  $i$ , let  $\hat{f}_i : \hat{M} \rightarrow \hat{M}$  be the map sending  $(x_{n_1, \dots, n_i, \dots, n_q})_{n_1, \dots, n_q \leq 0}$  to  $(x_{n_1, \dots, n_i+1, \dots, n_q})_{n_1, \dots, n_q \leq 0}$ .

- (a) Prove that  $\hat{M}$  is a compact space. Moreover,  $\hat{M}$  is metrizable if  $M$  is metrizable.

- (b) Show that every  $\hat{f}_i : \hat{M} \rightarrow \hat{M}$  is a homeomorphism with  $\pi \circ \hat{f}_i = f_i \circ \pi$ . Moreover, these homeomorphisms commute.
  - (c) Prove that  $\pi$  is continuous and surjective. In particular,  $\hat{M}$  is non-empty.
- 2.4.6. Let  $M$  be a compact space and  $g_1, \dots, g_q : M \rightarrow M$  be commuting continuous transformations. Define  $M_g = \bigcap_{n=1}^{\infty} g_1^n \cdots g_q^n(M)$ .
- (a) Check that  $M_g = \bigcap_{n_1, \dots, n_q} g_1^{n_1} \cdots g_q^{n_q}(M)$ , where the intersection is over all  $q$ -tuples  $(n_1, \dots, n_q)$  with  $n_i \geq 1$  for every  $i$ .
  - (b) Show that  $g_i(M_g) \subset M_g$  and the restriction  $f_i = g_i | M_g$  is surjective, for every  $i$ .
- [Observation: It is clear that these restrictions  $f_i$  commute.]
- 2.4.7. Use the construction in Exercises 2.4.5 and 2.4.6 to extend the proof of Theorem 1.5.1 to the case when the transformations  $f_i$  are not necessarily invertible.

## 2.5 Arithmetic progressions

In this section we prove two fundamental results of combinatorial arithmetics, the theorem of van der Waerden and the theorem of Szemerédi, using the multiple recurrence theorems (Theorem 1.5.1 and Theorem 1.5.2) introduced in Section 1.5.

We call a *partition* of the set  $\mathbb{Z}$  of integers numbers any finite family of pairwise disjoint sets  $S_1, \dots, S_k \subset \mathbb{Z}$  whose union is the whole of  $\mathbb{Z}$ . Recall that a (finite) *arithmetic progression* is a sequence of the form

$$m + n, m + 2n, \dots, m + qn, \quad \text{with } m \in \mathbb{Z} \text{ and } n, q \geq 1.$$

The number  $q$  is called the *length* of the progression.

The next theorem was originally proven by the Dutch mathematician Bartel van der Waerden [vdW27] in the 1920's:

**Theorem 2.5.1** (van der Waerden). *Given any partition  $\{S_1, \dots, S_l\}$  of  $\mathbb{Z}$ , there exists  $j \in \{1, \dots, l\}$  such that  $S_j$  contains arithmetic progressions of every length. In other words, for every  $q \geq 1$  there exist  $m \in \mathbb{Z}$  and  $n \geq 1$  such that  $m + in \in S_j$  for every  $1 \leq i \leq q$ .*

Some time afterwards, the Hungarian mathematicians Pál Erdős and Pál Turan [ET36] conjectured the following statement, which is stronger than the theorem of van der Waerden: *any set  $S \subset \mathbb{Z}$  whose upper density is positive contains arithmetic progressions of every length*. This was proven by another Hungarian mathematician, Endre Szemerédi [Sze75], almost four decades later. To state the theorem of Szemerédi precisely, we need to define the notion of upper density of a subset of  $\mathbb{Z}$ .

We call an *interval* of the set  $\mathbb{Z}$  any subset  $I$  of the form  $\{n \in \mathbb{Z} : a \leq n < b\}$  with  $a \leq b$ . The *cardinal* of an interval  $I$  is the number  $\#I = b - a$ . The *upper*



density of a subset  $S$  of  $\mathbb{Z}$  is the number

$$D_u(S) = \limsup_{\#I \rightarrow \infty} \frac{\#(S \cap I)}{\#I},$$

where  $I$  represents any interval of  $\mathbb{Z}$ . The lower density  $D_l(S)$  of a subset  $S$  of  $\mathbb{Z}$  is defined analogously, just replacing limit superior with limit inferior. In other words,  $D_u(S)$  is the largest and  $D_l(S)$  is the smallest number  $D$  such that

$$\frac{\#(S \cap I_j)}{\#I_j} \rightarrow D \quad \text{for some sequence of intervals } I_j \subset \mathbb{Z} \text{ with } \#I_j \rightarrow \infty.$$

In the next lemma we collect some simple properties of the upper and lower densities. The proof is left as an exercise (Exercise 2.5.1).

**Lemma 2.5.2.** For any  $S \subset \mathbb{Z}$ ,

$$0 \leq D_l(S) \leq D_u(S) \leq 1 \quad \text{and} \quad D_l(S) = 1 - D_u(\mathbb{Z} \setminus S).$$

Moreover, if  $S_1, \dots, S_l$  is a partition of  $\mathbb{Z}$  then

$$D_l(S_1) + \dots + D_l(S_l) \leq 1 \leq D_u(S_1) + \dots + D_u(S_l).$$

**Example 2.5.3.** Let  $S$  be the set of even numbers. For any interval  $I \subset \mathbb{Z}$ , we have  $\#(S \cap I) = \#I/2$  if the cardinal of  $I$  is even and  $\#(S \cap I) = (\#I \pm 1)/2$  if the cardinal of  $I$  is odd; the sign  $\pm$  is positive if the smallest element of  $I$  is an even number and it is negative otherwise. It follows, immediately, that  $D_u(S) = D_l(S) = 1/2$ .

**Example 2.5.4.** Let  $S$  be the following subset of  $\mathbb{Z}$ :

$$\{1, 3, 4, 7, 8, 9, 13, 14, 15, 16, 21, 22, 23, 24, 25, 31, 32, 33, 34, 35, 36, 43, \dots\}.$$

That is, for each  $k \geq 1$  we include in  $S$  a block of  $k$  consecutive integers and then we omit the next  $k$  integer numbers. On the one hand,  $S$  contains intervals of every length. Consequently,  $D_u(S) = 1$ . On the other hand, the complement of  $S$  also contains intervals of every length. So,  $D_l(S) = 1 - D_u(\mathbb{Z} \setminus S) = 0$ .

Notice that, in both examples, the set  $S$  contains arithmetic progressions of every length. Actually, in Example 2.5.3 the set  $S$  even contains arithmetic progressions of infinite length. That is not true in Example 2.5.4, because in this case the complement of  $S$  contains arbitrarily long intervals.

**Theorem 2.5.5 (Szemerédi).** If  $S$  is a subset of  $\mathbb{Z}$  with positive upper density then it contains arithmetic progressions of every length.

The theorem of van der Waerden is an easy consequence of the theorem of Szemerédi. Indeed, it follows from Lemma 2.5.2 that if  $S_1, \dots, S_j$  is a partition of  $\mathbb{Z}$  then there exists  $j$  such that  $D_u(S_j) > 0$ . By Theorem 2.5.5, such an  $S_j$  contains arithmetic progressions of every length.

The original proofs of these results were combinatorial. Then, Furstenberg (see [Fur81]) observed that the two theorems could also be deduced from

ideas in ergodic theory: we will show in Section 2.5.1 how to obtain the theorem of van der Waerden from the multiple recurrence theorem of Birkhoff (Theorem 1.5.1); similar arguments yield the theorem of Szemerédi from the multiple recurrence theorem of Poincaré (Theorem 1.5.2), as we will see in Section 2.5.2.

The theory of Szemerédi remains a very active research area. In particular, alternative proofs of Theorem 2.5.5 have been given by other authors. Recently, this led to the following spectacular result of the British mathematician Ben Green and the Australian mathematician Terence Tao [GT08]: *the set of prime numbers contains arithmetic progressions of every length*. This is not a consequence of the theorem of Szemerédi, because the upper density of the set of prime numbers is zero, but the theorem of Szemerédi does have an important role in the proof. On the other hand, the Green–Tao theorem is a special case of yet another conjecture of Erdős: if  $S \subset \mathbb{N}$  is such that the sum of the inverses diverges, that is, such that

$$\sum_{n \in S} \frac{1}{n} = \infty,$$

then  $S$  contains arithmetic progressions of every length. This more general statement remains open.

### 2.5.1 Theorem of van der Waerden

In this section we prove Theorem 2.5.1. The idea of the proof is to reduce the conclusion of the theorem to a claim about the shift map

$$\sigma : \Sigma \rightarrow \Sigma, \quad (\alpha_n)_{n \in \mathbb{Z}} \mapsto (\alpha_{n+1})_{n \in \mathbb{Z}}$$

in the space  $\Sigma = \{1, 2, \dots, l\}^{\mathbb{Z}}$  of two-sided sequences with values in the set  $\{1, 2, \dots, l\}$ . This claim will then be proved using the multiple recurrence theorem of Birkhoff.

Observe that every partition  $\{S_1, \dots, S_l\}$  of  $\mathbb{Z}$  into  $l \geq 2$  subsets determines an element  $\underline{\alpha} = (\alpha_n)_{n \in \mathbb{Z}}$  of  $\Sigma$ , through  $\alpha_n = i \Leftrightarrow n \in S_i$ . Conversely, every  $\underline{\alpha} \in \Sigma$  determines a partition of  $\mathbb{Z}$  into subsets

$$S_i = \{n \in \mathbb{Z} : \alpha_n = i\}, \quad i = 1, \dots, l.$$

We are going to show that for every  $\underline{\alpha} \in \Sigma$  and every  $q \geq 1$ , there exist  $m \in \mathbb{Z}$  and  $n \geq 1$  such that

$$\alpha_{m+n} = \dots = \alpha_{m+qn}. \quad (2.5.1)$$

In view of what we have just observed, this means that for every partition  $\{S_1, \dots, S_l\}$  and every  $q \geq 1$  there exists  $i \in \{1, \dots, l\}$  such that  $S_i$  contains some arithmetic progression of length  $q$ . Since there are finitely many  $S_i$ , that implies that some  $S_j$  contains arithmetic progressions of arbitrarily large lengths. This is the same as saying that  $S_j$  contains arithmetic progressions of every length

because, clearly, every arithmetic progression of length  $q$  contains arithmetic progressions of every length smaller than  $q$ . This reduces the proof of the theorem to proving the claim in (2.5.1).

To that end, let us consider on  $\Sigma$  the distance defined by  $d(\underline{\beta}, \underline{\gamma}) = 2^{-N(\underline{\beta}, \underline{\gamma})}$ , where

$$N(\underline{\beta}, \underline{\gamma}) = \max \{ N \geq 0 : \beta_n = \gamma_n \text{ for every } n \in \mathbb{Z} \text{ with } |n| < N \}.$$

Note that

$$d(\underline{\beta}, \underline{\gamma}) < 1 \quad \text{if and only if } \beta_0 = \gamma_0. \tag{2.5.2}$$

Since the metric space  $(\Sigma, d)$  is compact, the closure  $Z = \overline{\{\sigma^n(\underline{\alpha}) : n \in \mathbb{Z}\}}$  of the trajectory of  $\underline{\alpha}$  is also compact. Moreover,  $Z$  is invariant under the shift map. Let us consider the transformations  $f_1 = \sigma, f_2 = \sigma^2, \dots, f_q = \sigma^q$  defined from  $Z$  to  $Z$ . It is clear from the definition that these transformations commute with each other. So, we may use Theorem 1.5.1 to conclude that there exist  $\underline{\theta} \in Z$  and a sequence  $(n_k)_k \rightarrow \infty$  such that

$$\lim_k f_i^{n_k}(\underline{\theta}) = \underline{\theta} \quad \text{for every } i = 1, 2, \dots, q.$$

Observe that  $f_i^{n_j} = \sigma^{in_j}$ . In particular, we may fix  $n = n_j$  such that the iterates  $\sigma^n(\underline{\theta}), \sigma^{2n}(\underline{\theta}), \dots, \sigma^{qn}(\underline{\theta})$  are all within a distance of less than  $1/2$  from the point  $\underline{\theta}$ . Consequently,

$$d(\sigma^{in}(\underline{\theta}), \sigma^{jn}(\underline{\theta})) < 1 \quad \text{for every } 1 \leq i, j \leq q.$$

Then, as  $\underline{\theta}$  is in the closure  $Z$  of the orbit of  $\underline{\alpha}$ , we may find  $m \in \mathbb{Z}$  such that  $\sigma^m(\underline{\alpha})$  is so close to  $\underline{\theta}$  that

$$d(\sigma^{m+in}(\underline{\alpha}), \sigma^{m+jn}(\underline{\alpha})) < 1 \quad \text{for every } 1 \leq i, j \leq q.$$

Taking into account the observation (2.5.2) and the definition of the shift map  $\sigma$ , this means that  $\alpha_{m+n} = \dots = \alpha_{m+qn}$ , as we wanted to prove. This completes the proof of the theorem of van der Waerden.

### 2.5.2 Theorem of Szemerédi

Now let us prove Theorem 2.5.5. We use the same kind of dictionary between partitions of  $\mathbb{Z}$  and sequences of integer numbers that was used in the previous section to prove the theorem of van der Waerden.

Let  $S$  be a subset of integer numbers with positive upper density, that is, such that there exist  $c > 0$  and intervals  $I_j = [a_j, b_j]$  of  $\mathbb{Z}$  satisfying

$$\lim_j \#I_j = \infty \quad \text{and} \quad \lim_j \frac{\#(S \cap I_j)}{\#I_j} \geq c.$$

Let us associate with  $S$  the sequence  $\underline{\alpha} = (\alpha_j)_{j \in \mathbb{Z}} \in \Sigma = \{0, 1\}^{\mathbb{Z}}$  defined by

$$\alpha_j = 1 \Leftrightarrow j \in S.$$

Consider the shift map  $\sigma : \Sigma \rightarrow \Sigma$  and the subset  $A = \{\underline{\alpha} \in \Sigma : \alpha_0 = 1\}$  of  $\Sigma$ . Note that both  $A$  and its complement are open cylinders of  $\Sigma$ . Thus,  $A$  is both open and closed in  $\Sigma$ . Moreover, for every  $j \in \mathbb{Z}$ ,

$$\sigma^j(\underline{\alpha}) \in A \Leftrightarrow \alpha_j = 1 \Leftrightarrow j \in S.$$

So, to prove the theorem of Szemerédi it suffices to show that for every  $k \in \mathbb{N}$  there exist  $m \in \mathbb{Z}$  and  $n \geq 1$  such that

$$\sigma^{m+n}(\underline{\alpha}), \sigma^{m+2n}(\underline{\alpha}), \dots, \sigma^{m+kn}(\underline{\alpha}) \in A. \tag{2.5.3}$$

For that, let us consider the sequence  $\mu_j$  of probability measures defined on  $\Sigma$  by

$$\mu_j = \frac{1}{\#I_j} \sum_{i \in I_j} \delta_{\sigma^i(\underline{\alpha})}. \tag{2.5.4}$$

Since the space  $\mathcal{M}_1(\Sigma)$  of all probability measures on  $\Sigma$  is compact (Theorem 2.1.5), up to replacing  $(\mu_j)_j$  by some subsequence we may suppose that it converges in the weak\* topology to some probability measure  $\mu$  on  $\Sigma$ .

Observe that  $\mu$  is a  $\sigma$ -invariant probability measure, for, given any continuous function  $\varphi : \Sigma \rightarrow \mathbb{R}$ ,

$$\begin{aligned} \int (\varphi \circ \sigma) d\mu_j &= \frac{1}{\#I_j} \sum_{i \in I_j} \varphi(\sigma^i(\underline{\alpha})) + \frac{1}{\#I_j} \left[ \varphi(\sigma^{b_j}(\underline{\alpha})) - \varphi(\sigma^{a_j}(\underline{\alpha})) \right] \\ &= \int \varphi d\mu_j + \frac{1}{\#I_j} \left[ \varphi(\sigma^{b_j}(\underline{\alpha})) - \varphi(\sigma^{a_j}(\underline{\alpha})) \right] \end{aligned}$$

and, taking the limit when  $j \rightarrow \infty$ , it follows that  $\int (\varphi \circ \sigma) d\mu = \int \varphi d\mu$ . Observe also that  $\mu(A) > 0$ . Indeed, since  $A$  is closed, Theorem 2.1.2 ensures that

$$\mu(A) \geq \limsup_j \mu_j(A) = \limsup_j \frac{\#(S \cap I_j)}{\#I_j} \geq c.$$

Given any  $k \geq 1$ , consider  $f_i = \sigma^i$  for  $i = 1, \dots, k$ . It is clear that these transformations commute with each other. So, we are in a position to apply Theorem 1.5.2 to conclude that there exists some  $n \geq 1$  such that

$$\mu(A \cap \sigma^{-n}(A) \cap \dots \cap \sigma^{-kn}(A)) > 0.$$

Since  $A$  is open, this implies (Theorem 2.1.2) that

$$\mu_l(A \cap \sigma^{-n}(A) \cap \dots \cap \sigma^{-kn}(A)) > 0$$

for every  $l$  sufficiently large. By the definition (2.5.4) of  $\mu_l$ , this means that there exists some  $m \in I_l$  such that

$$\sigma^m(\underline{\alpha}) \in A \cap \sigma^{-n}(A) \cap \dots \cap \sigma^{-kn}(A).$$

In particular,  $\sigma^{m+in}(\underline{\alpha}) \in A$  for every  $i = 1, \dots, k$ , as we wanted to prove.

### 2.5.3 Exercises

- 2.5.1. Prove Lemma 2.5.2.
- 2.5.2. Show that the conclusion of Theorem 2.5.1 remains valid for partitions of finite subsets of  $\mathbb{Z}$ , as long as they are sufficiently large. More precisely: given  $q, l \geq 1$  there exists  $N \geq 1$  such that, for any partition of the set  $\{1, 2, \dots, N\}$  into  $l$  subsets, at least one of these subsets contains arithmetic progressions of length  $q$ .
- 2.5.3. A point  $x \in M$  is said to be *super non-wandering* if, given any neighborhood  $U$  of  $x$  and any  $k \geq 1$ , there exists  $n \geq 1$  such that  $\bigcap_{j=0}^k f^{-jn}(U) \neq \emptyset$ . Show that the theorem of van der Waerden is equivalent to the following statement: every invertible transformation on a compact metric space has some super non-wandering point.
- 2.5.4. Prove the following generalization of the theorem of van der Waerden to arbitrary dimension, called the Grünwald theorem: given any partition  $\mathbb{N}^k = S_1 \cup \dots \cup S_l$  and any  $q \geq 1$ , there exist  $j \in \{1, \dots, l\}$ ,  $d \in \mathbb{N}$  and  $b \in \mathbb{N}^k$  such that

$$b + d(a_1, \dots, a_k) \in S_j \quad \text{for any } 1 \leq a_i \leq q \text{ and any } 1 \leq i \leq k.$$