

# 7

## Correlations

The models of dynamical systems that interest us the most, transformations and flows, are deterministic: the state of the system at any time determines the whole future trajectory; when the system is invertible, the past trajectory is equally determined. However, these systems may also present stochastic (that is, “partly random”) behavior: at some level coarser than that of individual trajectories, information about the past is gradually lost as the system is iterated. That is the subject of the present chapter.

In probability theory one calls the *correlation* between two random variables  $X$  and  $Y$  the number

$$C(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Note that the expression  $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$  is positive if  $X$  and  $Y$  are on the same side (either larger or smaller) of their respective means,  $\mathbb{E}[X]$  and  $\mathbb{E}[Y]$ , and it is negative otherwise. Therefore, the sign of  $C(X, Y)$  indicates whether the two variables exhibit, predominantly, the same behavior or opposite behaviors, relative to their means. Furthermore, correlation close to zero indicates that the two behaviors are little, if at all, related to each other.

Given an invariant probability measure  $\mu$  of a dynamical system  $f : M \rightarrow M$  and given measurable functions  $\varphi, \psi : M \rightarrow \mathbb{R}$ , we want to analyze the evolution of the correlations

$$C_n(\varphi, \psi) = C(\varphi \circ f^n, \psi)$$

when time  $n$  goes to infinity. We may think of  $\varphi$  and  $\psi$  as quantities that are measured in the system, such as temperature, acidity (pH), kinetic energy, and so forth. Then  $C_n(\varphi, \psi)$  measures how much the value of  $\varphi$  at time  $n$  is correlated with the value of  $\psi$  at time zero; to what extent one value “influences” the other.

For example, if  $\varphi = \mathcal{X}_A$  and  $\psi = \mathcal{X}_B$  are characteristic functions, then  $\psi(x)$  provides information on the position of the initial point  $x$ , whereas  $\varphi(f^n(x))$  informs on the position of its  $n$ -th iterate  $f^n(x)$ . If the correlation  $C_n(\varphi, \psi)$  is small, then the first information is of little use to make predictions about the

second one. That kind of behavior, where correlations approach zero as time  $n$  increases, is quite common in important models, as we are going to see.

We start by introducing the notions of (strong) mixing and weak mixing systems, and by studying their basic properties (Section 7.1). In Sections 7.2 and 7.3 we discuss these notions in the context of Markov shifts, which generalize Bernoulli shifts, and of interval exchanges, which are an extension of the class of circle rotations. In Section 7.4 we analyze, in quantitative terms, the speed of decay of correlations for certain classes of functions.

## 7.1 Mixing systems

Let  $f : M \rightarrow M$  be a measurable transformation and  $\mu$  be an invariant probability measure. The *correlations sequence* of two measurable functions  $\varphi, \psi : M \rightarrow \mathbb{R}$  is defined by

$$C_n(\varphi, \psi) = \int (\varphi \circ f^n) \psi \, d\mu - \int \varphi \, d\mu \int \psi \, d\mu, \quad n \in \mathbb{N}. \quad (7.1.1)$$

We say that the system  $(f, \mu)$  is *mixing* if

$$\lim_n C_n(\mathcal{X}_A, \mathcal{X}_B) = \lim_n \mu(f^{-n}(A) \cap B) - \mu(A)\mu(B) = 0, \quad (7.1.2)$$

for any measurable sets  $A, B \subset M$ . In other words, when  $n$  grows the probability of the event  $\{x \in B \text{ and } f^n(x) \in A\}$  converges to the product of the probabilities of the events  $\{x \in B\}$  and  $\{f^n(x) \in A\}$ .

Analogously, given a flow  $f^t : M \rightarrow M$ ,  $t \in \mathbb{R}$  and an invariant probability measure  $\mu$ , we define

$$C_t(\varphi, \psi) = \int (\varphi \circ f^t) \psi \, d\mu - \int \varphi \, d\mu \int \psi \, d\mu, \quad t \in \mathbb{R} \quad (7.1.3)$$

and we say that the system  $(f^t, \mu)$  is *mixing* if

$$\lim_{t \rightarrow +\infty} C_t(\mathcal{X}_A, \mathcal{X}_B) = \lim_{t \rightarrow +\infty} \mu(f^{-t}(A) \cap B) - \mu(A)\mu(B) = 0, \quad (7.1.4)$$

for any measurable sets  $A, B \subset M$ .

### 7.1.1 Properties

A mixing system is necessarily ergodic. Indeed, suppose that there exists some invariant set  $A \subset M$  with  $0 < \mu(A) < 1$ . Taking  $B = A^c$ , we get  $f^{-n}(A) \cap B = \emptyset$  for every  $n$ . Then,  $\mu(f^{-n}(A) \cap B) = 0$  for every  $n$ , whereas  $\mu(A)\mu(B) \neq 0$ . In particular,  $(f, \mu)$  is not mixing. The example that follows shows that ergodicity is strictly weaker than mixing:

**Example 7.1.1.** Let  $\theta \in \mathbb{R}$  be an irrational number. As we have seen in Section 4.2.1, the rotation  $R_\theta : S^1 \rightarrow S^1$  is ergodic with respect to the Lebesgue measure  $m$ . However,  $(R_\theta, m)$  is not mixing. Indeed, if  $A, B \subset S^1$  are two small intervals then  $R_\theta^{-n}(A) \cap B$  is empty and, thus,  $m(R_\theta^{-n}(A) \cap B) = 0$  for infinitely

many values of  $n$ . Since  $m(A)m(B) \neq 0$ , it follows that the condition in (7.1.2) does not hold.

It is clear from the definition (7.1.2) that if  $(f, \mu)$  is mixing then  $(f^k, \mu)$  is mixing for every  $k \in \mathbb{N}$ . The corresponding statement for ergodicity is false: the map  $f(x) = 1 - x$  on the set  $\{0, 1\}$  is ergodic with respect to the measure  $(\delta_0 + \delta_1)/2$  but the second iterate  $f^2$  is not.

**Lemma 7.1.2.** *Assume that  $\lim_n \mu(f^{-n}(A) \cap B) = \mu(A)\mu(B)$  for every pair of sets  $A$  and  $B$  in an algebra  $\mathcal{A}$  that generates the  $\sigma$ -algebra of measurable sets. Then  $(f, \mu)$  is mixing.*

*Proof.* Let  $\mathcal{C}$  be the family of all measurable sets  $A$  such that  $\mu(f^{-n}(A) \cap B)$  converges to  $\mu(A)\mu(B)$  for every  $B \in \mathcal{A}$ . By assumption,  $\mathcal{C}$  contains  $\mathcal{A}$ . We claim that  $\mathcal{C}$  is a monotone class. Indeed, let  $A = \bigcup_k A_k$  be the union of an increasing sequence  $A_1 \subset \dots \subset A_k \subset \dots$  of elements of  $\mathcal{C}$ . Given  $\varepsilon > 0$ , there exists  $k_0 \geq 1$  such that

$$\mu(A) - \mu(A_k) = \mu(A \setminus A_k) < \varepsilon$$

for every  $k \geq k_0$ . Moreover, for every  $n \geq 1$ ,

$$\begin{aligned} \mu(f^{-n}(A) \cap B) - \mu(f^{-n}(A_k) \cap B) &= \mu(f^{-n}(A \setminus A_k) \cap B) \\ &\leq \mu(f^{-n}(A \setminus A_k)) = \mu(A \setminus A_k) < \varepsilon. \end{aligned}$$

For each fixed  $k \geq k_0$ , the fact that  $A_k \in \mathcal{C}$  ensures that there exists  $n(k) \geq 1$  such that

$$|\mu(f^{-n}(A_k) \cap B) - \mu(A_k)\mu(B)| < \varepsilon \quad \text{for every } n \geq n(k).$$

Adding these three inequalities we conclude that

$$|\mu(f^{-n}(A) \cap B) - \mu(A)\mu(B)| < 3\varepsilon \quad \text{for every } n \geq n(k_0).$$

Since  $\varepsilon > 0$  is arbitrary, this shows that  $A \in \mathcal{C}$ . In the same way, one proves that the intersection of any decreasing sequence of elements of  $\mathcal{C}$  is still an element of  $\mathcal{C}$ . So,  $\mathcal{C}$  is indeed a monotone class. By the monotone class theorem (Theorem A.1.18), it follows that  $\mathcal{C}$  contains every measurable set: for every measurable set  $A$  one has

$$\lim_n \mu(f^{-n}(A) \cap B) = \mu(A)\mu(B) \quad \text{for every } B \in \mathcal{A}.$$

All that is left to do is to deduce that this property holds for every measurable set  $B$ . This follows from precisely the same kind of arguments as we have just detailed, as the reader may readily check.

**Example 7.1.3.** Every Bernoulli shift (recall Section 4.2.3) is mixing. Indeed, given any two cylinders  $A = [p; A_p, \dots, A_q]$  and  $B = [r; B_r, \dots, B_s]$ ,

$$\begin{aligned} \mu(f^{-n}(A) \cap B) &= \mu([r; B_r, \dots, B_s, X, \dots, X, A_p, \dots, A_q]) \\ &= \mu([r; B_r, \dots, B_s])\mu([p; A_p, \dots, A_q]) = \mu(A)\mu(B) \end{aligned}$$

for every  $n > s - p$ . Let  $\mathcal{A}$  be the algebra generated by the cylinders: its elements are the finite pairwise disjoint unions of cylinders. It follows from what we have just said that  $\mu(f^{-n}(A) \cap B) = \mu(A)\mu(B)$  for every pair of sets  $A, B \in \mathcal{A}$  and every  $n$  sufficiently large. Since  $\mathcal{A}$  generates the  $\sigma$ -algebra of measurable sets, we may use Lemma 7.1.2 to conclude that the system is mixing, as stated.

**Example 7.1.4.** Let  $g : S^1 \rightarrow S^1$  be defined by  $g(x) = kx$ , where  $k \geq 2$  is an integer number, and let  $m$  be the Lebesgue measure on the circle. The system  $(g, m)$  is equivalent to a Bernoulli shift, in the following sense. Let  $X = \{0, 1, \dots, k - 1\}$  and let  $f : M \rightarrow M$  be the shift map in  $M = X^{\mathbb{N}}$ . Consider the product measure  $\mu = \nu^{\mathbb{N}}$  in  $M$ , where  $\nu$  is the probability measure defined by  $\nu(A) = \#A/k$  for every  $A \subset X$ . The map

$$h : M \rightarrow S^1, \quad h((a_n)_n) = \sum_{n=1}^{\infty} \frac{a_{n-1}}{k^n}$$

is a bijection, restricted to a full measure subset, and both  $h$  and its inverse are measurable. Moreover,  $h_*\mu = m$  and  $h \circ f = g \circ h$  at almost every point. We say that  $h$  is an *ergodic equivalence* between  $(g, m)$  and  $(f, \mu)$ . Through it, properties of one system may be translated to corresponding properties for the other system. In particular, recalling Example 7.1.3, we get that  $(g, m)$  is mixing: given any measurable sets  $A, B \subset S^1$ ,

$$\begin{aligned} m(g^{-n}(A) \cap B) &= \mu\left(h^{-1}(g^{-n}(A) \cap B)\right) = \mu\left(f^{-n}(h^{-1}(A)) \cap h^{-1}(B)\right) \\ &\rightarrow \mu(h^{-1}(A))\mu(h^{-1}(B)) = m(A)m(B) \quad \text{when } n \rightarrow \infty. \end{aligned}$$

**Example 7.1.5.** For surjective endomorphisms of the torus (Section 4.2.5) mixing and ergodicity are equivalent properties: the system  $(f_A, m)$  is mixing if and only if no eigenvalue of the matrix  $A$  is a root of unity (compare Theorem 4.2.14). In Exercise 7.1.4 we invite the reader to prove this fact; a stronger statement will appear in Exercise 8.4.2. More generally, relative to the Haar measure, a surjective endomorphism of a compact group is mixing if and only if it is ergodic. In fact, even stronger statements are true, as we will comment upon in Section 9.5.3.

Let us also discuss the topological version of the notion of a mixing system. For that, take the ambient  $M$  to be a topological space. A transformation  $f : M \rightarrow M$  is said to be *topologically mixing* if, given any non-empty open sets  $U, V \subset M$ , there exists  $n_0 \in \mathbb{N}$  such that  $f^{-n}(U) \cap V$  is non-empty for every  $n \geq n_0$ . This is similar to but strictly stronger than the hypothesis of Lemma 4.3.4: in the lemma we asked  $f^{-n}(U)$  to intersect  $V$  for *some* value of  $n$ , whereas now we request that to happen for every  $n$  sufficiently large.

**Proposition 7.1.6.** *If  $(f, \mu)$  is mixing then the restriction of  $f$  to the support of  $\mu$  is topologically mixing.*

*Proof.* Denote  $X = \text{supp}(\mu)$ . Let  $A, B \subset X$  be open sets. By the definition of support of a measure,  $\mu(A) > 0$  and  $\mu(B) > 0$ . Hence, since  $\mu$  is mixing, there exists  $n_0$  such that  $\mu(f^{-n}(A) \cap B) > \mu(A)\mu(B)/2 > 0$  for every  $n \geq n_0$ . In particular,  $\mu(f^{-n}(A) \cap B) \neq \emptyset$  for every  $n \geq n_0$ .

It follows directly from this proposition that if  $f$  admits some invariant probability measure  $\mu$  that is mixing and positive on open sets, then  $f$  is topologically mixing. For example, given any finite set  $X = \{1, \dots, d\}$ , the shift map

$$f : X^{\mathbb{Z}} \rightarrow X^{\mathbb{Z}} \quad (\text{or } f : X^{\mathbb{N}} \rightarrow X^{\mathbb{N}})$$

is topologically mixing. Indeed, for any probability measure  $\nu$  supported on the whole of  $X$ , the Bernoulli measure  $\mu = \nu^{\mathbb{Z}}$  (or  $\mu = \nu^{\mathbb{N}}$ ) is mixing and positive on open sets, as we have seen in Example 7.1.3. Analogously, by Example 7.1.4, every transformation  $f : S^1 \rightarrow S^1$  of the form  $f(x) = kx$  with  $k \geq 2$  is topologically mixing.

**Example 7.1.7.** Translations in a metrizable group  $G$  are never topologically mixing. Indeed, consider any left-translation  $L_g$  (the case of right-translations is analogous). We may suppose that  $g$  is not the unit element  $e$  since otherwise it is obvious that  $L_g$  is not topologically mixing. Fix some distance  $d$  invariant under all the translations of the group  $G$  (recall Lemma 6.3.6) and let  $\alpha = d(e, g^{-1})$ . Consider  $U = V = \text{ball of radius } \alpha/4 \text{ around } e$ . Every  $L_g^{-n}(U)$  is a ball of radius  $\alpha/4$ . Assume that  $L_g^{-n}(U)$  intersects  $V$ . Then  $L_g^{-n}(U)$  is contained in the ball of radius  $3\alpha/4$  and, thus,  $L_g^{-n-1}(U)$  is contained in the ball of radius  $3\alpha/4$  around  $g^{-1}$ . Consequently,  $L_g^{-n-1}(U)$  does not intersect  $V$ . Since  $n$  is arbitrary, this shows that  $L_g$  is not topologically mixing.

### 7.1.2 Weak mixing

A system  $(f, \mu)$  is *weak mixing* if, given any measurable sets  $A, B \subset M$ ,

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} |C_j(\mathcal{X}_A, \mathcal{X}_B)| = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} |\mu(f^{-j}(A) \cap B) - \mu(A)\mu(B)| = 0. \quad (7.1.5)$$

It is clear from the definition that every mixing system is also weak mixing. On the other hand, every weak mixing system is ergodic. Indeed, if  $A \subset M$  is an invariant set then

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} |C_j(\mathcal{X}_A, \mathcal{X}_{A^c})| = \mu(A)\mu(A^c)$$

and, hence, the hypothesis implies that  $\mu(A) = 0$  or  $\mu(A^c) = 0$ .

**Example 7.1.8.** Translations in metrizable compact groups are never weak mixing with respect to the Haar measure  $\mu$  (or any other invariant measure

positive on open sets). Indeed, as observed in Example 7.1.7, it is always possible to choose open sets  $U$  and  $V$  such that  $f^{-n}(U) \cap V$  is empty for at least one in every two consecutive values  $n$ . Then,

$$\liminf_n \frac{1}{n} \sum_{j=0}^{n-1} |\mu(f^{-j}(U) \cap V) - \mu(U)\mu(V)| \geq \frac{1}{2} \mu(U)\mu(V) > 0.$$

In this way we get several examples of ergodic systems, even uniquely ergodic ones, that are not weak mixing.

We are going to see in Section 7.3.2 that the family of interval exchanges contains many systems that are weak mixing (and uniquely ergodic) but are not mixing.

The proof of the next result is analogous to the proof of Lemma 7.1.2 and is left to the reader:

**Lemma 7.1.9.** *Assume that  $\lim_n (1/n) \sum_{j=0}^{n-1} |\mu(f^{-j}(A) \cap B) - \mu(A)\mu(B)| = 0$  for every pair of sets  $A$  and  $B$  in some algebra  $\mathcal{A}$  that generates the  $\sigma$ -algebra of measurable sets. Then  $(f, \mu)$  is weak mixing.*

**Example 7.1.10.** Given a system  $(f, \mu)$ , let us consider the product transformation  $f_2 : M \times M \rightarrow M \times M$  given by  $f_2(x, y) = (f(x), f(y))$ . It is easy to see that  $f_2$  preserves the product measure  $\mu_2 = \mu \times \mu$ . If  $(f_2, \mu_2)$  is ergodic then  $(f, \mu)$  is ergodic: just note that if  $A \subset M$  is invariant under  $f$  and  $\mu(A) \in (0, 1)$  then  $A \times A$  is invariant under  $f_2$  and  $\mu_2(A \times A) \in (0, 1)$ .

The converse is not true in general, that is,  $(f_2, \mu_2)$  may not be ergodic even if  $(f, \mu)$  is ergodic. For example, if  $f : S^1 \rightarrow S^1$  is an irrational rotation and  $d$  is a distance invariant under rotations, then any neighborhood  $\{(x, y) : d(x, y) < r\}$  of the diagonal is invariant under  $f_2$ .

The next result shows that this type of phenomenon cannot occur in the category of weak mixing systems:

**Proposition 7.1.11.** *The following conditions are equivalent:*

- (i)  $(f, \mu)$  is weak mixing;
- (ii)  $(f_2, \mu_2)$  is weak mixing;
- (iii)  $(f_2, \mu_2)$  is ergodic.

*Proof.* To prove that (i) implies (ii), consider any measurable sets  $A, B, C, D$  in  $M$ . Then

$$\begin{aligned} & \left| \mu_2(f_2^{-j}(A \times B) \cap (C \times D)) - \mu_2(A \times B)\mu_2(C \times D) \right| \\ &= \left| \mu(f^{-j}(A) \cap C)\mu(f^{-j}(B) \cap D) - \mu(A)\mu(B)\mu(C)\mu(D) \right| \\ &\leq \left| \mu(f^{-j}(A) \cap C) - \mu(A)\mu(C) \right| + \left| \mu(f^{-j}(B) \cap D) - \mu(B)\mu(D) \right|. \end{aligned}$$

Therefore, the hypothesis (i) implies that

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} |\mu_2(f_2^{-j}(A \times B) \cap (C \times D)) - \mu_2(A \times B)\mu_2(C \times D)| = 0.$$

It follows that

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} |\mu_2(f_2^{-j}(X) \cap Y) - \mu_2(X)\mu_2(Y)| = 0$$

for any  $X, Y$  in the algebra generated by the products  $E \times F$  of measurable subsets of  $M$ , that is, the algebra formed by the finite pairwise disjoint unions of such products. Since this algebra generates the  $\sigma$ -algebra of measurable subsets of  $M \times M$ , we may use Lemma 7.1.9 to conclude that  $(f_2, \mu_2)$  is weak mixing.

It is immediate that (i) implies (iii). To prove that (iii) implies (i), observe that

$$\begin{aligned} & \frac{1}{n} \sum_{j=0}^{n-1} [\mu(f^{-j}(A) \cap B) - \mu(A)\mu(B)]^2 \\ &= \frac{1}{n} \sum_{j=0}^{n-1} [\mu(f^{-j}(A) \cap B)^2 - 2\mu(A)\mu(B)\mu(f^{-j}(A) \cap B) + \mu(A)^2\mu(B)^2]. \end{aligned}$$

The right-hand side may be rewritten as

$$\begin{aligned} & \frac{1}{n} \sum_{j=0}^{n-1} [\mu_2(f_2^{-j}(A \times A) \cap (B \times B)) - \mu_2(A \times A)\mu_2(B \times B)] \\ & \quad - 2\mu(A)\mu(B) \frac{1}{n} \sum_{j=0}^{n-1} [\mu(f^{-j}(A) \cap B) - \mu(A)\mu(B)]. \end{aligned}$$

Since  $(f_2, \mu_2)$  is ergodic and, consequently,  $(f, \mu)$  is also ergodic, part (ii) of Proposition 4.1.4 gives that both terms in this expression converge to zero. In this way, we conclude that

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} [\mu(f^{-j}(A) \cap B) - \mu(A)\mu(B)]^2 = 0$$

for any measurable sets  $A, B \subset M$ . Using Exercise 7.1.2, we deduce that  $(f, \mu)$  is weak mixing.

### 7.1.3 Spectral characterization

In this section we discuss equivalent formulations of the notions of mixing and weak mixing systems, in terms of the Koopman operator.

**Proposition 7.1.12.** *The following conditions are equivalent:*

- (i)  $(f, \mu)$  is mixing.
- (ii) There exist  $p, q \in [1, \infty]$  with  $1/p + 1/q = 1$  such that  $C_n(\varphi, \psi) \rightarrow 0$  for any  $\varphi \in L^p(\mu)$  and  $\psi \in L^q(\mu)$ .
- (iii) The condition in part (ii) holds for  $\varphi$  in some dense subset of  $L^p(\mu)$  and  $\psi$  in some dense subset of  $L^q(\mu)$ .

*Proof.* Condition (i) is the special case of (ii) for characteristic functions. Since the correlations  $(\varphi, \psi) \mapsto C_n(\varphi, \psi)$  are bilinear functions, condition (i) implies that  $C_n(\varphi, \psi) \rightarrow 0$  for any simple functions  $\varphi$  and  $\psi$ . This implies (iii), since the simple functions form a dense subset of  $L^r(\mu)$  for any  $r \geq 1$ .

To show that (iii) implies (ii), let us begin by observing that as correlations  $C_n(\varphi, \psi)$  are equicontinuous functions of  $\varphi$  and  $\psi$ . Indeed, given  $\varphi_1, \varphi_2 \in L^p(\mu)$  and  $\psi_1, \psi_2 \in L^q(\mu)$ , the Hölder inequality (Theorem A.5.5) gives that

$$\left| \int (\varphi_1 \circ f^n) \psi_1 d\mu - \int (\varphi_2 \circ f^n) \psi_2 d\mu \right| \leq \|\varphi_1 - \varphi_2\|_p \|\psi_1\|_q + \|\varphi_2\|_p \|\psi_1 - \psi_2\|_q.$$

Moreover,

$$\left| \int \varphi_1 d\mu \int \psi_1 d\mu - \int \varphi_2 d\mu \int \psi_2 d\mu \right| \leq \|\varphi_1 - \varphi_2\|_1 \|\psi_1\|_1 + \|\varphi_2\|_1 \|\psi_1 - \psi_2\|_1.$$

Adding these inequalities, and noting that  $\|\cdot\|_1 \leq \|\cdot\|_r$  for every  $r \geq 1$ , we get that

$$|C_n(\varphi_1, \psi_1) - C_n(\varphi_2, \psi_2)| \leq 2\|\varphi_1 - \varphi_2\|_p \|\psi_1\|_q + 2\|\varphi_2\|_p \|\psi_1 - \psi_2\|_q \tag{7.1.6}$$

for every  $n \geq 1$ . Then, given  $\varepsilon > 0$  and any  $\varphi \in L^p(\mu)$  and  $\psi \in L^q(\mu)$ , we may take  $\varphi'$  and  $\psi'$  in the dense subsets mentioned in the hypothesis such that

$$\|\varphi - \varphi'\|_p < \varepsilon \quad \text{and} \quad \|\psi - \psi'\|_q < \varepsilon.$$

In particular,  $\|\varphi'\|_p < \|\varphi\|_p + \varepsilon$  and  $\|\psi'\|_q < \|\psi\|_q + \varepsilon$ . Then, (7.1.6) gives that

$$|C_n(\varphi, \psi)| \leq |C_n(\varphi', \psi')| + 2\varepsilon(\|\varphi\|_p + \|\psi\|_q + 2\varepsilon) \quad \text{for every } n.$$

Moreover, by hypothesis,  $|C_n(\varphi', \psi')| < \varepsilon$  for every  $n$  sufficiently large. Since  $\varepsilon$  is arbitrary, these two inequalities imply that  $C_n(\varphi, \psi)$  converges to zero when  $n \rightarrow \infty$ . This proves property (ii).

The same argument proves the following version of Proposition 7.1.12 for the weak mixing property:

**Proposition 7.1.13.** *The following conditions are equivalent:*

- (i)  $(f, \mu)$  is weak mixing.
- (ii) There exist  $p, q \in [1, \infty]$  with  $1/p + 1/q = 1$  such that  $(1/n) \sum_{j=1}^n |C_j(\varphi, \psi)|$  converges to 0 for any  $\varphi \in L^p(\mu)$  and  $\psi \in L^q(\mu)$ .
- (iii) The condition in part (ii) holds for  $\varphi$  in some dense subset of  $L^p(\mu)$  and  $\psi$  in some dense subset of  $L^q(\mu)$ .



In the case  $p = q = 2$ , we may express the correlations in terms of the inner product  $\cdot$  in the Hilbert space  $L^2(\mu)$ :

$$C_n(\varphi, \psi) = [U_f^n \varphi - (\varphi \cdot 1)] \cdot \psi \quad \text{for every } \varphi, \psi \in L^2(\mu).$$

Therefore, Proposition 7.1.12 gives that  $(f, \mu)$  is mixing if and only if

$$\lim_n [U_f^n \varphi - (\varphi \cdot 1)] \cdot \psi = 0 \quad \text{for every } \varphi, \psi \in L^2(\mu), \tag{7.1.7}$$

and Proposition 7.1.13 gives that  $(f, \mu)$  is weak mixing if and only if

$$\lim_n \frac{1}{n} \sum_{j=1}^n |[U_f^j \varphi - (\varphi \cdot 1)] \cdot \psi| = 0 \quad \text{for every } \varphi, \psi \in L^2(\mu). \tag{7.1.8}$$

The condition (7.1.7) means that  $U_f^n \varphi$  converges *weakly* to  $\varphi \cdot 1 = \int \varphi d\mu$ , while (7.1.8) is a Cesàro version of that assertion. Compare both conditions with the characterization of ergodicity in (4.1.7).

**Corollary 7.1.14.** *Let  $f : M \rightarrow M$  be a mixing transformation relative to some invariant probability measure  $\mu$ . Let  $\nu$  be any probability measure on  $M$ , absolutely continuous with respect to  $\mu$ . Then  $f_*^n \nu$  converges pointwise to  $\mu$ , that is,  $\nu(f^{-n}(B)) \rightarrow \mu(B)$  for every measurable set  $B \subset M$ .*

*Proof.* Let  $\varphi = \chi_B$  and  $\psi = d\nu/d\mu$ . Note that  $\varphi \in L^\infty(\mu)$  and  $\psi \in L^1(\mu)$ . Hence, by Proposition 7.1.12,

$$\int (\chi_B \circ f^n) \frac{d\nu}{d\mu} d\mu = \int (U_f^n \varphi) \psi d\mu \rightarrow \int \varphi d\mu \int \psi d\mu = \int \chi_B d\mu \int \frac{d\nu}{d\mu} d\mu.$$

The sequence on the left-hand side coincides with  $\int (\chi_B \circ f^n) d\nu = \nu(f^{-n}(B))$ . The right-hand side is equal to  $\mu(B) \int 1 d\nu = \mu(B)$ .

### 7.1.4 Exercises

- 7.1.1. Show that  $(f, \mu)$  is mixing if and only if  $\mu(f^{-n}(A) \cap A) \rightarrow \mu(A)^2$  for every measurable set  $A$ .
- 7.1.2. Let  $(a_n)_n$  be a bounded sequence of real numbers. Prove that

$$\lim_n \frac{1}{n} \sum_{j=1}^n |a_j| = 0$$

if and only if there exists  $E \subset \mathbb{N}$  with density zero at infinity (that is, with  $\lim_n (1/n) \#(E \cap \{0, \dots, n-1\}) = 0$ ) such that the restriction of  $(a_n)_n$  to the complement of  $E$  converges to zero when  $n \rightarrow \infty$ . Deduce that

$$\lim_n \frac{1}{n} \sum_{j=1}^n |a_j| = 0 \quad \Leftrightarrow \quad \lim_n \frac{1}{n} \sum_{j=1}^n (a_j)^2 = 0.$$

- 7.1.3. Prove that if  $\mu$  is weak mixing for  $f$  then  $\mu$  is weak mixing for every iterate  $f^k$ ,  $k \geq 1$ .

- 7.1.4. Show that if no eigenvalue of  $A \in \text{SL}(d, \mathbb{R})$  is a root of unity then the linear endomorphism  $f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$  induced by  $A$  is mixing, with respect to the Haar measure.
- 7.1.5. Let  $f : M \rightarrow M$  be a measurable transformation in a metric space. Check that an invariant probability measure  $\mu$  is mixing if and only if  $(f_*^n \eta)_n$  converges to  $\mu$  in the weak\* topology for every probability measure  $\eta$  absolutely continuous with respect to  $\mu$ .
- 7.1.6. (Multiple von Neumann ergodic theorem). Show that if  $(f, \mu)$  is weak mixing then

$$\frac{1}{N} \sum_{n=0}^{N-1} (\varphi_1 \circ f^n) \cdots (\varphi_k \circ f^{kn}) \rightarrow \int \varphi_1 d\mu \cdots \int \varphi_k d\mu \quad \text{in } L^2(\mu),$$

for any bounded measurable functions  $\varphi_1, \dots, \varphi_k$ .

## 7.2 Markov shifts

In this section we introduce an important class of systems that generalizes the notion of Bernoulli shift. As explained previously, Bernoulli shifts model sequences of identical experiments such that the outcome of each experiment is *independent* of all the others. In the definition of Markov shifts we weaken this independence condition: we allow each outcome to depend on the preceding one, but not the others. More generally, Markov shifts may be used to model the so-called *finite memory* processes, that is, sequences of experiments for which there exists  $k \geq 1$  such that the outcome of each experiment depends only on the outcomes of the  $k$  previous experiments. In this regard, see Exercise 7.2.4.

To define a Markov shift, let  $(X, \mathcal{A})$  be a measurable space and  $\Sigma = X^{\mathbb{N}}$  (or  $\Sigma = X^{\mathbb{Z}}$ ) be the space of all sequences in  $X$ , endowed with the product  $\sigma$ -algebra. Let us consider the shift map

$$\sigma : \Sigma \rightarrow \Sigma, \quad \sigma((x_n)_n) = (x_{n+1})_n.$$

Let us be given a family  $\{P(x, \cdot) : x \in X\}$  of probability measures on  $X$  that depend measurably on the point  $x$ . They will be called *transition probabilities*: for each measurable set  $E \subset X$ , the number  $P(x, E)$  is meant to represent the probability that  $x_{n+1} \in E$ , given that  $x_n = x$ . A probability measure  $p$  in  $X$  is called a *stationary measure*, relative to the family of transition probabilities, if it satisfies

$$\int P(x, E) dp(x) = p(E), \quad \text{for every measurable set } E \subset X. \quad (7.2.1)$$

Heuristically, this means that, relative to  $p$ , a probability of the event  $x_{n+1} \in E$  is equal to the probability of the event  $x_n \in E$ .

Fix any stationary measure  $p$  (assuming it exists) and then define

$$\begin{aligned} \mu([m; A_m, \dots, A_n]) &= \int_{A_m} dp(x_m) \int_{A_{m+1}} dP(x_m, x_{m+1}) \cdots \int_{A_n} dP(x_{n-1}, x_n) \end{aligned} \tag{7.2.2}$$

for every cylinder  $[m; A_m, \dots, A_n]$  of  $\Sigma$ . One can show (check Exercise 7.2.1) that this function extends to a probability measure in the  $\sigma$ -algebra generated by the cylinders. This probability measure is invariant under the shift map  $\sigma$ , because the right-hand side of (7.2.2) does not depend on  $m$ . Every probability measure  $\mu$  obtained in this way is called a *Markov measure*; moreover, the system  $(\sigma, \mu)$  is called a *Markov shift*.

**Example 7.2.1** (Bernoulli measure). Suppose that  $P(x, \cdot)$  does not depend on  $x$ , that is, that there exists a probability measure  $\nu$  on  $X$  such that  $P(x, \cdot) = \nu$  for every  $x \in X$ . Then

$$\int P(x, E) dp(x) = \int \nu(E) dp(x) = \nu(E)$$

for every probability measure  $p$  and every measurable set  $E \subset X$ . Therefore, there exists exactly one stationary measure, namely  $p = \nu$ . The definition in (7.2.2) gives

$$\begin{aligned} \mu([m; A_m, \dots, A_n]) &= \int_{A_m} d\nu(x_m) \int_{A_{m+1}} d\nu(x_{m+1}) \cdots \int_{A_n} d\nu(x_n) \\ &= \nu(A_m)\nu(A_{m+1}) \cdots \nu(A_n). \end{aligned}$$

**Example 7.2.2.** Suppose that the set  $X$  is finite, say  $X = \{1, \dots, d\}$  for some  $d \geq 2$ . Any family of transition probabilities  $P(x, \cdot)$  on  $X$  is completely characterized by the values

$$P_{i,j} = P(i, \{j\}), \quad 1 \leq i, j \leq d. \tag{7.2.3}$$

Moreover, a measure  $p$  on the set  $X$  is completely characterized by the values  $p_i = p(\{i\})$ ,  $1 \leq i \leq d$ . With these notations, the definition (7.2.1) translates to

$$\sum_{i=1}^d p_i P_{i,j} = p_j, \quad \text{for every } 1 \leq j \leq d. \tag{7.2.4}$$

Moreover, a Markov measure  $\mu$  is determined by

$$\mu([m; a_m, \dots, a_n]) = p_{a_m} P_{a_m, a_{m+1}} \cdots P_{a_{n-1}, a_n}. \tag{7.2.5}$$

In the remainder of this book we always restrict ourselves to *finite* Markov shifts, that is, to the context of Example 7.2.2. We take the set  $X$  endowed with the discrete topology and the corresponding Borel  $\sigma$ -algebra. Observe that the matrix

$$P = (P_{i,j})_{1 \leq i, j \leq d}$$

defined by (7.2.3) satisfies the following conditions:

- (i)  $P_{ij} \geq 0$  for every  $1 \leq i, j \leq d$ ;  
 (ii)  $\sum_{j=1}^d P_{ij} = 1$  for every  $1 \leq i \leq d$ .

We say that  $P$  is a *stochastic matrix*. Conversely, any matrix that satisfies (i) and (ii) defines a family of transition probabilities on the set  $X$ . Observe also that, denoting  $p = (p_1, \dots, p_d)$ , the relation (7.2.4) corresponds to

$$P^*p = p, \quad (7.2.6)$$

where  $P^*$  denotes the transpose of the matrix  $P$ . In other words, the stationary measures correspond precisely to the eigenvectors of the transposed matrix for the eigenvalue 1. Using the following classical result, one can show that such eigenvalues always exist:

**Theorem 7.2.3** (Perron–Frobenius). *Let  $A$  be a  $d \times d$  matrix with non-negative coefficients. Then there exists  $\lambda \geq 0$  and some vector  $v \neq 0$  with non-negative coefficients such that  $Av = \lambda v$  and  $\lambda \geq |\gamma|$  for every eigenvalue  $\gamma$  of  $A$ .*

*If  $A$  has some power whose coefficients are all positive then  $\lambda > 0$  and it has some eigenvector  $v$  whose coefficients are all positive. Indeed,  $\lambda > |\gamma|$  for any other eigenvalue  $\gamma$  of  $A$ . Moreover, the eigenvalue  $\lambda$  has multiplicity 1 and it is the only eigenvalue of  $A$  having some eigenvector with non-negative coefficients.*

A proof of the Perron–Frobenius theorem may be found in Meyers [Mey00], for example. Applying this theorem to the matrix  $A = P^*$ , we conclude that there exist  $\lambda \geq 0$  and  $p \neq 0$  with  $p_i \geq 0$  for every  $i$ , such that

$$\sum_{i=1}^d p_i P_{ij} = \lambda p_j, \quad \text{for every } 1 \leq j \leq d.$$

Adding over  $j = 1, \dots, d$  we get that

$$\sum_{j=1}^d \sum_{i=1}^d p_i P_{ij} = \lambda \sum_{j=1}^d p_j.$$

Using property (ii) of the stochastic matrix, the left-hand side of this equality may be written as

$$\sum_{i=1}^d p_i \sum_{j=1}^d P_{ij} = \sum_{i=1}^d p_i.$$

Comparing the last two equalities and recalling that the sum of the coefficients of  $p$  is a positive number, we conclude that  $\lambda = 1$ . This proves our claim that there always exist vectors  $p \neq 0$  satisfying (7.2.6).

When  $P^n$  has positive coefficients for some  $n \geq 1$ , it follows from Theorem 7.2.3 that the eigenvector is unique up to scaling, and it may be chosen with positive coefficients.

**Example 7.2.4.** In general,  $p$  is not unique and it may also happen that there is no eigenvalue with positive coefficients. For example, consider:

$$P = \begin{pmatrix} 1-a & a & 0 & 0 & 0 \\ b & 1-b & 0 & 0 & 0 \\ 0 & 0 & 1-c & c & 0 \\ 0 & 0 & d & 1-d & 0 \\ e & 0 & 0 & 0 & 1-e \end{pmatrix}$$

where  $a, b, c, d, e \in (0, 1)$ . A vector  $p = (p_1, p_2, p_3, p_4, p_5)$  satisfies  $P^*p = p$  if and only if  $ap_1 = bp_2$  and  $cp_3 = dp_4$  and  $p_5 = 0$ . Therefore, the eigenspace has dimension 2 and no eigenvector has positive coefficients.

On the other hand, suppose that  $p$  is such that  $p_i = 0$  for some  $i$ . Let  $\mu$  be the corresponding Markov measure and let  $\Sigma_i = (X \setminus \{i\})^{\mathbb{N}}$  (or  $\Sigma_i = (X \setminus \{i\})^{\mathbb{Z}}$ ). Then  $\mu(\Sigma_i) = 1$ , since  $\mu([n; i]) = p_i = 0$  for every  $n$ . This means that we may eliminate the symbol  $i$ , and still have a system that is equivalent to the original one. Therefore, up to removing from the set  $X$  a certain number of superfluous symbols, we may always take the eigenvector  $p$  to have positive coefficients.

Denote by  $\Sigma_P$  the set of all sequences  $(x_n)_n \in \Sigma$  satisfying

$$P_{x_n x_{n+1}} > 0 \quad \text{for every } n, \tag{7.2.7}$$

that is, such that all the transitions are ‘‘allowed’’ by  $P$ . It is clear from the definition that  $\Sigma_P$  is invariant under the shift map  $\sigma$ . The transformations  $\sigma : \Sigma_P \rightarrow \Sigma_P$  constructed in this way are called *shifts of finite type* and will be studied in more detail in Section 10.2.2.

**Lemma 7.2.5.** *The set  $\Sigma_P$  is closed in  $\Sigma$  and, given any solution of  $P^*p = p$  with positive coefficients, the support of the corresponding Markov measure  $\mu$  coincides with  $\Sigma_P$ .*

*Proof.* Let  $x^k = (x_n^k)_n$ ,  $k \in \mathbb{N}$  be any sequence in  $\Sigma_P$  and suppose that it converges in  $\Sigma$  to some  $x = (x_n)_n$ . By the definition of the topology in  $\Sigma$ , this means that for every  $n$  there exists  $k_n \geq 1$  such that  $x_n^k = x_n$  for every  $k \geq k_n$ . So, given any  $n$ , taking  $k \geq \max\{k_n, k_{n+1}\}$  we conclude that  $P_{x_n, x_{n+1}} = P_{x_n^k, x_{n+1}^k} > 0$ . This shows that  $x \in \Sigma_P$  and that proves the first part of the lemma.

To prove the second part, recall that the cylinders  $[m; x_m, \dots, x_n]$  form a basis of neighborhoods of any  $x = (x_n)_n$  in  $\Sigma$ . If  $x \in \Sigma_P$  then

$$\mu([m; x_m, \dots, x_n]) = p_{x_m} P_{x_m x_{m+1}} \cdots P_{x_{n-1} x_n} > 0$$

for every cylinder and, thus,  $x \in \text{supp } \mu$ . If  $x \notin \Sigma_P$  then there exists  $n$  such that  $P_{x_n, x_{n+1}} = 0$ . In that case,  $\mu([n; x_n, x_{n+1}]) = 0$  and so  $x \notin \text{supp } \mu$ .

**Example 7.2.6.** There are three possibilities for the support of a Markov measure in Example 7.2.4. If  $p = (p_1, p_2, 0, 0, 0)$  with  $p_1, p_2 > 0$  then we may eliminate the symbols 3, 4, 5. All the sequences of symbols 1, 2 are admissible.

Hence  $\text{supp } \mu = \{1, 2\}^{\mathbb{N}}$ . Analogously, if  $p = (0, 0, p_3, p_4, 0)$  with  $p_3, p_4 > 0$  then  $\text{supp } \mu = \{3, 4\}^{\mathbb{N}}$ . In all the other cases,  $p = (p_1, p_2, p_3, p_4, 0)$  with  $p_1, p_2, p_3, p_4 > 0$ . Eliminating the symbol 5, we get that the set of admissible sequences is

$$\Sigma_P = \{1, 2\}^{\mathbb{N}} \cup \{3, 4\}^{\mathbb{N}}.$$

Both sets in this union are invariant and have positive measure. So, in this case the Markov shift  $(\sigma, \mu)$  is not ergodic. But it follows from the theory presented in the next section that in the previous two cases the system  $(\sigma, \mu)$  is indeed ergodic.

In the next lemma we collect some simple properties of stochastic matrices that will be useful in what follows:

**Lemma 7.2.7.** *Let  $P$  be a stochastic matrix and  $p = (p_1, \dots, p_d)$  be a solution of  $P^*p = p$ . For every  $n \geq 0$ , denote by  $P_{ij}^n$ ,  $1 \leq i, j \leq d$  the coefficients of the matrix  $P^n$ . Then:*

- (i)  $\sum_{j=1}^d P_{ij}^n = 1$  for every  $1 \leq i \leq d$  and every  $n \geq 1$ ;
- (ii)  $\sum_{i=1}^d p_i P_{ij}^n = p_j$  for every  $1 \leq j \leq d$  and every  $n \geq 1$ ;
- (iii) the hyperplane  $H = \{(h_1, \dots, h_d) : h_1 + \dots + h_d = 0\}$  is invariant under the matrix  $P^*$ .

*Proof.* Condition (ii) in the definition of stochastic matrix may be written as  $Pu = u$ , with  $u = (1, \dots, 1)$ . Then  $P^n u = u$  for every  $n \geq 1$ . This is just another way of writing claim (i). Analogously,  $P^*p = p$  implies that  $(P^*)^n p = p$  for every  $n \geq 1$ , which is another way of writing claim (ii). Observe that  $H$  is the orthogonal complement of vector  $u$ . Since  $u$  is invariant under  $P$ , it follows that  $H$  is invariant under the transposed matrix  $P^*$ , as claimed in (iii).

### 7.2.1 Ergodicity

In this section we always take  $p = (p_1, \dots, p_d)$  to be a solution of  $P^*p = p$  with  $p_i > 0$  for every  $i$ , normalized in such a way that  $\sum_i p_i = 1$ . Let  $\mu$  be the corresponding Markov measure. We want to understand which conditions the stochastic matrix  $P$  must satisfy for the system  $(\sigma, \mu)$  to be ergodic.

We say that a stochastic matrix  $P$  is *irreducible* if for every  $1 \leq i, j \leq d$  there exists  $n \geq 0$  such that  $P_{ij}^n > 0$ . In other words,  $P$  is irreducible if any outcome  $i$  may be followed by any outcome  $j$ , after a certain number  $n$  of steps which may depend on  $i$  and  $j$ .

**Theorem 7.2.8.** *The Markov shift  $(\sigma, \mu)$  is ergodic if and only if the matrix  $P$  is irreducible.*

The remainder of the present section is dedicated to the proof of this theorem. We start by proving the following useful estimate:

**Lemma 7.2.9.** Let  $A = [m; a_m, \dots, a_q]$  and  $B = [r; b_r, \dots, b_s]$  be cylinders of  $\Sigma$  with  $r > q$ . Then

$$\mu(A \cap B) = \mu(A)\mu(B) \frac{P_{a_q, b_r}^{r-q}}{p_{b_r}}.$$

*Proof.* We may write  $A \cap B$  as a disjoint union

$$A \cap B = \bigcup_x [m; a_m, \dots, a_q, x_{q+1}, \dots, x_{r-1}, b_r, \dots, b_s],$$

over all  $x = (x_{q+1}, \dots, x_{r-1}) \in X^{r-q-1}$ . Then,

$$\begin{aligned} \mu(A \cap B) &= \sum_x p_{a_m} P_{a_m, a_{m+1}} \dots P_{a_{q-1}, a_q} P_{a_q, x_{q+1}} \dots P_{x_{r-1}, b_r} P_{b_r, b_{r+1}} \dots P_{b_{s-1}, b_s} \\ &= \mu(A) \sum_x P_{a_q, x_{q+1}} \dots P_{x_{r-1}, b_r} \frac{1}{p_{b_r}} \mu(B). \end{aligned}$$

The sum in this last expression is equal to  $P_{a_q, b_r}^{r-q}$ . Therefore,

$$\mu(A \cap B) = \mu(A)\mu(B) P_{a_q, b_r}^{r-q} / p_{b_r},$$

as stated.

**Lemma 7.2.10.** A stochastic matrix  $P$  is irreducible if and only if

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} P_{ij}^l = p_j \quad \text{for every } 1 \leq i, j \leq d. \tag{7.2.8}$$

*Proof.* Assume that (7.2.8) holds. Recall that  $p_j > 0$  for every  $j$ . Then, given any  $1 \leq i, j \leq d$ , we have  $P_{ij}^l > 0$  for infinitely many values of  $l$ . In particular,  $P$  is irreducible.

To prove the converse, consider  $A = [0; i]$  and  $B = [0; j]$ . By Lemma 7.2.9,

$$\frac{1}{n} \sum_{l=0}^{n-1} \mu(A \cap \sigma^{-l}(B)) = \frac{1}{p_j} \mu(A)\mu(B) \frac{1}{n} \sum_{l=0}^{n-1} P_{ij}^l.$$

According to Exercise 4.1.2, the left-hand side converges when  $n \rightarrow \infty$ . Therefore,

$$Q_{ij} = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P_{ij}^l$$

exists for every  $1 \leq i, j \leq d$ . Consider the matrix  $Q = (Q_{ij})_{i,j}$ , that is,

$$Q = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P^l. \tag{7.2.9}$$

Using Lemma 7.2.7(ii) and taking the limit when  $n \rightarrow \infty$ , we get that

$$\sum_{i=1}^d p_i Q_{ij} = p_j \quad \text{for every } 1 \leq j \leq d. \tag{7.2.10}$$

Observe also that, given any  $k \geq 1$ ,

$$P^k Q = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P^{k+l} = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P^l = Q. \tag{7.2.11}$$

It follows that  $Q_{i,j}$  does not depend on  $i$ . Indeed, suppose that there exist  $r$  and  $s$  such that  $Q_{r,j} < Q_{s,j}$ . Of course, we may choose  $s$  in such a way that the right-hand side of this inequality is larger. Since  $P$  is irreducible, there exists  $k$  such that  $P^k_{s,r} > 0$ . Hence, using (7.2.11) followed by Lemma 7.2.7(i),

$$Q_{s,j} = \sum_{i=1}^d P^k_{s,i} Q_{i,j} < \left( \sum_{i=1}^d P^k_{s,i} \right) Q_{s,j} = Q_{s,j},$$

which is a contradiction. This contradiction proves that  $Q_{i,j}$  does not depend on  $i$ , as claimed. Write  $Q_j = Q_{i,j}$  for any  $i$ . The property (7.2.10) gives that

$$p_j = \sum_{i=1}^d Q_{i,j} p_i = Q_j \left( \sum_{i=1}^d p_i \right) = Q_j,$$

for every  $j$ . This finishes the proof of the lemma.

*Proof of Theorem 7.2.8.* Suppose that  $\mu$  is ergodic. Let  $A = [0; i]$  and  $B = [0; j]$ . By Proposition 4.1.4,

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) = p_i p_j. \tag{7.2.12}$$

On the other hand, by Lemma 7.2.9, we have that  $\mu(A \cap \sigma^{-l}(B)) = p_i P^l_{i,j}$ . Using this identity in (7.2.12) and dividing both sides by  $p_i$  we find that

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} P^l_{i,j} = p_j.$$

Note that  $j$  is arbitrary. So, by Lemma 7.2.10, this proves that  $P$  is irreducible.

Now suppose that the matrix  $P$  is irreducible. We want to conclude that  $\mu$  is ergodic. According to Corollary 4.1.5, it is enough to prove that

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) \tag{7.2.13}$$

for any  $A$  and  $B$  in the algebra generated by the cylinders. Since the elements of this algebra are the finite pairwise disjoint unions of cylinders, it suffices to consider the case when  $A$  and  $B$  are cylinders, say  $A = [m; a_m, \dots, a_q]$  and  $B = [r; b_r, \dots, b_s]$ . Observe also that the validity of (7.2.13) is not affected if one replaces  $B$  by some pre-image  $\sigma^{-j}(B)$ . So, it is no restriction to suppose that  $r > q$ . Then, by Lemma 7.2.9,

$$\frac{1}{n} \sum_{l=0}^{n-1} \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) \frac{1}{p_{b_r}} \frac{1}{n} \sum_{l=0}^{n-1} P^{r-q+l}_{a_q, b_r}$$



for every  $n$ . By Lemma 7.2.10,

$$\lim_n \frac{1}{n} \sum_{l=0}^{n-1} P_{a_q, b_r}^{r-q+l} = \lim_n \frac{1}{n} \sum_{l=0}^{n-1} P_{a_q, b_r}^l = p_{b_r}.$$

This proves the property (7.2.13) for the cylinders  $A$  and  $B$ .

### 7.2.2 Mixing

In this section we characterize the Markov shifts that are mixing, in terms of the corresponding stochastic matrix  $P$ . As before, we take  $p$  to be a normalized solution of  $P^*p = p$  with positive coefficients and  $\mu$  to be the corresponding Markov measure.

We say that a stochastic matrix  $P$  is *aperiodic* if there exists  $n \geq 1$  such that  $P_{ij}^n > 0$  for every  $1 \leq i, j \leq d$ . In other words,  $P$  is aperiodic if some power  $P^n$  has only positive coefficients. The relation between the notions of aperiodicity and irreducibility is analyzed in Exercise 7.2.6.

**Theorem 7.2.11.** *The Markov shift  $(\sigma, \mu)$  is mixing if and only if the matrix  $P$  is aperiodic.*

For the proof of Theorem 7.2.11 we need the following fact:

**Lemma 7.2.12.** *A stochastic matrix  $P$  is aperiodic if and only if*

$$\lim_l P_{ij}^l = p_j \quad \text{for every } 1 \leq i, j \leq d. \tag{7.2.14}$$

*Proof.* Since we assume that  $p_j > 0$  for every  $j$ , it is clear that (7.2.14) implies that  $P_{ij}^l > 0$  for every  $i, j$  and every  $l$  sufficiently large.

Now suppose that  $P$  is aperiodic. Then we may apply the theorem of Perron–Frobenius (Theorem 7.2.3) to the matrix  $A = P^*$ . Since  $p$  is an eigenvector of  $A$  with positive coefficients, we get that  $\lambda = 1$  and all the other eigenvalues of  $A$  are smaller than 1 in absolute value. By Lemma 7.2.7(iii), the hyperplane  $H$  formed by the vectors  $(h_1, \dots, h_d)$  with  $h_1 + \dots + h_d = 0$  is invariant under  $A$ . It is clear that  $H$  is transverse to the direction of  $p$ . Then the decomposition

$$\mathbb{R}^d = \mathbb{R}p \oplus H \tag{7.2.15}$$

is invariant under  $A$  and the restriction of  $A$  to the hyperplane  $H$  is a contraction, meaning that its spectral radius is smaller than 1. It follows that the sequence  $(A^l)_l$  converges to the projection on the first coordinate of (7.2.15), that is, to the matrix  $B$  characterized by  $Bp = p$  and  $Bh = 0$  for every  $h \in H$ . In other words,  $(P^l)_l$  converges to  $B^*$ . Observe that

$$B_{ij} = p_i \quad \text{for every } 1 \leq i, j \leq d.$$

Therefore,  $\lim_n P_{ij}^n = B_{j,i} = p_j$  for every  $i, j$ .

*Proof of Theorem 7.2.11.* Suppose that the measure  $\mu$  is mixing. Let  $A = [0; i]$  and  $B = [0; j]$ . By Lemma 7.2.9, we have that  $\mu(A \cap \sigma^{-l}(B)) = p_i P_{ij}^l$  for every  $l$ . Therefore,

$$p_i \lim_l P_{ij}^l = \lim_l \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) = p_i p_j.$$

Dividing both sides by  $p_i$  we get that  $\lim_l P_{ij}^l = p_j$ . According to Lemma 7.2.12, this proves that  $P$  is aperiodic.

Now suppose that the matrix  $P$  is aperiodic. We want to conclude that  $\mu$  is mixing. According to Lemma 7.1.2, it is enough to prove that

$$\lim_l \mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) \tag{7.2.16}$$

for any  $A$  and  $B$  in the algebra generated by the cylinders. Since the elements of this algebra are the finite pairwise disjoint unions of cylinders, it suffices to treat the case when  $A$  and  $B$  are cylinders, say  $A = [m; a_m, \dots, a_q]$  and  $B = [r; b_r, \dots, b_s]$ . By Lemma 7.2.9,

$$\mu(A \cap \sigma^{-l}(B)) = \mu(A)\mu(B) \frac{1}{p_{b_r}} P_{a_q, b_r}^{r-q+l}$$

for every  $l > q - r$ . Then, using Lemma 7.2.12,

$$\begin{aligned} \lim_l \mu(A \cap \sigma^{-l}(B)) &= \mu(A)\mu(B) \frac{1}{p_{b_r}} \lim_l P_{a_q, b_r}^{r-q+l} \\ &= \mu(A)\mu(B) \frac{1}{p_{b_r}} \lim_l P_{a_q, b_r}^l = \mu(A)\mu(B) \end{aligned}$$

This proves the property (7.2.16) for cylinders  $A$  and  $B$ .

**Example 7.2.13.** In Example 7.2.4 we found different types of Markov measures, depending on the choice of the probability eigenvector  $p$ . In the first case,  $p = (p_1, p_2, 0, 0, 0)$  and the measure  $\mu$  is supported on  $\{1, 2\}^{\mathbb{N}}$ . Once the superfluous symbols 3, 4, 5 have been removed, the stochastic matrix reduces to

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}.$$

Since this matrix is aperiodic, the Markov measure  $\mu$  is mixing. The second case is entirely analogous. In the third case,  $p = (p_1, p_2, p_3, p_4, 0)$  and, after removing the superfluous symbol 5, the stochastic matrix reduces to

$$P = \begin{pmatrix} 1-a & a & 0 & 0 \\ b & 1-b & 0 & 0 \\ 0 & 0 & 1-c & c \\ 0 & 0 & d & 1-d \end{pmatrix}.$$

This matrix is not irreducible and, hence, the Markov measures that one finds in this case are not ergodic (recall also Example 7.2.6).

**Example 7.2.14.** It is not difficult to find examples of irreducible matrices that are not aperiodic:

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

Indeed, we see that  $P^n_{ij} > 0$  if and only if  $n$  has the same parity as  $i - j$ . Note that

$$P^2 = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}.$$

Exercise 7.2.6 shows that every irreducible matrix has a form of this type.

### 7.2.3 Exercises

7.2.1. Let  $X = \{1, \dots, d\}$  and  $P = (P_{ij})_{i,j}$  be a stochastic matrix and  $p = (p_i)_i$  be a probability vector such that  $P^*p = p$ . Show that the function defined on the set of all cylinders by

$$\mu([m; a_m, \dots, a_n]) = p_{a_m} P_{a_m, a_{m+1}} \cdots P_{a_{n-1}, a_n}$$

extends to a measure on the Borel  $\sigma$ -algebra of  $\Sigma = X^{\mathbb{N}}$  (or  $\Sigma = X^{\mathbb{Z}}$ ), invariant under the shift map  $\sigma : \Sigma \rightarrow \Sigma$ .

7.2.2. Prove that every weak mixing Markov shift is actually mixing.

7.2.3. Let  $X = \{1, \dots, d\}$  and let  $\mu$  be a Markov measure for the shift map  $\sigma : X^{\mathbb{Z}} \rightarrow X^{\mathbb{Z}}$ . Does it follow that  $\mu$  is also a Markov measure for the inverse  $\sigma^{-1} : \Sigma \rightarrow \Sigma$ ?

7.2.4. Let  $X$  be a finite set and  $\Sigma = X^{\mathbb{Z}}$  (or  $\Sigma = X^{\mathbb{N}}$ ). Let  $\mu$  be a probability measure on  $\Sigma$ , invariant under the shift map  $\sigma : \Sigma \rightarrow \Sigma$ . Given  $k \geq 0$ , we say that  $\mu$  has *memory*  $k$  if

$$\frac{\mu([m-l; a_{m-l}, \dots, a_{m-1}, a_m])}{\mu([m-l; a_{m-l}, \dots, a_{m-1}])} = \frac{\mu([m-k; a_{m-k}, \dots, a_{m-1}, a_m])}{\mu([m-k; a_{m-k}, \dots, a_{m-1}])}$$

for every  $l \geq k$ , every  $m$  and every  $(a_n)_n \in \Sigma$  (by convention, the equality holds whenever at least one of the denominators is zero). Check that the measures with memory 0 are the Bernoulli measures and the measures with memory 1 are the Markov measures. Show that every measure with memory  $k \geq 2$  is equivalent to a Markov measure in the space  $\tilde{\Sigma} = \tilde{X}^{\mathbb{Z}}$  (or  $\tilde{\Sigma} = \tilde{X}^{\mathbb{N}}$ ), where  $\tilde{X} = X^k$ .

7.2.5. The goal is to show that the set of all measures with finite memory is dense in the space  $\mathcal{M}_1(\sigma)$  of all probability measures invariant under the shift map  $\sigma : \Sigma \rightarrow \Sigma$ . Given any invariant probability measure  $\mu$  and any  $k \geq 1$ , consider the function  $\mu_k$  defined on the set of all cylinders by

- $\mu_k = \mu$  for cylinders with length less than or equal to  $k$ ;
- for every  $l \geq k$ , every  $m$  and every  $(a_n)_n \in \Sigma$ ,

$$\frac{\mu_k([m-l; a_{m-l}, \dots, a_{m-1}, a_m])}{\mu_k([m-l; a_{m-l}, \dots, a_{m-1}])} = \frac{\mu([m-k; a_{m-k}, \dots, a_{m-1}, a_m])}{\mu([m-k; a_{m-k}, \dots, a_{m-1}])}.$$

Show that  $\mu_k$  extends to a probability measure on the Borel  $\sigma$ -algebra of  $\Sigma$ , invariant under the shift map and with memory  $k$ . Check that  $\lim_k \mu_k = \mu$  in the weak\* topology.

7.2.6. Let  $P$  be an irreducible stochastic matrix. The goal is to show that there exist  $\kappa \geq 1$  and a partition of  $X$  into  $\kappa$  subsets such that the restriction of  $P^\kappa$  to each of these subsets is aperiodic. To do so:

- (1) For every  $i \in X$ , define  $R(i) = \{n \geq 1 : P^n_{i,i} > 0\}$ . Show that  $R(i)$  is closed under addition: if  $n_1, n_2 \in R(i)$  then  $n_1 + n_2 \in R(i)$ .
- (2) Let  $R \subset \mathbb{N}$  be closed under addition and let  $\kappa \geq 1$  be the greatest common divisor of its elements. Show that there exists  $m \geq 1$  such that  $R \cap [m, \infty) = \kappa\mathbb{N} \cap [m, \infty)$ .
- (3) Show that the greatest common divisor  $\kappa$  of the elements of  $R(i)$  does not depend on  $i \in X$  and that  $P$  is aperiodic if and only if  $\kappa = 1$ .
- (4) Assume that  $\kappa \geq 2$ . Find a partition  $\{X_r : 0 \leq r < \kappa\}$  of  $X$  such that the restriction of  $P^\kappa$  to each  $X_r$  is aperiodic.

### 7.3 Interval exchanges

By definition, an *interval exchange* is a bijection of the interval  $[0, 1)$  with a finite number of discontinuities and whose restriction to every continuity subinterval is a translation. Figure 7.1 describes an example with four continuity subintervals. To fix ideas, we always take the transformation to be continuous on the right, that is, we take all continuity subintervals to be closed on the left and open on the right.

As a direct consequence of the definition, every interval exchange preserves the Lebesgue measure on  $[0, 1)$ . These transformations exhibit a very rich dynamical behavior and they also have important connections with many other systems, such as polygonal billiards, conservative flows on surfaces and Teichmüller flows. For example, the construction that we sketch next shows that interval exchanges arise naturally as Poincaré return maps of conservative vector fields on surfaces.

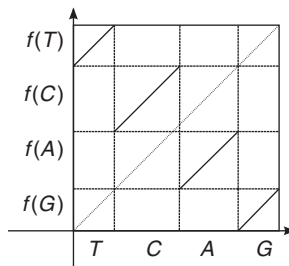


Figure 7.1. An interval exchange

**Example 7.3.1.** Let  $S$  be an orientable surface and  $\omega$  be an area form in  $S$ , that is, a differential 2-form that is non-degenerate at every point. We may associate with every vector field  $X$  a differential 1-form  $\beta$ , defined by

$$\beta_x(v) = \omega_x(X(x), v) \quad \text{for every vector } v \in T_xS.$$

Observe that  $X$  and  $\beta$  have the same zeros. Moreover, at all other points the kernel of  $\beta$  coincides with the direction of the vector field. The 1-form  $\beta$  permits the definition of the notion of “transverse arc-length” of curves  $c : [a, b] \rightarrow S$ , as follows:

$$\ell(c) = \int_c \beta = \int_a^b \beta_{c(t)}(\dot{c}(t)) dt.$$

Note that the flow trajectories have transverse arc-length zero. However, for curves transverse to the flow, the measure  $\ell$  is equivalent to the usual arc-length measure, in the sense that they have the same zero measure sets. We can show (see Exercise 7.3.1) that the 1-form  $\beta$  is closed if and only if  $X$  preserves area. Then, using the theorem of Green, the Poincaré maps of the flow preserve the transverse length. With an additional hypothesis on the zeros of  $X$ , the first-return map  $f : \Sigma \rightarrow \Sigma$  to any cross-section  $\Sigma$  is well defined and is continuous outside a finite subset of  $\Sigma$ . Then, parameterizing  $\Sigma$  by transverse arc length,  $f$  is an interval exchange.

An interval exchange is determined by two ingredients. The first one, of a combinatorial nature, concerns the number of continuity subintervals, the order of these subintervals and the order of their images inside the interval  $[0, 1]$ . This may be informed by assigning a label (for example, a letter) to each continuity subinterval and to its image, and by listing these labels in their corresponding orders, in two horizontal rows. For example, in the case described in Figure 7.1, we obtain

$$\pi = \begin{pmatrix} T & C & A & G \\ G & A & C & T \end{pmatrix}.$$

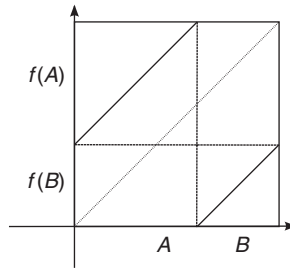
Note that the choice of the labels is arbitrary. We denote by  $\mathcal{A}$ , and call *alphabet*, the set of all labels.

The second ingredient, of a metric nature, concerns the lengths of the subintervals. This may be expressed through a vector with positive coefficients, indexed by the alphabet: each coefficient determines the length of the corresponding continuity subinterval (and of its image). In the case of Figure 7.1 this *length vector* has the form

$$\lambda = (\lambda_T, \lambda_C, \lambda_A, \lambda_G).$$

The sum of the coefficients of a length vector is always equal to 1.

Then, the interval exchange  $f : [0, 1] \rightarrow [0, 1]$  associated with each pair  $(\pi, \lambda)$  is defined as follows. For every label  $\alpha \in \mathcal{A}$ , denote by  $I_\alpha$  the corresponding continuity subinterval and define  $w_\alpha = v_1 - v_0$ , where  $v_0$  is the sum of the



**Figure 7.2.** Rotation viewed as an exchange of two intervals

lengths  $\lambda_\beta$  corresponding to all labels  $\beta$  to the left of  $\alpha$  on the top row of  $\pi$  and  $v_1$  is the sum of the lengths  $\lambda_\gamma$  corresponding to all the labels  $\gamma$  to the left of  $\alpha$  on the bottom row of  $\pi$ . Then

$$f(x) = x + w_\alpha \quad \text{for every } x \in I_\alpha.$$

The vector  $w = (w_\alpha)_{\alpha \in \mathcal{A}}$  is called the *translation vector*. Clearly, for each fixed  $\pi$ , the translation vector is a linear function of the length vector  $\lambda = (\lambda_\alpha)_{\alpha \in \mathcal{A}}$ .

**Example 7.3.2.** The simplest interval exchanges have only two continuity subintervals. See Figure 7.2. Choosing the alphabet  $\mathcal{A} = \{A, B\}$ , we get

$$\pi = \begin{pmatrix} A & B \\ B & A \end{pmatrix} \quad \text{and} \quad f(x) = \begin{cases} x + \lambda_B & \text{for } x \in I_A \\ x - \lambda_A = x + \lambda_B - 1 & \text{for } x \in I_B. \end{cases}$$

This transformation corresponds precisely to the rotation  $R_{\lambda_B}$  if we identify  $[0, 1)$  with the circle  $S^1$  in the natural way. In this sense, the class of interval exchanges are a generalization of the family of circle rotations.

### 7.3.1 Minimality and ergodicity

As we saw previously, a circle rotation  $R_\theta$  is minimal if and only if  $\theta$  is irrational. Moreover, in that case  $R_\theta$  is also uniquely ergodic. Given that almost every number is irrational, this means that minimality and unique ergodicity are typical in the family of circle rotations. In this section we discuss the two properties in the broader context of interval exchanges.

Let us start with an observation that has no analogue for rotations. We say that the combinatorics  $\pi$  of an interval exchange *reducible* if there exists some position such that the labels to the left of that position in the two rows of  $\pi$  are exactly the same. For example,

$$\pi = \begin{pmatrix} B & X & O & L & F & D \\ X & O & B & F & D & L \end{pmatrix}$$

is reducible, as the labels to the left of the fourth position are the same in both rows:  $B, O$  and  $X$ . As a consequence, for any length vector  $\lambda$ , the interval

exchange  $f$  defined by  $(\pi, \lambda)$  leaves the subinterval  $I_B \cup I_O \cup I_X$  invariant. In particular,  $f$  cannot be minimal, not even transitive. In what follows we always assume the combinatorics  $\pi$  to be irreducible.

It is natural to ask whether the interval exchange is minimal whenever the length vector  $\lambda = (\lambda_\alpha)_{\alpha \in \mathcal{A}}$  is rationally independent, that is, whenever

$$\sum_{\alpha \in \mathcal{A}} n_\alpha \lambda_\alpha \neq 0$$

for every non-zero vector  $(n_\alpha)_{\alpha \in \mathcal{A}}$  with integer coefficients. This turns out to be true but, in fact, the hypothesis of rational independence is a bit too strong: we are going to present a somewhat more general condition that still implies minimality.

We denote by  $\partial I_\alpha$  the left endpoint of each subinterval  $I_\alpha$ . We say that a pair  $(\pi, \lambda)$  satisfies the *Keane condition* if the trajectories of these points are disjoint:

$$f^m(\partial I_\alpha) \neq \partial I_\beta \quad \text{for every } m \geq 1 \text{ and any } \alpha, \beta \in \mathcal{A} \text{ with } \partial I_\beta \neq 0 \quad (7.3.1)$$

(note that there always exist  $\bar{\alpha}$  and  $\bar{\beta}$  such that  $f(\partial I_{\bar{\alpha}}) = 0 = \partial I_{\bar{\beta}}$ ). We leave the proof of the next lemma as an exercise (Exercise 7.3.2):

**Lemma 7.3.3.** (1) *If the pair  $(\pi, \lambda)$  satisfies the Keane condition then the combinatorics matrix  $\pi$  is irreducible.*

(2) *If  $\pi$  is irreducible and  $\lambda$  is rationally independent then the pair  $(\pi, \lambda)$  satisfies the Keane condition.*

Since the subset of rationally independent vectors has full Lebesgue measure, it follows that the Keane condition is satisfied for almost every length vector  $\lambda$ , if  $\pi$  is irreducible.

**Example 7.3.4.** In the case of two subintervals (recall Example 7.3.2), the interval exchange has the form  $f^m(x) = x + m\lambda_B \pmod{\mathbb{Z}}$ . Then, the Keane condition means that

$$m\lambda_B \neq \lambda_A + n \quad \text{and} \quad \lambda_A + m\lambda_B \neq \lambda_A + n$$

for every  $m \in \mathbb{N}$  and  $n \in \mathbb{Z}$ . It is clear that this holds if and only if the vector  $(\lambda_A, \lambda_B)$  is rationally independent.

**Example 7.3.5.** For exchanges of three or more intervals, the Keane condition is strictly weaker than the rational independence of the length vector. Consider, for example,

$$\pi = \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}.$$

Then  $f^m(x) = x + m\lambda_C \pmod{\mathbb{Z}}$  and, thus, the Keane condition means that

$$\{m\lambda_C, \lambda_A + m\lambda_C, \lambda_A + \lambda_B + m\lambda_C\} \quad \text{and} \quad \{\lambda_A + n, \lambda_A + \lambda_B + n\}$$

are disjoint for every  $m \in \mathbb{N}$  and  $n \in \mathbb{Z}$ . Equivalently,

$$p\lambda_C \notin \{q, \lambda_A + q\} \quad \text{for every } p \in \mathbb{Z} \text{ and } q \in \mathbb{Z}.$$

This may hold even when  $(\lambda_A, \lambda_B, \lambda_C)$  is rationally dependent.

The following result was proved by Michael Keane [Kea75]:

**Theorem 7.3.6** (Keane). *If  $(\pi, \lambda)$  satisfies the Keane condition then the interval exchange  $f$  is minimal.*

**Example 7.3.7.** The Keane condition is not necessary for minimality. For example, consider the interval exchange defined by  $(\pi, \lambda)$ , where

$$\pi = \begin{pmatrix} A & B & C & D \\ D & C & B & A \end{pmatrix},$$

$\lambda_A = \lambda_C$ ,  $\lambda_B = \lambda_D$  and  $\lambda_A/\lambda_B = \lambda_C/\lambda_D$  is irrational. Then  $(\pi, \lambda)$  does not satisfy the Keane condition and yet  $f$  is minimal.

As observed previously, every minimal circle rotation is also uniquely ergodic. This is still true for exchanges of three intervals, but not in general. Indeed, Keane gave an example of an exchange of four intervals exhibiting two ergodic probability measures, notwithstanding the fact that the combinatorics matrix  $\pi$  is irreducible and the length vector  $\lambda$  is rationally independent.

Keane conjectured that, nevertheless, it should be true that almost every interval exchange is uniquely ergodic. The following remarkable result, obtained independently by Howard Masur [Mas82] and William Veech [Vee82], established this conjecture:

**Theorem 7.3.8** (Masur, Veech). *Assume that  $\pi$  is irreducible. Then, for Lebesgue-almost every length vector  $\lambda$ , the interval exchange defined by  $(\pi, \lambda)$  is uniquely ergodic.*

Earlier, Michael Keane and Gérard Rauzy [KR80] had shown that unique ergodicity holds for a residual (Baire second category) subset of length vectors whenever the combinatorics is irreducible.

### 7.3.2 Mixing

The interval exchanges provide many examples of systems that are uniquely ergodic and weak mixing but not (strongly) mixing.

Indeed, the theorem of Masur–Veech (Theorem 7.3.8) asserts that almost every interval exchange is uniquely ergodic. Another deep theorem, due to Artur Avila and Giovanni Forni [AF07], states that, circle rotations (more precisely, interval exchanges with a unique discontinuity point) excluded, almost every interval exchange is weak mixing. The topological version of this fact had been proved by Arnaldo Nogueira and Donald Rudolph [NR97].



On the other hand, a result of Anatole Katok [Kat80] that we are going to discuss below asserts that interval exchanges are never mixing:

**Theorem 7.3.9.** *Let  $f : [0, 1) \rightarrow [0, 1)$  be an interval exchange and  $\mu$  be an invariant probability measure. Then  $(f, \mu)$  is not mixing.*

*Proof.* We may take  $\mu$  to be ergodic, for otherwise the conclusion is obvious. If  $\mu$  has some atom then its support is a periodic orbit and, thus,  $\mu$  cannot be mixing. Hence, we may also take  $\mu$  to be non-atomic. Denote by  $m$  the Lebesgue measure on the interval and consider the map

$$h : [0, 1) \rightarrow [0, 1), \quad h(x) = \mu([0, x]).$$

Then  $h$  is a homeomorphism and satisfies  $h_*\mu = m$ . Consequently, the map  $g = h \circ f \circ h^{-1} : [0, 1) \rightarrow [0, 1)$  has finitely many discontinuity points and preserves the Lebesgue measure. In particular, the restriction of  $g$  to each continuity subinterval is a translation. Therefore,  $g$  is also an interval exchange. It is clear that  $(f, \mu)$  is mixing if and only if  $(g, m)$  is mixing. Therefore, to prove Theorem 7.3.9 it is no restriction to suppose that  $\mu$  is the Lebesgue measure  $m$ . We do that from now on.

Our goal is to find a measurable set  $X$  such that  $m(X \cap f^{-n}(X))$  does not converge to  $m(X)^2$  when  $n \rightarrow \infty$ . Let  $d = \#\mathcal{A}$ .

**Lemma 7.3.10.** *Every interval  $J = [a, b)$  contained in some  $I_\beta$  admits a partition  $\{J_1, \dots, J_s\}$  into no more than  $d + 2$  subintervals of the form  $J_i = [a_i, b_i)$  and admits natural numbers  $t_1, \dots, t_s \geq 1$  such that*

- (i)  $f^n(J_i) \cap J = \emptyset$  for every  $0 < n < t_i$  and  $1 \leq i \leq s$ ;
- (ii)  $f^{t_i} \upharpoonright J_i$  is a translation for every  $1 \leq i \leq s$ ;
- (iii)  $\{f^{t_1}(J_1), \dots, f^{t_s}(J_s)\}$  is a partition of  $J$ ;
- (iv) the intervals  $f^n(J_i)$ ,  $1 \leq i \leq s$ ,  $0 \leq n < t_i$  are pairwise disjoint;
- (v)  $\bigcup_{n=0}^\infty f^n(J) = \bigcup_{i=1}^s \bigcup_{n=0}^{t_i-1} f^n(J_i)$ .

*Proof.* Let  $B$  be the set formed by the endpoints  $a$  and  $b$  of  $J$  together with the endpoints  $\partial I_\alpha$ ,  $\alpha \in \mathcal{A}$  minus the origin. Then  $\#B \leq d + 1$ . Let  $B_J \subset J$  be the set of points  $x \in J$  for which there exists  $m \geq 1$  such that  $f^m(x) \in B$  and  $f^n(x) \notin J$  for every  $0 < n < m$ . The fact that  $f$  is injective, together with the definition of  $m$ , implies that the map

$$B_J \rightarrow B, \quad x \mapsto f^m(x)$$

is injective. In particular,  $\#B_J \leq \#B \leq d + 1$ . Consider the partition of  $J$  into subintervals  $J_i = [a_i, b_i)$  with endpoints  $a_i, b_i$  in the set  $B_J \cup \{a, b\}$ . This partition has at most  $d + 2$  elements. By the Poincaré recurrence theorem, for each  $i$  there exists  $t_i \geq 1$  such that  $f^{t_i}(J_i)$  intersects  $J$ . Take  $t_i$  minimum with this property. Part (i) of the lemma is an immediate consequence. By the definition of  $B_J$ , the restriction of  $f^{t_i}$  to the interval  $J_i$  is a translation, as stated in part (ii), and

its image is contained in  $J$ . Moreover, the images  $f^{t_i}(J_i)$ ,  $1 \leq i \leq s$  are pairwise disjoint, since  $f$  is injective and the  $t_i$  are the first-return times to  $J$ . In particular,

$$\sum_{i=1}^s m(f^{t_i}(J_i)) = \sum_{i=1}^s m(J_i) = m(J)$$

and so  $\bigcup_{i=1}^s f^{t_i}(J_i) = J$ . This proves part (iii). Part (iv) also follows directly from the fact that  $f$  is injective and the  $t_i$  are the first-return times to  $J$ . Finally, part (v) is a direct consequence of part (iii).

Consider any interval  $J$  contained in some  $I_\beta$ . By ergodicity, the invariant set  $\bigcup_{n=0}^\infty f^n(J)$  has full measure. By part (v) of Lemma 7.3.10, this set is a finite union of intervals closed on the left and open on the right. Therefore,

$$\bigcup_{n=0}^\infty f^n(J) = \bigcup_{i=1}^s \bigcup_{n=0}^{t_i-1} f^n(J_i) = I.$$

So, by Lemma 7.3.10(iv), the family  $\mathcal{P}_J = \{f^n(J_i) : 1 \leq i \leq s \text{ and } 0 \leq n < t_i\}$  is a partition of  $I$ .

**Lemma 7.3.11.** *Given  $\delta > 0$  and  $N \geq 1$ , we may choose the interval  $J$  in such a way that  $\text{diam } \mathcal{P}_J < \delta$  and  $t_i \geq N$  for every  $i$ .*

*Proof.* It is clear that  $\text{diam } f^n(J_i) = \text{diam } J_i \leq \text{diam } J$  for every  $i$  and every  $n$ . Hence,  $\text{diam } \mathcal{P}_J < \delta$  as long as we pick  $J$  with diameter smaller than  $\delta$ . To get the second property in the statement, take any point  $x \in I$  such that  $f^n(x) \neq \partial I_\alpha$  for every  $0 \leq n < N$  and  $\alpha \in \mathcal{A}$ . We claim that  $f^n(x) \neq x$  for every  $0 < n < N$ . Otherwise, since  $f^n$  is a translation in the neighborhood of  $x$ , we would have  $f^n(y) = y$  for every point  $y$  in that neighborhood, which would contradict the hypothesis that  $(f, m)$  is ergodic. This proves our claim. Now it suffices to take  $J = [x, x + \varepsilon)$  with  $\varepsilon < \min_{0 < n < N} d(x, f^n(x))$  to ensure that  $t_i \geq N$  for every  $i$ .

**Lemma 7.3.12.** *For every  $1 \leq i \leq s$  there exist  $s_i \leq d + 2$  and natural numbers  $\{t_{i,1}, \dots, t_{i,s_i}\}$  such that  $t_{i,j} \geq t_i$  and, given any set  $A$  in the algebra  $\mathcal{A}_J$  generated by  $\mathcal{P}_J$ , there exists  $t_{i,j}$  such that*

$$m(A \cap f^{-t_{i,j}}(A)) \geq \frac{1}{(d+2)^2} m(A). \tag{7.3.2}$$

*Proof.* Applying Lemma 7.3.10 to each of the intervals  $J_i$ ,  $1 \leq i \leq s$  we find  $s_i \leq d + 2$ , a partition  $\{J_{i,j} : 1 \leq j \leq s_i\}$  of the interval  $J_i$  and natural numbers  $t_{i,j}$  such that each  $t_{i,j}$  is the first-return time of the points of  $J_{i,j}$  to  $J_i$ . It is clear that  $t_{i,j} \geq t_i$ , since  $t_i$  is the first-return time of any point of  $J_i$  to the interval  $J$ . The fact that  $J_{i,j} \subset f^{-t_{i,j}}(J_i)$  implies that

$$f^n(J_i) = \bigcup_{j=1}^{s_i} f^n(J_{i,j}) \subset \bigcup_{j=1}^{s_i} f^{-t_{i,j}}(f^n(J_i)) \quad \text{for every } n \geq 0.$$

Since the algebra  $\mathcal{A}_J$  is formed by the finite pairwise disjoint unions of intervals  $f^n(J_i)$ ,  $0 \leq n < t_i$ , it follows that

$$A \subset \bigcup_{i=1}^s \bigcup_{j=1}^{s_i} f^{-t_{ij}}(A) \quad \text{for every } A \in \mathcal{A}_J.$$

In particular,  $m(A) \leq \sum_{i=1}^s \sum_{j=1}^{s_i} m(A \cap f^{-t_{ij}}(A))$ . Recalling that  $s \leq d + 2$  and  $s_i \leq d + 2$  for every  $i$ , this implies (7.3.2).

We are ready to conclude the proof of Theorem 7.3.9. For that, let us fix a measurable set  $X \subset [0, 1)$  with

$$0 < m(X) < \frac{1}{4(d+2)^2}.$$

By Lemma 7.3.11, given any  $N \geq 1$  we may find an interval  $J \subset [0, 1)$  such that all the first-return times  $t_i \geq N$  and there exists  $A \in \mathcal{A}_J$  such that

$$m(X \Delta A) < \frac{1}{4}m(X)^2. \tag{7.3.3}$$

Applying Lemma 7.3.12, we get that there exists  $t_{i,j} \geq t_i \geq N$  such that:

$$\begin{aligned} m(X \cap f^{-t_{ij}}(X)) &\geq m(A \cap f^{-t_{ij}}(A)) - 2m(X \Delta A) \\ &\geq \frac{1}{(d+2)^2}m(A) - \frac{1}{2}m(X)^2. \end{aligned}$$

The relation (7.3.3) implies that  $m(A) \geq (3/4)m(X)$ . Therefore,

$$\begin{aligned} m(X \cap f^{-t_{ij}}(X)) &\geq \frac{3}{4} \frac{1}{(d+2)^2}m(X) - \frac{1}{2}m(X)^2 \\ &\geq 3m(X)^2 - \frac{1}{2}m(X)^2 > 2m(X)^2. \end{aligned}$$

This proves that  $\limsup_n m(X \cap f^{-n}(X)) \geq 2m(X)^2$ , and so the system  $(f, m)$  cannot be mixing.

### 7.3.3 Exercises

- 7.3.1. Let  $\omega$  be an area form on a surface. Let  $X$  be a differentiable vector field on  $S$  and  $\beta$  be the differential 1-form defined on  $S$  by  $\beta_x = \omega_x(X(x), \cdot)$ . Show that  $\beta$  is closed if and only if  $X$  preserves the area measure.
- 7.3.2. Prove Lemma 7.3.3.
- 7.3.3. Show that if  $(\pi, \lambda)$  satisfies the Keane condition then  $f$  has no periodic points. [Observation: This is a step in the proof of Theorem 7.3.6.]
- 7.3.4. Let  $f : [0, 1) \rightarrow [0, 1)$  be an irreducible interval exchange and let  $a \in (0, 1)$  be the largest of all the discontinuity points of  $f$  and  $f^{-1}$ . The *Rauzy–Veech renormalization*  $R(f) : [0, 1) \rightarrow [0, 1)$  is defined by  $R(f)(x) = g(ax)/a$ , where  $g$  is the first-return map of  $f$  to the interval  $[0, a)$ . Check that  $R(f)$  is an interval exchange with the same number of continuity subintervals as  $f$ , or less. If  $f$  is described by the data  $(\pi, \lambda)$ , how can we describe  $R(f)$ ?

7.3.5. Given  $d \geq 2$  and a bijection  $\sigma : \mathbb{N} \rightarrow \mathbb{N}$  without periodic points, consider the transformation  $f : [0, 1] \rightarrow [0, 1]$  where each  $f(x)$  is obtained by permuting the digits of the base  $d$  expansion of  $x$  as prescribed by  $\sigma$ . More precisely, if  $x = \sum_{n=1}^{\infty} a_n d^{-n}$  with  $a_n \in \{0, \dots, d-1\}$  and infinitely many values of  $n$  such that  $a_n < d-1$ , then  $f(x) = \sum_{n=1}^{\infty} a_{\sigma(n)} d^{-n}$ . Show that  $f$  preserves the Lebesgue measure  $m$  in the interval and that  $(f, m)$  is mixing.

### 7.4 Decay of correlations

In this section we discuss how quickly the correlations sequence  $C_n(\varphi, \psi)$  decays to zero in a mixing system. Since we are dealing with deterministic systems, we cannot expect interesting estimates to hold for arbitrary functions. However, as we are going to see, such estimates do exist in many important cases, if we restrict  $\varphi$  and  $\psi$  to suitable subsets of functions. Given that the correlations  $(\varphi, \psi) \mapsto C_n(\varphi, \psi)$  are bilinear functions, it is natural to consider subsets that are vector subspaces.

We say that  $(f, \mu)$  has *exponential decay of correlations* on a given vector space  $\mathcal{V}$  if there exists  $\lambda < 1$  and for every  $\varphi, \psi \in \mathcal{V}$  there exists  $A(\varphi, \psi) > 0$  such that

$$|C_n(\varphi, \psi)| \leq A(\varphi, \psi)\lambda^n \quad \text{for every } n \geq 1. \tag{7.4.1}$$

There are similar notions (polynomial decay, for instance) where the exponential  $\lambda^n$  is replaced by some other sequence converging to zero.

To illustrate the theory, let us analyze the issue of decay of correlations in the context of one-sided Markov shifts. That will also allow us to introduce several ideas that will be useful later in more general situations. Let  $f : M \rightarrow M$  be the shift map in  $M = X^{\mathbb{N}}$ , where  $X = \{1, \dots, d\}$  is a finite set. Let  $P = (P_{ij})_{i,j}$  be an aperiodic stochastic matrix and  $p = (p_i)_i$  be the positive eigenvector of  $P^*$ , normalized by  $p_1 + \dots + p_d = 1$ . Let  $\mu$  be the Markov measure defined in  $M$  by (7.2.2).

Consider  $L = G^{-1}P^*G$ , where  $G$  is the diagonal matrix whose coefficients are  $p_1, \dots, p_d$ . The coefficients of  $L$  are given by

$$L_{i,j} = \frac{p_j}{p_i} P_{j,i} \quad \text{for each } 1 \leq i, j \leq d.$$

Recall that we denote  $u = (1, \dots, 1)$  and  $H = \{(h_1, \dots, h_d) : h_1 + \dots + h_d = 0\}$ . Let

$$V = \{(v_1, \dots, v_d) : p_1 v_1 + \dots + p_d v_d = 0\}.$$

Then  $G(u) = p$  and  $G(V) = H$ . Recalling (7.2.15), it follows that the decomposition

$$\mathbb{R}^d = \mathbb{R}u \oplus V \tag{7.4.2}$$

is invariant under  $L$  and all the eigenvalues of the restriction of  $L$  to  $V$  are smaller than 1 in absolute value. We say that  $L$  has the *spectral gap property* if

the largest eigenvalue is simple and all the rest of the spectrum is contained in a closed disk with strictly smaller radius.

The transfer operator of the shift map  $f$  is the linear operator  $\mathcal{L}_f$  mapping each function  $\psi : M \rightarrow \mathbb{R}$  to the function  $\mathcal{L}_f \psi : M \rightarrow \mathbb{R}$  defined by

$$\mathcal{L}_f \psi(x_1, \dots, x_n, \dots) = \sum_{x_0=1}^d L_{x_1, x_0} \psi(x_0, x_1, \dots, x_n, \dots). \tag{7.4.3}$$

The transfer operator is dual to the Koopman operator  $U_f$ , in the following sense:

$$\int \varphi(\mathcal{L}_f \psi) d\mu = \int (U_f \varphi) \psi d\mu \tag{7.4.4}$$

for any bounded measurable functions  $\varphi, \psi$ . Let us prove this fact.

We call a function  $\varphi : M \rightarrow \mathbb{R}$  *locally constant* if there is  $n \geq 0$  such that every  $\varphi(x)$  depends only on the first  $n$  coordinates  $x_0, \dots, x_{n-1}$  of the point  $x$ . For example, characteristic functions of cylinders are locally constant functions. Since every bounded measurable function is a uniform limit of linear combinations of characteristic functions of cylinders, it follows that every bounded measurable function is the uniform limit of some sequence of locally constant functions. Thus, to prove (7.4.4) it suffices to consider the case when  $\varphi$  and  $\psi$  are both locally constant.

Then, consider functions  $\varphi$  and  $\psi$  that depend only on the first  $n$  coordinates. By the definition of Markov measure,

$$\int \varphi(\mathcal{L}_f \psi) d\mu = \sum_{a_1, \dots, a_n} p_{a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} \varphi(a_1, \dots, a_n) \mathcal{L}_f \psi(a_1, \dots, a_n).$$

Using the definition of the transfer operator, the right-hand side of this expression is equal to

$$\sum_{a_0, a_1, \dots, a_n} p_{a_0} P_{a_0, a_1} P_{a_1, a_2} \cdots P_{a_{n-1}, a_n} \varphi(a_1, \dots, a_n) \psi(a_0, a_1, \dots, a_n).$$

Observe that  $\varphi(a_1, \dots, a_n) = U_f \varphi(a_0, a_1, \dots, a_n)$ . So, using once more the definition of the Markov measure, this last expression is equal to  $\int (U_f \varphi) \psi d\mu$ . This proves the duality property (7.4.4).

As a consequence, we may write the correlations sequence in terms of the iterates of the transfer operator:

$$C_n(\varphi, \psi) = \int (U_f^n \varphi) \psi - \int \varphi d\mu \int \psi d\mu = \int \varphi (\mathcal{L}_f^n \psi - \int \psi d\mu) d\mu. \tag{7.4.5}$$

The property  $Lu = u$  means that  $\sum_j L_{i,j} = 1$  for every  $i$ . This has the following useful consequence:

$$\sup |\mathcal{L}_f^n \psi| \leq \sup |\psi| \quad \text{for every } \psi. \tag{7.4.6}$$

Taking  $\varphi \equiv 1$  in (7.4.4) we get the following special case, which will also be useful later:

$$\int \mathcal{L}_f \psi \, d\mu = \int \psi \, d\mu \quad \text{for every } \psi. \tag{7.4.7}$$

Now let us denote by  $\mathcal{E}_0$  the subset of functions  $\psi$  that depend only on the first coordinate. The map  $\psi \mapsto (\psi(1), \dots, \psi(d))$  is an isomorphism between  $\mathcal{E}_0$  and the Euclidean space  $\mathbb{R}^d$ . Moreover, the definition

$$\mathcal{L}_f \psi(x_1) = \sum_{x_0=1}^d L_{x_1, x_0} \psi(x_0)$$

shows that the restriction of the transfer operator to  $\mathcal{E}_0$  corresponds precisely to the operator  $L : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . Note also that the hyperplane  $V \subset \mathbb{R}^d$  corresponds to the subset of  $\psi \in \mathcal{E}_0$  such that  $\int \psi \, d\mu = 0$ . Consider in  $\mathcal{E}_0$  the norm defined by  $\|\psi\|_0 = \sup |\psi|$ .

Fix any number  $\lambda$  between 1 and the spectral radius of  $L$  restricted to  $V$ . Every function  $\psi \in \mathcal{E}_0$  may be written:

$$\psi = c + v \quad \text{with } c = \int \psi \, d\mu \in \mathbb{R}u \quad \text{and } v = \psi - \int \psi \, d\mu \in V.$$

Then the spectral gap property implies that there exists  $B > 1$  such that

$$\sup |\mathcal{L}_f^n \psi - \int \psi \, d\mu| \leq B \|\psi\|_0 \lambda^n \quad \text{for every } n \geq 1. \tag{7.4.8}$$

Using (7.4.5), it follows that

$$|C_n(\varphi, \psi)| \leq B \|\varphi\|_0 \|\psi\|_0 \lambda^n \quad \text{for every } n \geq 1.$$

In this way, we have shown that *every aperiodic Markov shift has exponential decay of correlations in  $\mathcal{E}_0$* .

With a little more effort, one can improve this result, by extending the conclusion to a much larger space of functions. Consider in  $M$  the distance defined by

$$d((x_n)_n, (y_n)_n) = 2^{-N(x,y)} \quad \text{where } N(x,y) = \min\{n \geq 0 : x_n \neq y_n\}.$$

Fix any  $\theta > 0$  and denote by  $\mathcal{E}$  the set of functions  $\varphi$  that are  $\theta$ -Hölder, that is, such that

$$K_\theta(\varphi) = \sup \left\{ \frac{|\varphi(x) - \varphi(y)|}{d(x,y)^\theta} : x \neq y \right\} \quad \text{is finite.}$$

It is clear that  $\mathcal{E}$  contains all the locally constant functions. We claim:

**Theorem 7.4.1.** *Every aperiodic Markov shift  $(f, \mu)$  has exponential decay of correlations in the space  $\mathcal{E}$  of  $\theta$ -Hölder functions, for any  $\theta > 0$ .*

Observe that  $\mathcal{L}_f(\mathcal{E}) \subset \mathcal{E}$ . The function  $\|\psi\| = \sup |\psi| + K_\theta(\psi)$  is a complete norm in  $\mathcal{E}$  and the linear operator  $\mathcal{L}_f : \mathcal{E} \rightarrow \mathcal{E}$  is continuous relative to this

norm (Exercise 7.4.1). One way to prove the theorem is by showing that this operator has the spectral gap property, with invariant decomposition

$$\mathcal{E} = \mathbb{R}u \oplus \left\{ \psi \in \mathcal{E} : \int \psi d\mu = 0 \right\}.$$

Once that is done, exactly the same argument that we used before for  $\mathcal{E}_0$  proves the exponential decay of correlations in  $\mathcal{E}$ . We do not present the details here (but we will come back to this theme, in a much more general context, near the end of Section 12.3). Instead, we give a direct proof that (7.4.8) may be extended to the space  $\mathcal{E}$ .

Given  $\psi \in \mathcal{E}$  and  $x = (x_1, \dots, x_n, \dots) \in M$ , we have

$$\mathcal{L}_f^k \psi(x) = \sum_{a_1, \dots, a_k=1}^d L_{x_1, a_k} \cdots L_{a_2, a_1} \psi(a_1, \dots, a_k, x_1, \dots, x_n, \dots)$$

for every  $k \geq 1$ . Then, given  $y = (y_1, \dots, y_n, \dots)$  with  $x_1 = y_1 = j$ ,

$$|\mathcal{L}_f^k \psi(x) - \mathcal{L}_f^k \psi(y)| \leq \sum_{a_1, \dots, a_k=1}^d L_{j, a_k} \cdots L_{a_2, a_1} K_\theta(\psi) 2^{-k\theta} d(x, y)^\theta.$$

Using the property  $\sum_{i=1}^d L_{j,i} = 1$ , we conclude that

$$|\mathcal{L}_f^k \psi(x) - \mathcal{L}_f^k \psi(y)| \leq K_\theta(\psi) 2^{-k\theta} d(x, y)^\theta \leq K_\theta(\psi) 2^{-k\theta}. \tag{7.4.9}$$

Given any function  $\varphi$ , denote by  $\pi\varphi$  the function that depends only on the first coordinate and coincides with the mean of  $\varphi$  on each cylinder  $[0; i]$ :

$$\pi\varphi(i) = \frac{1}{p_i} \int_{[0; i]} \varphi d\mu.$$

It is clear that  $\sup |\pi\varphi| \leq \sup |\varphi|$  and  $\int \pi\varphi d\mu = \int \varphi d\mu$ . The inequality (7.4.9) implies that

$$\sup |\mathcal{L}_f^k \psi - \pi(\mathcal{L}_f^k \psi)| \leq K_\theta(\psi) 2^{-k\theta} \quad \text{for every } k \geq 1.$$

Then, using the property (7.4.6),

$$\sup |\mathcal{L}_f^{k+l} \psi - \mathcal{L}_f^l \pi(\mathcal{L}_f^k \psi)| \leq K_\theta(\psi) 2^{-k\theta} \quad \text{for every } k, l \geq 1. \tag{7.4.10}$$

Moreover, the properties (7.4.6) and (7.4.7) imply that

$$\sup |\pi(\mathcal{L}_f^k \psi)| \leq \sup |\psi| \quad \text{and} \quad \int \pi(\mathcal{L}_f^k \psi) d\mu = \int \psi d\mu.$$

Therefore, the property (7.4.8) gives that

$$\sup \left| \mathcal{L}_f^l \pi(\mathcal{L}_f^k \psi) - \int \psi d\mu \right| \leq B \sup |\psi| \lambda^l \quad \text{for every } l \geq 1. \tag{7.4.11}$$

Adding (7.4.10) and (7.4.11), we get that

$$\sup \left| \mathcal{L}_f^{k+l} \psi - \int \psi d\mu \right| \leq K_\theta(\psi) 2^{-k\theta} + B \sup |\psi| \lambda^l \quad \text{for every } k, l \geq 1.$$

Fix  $\sigma < 1$  such that  $\sigma^2 \geq \max\{2^{-\theta}, \lambda\}$ . Then the previous inequality (with  $l \approx n/2 \approx k$ ) gives

$$\sup \left| \mathcal{L}_f^n \psi - \int \psi d\mu \right| \leq B \|\psi\| \sigma^{n-1} \quad \text{for every } n. \tag{7.4.12}$$

Now Theorem 7.4.1 follows from the same argument that we used before for  $\mathcal{E}_0$ , with (7.4.12) in the place of (7.4.8).

### 7.4.1 Exercises

- 7.4.1. Show that  $\|\varphi\| = \sup|\varphi| + K_\theta(\varphi)$  defines a complete norm in the space  $\mathcal{E}$  of  $\theta$ -Hölder functions and the transfer operator  $\mathcal{L}_f$  is continuous relative to this norm.
- 7.4.2. Let  $f : M \rightarrow M$  be a local diffeomorphism on a compact manifold  $M$  and  $d \geq 2$  be the degree of  $f$ . Assume that there exists  $\sigma > 1$  such that  $\|Df(x)v\| \geq \sigma \|v\|$  for every  $x \in M$  and every vector  $v$  tangent to  $M$  at the point  $x$ . Fix  $\theta > 0$  and let  $\mathcal{E}$  be the space of  $\theta$ -Hölder functions  $\varphi : M \rightarrow \mathbb{R}$ . For every  $\varphi \in \mathcal{E}$ , define

$$\mathcal{L}_f \varphi : M \rightarrow \mathbb{R}, \quad \mathcal{L}_f \varphi(y) = \frac{1}{d} \sum_{x \in f^{-1}(y)} \varphi(x).$$

- (a) Show that  $\inf \varphi \leq \inf \mathcal{L}_f \varphi \leq \sup \mathcal{L}_f \varphi \leq \sup \varphi$  and  $K_\theta(\mathcal{L}_f \varphi) \leq \sigma^{-\theta} K_\theta(\varphi)$  for every  $\varphi \in \mathcal{E}$ .
- (b) Conclude that  $\mathcal{L}_f : E \rightarrow E$  is a continuous linear operator (relative to the norm defined in Exercise 7.4.1) with  $\|\mathcal{L}_f\| = 1$ .
- (c) Show that, for every  $\varphi \in E$ , the sequence  $(\mathcal{L}_f^n \varphi)_n$  converges to a constant  $\nu_\varphi \in \mathbb{R}$  when  $n \rightarrow \infty$ . Moreover, there exists  $C > 0$  such that

$$\|\mathcal{L}_f^n \varphi - \nu_\varphi\| \leq C \sigma^{-n\theta} \|\varphi\| \quad \text{for every } n \text{ and every } \varphi \in \mathcal{E}.$$

- (d) Conclude that the operator  $\mathcal{L}_f : \mathcal{E} \rightarrow \mathcal{E}$  has the spectral gap property.
- (e) Show that the map  $\varphi \mapsto \nu_\varphi$  extends to a Borel probability measure on  $M$  (recall Theorem A.3.12).