

# 3

## Ergodic theorems

In this chapter we present the fundamental results of ergodic theory. To motivate the kind of statements that we are going to discuss, let us consider a measurable set  $E \subset M$  with positive measure and an arbitrary point  $x \in M$ . We want to analyze the set of iterates of  $x$  that visit  $E$ , that is,

$$\{j \geq 0 : f^j(x) \in E\}.$$

For example, the Poincaré recurrence theorem states that this set is infinite, for almost every  $x \in E$ . We would like to have more precise quantitative information. Let us call the *mean sojourn time* of  $x$  to  $E$  the value of

$$\tau(E, x) = \lim_{n \rightarrow \infty} \frac{1}{n} \#\{0 \leq j < n : f^j(x) \in E\}. \quad (3.0.1)$$

There is an analogous notion for flows, defined by

$$\tau(E, x) = \lim_{T \rightarrow \infty} \frac{1}{T} m(\{0 \leq t \leq T : f^t(x) \in E\}), \quad (3.0.2)$$

where  $m$  is the Lebesgue measure on the real line. It would be interesting to know, for example, under which conditions the mean sojourn time is positive. But before tackling this problem one must answer an even more basic question: when do the limits in (3.0.1)–(3.0.2) exist?

These questions go back to the work of the Austrian physicist Ludwig Boltzmann (1844–1906), who developed the kinetic theory of gases. Boltzmann was an emphatic supporter of the atomic theory, according to which gases are formed by a large number of small moving particles, constantly colliding with each other, at a time when this theory was still highly controversial. In principle, it should be possible to explain the behavior of a gas by applying the laws of classical mechanics to each one of these particles (molecules). In practice, this is not realistic because the number of molecules is huge.

The proposal of the kinetic theory was, then, to try and explain the behavior of gases at a macroscopic scale as the statistical combination of the motions of all its molecules. To formulate the theory in precise mathematical terms,

Boltzmann was forced to make an assumption that became known as the *ergodic hypothesis*. In modern language, the ergodic hypothesis claims that, for the kind of systems (Hamiltonian flows) that describe the motions of particles of a gas, *the mean sojourn time to any measurable set  $E$  exists and is equal to the measure of  $E$ , for almost every point  $x$ .*

Efforts to validate (or not) this hypothesis led to important developments, in mathematics (ergodic theory, dynamical systems) as well as in physics (statistical mechanics). In this chapter we concentrate on results concerning the *existence* of the mean sojourn time. The question of whether  $\tau(E, x) = \mu(E)$  for almost every  $x$  is the subject of Chapter 4.

Denoting by  $\varphi$  the characteristic function of the set  $E$ , we may rewrite the expression on the right-hand side of (3.0.1) as

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)). \quad (3.0.3)$$

This suggests a natural generalization of the original question: does the limit in (3.0.3) exist for more general functions  $\varphi$ , for example, for all integrable functions?

The ergodic theorem of von Neumann (Theorem 3.1.6) states that the limit in (3.0.3) does exist, in the space  $L^2(\mu)$ , for every function  $\varphi \in L^2(\mu)$ . The ergodic theorem of Birkhoff (Theorem 3.2.3) goes a lot further, by asserting that the convergence holds at  $\mu$ -almost every point, for every  $\varphi \in L^1(\mu)$ . In particular, the limit in (3.0.1) is well defined for  $\mu$ -almost every  $x$  (Theorem 3.2.1).

We give a direct proof of the theorem of von Neumann and we also show how it can be deduced from the theorem of Birkhoff. Concerning the latter, we are going to see that it can be obtained as a special case of an even stronger result, the subadditive ergodic theorem of Kingman (Theorem 3.3.3). This theorem asserts that  $\psi_n/n$  converges almost everywhere, for any sequence of functions  $\psi_n$  such that  $\psi_{m+n} \leq \psi_m + \psi_n \circ f^m$  for every  $m, n$ .

All these results remain valid for flows, as we comment upon in Section 3.4.

## 3.1 Ergodic theorem of von Neumann

In this section we state and prove the ergodic theorem of von Neumann. We begin by reviewing some general ideas concerning isometries in Hilbert spaces. See Appendices A.6 and A.7 for more information on this topic.

### 3.1.1 Isometries in Hilbert spaces

Let  $H$  be a Hilbert space and  $F$  be a closed subspace of  $H$ . Then,

$$H = F \oplus F^\perp, \quad (3.1.1)$$

where  $F^\perp = \{w \in H : v \cdot w = 0 \text{ for every } v \in F\}$  is the orthogonal complement of  $F$ . The projection  $P_F : H \rightarrow F$  associated with the decomposition (3.1.1) is called the *orthogonal projection* to  $F$ . It is uniquely characterized by

$$\|x - P_F(x)\| = \min\{\|x - v\| : v \in F\}.$$

Observe that  $P_F(v) = v$  for every  $v \in F$  and, consequently,  $P_F^2 = P_F$ .

**Example 3.1.1.** Consider the Hilbert space  $L^2(\mu)$ , with the inner product

$$\varphi \cdot \psi = \int \varphi \bar{\psi} d\mu.$$

Let  $F$  be the subspace of constant functions. Given any  $\varphi \in L^2(\mu)$ , we have that  $(P_F(\varphi) - \varphi) \cdot 1 = 0$ , that is,

$$P_F(\varphi) \cdot 1 = \varphi \cdot 1.$$

Since  $P_F(\varphi)$  is a constant function, the expression on the left-hand side is equal to  $P_F(\varphi)$ . The expression on the right-hand side is equal to  $\int \varphi d\mu$ . Therefore, the orthogonal projection to the subspace  $F$  is given by

$$P_F(\varphi) = \int \varphi d\mu.$$

Recall that the adjoint operator  $U^* : H \rightarrow H$  of a continuous linear operator  $U : H \rightarrow H$  is defined by the relation

$$U^*u \cdot v = u \cdot Uv \quad \text{for every } u, v \in H. \quad (3.1.2)$$

The operator  $U$  is said to be an *isometry* if it preserves the inner product:

$$Uu \cdot Uv = u \cdot v \quad \text{for every } u, v \in H. \quad (3.1.3)$$

This is equivalent to saying that  $U$  preserves the norm of  $H$  (see Exercise A.6.9). Another equivalent condition is  $U^*U = \text{id}$ . Indeed,

$$Uu \cdot Uv = u \cdot v \text{ for every } u, v \quad \Leftrightarrow \quad U^*Uu \cdot v = u \cdot v \text{ for every } u, v.$$

The property  $U^*U = \text{id}$  implies that  $U$  is injective. In general, an isometry need not be surjective. See Exercises 2.3.5 and 2.3.6. If an isometry is surjective then it is an isomorphism; such isometries are also called unitary operators.

**Example 3.1.2.** If  $f : M \rightarrow M$  preserves a measure  $\mu$  then, as we saw in Section 2.3.2, the Koopman operator  $U_f : L^2(\mu) \rightarrow L^2(\mu)$  is an isometry. If  $f$  is invertible then  $U_f$  is a unitary operator.

We call the *set of invariant vectors* of a continuous linear operator  $U : H \rightarrow H$  the subspace

$$I(U) = \{v \in H : Uv = v\}.$$

Observe that  $I(U)$  is a closed vector subspace, since  $U$  is continuous and linear. When  $U$  is an isometry, we have that  $I(U) = I(U^*)$ :

**Lemma 3.1.3.** *If  $U : H \rightarrow H$  is an isometry then  $Uv = v$  if and only if  $U^*v = v$ .*

*Proof.* Since  $U^*U = \text{id}$ , it is clear that  $Uv = v$  implies  $U^*v = v$ . Now assume that  $U^*v = v$ . Then,  $Uv \cdot v = v \cdot U^*v = v \cdot v = \|v\|^2$ . So, using the fact that  $U$  preserves the norm of  $H$ ,

$$\|Uv - v\|^2 = (Uv - v) \cdot (Uv - v) = \|Uv\|^2 - Uv \cdot v - v \cdot Uv + \|v\|^2 = 0.$$

This means that  $Uv = v$ .

To close this brief digression, let us quote a classical result from functional analysis, due to Marshall H. Stone, that permits the reduction of the study of Koopman operators of continuous time systems to the discrete case.

Let  $U_t : H \rightarrow H, t \in \mathbb{R}$  be a 1-parameter group of linear operators on a Banach space: by this we mean that  $U_0 = \text{id}$  and  $U_{t+s} = U_t U_s$  for every  $t, s \in \mathbb{R}$ . We say that the group is *strongly continuous* if

$$\lim_{t \rightarrow t_0} U_t v = U_{t_0} v, \quad \text{for every } t_0 \in \mathbb{R} \text{ and } v \in H.$$

**Theorem 3.1.4** (Stone). *If  $U_t : H \rightarrow H, t \in \mathbb{R}$  is a strongly continuous 1-parameter group of unitary operators on a complex Hilbert space then there exists a self-adjoint operator  $A$ , defined on a dense subspace  $D(A)$  of  $H$ , such that  $U_t |_{D(A)} = e^{itA}$  for every  $t \in \mathbb{R}$ .*

A proof may be found in Yosida [Yos68, § IX.9] and a simple application is given in Exercise 3.1.5. The operator  $iA$  is called the *infinitesimal generator* of the group. It may be retrieved through

$$iAv = \lim_{t \rightarrow 0} \frac{1}{t} (U_t v - v). \tag{3.1.4}$$

See Yosida [Yos68, § IX.3] for a proof of the fact that the limit on the right-hand side exists for every  $v$  in a dense subspace of  $H$ .

**Example 3.1.5.** Let  $H$  be the Banach space of continuous functions  $\varphi : S^1 \rightarrow \mathbb{C}$ , with the norm of uniform convergence. Define  $U_t(\varphi)(x) = \varphi(x+t)$  for every function  $\varphi \in H$ . Observe that  $(U_t)_t$  is a strongly continuous 1-parameter group of isometries of  $H$ . The infinitesimal generator is given by

$$iA\varphi(x) = \lim_{t \rightarrow 0} \frac{1}{t} (U_t \varphi(x) - \varphi(x)) = \lim_{t \rightarrow 0} \frac{1}{t} (\varphi(x+t) - \varphi(x)) = \varphi'(x).$$

Its domain is the subset of functions of class  $C^1$ , which is well known to be dense in  $H$ .

### 3.1.2 Statement and proof of the theorem

Our first ergodic theorem is:

**Theorem 3.1.6** (von Neumann). *Let  $U : H \rightarrow H$  be an isometry in a Hilbert space  $H$  and  $P$  be the orthogonal projection to the subspace  $I(U)$  of invariant*

vectors of  $U$ . Then,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} U^j v = Pv \quad \text{for every } v \in H. \tag{3.1.5}$$

*Proof.* Let  $L(U)$  be the set of vectors  $v \in H$  of the form  $v = Uu - u$  for some  $u \in H$  and let  $\bar{L}(U)$  be its closure. We claim that

$$I(U) = \bar{L}(U)^\perp. \tag{3.1.6}$$

This can be checked as follows. Consider any  $v \in I(U)$  and  $w \in \bar{L}(U)$ . By Lemma 3.1.3, we have that  $v \in I(U^*)$ , that is,  $U^*v = v$ . Moreover, by definition of  $\bar{L}(U)$ , there are  $u_n \in H$ ,  $n \geq 1$  such that  $(Uu_n - u_n)_n \rightarrow w$ . Since

$$v \cdot (Uu_n - u_n) = v \cdot Uu_n - v \cdot u_n = U^*v \cdot u_n - v \cdot u_n = 0$$

for every  $n$ , we conclude that  $v \cdot w = 0$ . This proves that  $I(U) \subset \bar{L}(U)^\perp$ . Next, consider any  $v \in \bar{L}(U)^\perp$ . Then, in particular,

$$v \cdot (Uu - u) = 0 \quad \text{or, equivalently,} \quad U^*v \cdot u - v \cdot u = 0$$

for every  $u \in H$ . This means that  $U^*v = v$ . Using Lemma 3.1.3 once more, we deduce that  $v \in I(U)$ . This shows that  $\bar{L}(U)^\perp \subset I(U)$ , which completes the proof of (3.1.6). As a consequence, using (3.1.1),

$$H = I(U) \oplus \bar{L}(U). \tag{3.1.7}$$

Now we prove the identity (3.1.5), successively, for  $v \in I(U)$ , for  $v \in \bar{L}(U)$  and for any  $v \in H$ . Begin by supposing that  $v \in I(U)$ . On the one hand,  $Pv = v$ . On the other hand,

$$\frac{1}{n} \sum_{j=0}^{n-1} U^j v = \frac{1}{n} \sum_{j=0}^{n-1} v = v$$

for every  $n$ , and so this sequence converges to  $v$  when  $n \rightarrow \infty$ . Combining these two observations we get (3.1.5) in this case.

Next, suppose that  $v \in L(U)$ . Then, by definition, there exists  $u \in H$  such that  $v = Uu - u$ . It is clear that

$$\frac{1}{n} \sum_{j=0}^{n-1} U^j v = \frac{1}{n} \sum_{j=0}^{n-1} (U^{j+1}u - U^j u) = \frac{1}{n} (U^n u - u).$$

The norm of this last expression is bounded by  $2\|u\|/n$  and, consequently, converges to zero when  $n \rightarrow \infty$ . This shows that

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} U^j v = 0 \quad \text{for every } v \in L(U). \tag{3.1.8}$$

More generally, suppose that  $v \in \bar{L}(U)$ . Then, there exist vectors  $v_k \in L(U)$  converging to  $v$  when  $k \rightarrow \infty$ . Observe that

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} U^j v - \frac{1}{n} \sum_{j=0}^{n-1} U^j v_k \right\| \leq \frac{1}{n} \sum_{j=0}^{n-1} \|U^j(v - v_k)\| \leq \|v - v_k\|$$

for every  $n$  and every  $k$ . Together with (3.1.8), this implies that

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} U^j v = 0 \quad \text{for every } v \in \bar{L}(U). \tag{3.1.9}$$

Since (3.1.6) implies that  $Pv = 0$  for every  $v \in \bar{L}(U)$ , this shows that (3.1.5) holds also when  $v \in \bar{L}(U)$ .

The general case of (3.1.5) follows immediately, as  $H = I(U) \oplus \bar{L}(U)$ .

### 3.1.3 Convergence in $L^2(\mu)$

Given a measurable transformation  $f : M \rightarrow M$  and an invariant probability measure  $\mu$  on  $M$ , we say that a measurable function  $\psi : M \rightarrow \mathbb{R}$  is *invariant* if  $\psi \circ f = \psi$  at  $\mu$ -almost every point. The following result is a special case of Theorem 3.1.6:

**Theorem 3.1.7.** *Given any  $\varphi \in L^2(\mu)$ , let  $\tilde{\varphi}$  be the orthogonal projection of  $\varphi$  to the subspace of invariant functions. Then the sequence*

$$\frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j \tag{3.1.10}$$

*converges to  $\tilde{\varphi}$  in the space  $L^2(\mu)$ . If  $f$  is invertible, then the sequence*

$$\frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^{-j} \tag{3.1.11}$$

*also converges to  $\tilde{\varphi}$  in  $L^2(\mu)$ .*

*Proof.* Let  $U = U_f : L^2(\mu) \rightarrow L^2(\mu)$  be the Koopman operator of  $(f, \mu)$ . Note that a function  $\psi$  is in the subspace  $I(U)$  of invariant functions if and only if  $\psi \circ f = \psi$  at  $\mu$ -almost every point. By Theorem 3.1.6, the sequence in (3.1.10) converges in the space  $L^2(\mu)$  to the orthogonal projection  $\tilde{\varphi}$  of  $\varphi$  to the subspace  $I(U)$ . This proves the first claim.

The second one is analogous, taking instead  $U = U_{f^{-1}}$ , which is the inverse of  $U_f$ . We get that the sequence in (3.1.11) converges in  $L^2(\mu)$  to the orthogonal projection of  $\varphi$  to the subspace  $I(U_{f^{-1}})$ . Observing that  $I(U_{f^{-1}}) = I(U_f)$ , we conclude that the limit of this sequence is just the same function  $\tilde{\varphi}$  as before.

### 3.1.4 Exercises

3.1.1. Show that under the hypotheses of the von Neumann ergodic theorem one has the following stronger conclusion:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j \rightarrow P(\varphi).$$

3.1.2. Use the previous exercise to show that, given any  $A \subset M$  with  $\mu(A) > 0$ , the set of values of  $n \in \mathbb{N}$  such that  $\mu(A \cap f^{-n}(A)) > 0$  is syndetic. [Observation: We have seen a different proof of this fact in Exercise 1.2.5.]

3.1.3. Prove that the set  $F = \{\varphi \in L^1(\mu) : \varphi \text{ is } f\text{-invariant}\}$  is a closed subspace of  $L^1(\mu)$ .

3.1.4. State and prove a version of the von Neumann ergodic theorem for flows.

3.1.5. Let  $f_t : M \rightarrow M$ ,  $t \in \mathbb{R}$  be a continuous flow on a compact metric space  $M$  and  $\mu$  be an invariant probability measure. Check that the 1-parameter group  $U_t : L^2(\mu) \rightarrow L^2(\mu)$ ,  $t \in \mathbb{R}$  of Koopman operators  $\varphi \mapsto U_t \varphi = \varphi \circ f_t$  is strongly continuous. Show that  $\mu$  is ergodic if and only if 0 is a simple eigenvalue of the infinitesimal generator of the group.

## 3.2 Birkhoff ergodic theorem

The theorem that we present in this section was proven by George David Birkhoff,<sup>1</sup> the prominent American mathematician of his generation and author of many other fundamental contributions to dynamics. It is a substantial improvement of the von Neumann ergodic theorem, because its conclusion is stated in terms of convergence at  $\mu$ -almost every point, which in this context is a stronger property than convergence in  $L^2(\mu)$ , as explained in Section 3.2.3.

### 3.2.1 Mean sojourn time

We start by stating the version of the theorem for mean sojourn times:

**Theorem 3.2.1** (Birkhoff). *Let  $f : M \rightarrow M$  be a measurable transformation and  $\mu$  be a probability measure invariant under  $f$ . Given any measurable set  $E \subset M$ , the mean sojourn time*

$$\tau(E, x) = \lim_n \frac{1}{n} \#\{j = 0, 1, \dots, n-1 : f^j(x) \in E\}$$

*exists at  $\mu$ -almost every point  $x \in M$ . Moreover,  $\int \tau(E, x) d\mu(x) = \mu(E)$ .*

Observe that if  $\tau(E, x)$  exists for some  $x \in M$  then

$$\tau(E, f(x)) = \tau(E, x). \quad (3.2.1)$$

<sup>1</sup> His son Garret Birkhoff was also a mathematician, and is well known for his work in algebra. The notion of projective distance that we use in Section 12.3 was due to him.

Indeed, by definition,

$$\begin{aligned} \tau(E, f(x)) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathcal{X}_E(f^j(x)) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) - \frac{1}{n} [\mathcal{X}_E(x) - \mathcal{X}_E(f^n(x))] \\ &= \tau(E, x) - \lim_{n \rightarrow \infty} \frac{1}{n} [\mathcal{X}_E(x) - \mathcal{X}_E(f^n(x))]. \end{aligned}$$

Since the characteristic function is bounded, the last limit is equal to zero. This proves (3.2.1).

The next example shows that the mean sojourn time does not exist for every point, in general:

**Example 3.2.2.** Consider the number  $x \in (0, 1)$  defined by the decimal expansion  $x = 0.a_1a_2a_3\dots$ , where  $a_i = 1$  if  $2^k \leq i < 2^{k+1}$  with  $k$  even and  $a_i = 0$  if  $2^k \leq i < 2^{k+1}$  with  $k$  odd. In other words,

$$x = 0.100111100000000111111111111110\dots,$$

where the lengths of the alternating blocks of 0s and 1s are given by successive powers of 2. Let  $f : [0, 1] \rightarrow [0, 1]$  be the transformation defined in Section 1.3.1 and let  $E = [0, 1/10)$ . That is,  $E$  is the set of all points whose decimal expansion starts with the digit 0. It is easy to check that if  $n = 2^k - 1$  with  $k$  even then

$$\frac{1}{n} \sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) = \frac{2^1 + 2^3 + \dots + 2^{k-1}}{2^k - 1} = \frac{2}{3}.$$

On the other hand, if one takes  $n = 2^k - 1$  with  $k$  odd then

$$\frac{1}{n} \sum_{j=0}^{n-1} \mathcal{X}_E(f^j(x)) = \frac{2^1 + 2^3 + \dots + 2^{k-2}}{2^k - 1} = \frac{2^k - 2}{3(2^k - 1)} \rightarrow \frac{1}{3}$$

as  $k \rightarrow \infty$ . Thus, the mean sojourn time of  $x$  in the set  $E$  does not exist.

### 3.2.2 Time averages

As we observed previously,

$$\tau(E, x) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)), \quad \text{where } \varphi = \mathcal{X}_E.$$

The next statement extends Theorem 3.2.1 to the case when  $\varphi$  is any integrable function:



**Theorem 3.2.3** (Birkhoff). *Let  $f : M \rightarrow M$  be a measurable transformation and  $\mu$  be a probability measure invariant under  $f$ . Given any integrable function  $\varphi : M \rightarrow \mathbb{R}$ , the limit*

$$\tilde{\varphi}(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) \tag{3.2.2}$$

*exists at  $\mu$ -almost every point  $x \in M$ . Moreover, the function  $\tilde{\varphi}$  defined in this way is integrable and satisfies*

$$\int \tilde{\varphi}(x) d\mu(x) = \int \varphi(x) d\mu(x).$$

In a little while, we will obtain this theorem as a special case of a more general result, the subadditive ergodic theorem. The limit  $\tilde{\varphi}$  is called the *time average*, or *orbital average*, of  $\varphi$ . The next proposition shows that time averages are constant on the orbit of  $\mu$ -almost every point, which generalizes (3.2.1):

**Proposition 3.2.4.** *Let  $\varphi : M \rightarrow \mathbb{R}$  be an integrable function. Then,*

$$\tilde{\varphi}(f(x)) = \tilde{\varphi}(x) \quad \text{for } \mu\text{-almost every point } x \in M. \tag{3.2.3}$$

*Proof.* By definition,

$$\begin{aligned} \tilde{\varphi}(f(x)) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \varphi(f^j(x)) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) + \frac{1}{n} [\varphi(f^n(x)) - \varphi(x)] \\ &= \tilde{\varphi}(x) + \lim_{n \rightarrow \infty} \frac{1}{n} [\varphi(f^n(x)) - \varphi(x)]. \end{aligned}$$

We need the following lemma:

**Lemma 3.2.5.** *If  $\phi$  is an integrable function then  $\lim_n (1/n)\phi(f^n(x)) = 0$  for  $\mu$ -almost every point  $x \in M$ .*

*Proof.* Fix any  $\varepsilon > 0$ . Since  $\mu$  is invariant, we have that

$$\begin{aligned} \mu(\{x \in M : |\phi(f^n(x))| \geq n\varepsilon\}) &= \mu(\{x \in M : |\phi(x)| \geq n\varepsilon\}) \\ &= \sum_{k=n}^{\infty} \mu(\{x \in M : k \leq \frac{|\phi(x)|}{\varepsilon} < k+1\}). \end{aligned}$$

Adding these expressions over  $n \in \mathbb{N}$ , we obtain

$$\begin{aligned} \sum_{n=1}^{\infty} \mu(\{x \in M : |\phi(f^n(x))| \geq n\varepsilon\}) &= \sum_{k=1}^{\infty} k\mu(\{x \in M : k \leq \frac{|\phi(x)|}{\varepsilon} < k+1\}) \\ &\leq \int \frac{|\phi|}{\varepsilon} d\mu. \end{aligned}$$

Since  $\phi$  is integrable, by assumption, all these expressions are finite. That implies that the set  $B(\varepsilon)$  of all points  $x$  such that  $|\phi(f^n(x))| \geq n\varepsilon$  for infinitely many values of  $n$  has zero measure (check Exercise A.1.6). Now, the definition of  $B(\varepsilon)$  implies that for every  $x \notin B(\varepsilon)$  there exists  $p \geq 1$  such that  $|\phi(f^n(x))| < n\varepsilon$  for every  $n \geq p$ . Consider the set  $B = \bigcup_{i=1}^{\infty} B(1/i)$ . Then  $B$  has zero measure and  $\lim_n (1/n)\phi(f^n(x)) = 0$  for every  $x \notin B$ .

Applying Lemma 3.2.5 to the function  $\phi = \varphi$  we obtain the identity in (3.2.3). This completes the proof of Proposition 3.2.4.

In general, the total measure subset of points for which the limit in (3.2.2) exists depends on the function  $\varphi$  under consideration. However, in some situations it is possible to choose such a set independent of the function. A useful example of such a situation is:

**Theorem 3.2.6.** *Let  $M$  be a compact metric space and  $f : M \rightarrow M$  be a measurable map. Then there exists some measurable set  $G \subset M$  with  $\mu(G) = 1$  such that*

$$\frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) \rightarrow \tilde{\varphi}(x) \tag{3.2.4}$$

for every  $x \in G$  and every continuous function  $\varphi : M \rightarrow \mathbb{R}$ .

*Proof.* By the Birkhoff ergodic theorem, for every continuous function  $\varphi$  there exists  $G(\varphi) \subset M$  such that  $\mu(G(\varphi)) = 1$  and (3.2.4) holds for every  $x \in G(\varphi)$ . By Theorem A.3.13, the space  $C^0(M)$  of continuous functions admits some countable dense subset  $\{\varphi_k : k \in \mathbb{N}\}$ . Take

$$G = \bigcap_{k=1}^{\infty} G(\varphi_k).$$

Since the intersection is countable, it is clear that  $\mu(G) = 1$ . So, it suffices to prove that (3.2.4) holds for every continuous function  $\varphi$  whenever  $x \in G$ . This can be done as follows. Given  $\varphi \in C^0(M)$  and any  $\varepsilon > 0$ , take  $k \in \mathbb{N}$  such that

$$\|\varphi - \varphi_k\| = \sup \{ |\varphi(x) - \varphi_k(x)| : x \in M \} \leq \varepsilon.$$

Then, given any point  $x \in G$ ,

$$\begin{aligned} \limsup_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) &\leq \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi_k(f^j(x)) + \varepsilon = \tilde{\varphi}_k(x) + \varepsilon \\ \liminf_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) &\geq \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi_k(f^j(x)) - \varepsilon = \tilde{\varphi}_k(x) - \varepsilon. \end{aligned}$$

This implies that

$$\limsup_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) - \liminf_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) \leq 2\varepsilon.$$

Since  $\varepsilon$  is arbitrary, it follows that the limit  $\tilde{\varphi}(x)$  exists, as stated.

In general, one can not say anything about the speed of convergence in Theorem 3.2.3. For example, it follows from a theorem of Kakutani and Petersen (check pages 94 to 99 of Petersen [Pet83]) that if the measure  $\mu$  is ergodic<sup>2</sup> and non-atomic then, given any sequence  $(a_n)_n$  of positive real numbers with  $\lim_n a_n = 0$ , there exists some bounded measurable function  $\varphi$  with

$$\limsup_n \frac{1}{a_n} \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) - \int \varphi d\mu \right| = +\infty.$$

Another interesting observation is that there is no analogue of the Birkhoff ergodic theorem for *infinite* invariant measures. Indeed, suppose that  $\mu$  is a  $\sigma$ -finite, but infinite, invariant measure of a transformation  $f : M \rightarrow M$ . We say that a measurable set  $W \subset M$  is *wandering* if the pre-images  $f^{-i}(W)$ ,  $i \geq 0$  are pairwise disjoint. Suppose that  $\mu$  is ergodic and *conservative*, that is, such that every wandering set has zero measure. Then, given any sequence  $(a_n)_n$  of positive real numbers,

- 1. either, for every  $\varphi \in L^1(\mu)$ ,

$$\liminf_n \frac{1}{a_n} \sum_{j=0}^{n-1} \varphi \circ f^j = 0 \quad \text{at almost every point;}$$

- 2. or, there exists  $(n_k)_k \rightarrow \infty$  such that, for every  $\varphi \in L^1(\mu)$ ,

$$\lim_k \frac{1}{a_{n_k}} \sum_{j=0}^{n_k-1} \varphi \circ f^j = \infty \quad \text{at almost every point.}$$

This and other related facts about infinite measures are proved in Section 2.4 of Aaronson [Aar97].

### 3.2.3 Theorem of von Neumann and consequences

The theorem of von Neumann (Theorem 3.1.7) may also be deduced directly from the theorem of Birkhoff, as we are going to explain.

<sup>2</sup> We say that an invariant measure  $\mu$  is *ergodic* if  $f^{-1}(A) = A$  up to measure zero implies that either  $\mu(A) = 0$  or  $\mu(A^c) = 0$ . The study of ergodic measures will be the subject of the next chapter.

Consider any  $\varphi \in L^2(\mu)$  and let  $\tilde{\varphi}$  be the corresponding time average. We start by showing that  $\tilde{\varphi} \in L^2(\mu)$  and its norm satisfies  $\|\tilde{\varphi}\|_2 \leq \|\varphi\|_2$ . Indeed,

$$|\tilde{\varphi}| \leq \lim_n \frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \quad \text{and, hence,} \quad |\tilde{\varphi}|^2 \leq \lim_n \left( \frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \right)^2.$$

Then, by the Fatou lemma (Theorem A.2.10),

$$\left[ \int |\tilde{\varphi}|^2 d\mu \right]^{1/2} \leq \liminf_n \left[ \int \left( \frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \right)^2 d\mu \right]^{1/2}. \tag{3.2.5}$$

We can use the Minkowski inequality (Theorem A.5.3) to bound the sequence on the right-hand side from above:

$$\left[ \int \left( \frac{1}{n} \sum_{j=0}^{n-1} |\varphi \circ f^j| \right)^2 d\mu \right]^{1/2} \leq \frac{1}{n} \sum_{j=0}^{n-1} \left[ \int |\varphi \circ f^j|^2 d\mu \right]^{1/2}. \tag{3.2.6}$$

Since  $\mu$  is invariant under  $f$ , the expression on the right-hand side is equal to  $\left[ \int |\varphi|^2 d\mu \right]^{1/2}$ . So, (3.2.5) and (3.2.6) imply that  $\|\tilde{\varphi}\|_2 \leq \|\varphi\|_2 < \infty$ .

Now let us show that  $(1/n) \sum_{j=0}^{n-1} \varphi \circ f^j$  converges to  $\tilde{\varphi}$  in  $L^2(\mu)$ . Initially, suppose that the function  $\varphi$  is bounded, that is, there exists  $C > 0$  such that  $|\varphi| \leq C$ . Then,

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j \right| \leq C \quad \text{for every } n \quad \text{and} \quad |\tilde{\varphi}| \leq C.$$

Then we may use the dominated convergence theorem (Theorem A.2.11) to conclude that

$$\lim_n \int \left( \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j - \tilde{\varphi} \right)^2 d\mu = \int \left( \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j - \tilde{\varphi} \right)^2 d\mu = 0.$$

In other words,  $(1/n) \sum_{j=0}^{n-1} \varphi \circ f^j$  converges to  $\tilde{\varphi}$  in the space  $L^2(\mu)$ . We are left to extend this conclusion to arbitrary functions  $\varphi$  in  $L^2(\mu)$ . For that, let us consider some sequence  $(\varphi_k)$  of bounded functions such that  $(\varphi_k)_k$  converges to  $\varphi$ . For example:

$$\varphi_k(x) = \begin{cases} \varphi(x) & \text{if } |\varphi(x)| \leq k \\ 0 & \text{otherwise.} \end{cases}$$

Denote by  $\tilde{\varphi}_k$  the corresponding time averages. Given any  $\varepsilon > 0$ , let  $k_0$  be fixed such that  $\|\varphi - \varphi_k\|_2 < \varepsilon/3$  for every  $k \geq k_0$ . Note that  $\|(\varphi - \varphi_k) \circ f^j\|_2$  is equal to  $\|\varphi - \varphi_k\|_2$  for every  $j \geq 0$ , because the measure  $\mu$  is invariant. Thus,

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} (\varphi - \varphi_k) \circ f^j \right\|_2 \leq \|\varphi - \varphi_k\|_2 < \varepsilon/3 \quad \text{for every } n \geq 1 \text{ and } k \geq k_0. \tag{3.2.7}$$

Observe also that  $\tilde{\varphi} - \tilde{\varphi}_k$  is the time average of the function  $\varphi - \varphi_k$ . So, the argument in the previous paragraph gives that

$$\|\tilde{\varphi} - \tilde{\varphi}_k\|_2 \leq \|\varphi - \varphi_k\|_2 < \varepsilon/3 \quad \text{for every } k \geq k_0. \tag{3.2.8}$$

By assumption, for every  $k \geq 1$  there exists  $n_0(k) \geq 1$  such that

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} \varphi_k \circ f^j - \tilde{\varphi}_k \right\|_2 < \varepsilon/3 \quad \text{for every } n \geq n_0(k). \tag{3.2.9}$$

Adding (3.2.7), (3.2.8), (3.2.9) we get that

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j - \tilde{\varphi} \right\|_2 < \varepsilon \quad \text{for every } n \geq n_0(k_0).$$

This completes the proof of the theorem of von Neumann from the theorem of Birkhoff.

Exercise 3.2.5 contains an extension of these conclusions to any  $L^p(\mu)$  space.

**Corollary 3.2.7.** *The time average  $\tilde{\varphi}$  of any function  $\varphi \in L^2(\mu)$  coincides with the orthogonal projection  $P(\varphi)$  of  $\varphi$  to the subspace of invariant functions.*

*Proof.* On the one hand, Theorem 3.1.7 gives that  $(1/n) \sum_{j=0}^{n-1} \varphi \circ f^j$  converges to  $P(\varphi)$  in  $L^2(\mu)$ . On the other hand, we have just shown that this sequence converges to  $\tilde{\varphi}$  in the space  $L^2(\mu)$ . So, by uniqueness of the limit,  $P(\varphi) = \tilde{\varphi}$ .

**Corollary 3.2.8.** *If  $f : M \rightarrow M$  is invertible then the time averages of any function  $\varphi \in L^2(\mu)$  relative to  $f$  and to  $f^{-1}$  coincide at  $\mu$ -almost every point:*

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^{-j} = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j \quad \text{at } \mu\text{-almost every point.} \tag{3.2.10}$$

*Proof.* The limit on the left-hand side of (3.2.10) is the orthogonal projection of  $\varphi$  to the subspace of functions invariant under  $f^{-1}$ , whereas the limit on the right-hand side is the orthogonal projection of  $\varphi$  to the subspace of functions invariant under  $f$ . It is clear that these two subspaces are exactly the same. Thus, the two limits coincide in  $L^2(\mu)$ .

### 3.2.4 Exercises

3.2.1. Let  $X = \{x_1, \dots, x_r\}$  be a finite set and  $\sigma : X \rightarrow X$  be a permutation. We call  $\sigma$  a *cyclic permutation* if it admits a unique orbit (containing all  $r$  elements of  $X$ ).

1. Prove that, for any cyclic permutation  $\sigma$  and any function  $\varphi : X \rightarrow \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(\sigma^i(x)) = \frac{\varphi(x_1) + \dots + \varphi(x_r)}{r}.$$

2. More generally, prove that for any permutation  $\sigma$  and any function  $\varphi$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(\sigma^i(x)) = \frac{\varphi(x) + \varphi(\sigma(x)) + \dots + \varphi(\sigma^{p-1}(x))}{p},$$

where  $p \geq 1$  is the cardinality of the orbit of  $x$ .

3.2.2. Check that Lemma 3.2.5 can also be deduced from the Birkhoff ergodic theorem and then we may even weaken the hypothesis: it suffices to suppose that  $\phi$  is measurable and  $\psi = \phi \circ f - \phi$  is integrable.

3.2.3. A function  $\varphi : \mathbb{Z} \rightarrow \mathbb{R}$  is said to be *uniformly quasi-periodic* if for every  $\varepsilon > 0$  there exists  $L(\varepsilon) \in \mathbb{N}$  such that every interval  $\{n + 1, \dots, n + L(\varepsilon)\}$  in the set of integers contains some  $\tau$  such that  $|\varphi(k + \tau) - \varphi(k)| < \varepsilon$  for every  $k \in \mathbb{Z}$ . Any such  $\tau$  is called an  $\varepsilon$ -*quasi-period* of  $f$ .

(a) Prove that if  $\varphi$  is uniformly quasi-periodic then  $\varphi$  is bounded.

(b) Show that for every  $\varepsilon > 0$  there exists  $\rho \geq 1$  such that

$$\left| \frac{1}{\rho} \sum_{j=n\rho+1}^{(n+1)\rho} \varphi(j) - \frac{1}{\rho} \sum_{j=1}^{\rho} \varphi(j) \right| < 2\varepsilon \quad \text{for every } n \geq 1.$$

(c) Show that the sequence  $(1/n) \sum_{j=1}^n \varphi(j)$  converges to some real number when  $n \rightarrow \infty$ .

(d) More generally, prove that  $\lim_n (1/n) \sum_{k=1}^n \varphi(x + k)$  exists for every  $x \in \mathbb{Z}$  and is independent of  $x$ .

3.2.4. Prove that for Lebesgue-almost every  $x \in [0, 1]$ , the geometric mean of the integer numbers  $a_1, \dots, a_n, \dots$  in the continued fraction expansion of  $x$  converges to some real number: in other words, there exists  $b \in \mathbb{R}$  such that  $\lim_n (a_1 a_2 \dots a_n)^{1/n} = b$ . [Observation: Compare with Exercise 4.2.12.]

3.2.5. Let  $\varphi : M \rightarrow \mathbb{R}$  be an integrable function and  $\tilde{\varphi}$  be the corresponding time average, given by Theorem 3.2.3. Show that if  $\varphi \in L^p(\mu)$  for some  $p > 1$  then  $\tilde{\varphi} \in L^p(\mu)$  and  $\|\tilde{\varphi}\|_p \leq \|\varphi\|_p$ . Moreover,

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

converges to  $\tilde{\varphi}$  in the space  $L^p(\mu)$ .

3.2.6. Prove the Birkhoff ergodic theorem for flows: if  $\mu$  is a probability measure invariant under a flow  $f$  and  $\varphi \in L^1(\mu)$  then the function

$$\tilde{\varphi}(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \varphi(f^t(x)) dt$$

is defined at  $\mu$ -almost every point and  $\int \tilde{\varphi} d\mu = \int \varphi d\mu$ .

3.2.7. Prove that if a continuous transformation  $f : M \rightarrow M$  of a compact metric space  $M$  admits exactly one invariant probability measure  $\mu$ , and this measure is such that  $\mu(A) > 0$  for every non-empty open set  $A \subset M$ , then every orbit of  $f$  is dense in  $M$ .

### 3.3 Subadditive ergodic theorem

A sequence of functions  $\varphi_n : M \rightarrow \mathbb{R}$  is said to be *subadditive* for a transformation  $f : M \rightarrow M$  if

$$\varphi_{m+n} \leq \varphi_m + \varphi_n \circ f^m \quad \text{for every } m, n \geq 1. \tag{3.3.1}$$

**Example 3.3.1.** A sequence  $\varphi_n : M \rightarrow \mathbb{R}$  is *additive* for the transformation  $f$  if  $\varphi_{m+n} = \varphi_m + \varphi_n \circ f^m$  for every  $m, n \geq 1$ . For example, the time sums

$$\varphi_n(x) = \sum_{j=0}^{n-1} \varphi(f^j(x))$$

of any function  $\varphi : M \rightarrow \mathbb{R}$  form an additive sequence. In fact, every additive sequence is of this form, with  $\varphi = \varphi_1$ . Of course, additive sequences are also subadditive.

For the next example we need the notion of the *norm* of a square matrix  $A$  of dimension  $d$ , which is defined as follows:

$$\|A\| = \sup \left\{ \frac{\|Av\|}{\|v\|} : v \in \mathbb{R}^d \setminus \{0\} \right\}. \tag{3.3.2}$$

Compare with (2.3.1). It follows directly from the definition that the norm of the product of two matrices is less than or equal to the product of the norms of those matrices:

$$\|AB\| \leq \|A\| \|B\|. \tag{3.3.3}$$

**Example 3.3.2.** Let  $A : M \rightarrow GL(d)$  be a measurable function with values in the *linear group*, that is, the set  $GL(d)$  of invertible square matrices of dimension  $d$ . Define

$$\phi^n(x) = A(f^{n-1}(x)) \cdots A(f(x))A(x)$$

for every  $n \geq 1$  and  $x \in M$ . Then the sequence  $\varphi_n(x) = \log \|\phi^n(x)\|$  is subadditive. Indeed,

$$\phi^{m+n}(x) = \phi^n(f^m(x))\phi^m(x).$$

and so, using (3.3.3),

$$\begin{aligned} \varphi_{m+n}(x) &= \log \|\phi^n(f^m(x))\phi^m(x)\| \\ &\leq \log \|\phi^m(x)\| + \log \|\phi^n(f^m(x))\| = \varphi_m(x) + \varphi_n(f^m(x)) \end{aligned}$$

for every  $m, n$  and  $x$ .

Recall that, given any function  $\varphi : M \rightarrow \mathbb{R}$ , we denote by  $\varphi^+ : M \rightarrow \mathbb{R}$  its positive part, which is defined by  $\varphi^+(x) = \max\{\varphi(x), 0\}$ .

**Theorem 3.3.3 (Kingman).** *Let  $\mu$  be a probability measure invariant under a transformation  $f : M \rightarrow M$  and let  $\varphi_n : M \rightarrow \mathbb{R}$ ,  $n \geq 1$  be a subadditive sequence of measurable functions such that  $\varphi_1^+ \in L^1(\mu)$ . Then  $(\varphi_n/n)_n$  converges at*

$\mu$ -almost every point to some function  $\varphi : M \rightarrow [-\infty, +\infty)$  that is invariant under  $f$ . Moreover,  $\varphi^+ \in L^1(\mu)$  and

$$\int \varphi d\mu = \lim_n \frac{1}{n} \int \varphi_n d\mu = \inf_n \frac{1}{n} \int \varphi_n d\mu \in [-\infty, +\infty).$$

The proof of Theorem 3.3.3 that we are going to present is due to Avila and Bochi [AB], who started from a proof of the Birkhoff ergodic theorem (Theorem 3.2.3) by Katznelson and Weiss [KW82]. An important observation is that Theorem 3.2.3 is *not* used in the arguments. This allows us to obtain the theorem of Birkhoff as a particular case of Theorem 3.3.3.

### 3.3.1 Preparing the proof

A sequence  $(a_n)_n$  in  $[-\infty, +\infty)$  is said to be *subadditive* if  $a_{m+n} \leq a_m + a_n$  for every  $m, n \geq 1$ .

**Lemma 3.3.4.** *If  $(a_n)_n$  is a subadditive sequence then*

$$\lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n} \in [-\infty, \infty). \tag{3.3.4}$$

*Proof.* If  $a_m = -\infty$  for some  $m$  then, by subadditivity,  $a_n = -\infty$  for every  $n > m$ . In that case, both sides of (3.3.4) are equal to  $-\infty$  and so the lemma holds. From now on let us assume that  $a_n \in \mathbb{R}$  for every  $n$ .

Let  $L = \inf_n (a_n/n) \in [-\infty, +\infty)$  and  $B$  be any real number larger than  $L$ . Then we may find  $k \geq 1$  such that

$$\frac{a_k}{k} < B.$$

For  $n > k$ , we may write  $n = kp + q$ , where  $p$  and  $q$  are integers such that  $p \geq 1$  and  $1 \leq q \leq k$ . Then, by subadditivity,

$$a_n \leq a_{kp} + a_q \leq pa_k + a_q \leq pa_k + \alpha,$$

where  $\alpha = \max\{a_i : 1 \leq i \leq k\}$ . Hence,

$$\frac{a_n}{n} \leq \frac{pk}{n} \frac{a_k}{k} + \frac{\alpha}{n}.$$

Observe that  $pk/n$  converges to 1 and  $\alpha/n$  converges to zero when  $n \rightarrow \infty$ . So, since  $a_k/k < B$ , we have that

$$L \leq \frac{a_n}{n} < B$$

for every  $n$  sufficiently large. Making  $B \rightarrow L$ , we conclude that

$$\lim_n \frac{a_n}{n} = L = \inf_n \frac{a_n}{n}.$$

This completes the argument.



Now let  $(\varphi_n)_n$  be as in Theorem 3.3.3. By subadditivity,

$$\varphi_n \leq \varphi_1 + \varphi_1 \circ f + \dots + \varphi_1 \circ f^{n-1}.$$

This relation remains valid if we replace  $\varphi_n$  and  $\varphi_1$  by their positive parts  $\varphi_n^+$  and  $\varphi_1^+$ . Hence, the hypothesis that  $\varphi_1^+ \in L^1(\mu)$  implies that  $\varphi_n^+ \in L^1(\mu)$  for every  $n$ . Moreover, the hypothesis that  $(\varphi_n)_n$  is subadditive implies that

$$a_n = \int \varphi_n d\mu, \quad n \geq 1$$

is a subadditive sequence in  $[-\infty, +\infty)$ . Therefore, by Lemma 3.3.4, the limit

$$L = \lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n} \in [-\infty, \infty)$$

exists. Define  $\varphi_- : M \rightarrow [-\infty, \infty]$  and  $\varphi_+ : M \rightarrow [-\infty, \infty]$  through

$$\varphi_-(x) = \liminf_n \frac{\varphi_n}{n}(x) \quad \text{and} \quad \varphi_+(x) = \limsup_n \frac{\varphi_n}{n}(x).$$

Clearly,  $\varphi_-(x) \leq \varphi_+(x)$  for every  $x \in M$ . We are going to prove that

$$\int \varphi_- d\mu \geq L \geq \int \varphi_+ d\mu, \tag{3.3.5}$$

as long as each function  $\varphi_n$  is bounded from below. Consequently, the two functions  $\varphi_-$  and  $\varphi_+$  coincide at  $\mu$ -almost every point and their integral is equal to  $L$ . Thus, the theorem will be proven in this case, with  $\varphi = \varphi_- = \varphi_+$  (the fact that  $\varphi$  is invariant under  $f$  is part of Exercise 3.3.2). At the end, we remove that boundedness assumption using a truncation trick.

### 3.3.2 Key lemma

In this section we assume that  $\varphi_- > -\infty$  at every point. Fix  $\varepsilon > 0$  and define, for each  $k \in \mathbb{N}$ ,

$$E_k = \{x \in M : \varphi_j(x) \leq j(\varphi_-(x) + \varepsilon) \text{ for some } j \in \{1, \dots, k\}\}.$$

It is clear that  $E_k \subset E_{k+1}$  for every  $k$ . Moreover, the definition of  $\varphi_-(x)$  implies that  $M = \bigcup_k E_k$ . Define also

$$\psi_k(x) = \begin{cases} \varphi_-(x) + \varepsilon & \text{if } x \in E_k \\ \varphi_1(x) & \text{if } x \in E_k^c. \end{cases}$$

It follows from the definition of  $E_k$  that  $\varphi_1(x) > \varphi_-(x) + \varepsilon$  for every  $x \in E_k^c$ . Combining this fact with the previous observations, we see that the sequence  $(\psi_k(x))_k$  is non-increasing and converges to  $\varphi_-(x) + \varepsilon$ , for every  $x \in M$ . In particular, by the monotone convergence theorem (Theorem A.2.9),

$$\int \psi_k d\mu \rightarrow \int (\varphi_- + \varepsilon) d\mu \quad \text{as } k \rightarrow \infty.$$

The crucial step in the proof of the theorem is the following estimate:

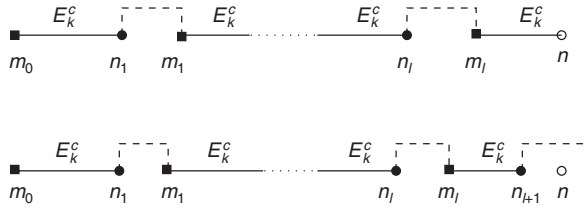


Figure 3.1. Decomposition of the trajectory of a point

**Lemma 3.3.5.** For every  $n > k \geq 1$  and  $\mu$ -almost every  $x \in M$ ,

$$\varphi_n(x) \leq \sum_{i=0}^{n-k-1} \psi_k(f^i(x)) + \sum_{i=n-k}^{n-1} \max\{\psi_k, \varphi_1\}(f^i(x)).$$

*Proof.* Take  $x \in M$  such that  $\varphi_-(x) = \varphi_-(f^j(x))$  for every  $j \geq 1$  (this holds at  $\mu$ -almost every point, according to Exercise 3.3.2). Consider the sequence, possibly finite, of integer numbers

$$m_0 \leq n_1 < m_1 \leq n_2 < m_2 \leq \dots \tag{3.3.6}$$

defined inductively as follows (see also Figure 3.1).

Define  $m_0 = 0$ . Let  $n_j$  be the smallest integer greater than or equal to  $m_{j-1}$  satisfying  $f^{n_j}(x) \in E_k$  (if it exists). Then, by the definition of  $E_k$ , there exists  $m_j$  such that  $1 \leq m_j - n_j \leq k$  and

$$\varphi_{m_j - n_j}(f^{n_j}(x)) \leq (m_j - n_j)(\varphi_-(f^{n_j}(x)) + \varepsilon). \tag{3.3.7}$$

This completes the definition of the sequence (3.3.6). Now, given  $n \geq k$ , let  $l \geq 0$  be the largest integer such that  $m_l \leq n$ . By subadditivity,

$$\varphi_{n_j - m_{j-1}}(f^{m_{j-1}}(x)) \leq \sum_{i=m_{j-1}}^{n_j-1} \varphi_1(f^i(x))$$

for every  $j = 1, \dots, l$  such that  $m_{j-1} \neq n_j$ , and analogously for  $\varphi_{n-m_l}(f^{m_l}(x))$ . Thus,

$$\varphi_n(x) \leq \sum_{i \in I} \varphi_1(f^i(x)) + \sum_{j=1}^l \varphi_{m_j - n_j}(f^{n_j}(x)) \tag{3.3.8}$$

where  $I = \bigcup_{j=1}^l [m_{j-1}, n_j) \cup [m_l, n)$ . Observe that

$$\varphi_1(f^i(x)) = \psi_k(f^i(x)) \quad \text{for every } i \in \bigcup_{j=1}^l [m_{j-1}, n_j) \cup [m_l, \min\{n_{l+1}, n\}),$$

since  $f^i(x) \in E_k^c$  in all these cases. Moreover, since  $\varphi_-$  is constant on orbits (see Exercise 3.3.2) and  $\psi_k \geq \varphi_- + \varepsilon$ , the relation (3.3.7) gives that

$$\varphi_{m_j - n_j}(f^{n_j}(x)) \leq \sum_{i=n_j}^{m_j-1} (\varphi_-(f^i(x)) + \varepsilon) \leq \sum_{i=n_j}^{m_j-1} \psi_k(f^i(x))$$

for every  $j = 1, \dots, l$ . In this way, using (3.3.8), we conclude that

$$\varphi_n(x) \leq \sum_{i=0}^{\min\{n_{l+1}, n\}-1} \psi_k(f^i(x)) + \sum_{i=n_{l+1}}^{n-1} \varphi_1(f^i(x)).$$

Since  $n_{l+1} > n - k$ , the lemma is proven.

### 3.3.3 Estimating $\varphi_-$

Towards establishing (3.3.5), in this section we prove the following lemma:

**Lemma 3.3.6.**  $\int \varphi_- d\mu = L$

*Proof.* Suppose for a while that  $\varphi_n/n$  is uniformly bounded from below, that is, that there exists  $\kappa > 0$  such that  $\varphi_n/n \geq -\kappa$  for every  $n$ . Applying the lemma of Fatou (Theorem A.2.10) to the sequence of non-negative functions  $\varphi_n/n + \kappa$ , we get that  $\varphi_-$  is integrable and

$$\int \varphi_- d\mu \leq \liminf_n \int \frac{\varphi_n}{n} d\mu = L.$$

To prove the opposite inequality, observe that Lemma 3.3.5 implies

$$\frac{1}{n} \int \varphi_n d\mu \leq \frac{n-k}{n} \int \psi_k d\mu + \frac{k}{n} \int \max\{\psi_k, \varphi_1\} d\mu. \tag{3.3.9}$$

Note that  $\max\{\psi_k, \varphi_1\} \leq \max\{\varphi_- + \varepsilon, \varphi_1^+\}$ , and this last function is integrable. So, the limit superior of the last term in (3.3.9) as  $n \rightarrow \infty$  is less than or equal to zero. So, making  $n \rightarrow \infty$  we get that  $L \leq \int \psi_k d\mu$  for every  $k$ . Then, making  $k \rightarrow \infty$ , we conclude that

$$L \leq \int \varphi_- d\mu + \varepsilon.$$

Finally, making  $\varepsilon \rightarrow 0$  we get that  $L \leq \int \varphi_- d\mu$ . This proves the lemma when  $\varphi_n/n$  is uniformly bounded from below.

We are left to remove this hypothesis. Define, for each  $\kappa > 0$ ,

$$\varphi_n^\kappa = \max\{\varphi_n, -\kappa n\} \quad \text{and} \quad \varphi_-^\kappa = \max\{\varphi_-, -\kappa\}.$$

The sequence  $(\varphi_n^\kappa)_n$  satisfies all the conditions of Theorem 3.3.3: indeed, it is subadditive and the positive part of  $\varphi_1^\kappa$  is integrable. Moreover, it is clear that  $\varphi_-^\kappa = \liminf_n (\varphi_n^\kappa/n)$ . So, the argument in the previous paragraph shows that

$$\int \varphi_-^\kappa d\mu = \inf_n \frac{1}{n} \int \varphi_n^\kappa d\mu. \tag{3.3.10}$$

By the monotone convergence theorem (Theorem A.2.9), we also have that

$$\int \varphi_n d\mu = \inf_\kappa \int \varphi_n^\kappa d\mu \quad \text{and} \quad \int \varphi_- d\mu = \inf_\kappa \int \varphi_-^\kappa d\mu. \tag{3.3.11}$$

Combining the relations (3.3.10) and (3.3.11), we get that

$$\int \varphi_- d\mu = \inf_{\kappa} \int \varphi_{-}^{\kappa} = \inf_{\kappa} \inf_n \frac{1}{n} \int \varphi_n^{\kappa} d\mu = \inf_n \frac{1}{n} \int \varphi_n d\mu = L.$$

This completes the proof of the lemma.

### 3.3.4 Bounding $\varphi_+$

To complete the proof of (3.3.5), we are now going to show that  $\int \varphi_+ d\mu \leq L$  as long as  $\inf_x \varphi_n(x)$  is finite for every  $n$ . Let us start by proving the following auxiliary result:

**Lemma 3.3.7.** *For any fixed  $k$ ,*

$$\limsup_n \frac{\varphi_{kn}}{n} = k \limsup_n \frac{\varphi_n}{n}.$$

*Proof.* The inequality  $\leq$  is clear, since  $\varphi_{kn}/kn$  is a subsequence of  $\varphi_n/n$ . To prove the opposite inequality, let us write  $n = kq_n + r_n$  with  $r_n \in \{1, \dots, k\}$ . By subadditivity,

$$\varphi_n \leq \varphi_{kq_n} + \varphi_{r_n} \circ f^{kq_n} \leq \varphi_{kq_n} + \psi \circ f^{kq_n},$$

where  $\psi = \max\{\varphi_1^+, \dots, \varphi_k^+\}$ . Observe that  $n/q_n \rightarrow k$  when  $n \rightarrow \infty$ . Moreover, as  $\psi \in L^1(\mu)$ , we may use Lemma 3.2.5 to see that  $\psi \circ f^{kq_n}/n$  converges to zero at  $\mu$ -almost every point. Hence, dividing all the terms in the previous relation by  $n$  and taking the limit superior as  $n \rightarrow \infty$ , we get that

$$\limsup_n \frac{1}{n} \varphi_n \leq \limsup_n \frac{1}{n} \varphi_{kq_n} + \limsup_n \frac{1}{n} \psi \circ f^{kq_n} = \frac{1}{k} \limsup_q \frac{1}{q} \varphi_{kq},$$

as stated in the lemma.

**Lemma 3.3.8.** *Suppose that  $\inf_x \varphi_n(x) > -\infty$  for every  $n$ . Then  $\int \varphi_+ d\mu \leq L$ .*

*Proof.* For each  $k$  and  $n \geq 1$ , consider  $\theta_n = -\sum_{j=0}^{n-1} \varphi_k \circ f^{jk}$ . Observe that

$$\int \theta_n d\mu = -n \int \varphi_k d\mu \quad \text{for every } n, \tag{3.3.12}$$

since  $f^k$  preserves the measure  $\mu$ . Since the sequence  $(\varphi_n)_n$  is subadditive,  $\theta_n \leq -\varphi_{kn}$  for every  $n$ . Hence, using Lemma 3.3.7,

$$\theta_- = \liminf_n \frac{\theta_n}{n} \leq -\limsup_n \frac{\varphi_{kn}}{n} = -k \limsup_n \frac{\varphi_n}{n} = -k\varphi_+,$$

and so

$$\int \theta_- d\mu \leq -k \int \varphi_+ d\mu. \tag{3.3.13}$$

Observe also that the sequence  $(\theta_n)_n$  is additive:  $\theta_{m+n} = \theta_m + \theta_n \circ f^{km}$  for every  $m, n \geq 1$ . Since  $\theta_1 = -\varphi_k$  is bounded from above by  $-\inf \varphi_k$ , we also have that

the function  $\theta_1^+$  is bounded and, consequently, integrable. Thus, we may apply Lemma 3.3.6, together with the equality (3.3.12), to conclude that

$$\int \theta_- d\mu = \inf_n \int \frac{\theta_n}{n} d\mu = - \int \varphi_k d\mu. \tag{3.3.14}$$

Putting (3.3.13) and (3.3.14) together we get that

$$\int \varphi_+ d\mu \leq \frac{1}{k} \int \varphi_k d\mu.$$

Finally, taking the infimum over  $k$  we get that  $\int \varphi_+ d\mu \leq L$ .

Lemmas 3.3.6 and 3.3.8 imply the relation (3.3.5) and, thus, Theorem 3.3.3 is proven when  $\inf \varphi_k > -\infty$  for every  $k$ . In the general case, consider

$$\varphi_n^\kappa = \max\{\varphi_n, -\kappa n\} \quad \text{and} \quad \varphi_-^\kappa = \max\{\varphi_-, -\kappa\} \quad \text{and} \quad \varphi_+^\kappa = \max\{\varphi_+, -\kappa\}$$

for every constant  $\kappa > 0$ . The previous arguments may be applied to the sequence  $(\varphi_n^\kappa)_n$  for each fixed  $\kappa > 0$ . Hence,  $\varphi_+^\kappa = \varphi_-^\kappa$  at  $\mu$ -almost every point for every  $\kappa > 0$ . Since  $\varphi_-^\kappa \rightarrow \varphi_-$  and  $\varphi_+^\kappa \rightarrow \varphi_+$  when  $\kappa \rightarrow \infty$ , it follows that  $\varphi_- = \varphi_+$  at  $\mu$ -almost every point. The proof of Theorem 3.3.3 is complete.

### 3.3.5 Lyapunov exponents

We have observed previously that every sequence of time sums

$$\varphi_n = \sum_{j=0}^{n-1} \varphi \circ f^j, \quad n \geq 1$$

is additive and, in particular, subadditive. Therefore, the ergodic theorem of Birkhoff (Theorem 3.2.3) is a particular case of Theorem 3.3.3. Another important consequence of the subadditive ergodic theorem is the theorem of Furstenberg–Kesten that we state next.

Let  $f : M \rightarrow M$  be a measurable transformation and  $\mu$  be an invariant probability measure. Consider any measurable function  $\theta : M \rightarrow \text{GL}(d)$  with values in the group  $\text{GL}(d)$ . The *cocycle* defined by  $\theta$  over  $f$  is the sequence of functions defined by

$$\phi^n(x) = \theta(f^{n-1}(x)) \cdots \theta(f(x))\theta(x) \text{ for } n \geq 1 \quad \text{and} \quad \phi^0(x) = \text{id}$$

for every  $x \in M$ . We leave it to the reader to check that

$$\phi^{m+n}(x) = \phi^n(f^m(x)) \cdot \phi^m(x) \quad \text{for every } m, n \in \mathbb{N} \text{ and } x \in M. \tag{3.3.15}$$

It is also easy to check that, conversely, any sequence  $(\phi^n)_n$  with this property is the cocycle defined by  $\theta = \phi^1$  over the transformation  $f$ .

**Theorem 3.3.9** (Furstenberg–Kesten). *If  $\log^+ \|\theta\| \in L^1(\mu)$  then*

$$\lambda_{\max}(x) = \lim_n \frac{1}{n} \log \|\phi^n(x)\|$$

exists at  $\mu$ -almost every point. Moreover,  $\lambda_{\max}^+ \in L^1(\mu)$  and

$$\int \lambda_{\max} d\mu = \lim_n \frac{1}{n} \int \log \|\phi^n\| d\mu = \inf_n \frac{1}{n} \int \log \|\phi^n\| d\mu.$$

If  $\log^+ \|\theta^{-1}\| \in L^1(\mu)$  then

$$\lambda_{\min}(x) = \lim_n -\frac{1}{n} \log \|\phi^n(x)^{-1}\|$$

exists at  $\mu$ -almost every point. Moreover,  $\lambda_{\min} \in L^1(\mu)$  and

$$\int \lambda_{\min} d\mu = \lim_n -\frac{1}{n} \int \log \|(\phi^n)^{-1}\| d\mu = \sup_n -\frac{1}{n} \int \log \|(\phi^n)^{-1}\| d\mu.$$

To deduce this result from Theorem 3.3.3 it suffices to note that the sequences

$$\varphi_n^{\max}(x) = \log \|\phi^n(x)\| \quad \text{and} \quad \varphi_n^{\min}(x) = \log \|\phi^n(x)^{-1}\|$$

are subadditive (recall Example 3.3.2).

The multiplicative ergodic theorem of Oseledets, which we are going to state in the following, provides a major refinement of the conclusion of the Furstenberg–Kesten theorem. It asserts that, under the same hypotheses as Theorem 3.3.9, for  $\mu$ -almost every  $x \in M$  there exist a positive integer  $k = k(x)$  and real numbers  $\lambda_1(x) > \dots > \lambda_k(x)$  and a *filtration* (that is, a decreasing sequence of vector subspaces)

$$\mathbb{R}^d = V_x^1 > \dots > V_x^k > V_x^{k+1} = \{0\} \tag{3.3.16}$$

such that, for every  $i \in \{1, \dots, k\}$  and  $\mu$ -almost every  $x \in M$ ,

- (a1)  $k(f(x)) = k(x)$  and  $\lambda_i(f(x)) = \lambda_i(x)$  and  $\theta(x) \cdot V_x^i = V_{f(x)}^i$ ;
- (b1)  $\lim_n \frac{1}{n} \log \|\phi^n(x)v\| = \lambda_i(x)$  for every  $v \in V_x^i \setminus V_x^{i+1}$ ;
- (c1)  $\lim_n \frac{1}{n} \log |\det \phi^n(x)| = \sum_{i=1}^k d_i(x) \lambda_i(x)$ , where  $d_i(x) = \dim V_x^i - \dim V_x^{i+1}$ .

Moreover, the numbers  $k(x)$  and  $\lambda_1(x), \dots, \lambda_k(x)$  and the subspaces  $V_x^1, \dots, V_x^k$  depend measurably on the point  $x$ .

The numbers  $\lambda_i(x)$  are called the *Lyapunov exponents* of  $\theta$  at the point  $x$ . They satisfy  $\lambda_1 = \lambda_{\max}$  and  $\lambda_k = \lambda_{\min}$ . For this reason, we also call  $\lambda_{\max}(x)$  and  $\lambda_{\min}(x)$  the *extremal Lyapunov exponents* of  $\theta$  at the point  $x$ . Each  $d_i(x)$  is called the *multiplicity* of the Lyapunov exponent  $\lambda_i(x)$ .

When  $f$  is invertible, we may extend the sequence  $\phi^n$  to the whole of  $\mathbb{Z}$ , through

$$\phi^{-n}(x) = \phi^n(f^{-n}(x))^{-1} \quad \text{for every } n \geq 1 \text{ and } x \in M.$$

Assuming also that  $\log^+ \|\theta^{-1}\| \in L^1(\mu)$ , one obtains a stronger conclusion than before: more than a filtration, there is a decomposition

$$\mathbb{R}^d = E_x^1 \oplus \dots \oplus E_x^k \tag{3.3.17}$$

such that, for every  $i = 1, \dots, k$ ,

(a2)  $\theta(x) \cdot E_x^i = E_{f(x)}^i$  and  $V_x^i = V_x^{i+1} \oplus E_x^i$ ; so,  $\dim E_x^i = \dim V_x^i - \dim V_x^{i+1}$ ;

(b2)  $\lim_{n \rightarrow \pm\infty} \frac{1}{n} \log \|\phi^n(x)v\| = \lambda_i(x)$  for every  $v \in E_x^i$  different from zero;

(c2)  $\lim_{n \rightarrow +\infty} \frac{1}{n} \log |\det \phi^n(x)| = \sum_{i=1}^k d_i(x) \lambda_i(x)$ , where  $d_i(x) = \dim E_x^i$ .

The reader will find a much more detailed discussion of these results, including proofs, in Chapter 4 of [Via14].

### 3.3.6 Exercises

3.3.1. Give a direct proof of the Birkhoff ergodic theorem (Theorem 3.2.3), using the approach in the proof of Theorem 3.3.3.

3.3.2. Given a subadditive sequence  $(\varphi_n)_n$  with  $\varphi_1^+ \in L^1(\mu)$ , show that the functions

$$\varphi_- = \liminf_n \frac{\varphi_n}{n} \quad \text{and} \quad \varphi_+ = \limsup_n \frac{\varphi_n}{n}$$

are  $f$ -invariant, that is, they satisfy  $\varphi_-(x) = \varphi_- \circ f(x)$  and  $\varphi_+(x) = \varphi_+ \circ f(x)$  for  $\mu$ -almost every  $x \in M$ .

3.3.3. State and prove the subadditive ergodic theorem for flows.

3.3.4. Let  $M$  be a compact manifold and  $f : M \rightarrow M$  be a diffeomorphism of class  $C^1$  that preserves the Lebesgue measure. Check that

$$\sum_{i=1}^{k(x)} d_i(x) \lambda_i(x) = 0 \quad \text{at } \mu\text{-almost every point } x \in M,$$

where  $\lambda_i(x)$ ,  $i = 1, \dots, k(x)$  are the Lyapunov exponents of  $Df$  at the point  $x$  and  $d_i(x)$ ,  $i = 1, \dots, k(x)$  are the corresponding multiplicities.

3.3.5. Let  $(\varphi_n)_n$  be a subadditive sequence of functions for some transformation  $f : M \rightarrow M$ . We call the *time constant* of  $(\varphi_n)_n$  the number

$$\lim_n \frac{1}{n} \int \varphi_n d\mu$$

when it exists. Assuming that the limit does exist and is finite, show that we may write  $\varphi_n = \psi_n + \gamma_n$  for each  $n$ , in such a way that  $(\psi_n)_n$  is an additive sequence and  $(\gamma_n)_n$  is a subadditive sequence with time constant equal to zero.

3.3.6. Under the assumptions of the Furstenberg–Kesten theorem, show that the sequence  $\psi_n = (1/n) \log \|\phi^n\|$  is *uniformly integrable*, in the following sense: for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that

$$\mu(E) < \delta \quad \Rightarrow \quad \int_E \psi_n^+ d\mu < \varepsilon \text{ for every } n.$$

3.3.7. Under the assumptions of the Furstenberg–Kesten theorem, let  $\Psi_k$  denote the time average of the function  $\psi_k = (1/k) \log \|\phi^k\|$  relative to the transformation  $f^k$ . Show that  $\lambda_{\max}(x) \leq \Psi_k(x)$  for every  $k$  and  $\mu$ -almost every  $x$ . Using Exercise 3.3.6, show that for every  $\rho > 0$  and  $\mu$ -almost every  $x$  there exists  $k$  such that  $\Psi_k(x) \leq \lambda_{\max}(x) + \rho$ .

### 3.4 Discrete time and continuous time

Most of the time we focus our presentation in the context of dynamical systems with discrete time. However, almost everything that was said so far extends, more or less straightforwardly, to systems with continuous time. One reason why the two theories are so similar is that one may relate systems of either kind to systems of the other kind, through certain constructions. That is the subject of the present section. For simplicity, we stick to the case of invertible systems. The general case may be handled using the notion of natural extension, which was described in Section 2.4.2.

#### 3.4.1 Suspension flows

Our first construction associates with every invertible map  $f : M \rightarrow M$  and every measurable function  $\tau : M \rightarrow (0, \infty)$  a flow  $g^t : N \rightarrow N, t \in \mathbb{R}$ , that we call the *suspension of  $f$  with return time  $\tau$* , whose recurrence properties are directly related to the recurrence properties of  $f$ . In particular, we associate a measure  $\nu$  invariant under this flow with every measure  $\mu$  invariant under  $f$ . For this construction we assume that the function  $\tau$  is such that

$$\sum_{j=1}^{\infty} \tau(f^j(x)) = \sum_{j=1}^{\infty} \tau(f^{-j}(x)) = +\infty \tag{3.4.1}$$

for every  $x \in M$ . That is the case, for instance, if  $\tau$  is bounded away from zero.

The first step is to construct the domain  $N$  of the suspension flow. Let us consider the transformation  $F : M \times \mathbb{R} \rightarrow M \times \mathbb{R}$  defined by

$$F(x, s) = (f(x), s - \tau(x)).$$

Note that  $F$  is invertible. Let  $\sim$  be the equivalence relation defined in  $M \times \mathbb{R}$  by

$$(x, s) \sim (\tilde{x}, \tilde{s}) \iff \text{there exists } n \in \mathbb{Z} \text{ such that } F^n(x, s) = (\tilde{x}, \tilde{s}).$$

We denote by  $N$  the set of equivalence classes and by  $\pi : M \times \mathbb{R} \rightarrow N$  the canonical projection associating with every  $(x, s) \in M \times \mathbb{R}$  the corresponding equivalence class.

Now consider the flow  $G^t : M \times \mathbb{R} \rightarrow M \times \mathbb{R}$  given by  $G^t(x, s) = (x, s + t)$ . It is clear that  $G^t \circ F = F \circ G^t$  for every  $t \in \mathbb{R}$ . This ensures that  $G^t, t \in \mathbb{R}$  induces a flow  $g^t, t \in \mathbb{R}$  in the quotient space  $N$ , given by

$$g^t(\pi(x, s)) = \pi(G^t(x, s)) \quad \text{for every } x \in M \text{ and } s, t \in \mathbb{R}. \tag{3.4.2}$$

Indeed, if  $\pi(x, s) = \pi(\tilde{x}, \tilde{s})$  then there exists  $n \in \mathbb{Z}$  such that  $F^n(x, s) = (\tilde{x}, \tilde{s})$ . Hence,

$$G^t(\tilde{x}, \tilde{s}) = G^t \circ F^n(x, s) = F^n \circ G^t(x, s)$$

and so  $\pi(G^t(x, s)) = \pi(G^t(\tilde{x}, \tilde{s}))$ . This shows that the flow  $g^t, t \in \mathbb{R}$  is well defined.



To better understand how this flow is related to the transformation  $f$ , we need to revisit the construction from a more concrete point of view. Let us consider  $D = \{(x, s) \in M \times \mathbb{R} : 0 \leq s < \tau(x)\}$ . We claim that  $D$  is a *fundamental domain* for the equivalence relation  $\sim$ , that is, it contains exactly one representative of each equivalence class. Uniqueness of the representative is immediate: just observe that if  $(x, s) \in D$  then  $F^n(x, s) = (x_n, s_n)$  with  $s_n < 0$  for every  $n > 0$  and  $s_n > \tau(f^n(x))$  for every  $n > 0$ . To prove existence, we need the condition (3.4.1): it ensures that the iterates  $(x_n, s_n) = F^n(x, s)$  of any  $(x, s)$  satisfy

$$\lim_{n \rightarrow +\infty} s_n = -\infty \quad \text{and} \quad \lim_{n \rightarrow -\infty} s_n = +\infty.$$

Then, taking  $n$  maximum such that  $s_n \geq 0$ , we find that  $(x_n, s_n) \in D$ . In this way, the claim is proved. Now observe that the claim means that the restriction of the projection  $\pi$  to domain  $D$  is a bijection over  $N$ . Thus, we may identify  $N$  with  $D$  and, in particular, we may consider  $g^t, t \in \mathbb{R}$  as a flow in  $D$ .

In just the same way, we may identify  $M$  with the subset  $\Sigma = \pi(M \times \{0\})$  of  $N$ . Observing that

$$g^{\tau(x)}(\pi(x, 0)) = \pi(x, \tau(x)) = \pi(f(x), 0), \tag{3.4.3}$$

we see that, through this identification, the transformation  $f : M \rightarrow M$  corresponds to the first-return map (or *Poincaré return map*) of the suspension flow to  $\Sigma$ . See Figure 3.2.

Now let  $\mu$  be a measure on  $M$  invariant under  $f$ . Let us denote by  $ds$  the Lebesgue measure on the real line  $\mathbb{R}$ . It is clear that the (infinite) measure  $\mu \times ds$  is invariant under the flow  $G^t, t \in \mathbb{R}$ . Moreover, it is invariant under the transformation  $F$ , since  $\mu$  is invariant under  $f$ . We call the *suspension of  $\mu$  with return time  $\tau$*  the measure  $\nu$  defined on  $N$  by

$$\nu = \pi_*(\mu \times ds \mid D). \tag{3.4.4}$$

In other words,  $\nu$  is the measure such that

$$\int \psi \, d\nu = \int \int_0^{\tau(x)} \psi(\pi(x, s)) \, ds \, d\mu(x)$$

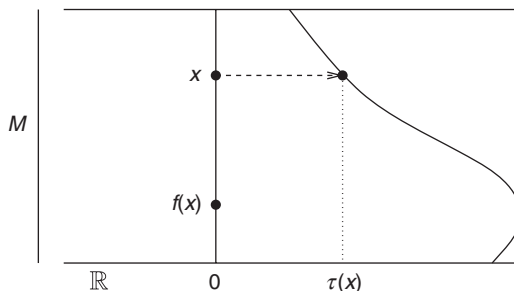


Figure 3.2. Suspension flow

for every bounded measurable function  $\psi : N \rightarrow (0, \infty)$ . In particular,

$$\nu(N) = \int 1 d\nu = \int \tau(x) d\mu(x) \tag{3.4.5}$$

is finite if and only if the function  $\tau$  is integrable with respect to  $\mu$ .

**Proposition 3.4.1.** *The flow  $g^t, t \in \mathbb{R}$  preserves the measure  $\nu$ .*

*Proof.* Let us fix  $t \in \mathbb{R}$ . Given any measurable set  $B \subset N$ , let  $\hat{B} = \pi^{-1}(B) \cap D$ . By the definition of  $\nu$ , we have that  $\nu(B) = (\mu \times ds)(\hat{B})$ . For each  $n \in \mathbb{Z}$ , let  $\hat{B}_n$  be the set of all pairs  $(x, s) \in \hat{B}$  such that  $G^{-t}(x, s) \in F^n(D)$  and let  $B_n = \pi(\hat{B}_n)$ . Since  $D$  is a fundamental domain,  $\{\hat{B}_n : n \in \mathbb{Z}\}$  is a partition of  $\hat{B}$  and  $\{B_n : n \in \mathbb{Z}\}$  is a partition of  $B$ . Moreover,  $\hat{B}_n = \pi^{-1}(B_n) \cap D$  and, consequently,  $\nu(B_n) = (\mu \times ds)(\hat{B}_n)$  for every  $n$ . The definition of the suspension flow gives that

$$\pi^{-1}(g^{-t}(B_n)) = G^{-t}(\pi^{-1}(B_n)) = G^{-t}\left(\bigcup_{k \in \mathbb{Z}} F^k(\hat{B}_n)\right) = \bigcup_{k \in \mathbb{Z}} F^k(G^{-t}(\hat{B}_n)).$$

Observing that  $F^{-n}(G^{-t}(\hat{B}_n)) \subset D$ , we conclude that

$$\nu(g^{-t}(B_n)) = (\mu \times ds)(\pi^{-1}(g^{-t}(B_n)) \cap D) = (\mu \times ds)(F^{-n}(G^{-t}(\hat{B}_n))).$$

As the measure  $\mu \times ds$  is invariant under both  $F$  and  $G^t$ , the last expression is equal to  $(\mu \times ds)(\hat{B}_n)$ . Therefore,

$$\nu(g^{-t}(B)) = \sum_{n \in \mathbb{Z}} \nu(g^{-t}(B_n)) = \sum_{n \in \mathbb{Z}} (\mu \times ds)(\hat{B}_n) = (\mu \times ds)(\hat{B}) = \nu(B).$$

This proves that  $\nu$  is invariant under the flow  $g^t, t \in \mathbb{R}$ .

In Exercise 3.4.2 we invite the reader to relate the recurrence properties of the systems  $(f, \mu)$  and  $(g^t, \nu)$ .

### 3.4.2 Poincaré maps

Next, we present a kind of inverse for the construction described in the previous section. Let  $g^t : N \rightarrow N, t \in \mathbb{R}$  be a measurable flow and  $\nu$  be an invariant measure. Let  $\Sigma \subset N$  be a *cross-section* of the flow, that is, a subset of  $N$  such that for every  $x \in \Sigma$  there exists  $\tau(x) \in (0, \infty]$  such that  $g^t(x) \notin \Sigma$  for every  $t \in (0, \tau(x))$  and  $g^{\tau(x)}(x) \in \Sigma$  whenever  $\tau(x)$  is finite. We call  $\tau(x)$  the *first-return time* of  $x$  to  $\Sigma$ . Our goal is to construct, starting from  $\nu$ , a measure  $\mu$  that is invariant under the *first-return map* (or *Poincaré return map*)

$$f : \{x \in \Sigma : \tau(x) < \infty\} \rightarrow \Sigma, \quad f(x) = g^{\tau(x)}(x).$$

Observe that this map is injective.

For each  $\rho > 0$ , denote  $\Sigma_\rho = \{x \in \Sigma : \tau(x) \geq \rho\}$ . Given  $A \subset \Sigma_\rho$  and  $\delta \in (0, \rho]$ , consider  $A_\delta = \{g^t(x) : x \in A \text{ and } 0 \leq t < \delta\}$ . Observe that the map  $(x, t) \mapsto g^t(x)$

is a bijection from  $A \times [0, \delta]$  to  $A_\delta$ . Assume that  $\Sigma$  is endowed with a  $\sigma$ -algebra of measurable subsets such that

1. the function  $\tau$  and the maps  $f$  and  $f^{-1}$  are measurable;
2. if  $A \subset \Sigma_\rho$  is measurable then  $A_\delta \subset N$  is measurable, for every  $\delta \in (0, \rho]$ .

**Lemma 3.4.2.** *Let  $A$  be a measurable subset of  $\Sigma_\rho$  for some  $\rho > 0$ . Then, the function  $\delta \mapsto \nu(A_\delta)/\delta$  is constant in the interval  $(0, \rho]$ .*

*Proof.* Consider any  $\delta \in (0, \rho]$  and  $l \geq 1$ . It is clear that

$$A_\delta = \bigcup_{i=0}^{l-1} g^{i\delta/l}(A_{\delta/l}),$$

and this is a disjoint union. Using that  $\nu$  is invariant under the flow  $g^t, t \in \mathbb{R}$ , we conclude that  $\nu(A_\delta) = l\nu(A_{\delta/l})$  for every  $\delta \in (0, \rho]$  and every  $l \geq 1$ . Then,  $\nu(A_{r\delta}) = r\nu(A_\delta)$  for every  $\delta \in (0, \rho]$  and every rational number  $r \in (0, 1)$ . Using, furthermore, the fact that both sides of this relation vary monotonically with  $r$ , we get that the equality remains true for every real number  $r \in (0, 1)$ . This implies the conclusion of the lemma.

Given any measurable subset  $A$  of  $\Sigma_\rho, \rho > 0$ , let us define

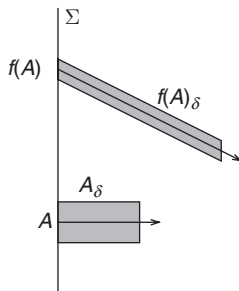
$$\mu(A) = \frac{\nu(A_\delta)}{\delta} \quad \text{for any } \delta \in (0, \rho]. \tag{3.4.6}$$

Next, given any measurable subset  $A$  of  $\Sigma$ , let

$$\mu(A) = \sup_\rho \mu(A \cap \Sigma_\rho). \tag{3.4.7}$$

See Figure 3.3. We leave it to the reader to check that  $\mu$  is a measure in  $\Sigma$  (Exercise 3.4.1). We call it the *flux* of  $\nu$  through  $\Sigma$  under the flow.

**Proposition 3.4.3.** *Suppose that the measure  $\nu$  is finite. Then the measure  $\mu$  in  $\Sigma$  is invariant under the Poincaré map  $f$ .*



**Figure 3.3.** Flux of a measure through a cross-section

*Proof.* Start by observing that the map  $f$  is essentially surjective: the complement of the image  $f(\Sigma)$  has measure zero. Indeed, suppose that there exists a set  $E$  with  $\mu(E) > 0$  contained in  $\Sigma \setminus f(\Sigma)$ . It is no restriction to assume that  $E \subset \Sigma_\rho$  for some  $\rho > 0$ . Then,  $\nu(E_\rho) > 0$ . Since  $\nu$  is finite, by assumption, we may apply the Poincaré recurrence theorem to the flow  $g^{-t}$ ,  $t \in \mathbb{R}$ . We get that there exists  $z \in E_\rho$  such that  $g^{-s}(z) \in E_\rho$  for arbitrarily large values of  $s > 0$ . By definition,  $z = g^t(y)$  for some  $y \in E$  and some  $t \in (0, \rho]$ . By construction, the backward trajectory of  $y$  intersects  $\Sigma$ . Hence, there exists  $x \in \Sigma$  such that  $f(x) = y$ . This contradicts the choice of  $E$ . Thus, the claim is proved.

Given any measurable set  $B \subset \Sigma$ , let us denote  $A = f^{-1}(B)$ . Moreover, given  $\varepsilon > 0$ , let us consider a countable partition of  $B$  into measurable subsets  $B^i$  satisfying the following conditions: for every  $i$  there is  $\rho_i > 0$  such that

1.  $B^i$  and  $A^i = f^{-1}(B^i)$  are contained in  $\Sigma_{\rho_i}$ ;
2.  $\sup(\tau | A^i) - \inf(\tau | A^i) < \varepsilon \rho_i$ .

Next, choose  $t_i < \inf(\tau | A^i) \leq \sup(\tau | A^i) < s_i$  such that  $s_i - t_i < \varepsilon \rho_i$ . Fix  $\delta_i = \rho_i/2$ . Then, using the fact that  $f$  is essentially surjective,

$$g^{t_i}(A^i_{\delta_i}) \supset B^i_{\delta_i - (s_i - t_i)} \quad \text{and} \quad g^{s_i}(A^i_{\delta_i}) \subset B^i_{\delta_i + (s_i - t_i)}.$$

Hence, using the hypothesis that  $\nu$  is invariant,

$$\begin{aligned} \nu(A^i_{\delta_i}) &= \nu(g^{t_i}(A^i_{\delta_i})) \geq \nu(B^i_{\delta_i - (s_i - t_i)}) \\ \nu(A^i_{\delta_i}) &= \nu(g^{s_i}(A^i_{\delta_i})) \leq \nu(B^i_{\delta_i + (s_i - t_i)}). \end{aligned}$$

Dividing by  $\delta_i$  we get that

$$\begin{aligned} \mu(A^i) &\geq 1 - \frac{(s_i - t_i)}{\delta} \mu(B^i) > (1 - 2\varepsilon)\mu(B^i) \\ \mu(A^i) &\leq 1 + \frac{(s_i - t_i)}{\delta} \mu(B^i) < (1 + 2\varepsilon)\mu(B^i). \end{aligned}$$

Finally, adding over all the values of  $i$ , we conclude that

$$(1 - 2\varepsilon)\mu(A) \leq \mu(B) \leq (1 + 2\varepsilon)\mu(A).$$

Since  $\varepsilon$  is arbitrary, this proves that the measure  $\mu$  is invariant under  $f$ .

### 3.4.3 Exercises

- 3.4.1. Check that the function  $\mu$  defined by (3.4.6)–(3.4.7) is a measure.
- 3.4.2. In the context of Section 3.4.1, suppose that  $M$  is a topological space and  $f : M \rightarrow M$  and  $\tau : M \rightarrow (0, \infty)$  are continuous. Let  $g^t : N \rightarrow N$  be the suspension flow and  $\nu$  be the suspension of some Borel measure  $\mu$  invariant under  $f$ .
  - (a) Show that if  $x \in M$  is recurrent for the transformation  $f$  then  $\pi(x, s) \in N$  is recurrent for the flow  $g^t$ , for every  $s \in \mathbb{R}$ .
  - (b) Show that if  $\pi(x, s) \in N$  is recurrent for the flow  $g^t$ , for some  $s \in \mathbb{R}$ , then  $x \in M$  is recurrent for  $f$ .

- (c) Conclude that the set of recurrent points of  $f$  has total measure for  $\mu$  if and only if the set of recurrent points of  $g^t, t \in \mathbb{R}$  has total measure for  $\nu$ . In particular, this happens if at least one of the measures  $\mu$  or  $\nu$  is finite.
- 3.4.3. Let  $g^t : N \rightarrow N, t \in \mathbb{R}$  be the flow defined by a vector field  $X$  of class  $C^1$  on a compact Riemannian manifold  $N$ . Assume that this flow preserves the volume measure  $\nu$  associated with the Riemannian metric. Let  $\Sigma$  be a hypersurface of  $N$  transverse to  $X$  and  $\nu_\Sigma$  be the volume measure on  $\Sigma$  associated with the restriction of the Riemannian metric. Define  $\phi : \Sigma \rightarrow (0, \infty)$  through  $\phi(y) = |X(y) \cdot n(y)|$ , where  $n(\cdot)$  is a unit vector field orthogonal to  $\Sigma$ . Show that the measure  $\eta = \phi\nu_\Sigma$  is invariant under the Poincaré map  $f : \Sigma \rightarrow \Sigma$  of the flow. Indeed,  $\eta$  coincides with the flux of  $\nu$  through  $\Sigma$ .
- 3.4.4. The following construction has a significant role in the theory of interval exchanges. Let  $\hat{N} \subset \mathbb{R}_+^4$  be the set of all 4-tuples  $(\lambda_1, \lambda_2, h_1, h_2)$  of positive real numbers, endowed with the standard volume measure  $\hat{\nu} = d\lambda_1 d\lambda_2 dh_1 dh_2$ . Define

$$F : \hat{N} \rightarrow \hat{N}, \quad F(\lambda_1, \lambda_2, h_1, h_2) = \begin{cases} (\lambda_1 - \lambda_2, \lambda_2, h_1, h_1 + h_2) & \text{if } \lambda_1 > \lambda_2 \\ (\lambda_1, \lambda_2 - \lambda_1, h_1 + h_2, h_2) & \text{if } \lambda_1 < \lambda_2. \end{cases}$$

( $F$  is not defined when  $\lambda_1 = \lambda_2$ .) Let  $N$  be the quotient of  $\hat{N}$  by the equivalence relation  $z \sim \tilde{z} \Leftrightarrow F^n(z) = \tilde{z}$  for some  $n \in \mathbb{Z}$  and let  $\pi : \hat{N} \rightarrow N$  be the canonical projection. Define

$$G^t : \hat{N} \rightarrow \hat{N}, t \in \mathbb{R}, \quad G^t(\lambda_1, \lambda_2, h_1, h_2) = (e^t \lambda_1, e^t \lambda_2, e^{-t} h_1, e^{-t} h_2).$$

Let  $\hat{a} : \hat{N} \rightarrow (0, \infty)$  be the functional given by  $\hat{a}(\lambda_1, \lambda_2, h_1, h_2) = \lambda_1 h_1 + \lambda_2 h_2$ . For each  $c > 0$ , let  $\hat{N}_c$  be the subset of all  $x \in \hat{N}$  such that  $\hat{a}(x) = c$ , let  $\hat{\nu}_c$  be the volume measure defined on  $\hat{N}_c$  by the restriction of the Riemannian metric of  $\mathbb{R}_+^4$  and let  $\hat{\eta}_c = \hat{\nu}_c / \|\text{grad } \hat{a}\|$ .

- (a) Show that  $F$  preserves the functional  $\hat{a}$  and so there exists a functional  $a : N \rightarrow (0, \infty)$  such that  $a \circ \pi = \hat{a}$ . Show that  $G^t$  commutes with  $F$  and preserves  $\hat{a}$ . Hence,  $(G^t)_t$  induces a flow  $(g^t)_t$  in the quotient space  $N$  that preserves the functional  $a$ . Check that  $F$  and  $(G^t)_t$  preserve  $\hat{\nu}$  and  $\hat{\eta}_c$  for every  $c$ .
- (b) Check that  $D = \{(\lambda_1, \lambda_2, h_1, h_2) : \lambda_1 + \lambda_2 \geq 1 > \max\{\lambda_1, \lambda_2\}\}$  is a fundamental domain for  $\sim$ . Consider the measure  $\nu = \pi_*(\hat{\nu} \mid D)$  on  $N$ . Check that the definition does not depend on the choice of the fundamental domain and show that  $\nu$  is invariant under the flow  $(g^t)_t$ . Is the measure  $\nu$  finite?
- (c) Check that  $\Sigma = \pi(\{(\lambda_1, \lambda_2, h_1, h_2) : \lambda_1 + \lambda_2 = 1\})$  is a cross-section of the flow  $(g^t)_t$ . Calculate the Poincaré map  $f : \Sigma \rightarrow \Sigma$  and the corresponding first-return time function  $\tau$ . Calculate the flux  $\mu$  of the measure  $\nu$  through  $\Sigma$ . Is the measure  $\mu$  finite?
- (d) For every  $c > 0$ , let  $N_c = \pi(\hat{N}_c)$  and  $\eta_c = \pi_*(\hat{\eta}_c \mid D)$ . Show that  $N_c$  and  $\eta_c$  are invariant under  $(g^t)_t$ , for every  $c > 0$ . Check that  $\eta_c(N_c) < \infty$  for every  $c$ . Conclude that  $\nu$ -almost every point is recurrent for the flow  $(g^t)_t$ .