

# Dynamical Systems, MA 760

## 1 Measure Theory (reminder)

1.1 DEFINITION. A  $\sigma$ -**algebra**  $\mathcal{B}$  for a set  $X$  is a collection of subsets of  $X$  such that

- (i)  $\emptyset, X \in \mathcal{B}$
- (ii)  $\{B_i\}_{i=1}^{\infty} \in \mathcal{B} \implies \bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$
- (iii)  $B \in \mathcal{B} \implies B^c \in \mathcal{B}$

A pair  $(X, \mathcal{B})$  is called a **measurable space**. Sets  $B \in \mathcal{B}$  are said to be **measurable**.

1.2 DEFINITION. A **measure**  $m$  on  $(X, \mathcal{B})$  is a function  $m : \mathcal{B} \rightarrow \mathbb{R} \cup \{+\infty\}$  such that

- (i)  $m(B) \geq 0$  for all  $B \in \mathcal{B}$
- (ii)  $m(\emptyset) = 0$
- (iii)  $\{B_i\}_{i=1}^{\infty} \in \mathcal{B}$  and  $B_i \cap B_j = \emptyset$  for  $i \neq j \implies m(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} m(B_i)$

The property (iii) is called  $\sigma$ -**additivity** or **countable additivity**. We use the obvious convention: if  $m(B_i) = \infty$  for some  $i$ , then  $\sum m(B_i) = \infty$ .  $\square$

1.3 EXERCISE. Show that if  $m(X) < \infty$ , then the clause (ii) of Definition 1.2 follows from (i) and (iii). Construct an example of a function  $m : \mathcal{B} \rightarrow \mathbb{R} \cup \{+\infty\}$  that satisfies (i) and (iii) but not (ii).

1.4 REMARK. We say that a  $\sigma$ -algebra  $\mathcal{B}$  is closed under countable (and hence, also finite) unions. It is easy to show that  $\mathcal{B}$  is also closed under countable (and finite) intersections, i.e.  $\{B_i\}_{i=1}^{\infty} \in \mathcal{B} \implies \bigcap_{i=1}^{\infty} B_i \in \mathcal{B}$  [to prove this, just use the formula  $\bigcap_{i=1}^{\infty} B_i = (\bigcup_{i=1}^{\infty} B_i^c)^c$ ]. Also,  $\mathcal{B}$  is closed under differences and symmetric differences, i.e.  $A, B \in \mathcal{B} \implies A \setminus B \in \mathcal{B}$  and  $A \Delta B \in \mathcal{B}$ , where  $A \Delta B = (A \setminus B) \cup (B \setminus A)$  [the proof is simple].

1.5 REMARK. Measures have the following simple properties:  $A \subset B \implies m(B \setminus A) = m(B) - m(A)$  and  $m(A) \leq m(B)$ . In particular,  $m(B) \leq m(X)$  for all  $B \in \mathcal{B}$ . If  $m(X) < \infty$ , then  $m$  is said to be **finite** (otherwise, it is called **infinite**). If  $m(X) = 1$ , then  $m$  is called a **probability measure** or just a **probability**. In this course, we will only deal with probability measures.

1.6 REMARK. For any set  $X$ , there are two trivial  $\sigma$ -algebras. One is *minimal*, it consists of the sets  $X$  and  $\emptyset$  only. The other is *maximal*, it contains all the subsets of  $X$ . The

latter one is denoted by  $2^X$ . (Note: if  $X$  is a finite set of  $n$  elements, then the maximal  $\sigma$ -algebra contains exactly  $2^n$  sets.)

### 1.7 EXAMPLES OF MEASURES:

(a) Let  $X = (a, b) \subset \mathbb{R}$ ,  $\mathcal{B}$  the Borel  $\sigma$ -algebra of  $X$ , and  $m$  the Lebesgue measure on  $X$ . When  $a = -\infty$  or  $b = \infty$  (or both), then  $m$  is infinite, otherwise  $m$  is finite.

(b) Let  $X = (a, b) \subset \mathbb{R}$ ,  $\mathcal{B}$  the Borel  $\sigma$ -algebra of  $X$ , again  $m$  the Lebesgue measure on  $X$ , and  $f : X \rightarrow \mathbb{R}$  is an integrable nonnegative function. Then

$$\mu(B) = \int_B f(x) dm(x) \quad \text{for } B \in \mathcal{B}$$

defines a measure  $\mu$  on  $X$ . The function  $f(x)$  is called the **density** of the measure  $\mu$ .

(c) Let  $(X, \mathcal{B})$  be an arbitrary measurable space and  $x \in X$  a selected point. The measure  $\delta_x$  defined by

$$\delta_x(B) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{otherwise} \end{cases}$$

is called a **delta-measure** or a **Dirac measure** (concentrated at  $x$ ).

(d) Let  $X$  be a finite or countable set, say,  $X = \{1, 2, \dots\}$ . Then any measure  $m$  on  $(X, 2^X)$  is determined by the numbers  $p_i = m(\{i\})$ ,  $i \in X$ , because

$$m(B) = \sum_{i \in B} p_i \quad \text{for any } B \subset X$$

**1.8 CONVENTION.** Whenever  $X$  is a finite or countable set, then we always consider the  $\sigma$ -algebra  $2^X$ . If  $X \subset \mathbb{R}$ , then we consider the Borel  $\sigma$ -algebra (unless otherwise stated).

**1.9 REMARK.**  $\sigma$ -algebras are not necessarily closed under uncountable unions or intersections. If they were, then the Borel  $\sigma$ -algebra for  $\mathbb{R}$  would contain all the subsets of  $\mathbb{R}$ , which we know is not the case.

**1.10 EXERCISE.** Let  $X$  be a finite set, say,  $X = \{1, 2, \dots, n\}$ . Describe all probability measures on  $X$ . Hint: use Example 1.7(d) and recall the notion of a *simplex* from geometry.

**1.11 REMARK.** Let  $m$  be a measure on  $(X, \mathcal{B})$  and  $c \geq 0$ . Then  $cm$  is a measure defined by  $(cm)(B) = c \cdot m(B)$  for all  $B \in \mathcal{B}$ . Let  $m_1$  and  $m_2$  be two measures on  $(X, \mathcal{B})$ . Then  $m_1 + m_2$  is a measure defined by  $(m_1 + m_2)(B) = m_1(B) + m_2(B)$  for all  $B \in \mathcal{B}$ . Hence, we can add measures and multiply them by nonnegative constants.

1.12 LEMMA. If  $m_1$  and  $m_2$  are two probability measures on  $(X, \mathcal{B})$ , then  $pm_1 + (1-p)m_2$  is a probability measure for every  $0 \leq p \leq 1$ . Hence, the set of all probability measures on  $(X, \mathcal{B})$  is *convex*.

1.13 DEFINITION. If  $m$  is a finite measure on  $(X, \mathcal{B})$  with  $m(X) > 0$ , then the measure  $m_1 = cm$ , where  $c = 1/m(X)$ , is a probability measure. The multiplication of  $m$  by  $1/m(X)$  is called the **normalization**, and  $m_1$  is called the **normalized measure**.

Note:  $\sigma$ -algebras are usually quite complicated and contain many “weird” sets. Fortunately, it is often enough to deal with certain “nice” sets that “represent” the entire  $\sigma$ -algebra.

1.14 DEFINITION. An **algebra**  $\mathcal{A}$  for a set  $X$  is a collection of subsets of  $X$  such that

- (i)  $\emptyset, X \in \mathcal{A}$
- (ii)  $\{A_i\}_{i=1}^n \in \mathcal{A} \implies \cup_{i=1}^n A_i \in \mathcal{A}$
- (iii)  $A \in \mathcal{A} \implies A^c \in \mathcal{A}$

[Note the difference from Definition 1.1: now only finite unions are required to belong in  $\mathcal{A}$ , not countable.]

1.15 EXAMPLES OF ALGEBRAS:

- (i) Let  $X = [a, b] \subset \mathbb{R}$ . Finite unions of subintervals<sup>1</sup> of  $X$  make an algebra.
- (ii) Let  $X = \mathbb{R}$ . Finite unions of intervals (including infinite intervals like  $(a, \infty)$  and  $(-\infty, b)$ ) make an algebra.

1.16 LEMMA. The intersection of any family of  $\sigma$ -algebras of a set  $X$  is always a  $\sigma$ -algebra of  $X$  (the family itself may be finite, countable or uncountable). The same property holds for algebras.

1.17 DEFINITION. Let  $\mathcal{J}$  be any collection of subsets of  $X$ . The intersection of all  $\sigma$ -algebras (algebras) containing  $\mathcal{J}$  is the *minimal*  $\sigma$ -algebra (resp., algebra) containing  $\mathcal{J}$ . It is called the  $\sigma$ -algebra (resp., algebra) **generated by**  $\mathcal{J}$  and denoted by  $\mathcal{B}(\mathcal{J})$  (resp., by  $\mathcal{A}(\mathcal{J})$ ).

A simple but useful fact: if  $\mathcal{J}$  is finite (countable), then  $\mathcal{A}(\mathcal{J})$  is also finite (countable).

---

<sup>1</sup>This includes open, closed, and semi-open intervals, like  $(c, d)$ ,  $[c, d]$ ,  $(c, d]$  and  $[c, d)$ .

1.18 DEFINITION. Let  $X$  be a topological space. Then the  $\sigma$ -algebra generated by the collection of all open sets is called the **Borel**  $\sigma$ -algebra of  $X$ . Sets in this  $\sigma$ -algebra are called **Borel sets**. Any measure defined on the Borel  $\sigma$ -algebra of  $X$  is called a **Borel measure**.

The following theorem is particularly helpful in many proofs:

1.19 APPROXIMATION THEOREM. Let  $m$  be a probability measure on  $(X, \mathcal{B})$  and let  $\mathcal{A}$  be an algebra which generates  $\mathcal{B}$ , i.e. such that  $\mathcal{B}(\mathcal{A}) = \mathcal{B}$ . Then for any  $B \in \mathcal{B}$  and any  $\varepsilon > 0$  there is an  $A \in \mathcal{A}$  such that  $m(A \Delta B) < \varepsilon$ .

That is, the sets of the  $\sigma$ -algebra  $\mathcal{B}$  can be approximated arbitrarily well by sets of the algebra  $\mathcal{A}$ .

For constructing measures, the next theorem can be very useful.

1.20 DEFINITION. Let  $\mathcal{A}$  be an algebra of  $X$ . A nonnegative function  $m_0 : \mathcal{A} \rightarrow \mathbb{R}$  is said to be  $\sigma$ -**additive** (or **countably additive**) if for any sequence  $\{A_i\}_{i=1}^{\infty}$  of disjoint sets  $A_i \in \mathcal{A}$  such that  $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$  we have  $m_0(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m_0(A_i)$ .

Note: we have to assume that  $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$ , since this does not automatically hold for algebras.

1.21 EXTENSION THEOREM. Let  $\mathcal{A}$  be an algebra of  $X$  and  $m_0 : \mathcal{A} \rightarrow \mathbb{R}$  a  $\sigma$ -additive nonnegative function. Then there is a unique finite measure  $m$  on  $(X, \mathcal{B}(\mathcal{A}))$  that coincides with  $m_0$  on  $\mathcal{A}$ . (We say that  $m$  **extends**  $m_0$  from  $\mathcal{A}$  to  $\mathcal{B}(\mathcal{A})$ .)

Therefore, to construct a measure on a  $\sigma$ -algebra, it is sufficient to construct a  $\sigma$ -additive function on an algebra that generates the  $\sigma$ -algebra.

1.22 THEOREM. Let  $(X, \mathcal{B})$  be a measurable space and  $\mathcal{J}$  a collection of subsets of  $X$  that generates  $\mathcal{B}$ , i.e. such that  $\mathcal{B}(\mathcal{J}) = \mathcal{B}$ . Suppose two measures,  $\mu_1$  and  $\mu_2$ , agree on  $\mathcal{J}$ , i.e.  $\mu_1(A) = \mu_2(A)$  for all  $A \in \mathcal{J}$ , and  $\mu_1(X) = \mu_2(X)$ . Then  $\mu_1 = \mu_2$ .

1.23 COROLLARY. If two Borel measures,  $\mu_1$  and  $\mu_2$ , on  $X = (a, b) \subset \mathbb{R}$  agree on subintervals of  $X$ , then  $\mu_1 = \mu_2$ . It is enough to require the agreement for all open intervals or for all closed intervals only.

1.24 COROLLARY. Let  $X$  be a topological space and  $\mathcal{B}$  its Borel  $\sigma$ -algebra. If two measures agree on the open sets, then they are equal.

Theorems 1.19, 1.21, and 1.22 are given without proofs here. Some were proved in Real Analysis. In any case, their proofs are beyond the scope of this course.

## 2 Simplest Examples and Basic Definitions

**2.1 CIRCLE ROTATION.** Let  $X$  be a circle and  $T : X \rightarrow X$  a transformation specified by rotating the circle  $X$  through some angle.

The circle  $X$  can be coordinatized by the (polar) angle  $\theta \in [0, 2\pi)$  and the map  $T$  specified by  $T(\theta) = \theta + \theta_0 \pmod{2\pi}$ , where  $\theta_0$  is the angle of rotation. Alternatively, we can use a complex variable  $z$  and define  $X = \{|z| = 1\}$  and  $T(z) = e^{i\theta_0}z$ .

However, we prefer to have a circle of unit length and use the coordinate  $x$  on  $X$  such that  $0 \leq x < 1$ . Equivalently,  $X$  can be thought of as a closed unit interval  $[0, 1]$  with the endpoints 0 and 1 identified. Then we set  $T(x) = x + a \pmod{1}$ , where the constant  $a$  plays the role of the angle of rotation.

**2.2 DOUBLING MAP.** Let  $X = [0, 1)$  and  $T : X \rightarrow X$  be a function defined by  $T(x) = 2x \pmod{1}$ . Again, one can think of  $X$  as the unit circle and  $x$  the angle measure, then  $T$  doubles angles. For this reason  $T$  is also called the **angle doubling map**.

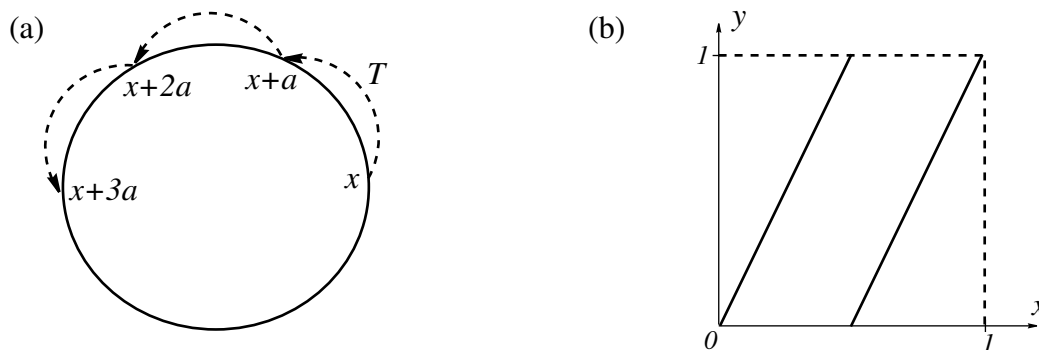


Figure 1: The circle rotation (a) and the doubling map (b).

**2.3 DEFINITION.** In dynamical systems, we deal with **iterates** of a given map  $T : X \rightarrow X$ , i.e. with the sequence of maps  $T^n : X \rightarrow X$ ,  $n \geq 1$ , defined by

$$T^n = \underbrace{T \circ T \circ \dots \circ T}_n$$

For any point  $x_0 \in X$ , the sequence

$$x_0, x_1 = T(x_0), x_2 = T^2(x_0), \dots, x_n = T^n(x_0), \dots$$

is called the **trajectory** of  $x_0$  (or, sometimes, the **orbit** of  $x_0$ ). We are usually interested in the overall behavior of the sequence  $\{x_n\}$  rather than in individual points  $x_1, x_2$ , etc. The variable  $n$  is called **time**<sup>2</sup>. We think of  $x_0$  as the **initial point** and of  $x_n = T^n(x_0)$

<sup>2</sup>Note:  $n$  only takes integral values. For this reason it is also called **discrete time**.

as its **image at time**  $n$ .

2.4. REMARKS. For the circle rotation,  $T^n(x) = x + na \pmod{1}$ , so  $T^n$  is the rotation through the angle  $na$ . For the doubling map,  $T^n(x) = 2^n x \pmod{1}$ , i.e. the graph of  $T^n$  consists of  $2^n$  branches, each has slope  $2^n$ .

2.5 QUESTION. Suppose  $A \subset X$  is some subset of interest. We want to see whether a trajectory  $\{x_n\}$  hits  $A$  at time  $n$ , i.e. whether  $x_n \in A$ . This happens whenever  $T^n(x_0) \in A$ , i.e. whenever  $x_0 \in (T^n)^{-1}(A)$ ; here  $(T^n)^{-1}$  (also denoted by  $T^{-n}$ ) is the inverse map.

2.6 DEFINITION. For any  $n \geq 1$  and any subset  $A \subset X$  the set

$$T^{-n}(A) = \{y \in X : T^n(y) \in A\}$$

is the **preimage** of  $A$  under  $T^n$ . For any point  $x \in X$  the set  $T^{-n}(x) = \{y \in X : T^n(y) = x\}$  is the **full preimage** of  $x$  under  $T^n$ . Any particular point  $y \in T^{-n}(x)$  is called **a preimage** of  $x$  under  $T^n$ . Note:  $T^{-n}$  is not necessarily a pointwise map on  $X$ , it takes points to sets (and sets to sets).

2.7 REMARKS. For the circle rotation,  $T^{-n}(x) = x - na \pmod{1}$  is the rotation through the angle  $-na$ . For the doubling map,  $T^{-n}(x)$  is a set consisting of  $2^n$  points  $\{(x+i)/2^n\}$ ,  $0 \leq i \leq 2^n - 1$ .

2.8 EXERCISE. Verify by direct inspection the following simple properties of  $T^{-n}$ :

- (a)  $T^{-m}(T^{-n}(A)) = T^{-(m+n)}(A)$  for all  $m, n \geq 1$ ;
- (b) if  $A \cap B = \emptyset$ , then  $T^{-n}(A) \cap T^{-n}(B) = \emptyset$ ;
- (c) for any  $A \subset X$  we have  $T^{-n}(A^c) = (T^{-n}(A))^c$ ;
- (d) for any  $A, B \subset X$  we have  $T^{-n}(A \cup B) = T^{-n}(A) \cup T^{-n}(B)$ ;
- (e) for any  $A, B \subset X$  we have  $T^{-n}(A \cap B) = T^{-n}(A) \cap T^{-n}(B)$ ;
- (f) for any  $A, B \subset X$  we have  $T^{-n}(A \setminus B) = T^{-n}(A) \setminus T^{-n}(B)$ ;

Therefore,  $T^{-n}$  neatly preserves all the set-theoretic operations. The properties (d) and (e) can be easily extended to countable unions and intersections. A curious remark: most of the above properties fail (!) for  $T^n$ . Give some counterexamples.

Next, if  $A \subset X$  is a “nice” (say, Borel) set, then we want  $T^{-n}(A)$  to be “nice” as well.

2.9 DEFINITION. Let  $X$  be a set with a  $\sigma$ -algebra  $\mathcal{B}$ . A transformation  $T : X \rightarrow X$  is called **measurable** if  $T^{-1}(B) \in \mathcal{B}$  for every  $B \in \mathcal{B}$ .

Note: this implies  $T^{-n}(B) \in \mathcal{B}$  for all  $B \in \mathcal{B}$  and  $n \geq 1$ . We do not require (and do not need) that  $T^n(B) \in \mathcal{B}$  for  $n \geq 1$ .

**2.10 LEMMA.** Let  $(X, \mathcal{B})$  be a measurable space and  $T : X \rightarrow X$  a transformation. Fix  $n \geq 1$ . Then the collection of sets  $\{T^{-n}(B) : B \in \mathcal{B}\}$  is a  $\sigma$ -algebra. It is denoted by  $T^{-n}(\mathcal{B})$ . Also, the collection of sets  $\{B \subset X : T^{-n}(B) \in \mathcal{B}\}$  is a  $\sigma$ -algebra.

*Proof:* This easily follows from 2.8.  $\square$

**2.11 THEOREM.** Let  $(X, \mathcal{B})$  be a measurable space and  $\mathcal{J}$  a collection of subsets of  $X$  that generates  $\mathcal{B}$ , i.e.  $\mathcal{B}(\mathcal{J}) = \mathcal{B}$ . Then a transformation  $T : X \rightarrow X$  is measurable iff  $T^{-1}(A) \in \mathcal{B}$  for all  $A \in \mathcal{J}$ .

*Proof:* The collection of subsets  $\mathcal{B}_1 = \{B \subset X : T^{-1}(B) \in \mathcal{B}\}$  is a  $\sigma$ -algebra (by 2.10), and it is assumed that  $\mathcal{J} \subset \mathcal{B}_1$ . Hence,  $\mathcal{B} \subset \mathcal{B}_1$ .  $\square$

**2.12 COROLLARY.** If  $X$  is a topological space and  $\mathcal{B}$  its Borel  $\sigma$ -algebra, then  $T : X \rightarrow X$  is measurable iff  $T^{-1}(B) \in \mathcal{B}$  for every open set  $B$ .

Note: every continuous transformation is measurable.

**2.13 COROLLARY.** If  $X = (a, b) \subset \mathbb{R}$  and  $\mathcal{B}$  is the Borel  $\sigma$ -algebra of  $X$ , then  $T : X \rightarrow X$  is measurable iff  $T^{-1}(B) \in \mathcal{B}$  for every interval  $B \subset X$ .

**2.14 EXAMPLES.** We can now easily check that the circle rotation and doubling map are measurable. Indeed, the preimage of any interval is an interval for the circle rotation and a union of two intervals for the doubling map.

**2.15 QUESTION 2.5 CONTINUED.** We now want to know “how many” points  $x_0 \in X$  hit  $A$  at time  $n$ . Since usually the number of those points is infinite, then we translate this question into the language of measures.

Suppose  $\mu$  is a fixed probability measure on  $X$ . Then we want to compute

$$\mu(\{x \in X : T^n(x) \in A\}) = \mu(T^{-n}(A))$$

The value of  $\mu(T^{-n}(A))$  is the “fraction” of points that hit  $A$  at time  $n$ , or the “chance” that a randomly selected point  $x$  hits  $A$  at time  $n$ .

**2.16 LEMMA.** Let  $\mu$  be a probability measure on  $(X, \mathcal{B})$  and  $T : X \rightarrow X$  a measurable transformation. Then, for every  $n \geq 1$ , the function  $\mu_n : \mathcal{B} \rightarrow \mathbb{R}$  given by

$$\mu_n(B) = \mu(T^{-n}(B)) \quad \forall B \in \mathcal{B}$$

is a probability measure on  $(X, \mathcal{B})$ .

*Proof:* This easily follows from 2.8.  $\square$

Note: one probability measure  $\mu$  determines another probability measure,  $\mu_n$ , for every  $n \geq 1$ .

**2.17 DEFINITION.** Let  $(X, \mathcal{B})$  be a measurable space. Denote by  $\mathcal{M} = \mathcal{M}(X)$  the set of all probability measures on  $X$ . Then a measurable map  $T : X \rightarrow X$  induces a map  $T : \mathcal{M} \rightarrow \mathcal{M}$ . For every  $\mu \in \mathcal{M}$  the measure  $T\mu$  is defined by

$$(T\mu)(B) = \mu(T^{-1}(B)) \quad \forall B \in \mathcal{B}$$

**2.18 REMARK.** In the notation of Lemma 2.16, we have  $\mu_1 = T\mu$ , hence  $\mu_n = T^n\mu$  for all  $n \geq 1$ . So, the iteration of  $T$  on  $X$  corresponds to the iteration of  $T$  on  $\mathcal{M}$ .

Note also that  $T$  is a *linear* map on  $\mathcal{M}$  in the sense that  $T(p\mu_1 + (1-p)\mu_2) = pT(\mu_1) + (1-p)T(\mu_2)$  for every  $\mu_1, \mu_2 \in \mathcal{M}$  and  $0 \leq p \leq 1$ .

It would be very convenient to have  $\mu_n = \mu$  for all  $n \geq 1$  in Lemma 2.16, so that one measure  $\mu$  would describe all the iterates of  $T$ . This requires  $T\mu = \mu$ .

**2.19 DEFINITION.** A probability measure  $\mu$  is said to be **invariant under  $T$** , or  **$T$ -invariant**, if  $T\mu = \mu$ . Equivalently,  $\mu(T^{-1}(B)) = \mu(B)$  for all  $B \in \mathcal{B}$ .

We also say that  $T$  **preserves** the measure  $\mu$ . A map  $T : X \rightarrow X$  that preserves a measure  $\mu$  is called a **measure-preserving map**. This is our notion of a dynamical system.

**2.20 THEOREM.** Let  $(X, \mathcal{B})$  be a measurable space and  $\mathcal{J}$  a collection of subsets of  $X$  that generates  $\mathcal{B}$ , i.e.  $\mathcal{B}(\mathcal{J}) = \mathcal{B}$ . Let  $T : X \rightarrow X$  be a measurable transformation. Then a probability measure  $\mu$  is  $T$ -invariant iff  $\mu(T^{-1}(A)) = \mu(A)$  for all  $A \in \mathcal{J}$ .

*Proof:* We need to show that the measures  $T\mu$  and  $\mu$  are equal. This easily follows from 1.22.  $\square$

**2.21 EXAMPLES.** Let  $m$  be the Lebesgue measure on the unit interval  $X = [0, 1)$ . Then  $m$  is invariant under the circle rotation and under the doubling map.

Indeed, for any interval  $A = (c, d) \subset X$  its preimage under the circle rotation is another interval with the same length. For the doubling map,  $T^{-1}A$  is a union of two intervals, one is  $(c/2, d/2)$  and the other  $((c+1)/2, (d+1)/2)$ . Their total length is  $d - c$ , which is  $m(A)$ . We are done.

The invariance of the Lebesgue measure  $m$  can be interpreted as follows: for any Borel set  $B \subset X$  the chance that a randomly selected point in  $X$  hits  $B$  at time  $n$  equals  $m(B)$  (and this chance does not depend on  $n$ ).



**2.22 DEFINITION.** If  $T(x) = x$ , then  $x$  is called a **fixed point** for the map  $T$ . If  $T^n(x) = x$  for some  $n \geq 1$ , then  $x$  is called a **periodic point** for the map  $T$ , and  $n$  is its period. The smallest such  $n$  is said to be the **minimal period** of  $x$ .

Note: if  $x$  is a periodic point with a minimal period  $n \geq 2$ , then the map  $T$  cyclically permutes  $n$  points  $x_0 = x, x_1 = T(x), \dots, x_{n-1} = T^{n-1}(x)$ . That is,  $T(x_i) = x_{i+1}$  and  $T(x_{n-1}) = x_0$ .

**2.23 REMARK.** If  $x$  is a fixed point for the map  $T$ , then the delta-measure  $\delta_x$  is invariant under  $T$ . If  $x$  is a periodic point with a minimal period  $n \geq 2$ , then the measure  $(\delta_{x_0} + \delta_{x_1} + \dots + \delta_{x_{n-1}})/n$  is  $T$ -invariant (we use the notation of 1.7(c) and 2.22). Check these two facts by direct inspection.

**2.24 EXERCISE.** A point  $x$  is called an **atom** for a measure  $\mu$  if  $\mu(\{x\}) > 0$ . Show that if  $x$  is an atom for a  $T$ -invariant measure, then  $x$  is a periodic point.

**2.25 MORE EXAMPLES.** Usually, a transformation  $T : X \rightarrow X$  has many invariant measures. If  $T : X \rightarrow X$  is the identity, i.e.  $T(x) = x$  for all  $x \in X$ , then every probability measure on  $X$  is invariant.

For the doubling map, the delta measure  $\delta_0$  concentrated at zero is invariant, since  $T(0) = 0$ . The measure  $\mu = 0.5\delta_{1/3} + 0.5\delta_{2/3}$  is also invariant (guess, why).

From the physics point of view, though, the most interesting and important invariant measures are those which are absolutely continuous with respect to the Lebesgue measure.

**2.26 EXERCISES** (some are rather challenging):

- (a) Let  $X = \mathbb{N}$  (the set of natural numbers) and  $T : X \rightarrow X$  defined by  $T(x) = x + 1$ . Show that  $T$  has no invariant measures.
- (b) Let  $X = \mathbb{R}$  and  $T : X \rightarrow X$  defined by  $T(x) = x + a$  with a constant  $a \neq 0$ . Show that  $T$  has no invariant measures<sup>3</sup>.
- (c) Let  $X = (0, 1)$  and  $T : X \rightarrow X$  defined by  $T(x) = x^2$ . Show that  $T$  has no invariant measures.
- (d) Let  $X = [0, 1]$  and  $T : X \rightarrow X$  defined by  $T(x) = x/2$  for  $x > 0$  and by  $T(0) = 1$ . Show that  $T$  has no invariant measures.
- (e) Let  $X = [0, 1]$  and  $T : X \rightarrow X$  defined by  $T(x) = x^2$ . Find all  $T$ -invariant measures.
- (f) Let  $X = \{1, 2, \dots, m\}$  be a finite set, and  $T : X \rightarrow X$  a permutation (i.e., a bijection of  $X$ ). Describe all  $T$ -invariant measures.

---

<sup>3</sup>Definition 2.19 can be extended to nonprobability measures: a finite or infinite measure  $\mu$  is  $T$ -invariant if  $\mu(T^{-1}(B)) = \mu(B)$  for all  $B \in \mathcal{B}$ . Under this extension, the Lebesgue measure on  $\mathbb{R}$  is invariant. In this course, though, we only deal with probability invariant measures.

**2.27 EXAMPLE.** The doubling map can be slightly generalized as follows. For an integer  $k \geq 2$  we define  $T(x) = kx \pmod{1}$  on the set  $X = [0, 1)$ . This map has many properties similar to those of the doubling map. In particular, the Lebesgue measure  $m$  is  $T$ -invariant.

Consider, more specifically, the map  $T$  with  $k = 10$ . For  $x \in X$  let  $x = 0.i_0i_1i_2\dots$  be the infinite decimal fraction representing  $x$ . Then  $10x = i_0.i_1i_2\dots$ , hence  $T(x) = 0.i_1i_2i_3\dots$ . It is just as easy to see that  $T^n(x) = 0.i_ni_{n+1}i_{n+2}\dots$  for all  $n \geq 1$ .

Why is this interesting? Consider the set  $A_r = [r/10, (r+1)/10)$  for some  $r = 0, 1, \dots, 9$ . The inclusion  $x \in A_r$  means that the decimal representation  $x = 0.i_0i_1i_2\dots$  starts with  $r$ , i.e.  $i_0 = r$ . Therefore,  $T^n(x) \in A_r$  means that the  $n$ -th digit in the decimal representation of  $x$  is  $r$ , i.e.  $i_n = r$ .

Let  $A_r(n)$  be the set of points  $x \in [0, 1)$  whose decimal representation has  $r$  at the  $n$ -th place ( $n \geq 0$ ). Note that  $A_r(0) = A_r$  and  $A_r(n) = T^{-n}(A_r)$  for all  $n \geq 0$ .

Since the Lebesgue measure  $m$  is  $T$ -invariant, we have  $m(A_r(n)) = m(T^{-n}(A_r)) = m(A_r) = 0.1$ . This means that the chance that for a randomly selected point  $x \in X$ , the  $n$ -th digit in its decimal representation of  $x$  is  $r$ , equals 0.1 (for any  $r$  and any  $n$ ). We will see more of this map later.

**2.28 THE DOUBLING MAP REVISITED.** The above discussion of the map  $T(x) = 10x \pmod{1}$  applies to the doubling map, provided one uses the binary number system. In the binary system, every point  $x \in [0, 1)$  has an infinite representation  $x = (0.i_0i_1i_2\dots)_2$  where  $i_n$ ,  $n \geq 0$ , are binary digits, i.e. zeroes and ones. Then  $2x = (i_0.i_1i_2\dots)_2$ , hence  $T(x) = (0.i_1i_2i_3\dots)_2$ . It is again easy to see that  $T^n(x) = (0.i_ni_{n+1}i_{n+2}\dots)_2$  for all  $n \geq 1$ .

**2.29 REMARK.** We note how  $T$  acts on the sequence of digits  $i_0i_1i_2\dots$  in both examples 2.27 and 2.28: the first (leftmost) digit is dropped and the rest of the sequence is moved (shifted) to the left, so that the second digit becomes the first, etc. We will see more of shift maps in Section 10.

### 3 More of Measure Theory

In this section,  $X$  is a compact metric space (or, at least, a compact metrisable topological space) and  $\mathcal{B}$  its Borel  $\sigma$ -algebra.

**3.1 STANDARD NOTATION AND FACTS.** The space of continuous functions  $f : X \rightarrow \mathbb{R}$  is denoted by  $C(X)$ . (Very rarely, we will need to consider complex-valued continuous functions  $f : X \rightarrow \mathbb{C}$ , then we shall indicate necessary changes.)

The space  $C(X)$  is a vector space (usually, infinite-dimensional). It has a norm  $\|f\| = \sup_x |f(x)|$  that makes it a metric space with distance between  $f, g \in C(X)$  given by  $\|f - g\|$ .

Probability measures  $\mu \in \mathcal{M}(X)$  can be identified with special maps  $J_\mu : C(X) \rightarrow \mathbb{R}$  defined by

$$J_\mu(f) = \int_X f d\mu$$

**3.2 FACT.** For each probability measure  $\mu \in \mathcal{M}(X)$  the map  $J_\mu : C(X) \rightarrow \mathbb{R}$  has three characteristic properties:

(J1) It is a linear and continuous map.

(J2) It is positive, i.e.  $J_\mu(f) \geq 0$  if  $f(x) \geq 0 \quad \forall x \in X$ .

(J3) It preserves unity, i.e.  $J_\mu(\mathbb{1}) = 1$ , where  $\mathbb{1}(x) = 1 \quad \forall x \in X$ .

*Proof* goes by a direct inspection.

**3.3 FACT.** If  $J_{\mu_1}(f) = J_{\mu_2}(f)$  for all  $f \in C(X)$ , then  $\mu_1 = \mu_2$ .

*Proof.* See Walters, pp. 147–148.

**3.4 FACT (RIESZ REPRESENTATION THEOREM).** If  $J : C(X) \rightarrow \mathbb{R}$  is a map with properties (J1)–(J3), then there is a measure  $\mu \in \mathcal{M}(X)$  such that  $J = J_\mu$ .

*Proof* was given in Real Analysis.

The representation of measures by integrals of continuous functions allows us to define a very useful topology on  $\mathcal{M}$ , called the weak\* topology.

**3.5 “DEFINITION”.** The **weak\* topology** on  $\mathcal{M}$  is defined so that, as  $n \rightarrow \infty$

$$\mu_n \rightarrow \mu \quad \Longleftrightarrow \quad \int_X f d\mu_n \rightarrow \int_X f d\mu \quad \forall f \in C(X)$$

The convergence of measures in the weak\* topology is called the **weak convergence**. As we see, it is equivalent to the convergence of integrals of continuous functions.

This is not a formal definition (see one below), but it is what one remembers and uses in practice.

**3.6 DEFINITION.** The weak\* topology can be defined formally. For any  $\mu_0 \in \mathcal{M}(X)$ , any finite collection of functions  $f_1, \dots, f_k \in C(X)$  and  $\varepsilon > 0$  the set

$$V_{\mu_0}(f_1, \dots, f_k; \varepsilon) = \left\{ \mu \in \mathcal{M}(X) : \left| \int_X f_i d\mu - \int_X f_i d\mu_0 \right| < \varepsilon, 1 \leq i \leq k \right\}$$

is open in the weak\* topology. These sets make a basis of the weak\* topology.

An exercise: check that 3.6 implies 3.5, indeed.

**3.8 REMARK.** The weak\* topology is metrisable. A metric on  $\mathcal{M}(X)$  that gives the weak\* topology can be defined as follows. Let  $\{f_n\}_{n=1}^\infty$  be a countable dense subset<sup>4</sup> of  $C(X)$ . Then for every  $\mu, \nu \in \mathcal{M}(X)$  we set

$$D(\mu, \nu) = \sum_{n=1}^{\infty} \frac{|\int f_n d\mu - \int f_n d\nu|}{2^n \|f_n\|}$$

Unfortunately, this metric depends on the choice of  $\{f_n\}$ , and there is no standard metric on  $\mathcal{M}(X)$  that gives the weak\* topology.

*Proof.* See Walters, pp. 148–149.

**3.9 REMARK.** There is a standard metric on  $\mathcal{M}(X)$ , defined by

$$D_{\text{var}}(\mu, \nu) = \text{total variation of } \mu - \nu$$

but it does not give the weak\* topology. It is, in a sense, *too strong*.

**3.10 EXAMPLE.** Let  $m$  be the Lebesgue measure on  $X = [0, 1]$ . For  $N \geq 1$ , let  $x_i = i/N$  for  $1 \leq i \leq N$ . Consider the measure

$$\mu^{(N)} = (\delta_{x_1} + \dots + \delta_{x_N})/N$$

This is a uniform measure on the finite set  $\{x_i\}$ ,  $1 \leq i \leq N$ . Each point  $x_i$  is an atom for  $\mu^{(N)}$ . Note that for  $f \in C(X)$

$$\int_X f d\mu^{(N)} = \frac{1}{N} (f(x_1) + \dots + f(x_N))$$

---

<sup>4</sup>A countable dense subset of  $C(X)$  exists whenever  $X$  is a metrisable compact Hausdorff space.

We know from Calculus I that, as  $N \rightarrow \infty$ ,

$$\int_X f d\mu^{(N)} \rightarrow \int_X f dm \quad \forall f \in C(X)$$

Hence  $\mu^{(N)} \rightarrow m$ , as  $N \rightarrow \infty$ , in the weak\* topology.

### 3.11 REMARKS.

- (i) The map  $X \rightarrow \mathcal{M}(X)$  given by  $x \mapsto \delta_x$  is continuous in the weak\* topology.
- (ii) The convergence  $\mu_n \rightarrow \mu$  in the weak\* topology is equivalent to

$$\limsup_n \mu_n(F) \leq \mu(F)$$

for every closed set  $F \subset X$ .

- (iii) The convergence  $\mu_n \rightarrow \mu$  in the weak\* topology is equivalent to

$$\liminf_n \mu_n(U) \geq \mu(U)$$

for every open set  $U \subset X$ .

- (iv) The convergence  $\mu_n \rightarrow \mu$  in the weak\* topology is equivalent to

$$\lim_n \mu_n(A) = \mu(A)$$

for every set  $A \subset X$  such that  $\mu(\partial A) = 0$ .

*Proof.* See Walters, pp. 149–150 and references therein.

**3.12 THEOREM (ALAOGLU).** The set  $\mathcal{M}(X)$  is compact in the weak\* topology. In particular, every sequence of measures  $\mu_n$  has a (weakly) convergent subsequence.

*Proof.* See Walters, pp. 150.

**3.13 DEFINITION.** Let  $N \geq 1$  and  $\{x_i\}$ ,  $1 \leq i \leq N$ , a finite collection of points in  $X$  (not necessarily distinct). We call

$$\mu^{(N)} = (\delta_{x_1} + \cdots + \delta_{x_N})/N$$

the **uniform atomic measure** supported on the points  $x_1, \dots, x_N$ .

We have seen in Example 3.10 that the Lebesgue measure  $m$  on  $X = [0, 1]$  can be approximated by uniform atomic measures.

**3.14 THEOREM.** Any measure  $\mu \in \mathcal{M}(X)$  can be approximated by uniform atomic measures, i.e. for every  $\mu \in \mathcal{M}(X)$  there is a sequence of such measures  $\mu^{(N)}$  that converges to  $\mu$  in the weak\* topology. (Proof is left as an exercise.)

3.15 A PHYSICAL/PHILOSOPHICAL ESSAY ABOUT MEASURES. It would be helpful if we could visualize a measure. One is used to think of the measure  $\mu(A)$  of a set  $A$  as the “size” of  $A$ . This is suitable for measure theory, but not for dynamical systems. Why not? Because here a transformation  $T$  acts on both  $X$  and  $\mathcal{M}(X)$ , so one has to deal with a sequence of measures  $\mu_n = T^n\mu$ . Thus,  $\mu_n(A)$  changes with  $n$ , while the set  $A$  does not (so its “size” should not change either).

There is a way to visualize measures that works for dynamical systems. Let  $\mu$  be a probability measure and  $\mu^{(N)}$  a uniform atomic measure approximating  $\mu$ , which is supported on some points  $x_1, \dots, x_N$ . If  $N$  is large enough, then for all practical purposes (and physical applications) the measures  $\mu$  and  $\mu^{(N)}$  are indistinguishable. We can say that the points  $x_1, \dots, x_N$  *represent* our measure  $\mu$ , i.e. we can “visualize”  $\mu$  by looking at the set of points  $x_1, \dots, x_N$ .

Now, how do we visualize the measures  $\mu_n = T^n\mu$ ? For each  $n \geq 1$  the measure  $\mu_n$  can be usually approximated by  $\mu_n^{(N)} = T^n\mu^{(N)}$ . The measure  $\mu_n^{(N)}$  is another uniform atomic measure, which is supported on the points  $\{T^n(x_i)\}$ ,  $1 \leq i \leq N$ . Hence, we can “visualize”  $\mu_n$  by looking at the set of points  $\{T^n(x_i)\}$ , the images of the original points  $\{x_i\}$ . Now  $\mu_n^{(N)}(A)$  can change with  $n$  depending on the balance of “incoming” and “outgoing” representative points that move in and out under  $T$ .

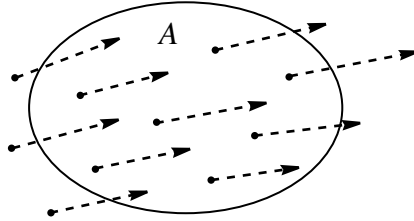


Figure 2: Three points leave the set  $A$  and three new points come in.

Now, what does a  $T$ -invariant measure  $\mu$  “look like”? If  $\mu$  is invariant, then  $T\mu = \mu$ , and then usually  $T\mu^{(N)} \approx \mu^{(N)}$ . This means that (loosely speaking) for a generic measurable set  $A$  we have  $\#\{i : x_i \in A\} \approx \#\{i : T(x_i) \in A\}$ . So, the number of points that leave the set  $A$  (go out) is about the same as the number of points that enter the set  $A$  (come in), see Fig. 2.

Suppose you first look at the set of points  $\{x_i\}$  representing the measure  $\mu$  (this is how you “visualize”  $\mu$ ). Now you apply  $T$  (turn the “switch” on), so that each point  $x_i$  instantaneously jumps to  $T(x_i)$ . You then look again at the newly obtained set of points  $\{T(x_i)\}$ . If  $\mu$  is invariant, then this set should look exactly the same as the old set  $\{x_i\}$ , i.e. you should not notice any difference in the general appearance of your set of points *before* you apply  $T$  and *after* you apply  $T$  (even though each individual point “jumps” somewhere). The preservation of that “general picture” (or general structure) of the set of representative points by  $T$  is exactly how physicists “see” the invariance of the measure  $\mu$ .

## 4 Measure-preserving Transformations

Here we begin a systematic study of measurable maps and their invariant measures.

In this section,  $(X, \mathcal{B})$  is a measurable space and  $T : X \rightarrow X$  is a measurable transformation. We denote by  $\mathcal{M} = \mathcal{M}(X)$  the set of all probability measures on  $X$  and by  $\mathcal{M}_{\text{inv}} = \mathcal{M}_{\text{inv}}(X, T)$  the set of all  $T$ -invariant probability measures on  $X$ .

**4.1 REMARK.**  $\mathcal{M}_{\text{inv}} \subset \mathcal{M}$ , and  $\mathcal{M}_{\text{inv}}$  is a convex set (i.e., if  $\mu_1, \mu_2 \in \mathcal{M}_{\text{inv}}$  then  $p\mu_1 + (1-p)\mu_2 \in \mathcal{M}_{\text{inv}}$  for all  $0 \leq p \leq 1$ ). The set  $\mathcal{M}_{\text{inv}}$  may be empty, see Exercises 2.26 (a), (b), (c), and (d).

**4.2 THEOREM (BOGOLYUBOV-KRYLOV).** If  $X$  is a compact metrisable topological space and  $T : X \rightarrow X$  a continuous transformation, then  $\mathcal{M}_{\text{inv}} \neq \emptyset$ , i.e.  $T$  has at least one invariant measure.

*Proof.* See Walters, pp. 151–152, and Pollicott, pp. 8–9.

**4.3 REMARKS.** Several interesting facts are involved in the proof of Theorem 4.2:

- (a) The map  $T : \mathcal{M} \rightarrow \mathcal{M}$  is continuous in the weak\* topology.
- (b) For any measure  $\mu \in \mathcal{M}$  every accumulation point of the sequence  $(\mu + T\mu + \cdots + T^{n-1}\mu)/n$  is a  $T$ -invariant measure. This is a very helpful method for constructing invariant measures, even in a broader context than that of Theorem 4.2.
- (c) A general fixed point theorem by Schauder (or Tychonov-Schauder) says that a continuous transformation of a compact convex set always has a fixed point. This provides an alternative proof of Theorem 4.2 – based on the continuity of  $T : \mathcal{M} \rightarrow \mathcal{M}$  and the Schauder theorem.

**4.4 REMARKS.** The assumption on the compactness of  $X$  in Theorem 4.2 cannot be dropped, see Examples 2.26(bc). The assumption on the continuity of  $T$  in Theorem 4.2 cannot be dropped either, see Examples 2.26(d).

We now return to general measurable maps on measurable spaces. Our next step is to learn how to use functions  $f : X \rightarrow \mathbb{R}$  (or, more generally,  $f : X \rightarrow \mathbb{C}$ ) in the study of invariant measures.

**4.5 NOTATION.** We denote by  $L^0(X)$  the set of all measurable functions  $f : X \rightarrow \mathbb{R}$ . Given a measure  $\mu \in \mathcal{M}$ , for any  $p > 0$  we denote

$$L^p_\mu(X) = \left\{ f \in L^0(X) : \int_X |f|^p d\mu < \infty \right\}$$

This is a vector space with norm  $\|f\|_p = \int_X |f|^p d\mu$ . In addition, the space  $L^2_\mu(X)$  has a scalar product<sup>5</sup> defined by

$$\langle f, g \rangle = \int_X f g d\mu$$

We denote by  $L^\infty(X)$  the set of all bounded functions on  $X$ . It is a linear space with norm  $\|f\|_\infty = \sup_x |f(x)|$ .

**4.6 LEMMA (CHARACTERIZING INVARIANT MEASURES).** A measure  $\mu \in \mathcal{M}$  is  $T$ -invariant if and only if

$$\int_X f \circ T d\mu = \int_X f d\mu \quad \forall f \in L^0(X)$$

(if one integral is infinite or does not exist, then the other has the same property).

*Proof.* See Walters, p. 25, and Pollicott, p. 6.

The lemma 4.6 is a particular case of a more general statement:

**4.7 LEMMA.** For any measure  $\mu \in \mathcal{M}$  its image  $\mu_1 = T\mu$  is characterized by

$$\int_X f \circ T d\mu = \int_X f d\mu_1 \quad \forall f \in L^0(X)$$

(again, if one integral is infinite or does not exist, then the other has the same property).

*Proof.* See Walters, p. 25. Also, Pollicott's argument on p. 6 applies with obvious changes.

**4.8 REMARK.** The above lemma is, in fact, a generalized change of variable formula. If  $x \in X$  and  $y = T(x)$ , then  $\int f(y) d\mu_1(y) = \int f(Tx) d\mu(x)$ .

**4.9 REMARK.** If  $X$  is a compact metrisable topological space and  $T : X \rightarrow X$  a continuous map (as in Section 3), then Lemma 4.6 can be slightly improved:  $\mu \in \mathcal{M}$  is  $T$ -invariant if and only if  $\int f \circ T d\mu = \int f d\mu$  for all  $f \in C(X)$ .

*Proof.* See Walters, p. 151.

**4.10 DEFINITION.** A measurable transformation  $T : X \rightarrow X$  induces a map  $U_T : L^0(X) \rightarrow L^0(X)$  defined by

$$(U_T f)(x) := (f \circ T)(x) = f(T(x))$$

**4.11 SIMPLE PROPERTIES.**

---

<sup>5</sup>If  $f$  and  $g$  are complex-valued functions, then we define  $\langle f, g \rangle = \int f \bar{g} d\mu$ .



- (a) For every  $n \geq 1$ , we have  $U_{T^n} = (U_T)^n$ .
- (b) The map  $U_T$  is linear.
- (c) The map  $U_T$  takes  $L^\infty(X)$  into itself and does not increase the norm  $\|\cdot\|_\infty$ , i.e.  $\|U_T f\|_\infty \leq \|f\|_\infty$ . Moreover, if  $T$  is *onto*, then  $\|U_T f\|_\infty = \|f\|_\infty$ .
- (d) For any  $T$ -invariant measure  $\mu$ , the map  $U_T$  takes  $L_\mu^p(X)$  into itself and preserves the norm  $\|\cdot\|_p$ , i.e.  $\|U_T f\|_p = \|f\|_p$ . It also preserves the scalar product in  $L_\mu^2(X)$ , i.e.  $\langle U_T f, U_T g \rangle = \langle f, g \rangle$ , i.e.  $U_T$  is a **unitary** operator. This explains the notation  $U_T$ .

4.12 REMARK. Given a function  $f : X \rightarrow \mathbb{R}$  and a point  $x \in X$ , the sequence

$$(U_T^n f)(x) = f(T^n x), \quad n \geq 0$$

consists of the values of the function  $f$  at times  $n \geq 0$  along the orbit of the point  $x$ . If  $f$  is a physical parameter (such as temperature), then this sequence consists of its measurements taken at successive time moments. It is also called the **time series**. That is what physicists (and other scientists) observe in experiments, so the behavior of the time series is just as important as that of the orbit  $\{T^n(x)\}$ .

## 5 More Examples

5.1 BAKER'S TRANSFORMATION. Let

$$X = \{(x, y) : 0 \leq x < 1, 0 \leq y < 1\}$$

be a unit square on the  $xy$  plane. The baker's map  $T : X \rightarrow X$  is defined by

$$T(x, y) = \begin{cases} (2x, y/2) & \text{if } 0 \leq x < 1/2 \\ (2x - 1, (y + 1)/2) & \text{if } 1/2 \leq x < 1 \end{cases}$$

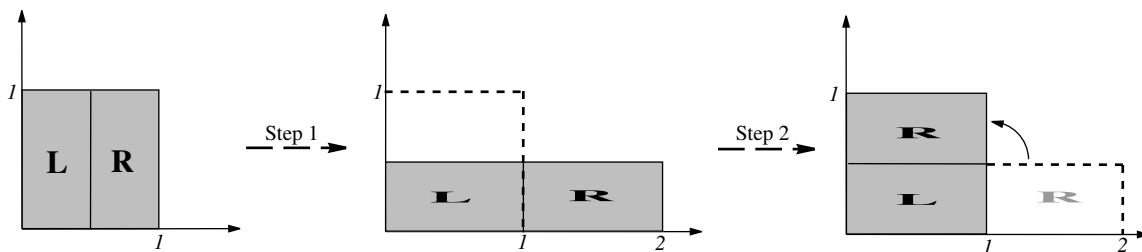


Figure 3: The baker's map constructed in two steps.

The action of  $T$  is shown on Fig. 3. At step 1, the square  $X$  is transformed into a rectangle by the linear map  $(x, y) \mapsto (2x, y/2)$  (so  $X$  is stretched in the  $x$  direction and squeezed in the  $y$  direction). At step 2, the rectangle is cut in half and its right half is placed atop its left half. This process resembles the way a baker kneads dough, hence the name **baker's map**.

5.2 REMARKS. The baker's transformation is a bijection of the unit square  $X$ . Its inverse,  $T^{-1} : X \rightarrow X$ , satisfies similar equations:

$$T^{-1}(x, y) = \begin{cases} (x/2, 2y) & \text{if } 0 \leq y < 1/2 \\ ((x + 1)/2, 2y - 1) & \text{if } 1/2 \leq y < 1 \end{cases}$$

Note that  $T$  is discontinuous on the line  $x = 1/2$  but continuous elsewhere. Such maps are said to be *piecewise continuous*. Moreover,  $T$  is *piecewise smooth* (and even *piecewise linear*). The map  $T^{-1}$  has the same properties, except it is discontinuous on another line,  $y = 1/2$ .

5.3 PROPOSITION. Let  $\mathcal{B}$  be the Borel  $\sigma$ -algebra and  $m$  the Lebesgue measure on the unit square<sup>6</sup>  $X$ . Then the baker's map  $T : X \rightarrow X$  is measurable and preserves  $m$ .

---

<sup>6</sup>Let us adopt a convention: whenever  $X \subset \mathbb{R}^d$ ,  $d \geq 2$ , then  $\mathcal{B}$  will denote the Borel  $\sigma$ -algebra and  $m$  the Lebesgue measure on  $X$ .

*Proof.* By 2.11 and 2.20, it is enough to find a collection of subsets  $\mathcal{J} \subset \mathcal{B}$  such that  $\mathcal{B}(\mathcal{J}) = \mathcal{B}$  and  $T^{-1}(A) \in \mathcal{B}$  and  $m(T^{-1}(A)) = m(A)$  for all  $A \in \mathcal{J}$ . Let  $\mathcal{J}$  consist of subrectangles  $A \subset X$  that do not intersect the line  $y = 1/2$  (in this case  $T^{-1}(A)$  will be a subrectangle as well). This will do it.  $\square$

5.4 REMARK. In the proof of Proposition 5.3, it is enough to restrict  $\mathcal{J}$  to rectangles of diameter  $< \varepsilon$ , where  $\varepsilon$  is an arbitrarily small positive number. Taking the limit as  $\varepsilon \rightarrow 0$ , the condition  $m(T^{-1}(A)) = m(A)$  will reduce to

$$|\det DT| = 1 \tag{1}$$

where  $DT$  is the matrix of partial derivatives of  $T$ . In our case

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

Hence,  $|\det DT| = 1$ , indeed.

5.5 “DIGITAL” REPRESENTATION OF THE BAKER’S MAP. Let  $(x, y) \in X$  be an arbitrary point. Let  $x = 0.i_0i_1i_2\dots$  be an infinite representation of  $x$  in the binary system, cf. 2.28. Let  $y = 0.j_1j_2j_3\dots$  be the binary representation of  $y$ . Then the image  $(x', y') = T(x, y)$  has representation

$$x' = 0.i_1i_2i_3\dots \quad \text{and} \quad y' = 0.i_0j_1j_2\dots$$

Note that it is remarkably simple: the first digit of  $x$  becomes the first digit of  $y$ , and the rest of the digits shift accordingly.

This suggests the following trick. Let us reverse the sequence  $j_1j_2j_3\dots$  and append it to the sequence  $i_0i_1i_2\dots$  on the left:

$$\dots j_3j_2j_1i_0i_1i_2\dots \tag{2}$$

thus obtaining *one* sequence of binary digits (0’s and 1’s), which is infinite in both directions (we call that a **double infinite** sequence). This sequence represents a pair of real numbers  $x$  and  $y$ , i.e. a point in the square  $X$ .

For convenience, let us denote  $\omega_k = i_k$  for  $k = 0, 1, 2, \dots$  and  $\omega_{-k} = j_k$  for  $k = 1, 2, \dots$ . Then the double infinite sequence (2) will look like

$$\underline{\omega} = (\dots \omega_{-3}\omega_{-2}\omega_{-1}\omega_0\omega_1\omega_2\dots) \tag{3}$$

The point  $\omega_0$  is the first digit of  $x$  here.

How does  $T$  act on double infinite sequences? If a point  $(x, y) \in X$  is represented by a sequence (3), then its image  $(x', y') = T(x, y)$  is represented by another sequence

$$\underline{\omega}' = (\dots \omega'_{-3}\omega'_{-2}\omega'_{-1}\omega'_0\omega'_1\omega'_2\dots)$$

that is obtained by the rule  $\omega'_k = \omega_{k+1}$  for all  $k \in \mathbb{Z}$ . Hence,  $T$  corresponds to the **left shift** on representing sequences.

**5.6 EXERCISES.** Let  $X$  be a unit square and  $T : X \rightarrow X$  a diffeomorphism (a bijection of  $X$  such that both  $T$  and  $T^{-1}$  are smooth everywhere).

- (a) Show that if  $|\det DT| \equiv 1$ , then  $T$  preserves the Lebesgue measure  $m$ .
- (b) Let  $\mu$  be a probability measure on  $X$  with density  $f(x, y)$ , cf. 1.7 (b), which means that for any Borel set  $A \subset X$

$$\mu(A) = \int_A f(x, y) dx dy = \int_A f dm$$

Then the measure  $\mu_1 = T\mu$  has density  $f_1(x, y)$  given by

$$f_1(x, y) = \frac{f(x_1, y_1)}{|\det DT(x_1, y_1)|}$$

where  $(x_1, y_1) = T^{-1}(x, y)$ .

- (c) A probability measure  $\mu$  on  $X$  with density  $f(x, y)$  is  $T$ -invariant iff

$$f(x, y) = \frac{f(x_1, y_1)}{|\det DT(x_1, y_1)|} \tag{4}$$

for  $m$ -almost all  $(x, y) \in X$ , here again  $(x_1, y_1) = T^{-1}(x, y)$ . Precisely, the set of points  $(x, y) \in X$  where (4) fails must have zero Lebesgue measure.

**5.7 REMARK.** So, we have a criterion of the invariance of the Lebesgue measure that is based on the identity (1). It can be modified and applied to the doubling map of the unit interval  $T(x) = 2x \pmod{1}$ , cf. 2.2. Let  $x \in X = [0, 1)$  and  $T^{-1}(x) = \{y_1, y_2\}$  be the full preimage of  $x$  (say,  $y_1 = x/2$  and  $y_2 = (x+1)/2$ ). Then the  $T$ -invariance of the Lebesgue measure is “equivalent” to

$$\sum_{i=1}^2 \frac{1}{|T'(y_i)|} = 1$$

This is obviously true since  $T'(y) \equiv 2$ .

**5.8 EXERCISES.** Let  $X = [0, 1]$  be the unit interval (which may be open, closed or semiopen). Let  $T : X \rightarrow X$  be a **piecewise monotonic map**, i.e. there are points  $0 = a_0 < a_1 < a_2 < \dots < a_{k-1} < a_k = 1$  such that  $T$  is strictly monotonic and differentiable on each of the  $k$  intervals  $(a_i, a_{i+1})$ ,  $0 \leq i \leq k-1$ .

(a) Show that the Lebesgue measure is  $T$ -invariant iff

$$\sum_{y \in T^{-1}(x)} \frac{1}{|T'(y)|} = 1$$

for almost every point  $x \in X$ . (Note: every point  $x$  has at most  $k$  preimages.)

(b) Let  $\mu$  be a probability measure on  $X$  with density  $f(x)$ . Then the measure  $\mu_1 = T\mu$  has density  $f_1(x)$  given by

$$f_1(x) = \sum_{y \in T^{-1}(x)} \frac{f(y)}{|T'(y)|}$$

(c) A probability measure  $\mu$  on  $X$  with density  $f(x)$  is  $T$ -invariant iff

$$f(x) = \sum_{y \in T^{-1}(x)} \frac{f(y)}{|T'(y)|} \quad (5)$$

for  $m$ -almost all  $x \in X$ . Precisely, the set of points  $x \in X$  where (5) fails must have zero Lebesgue measure.

Note: the sequence of problems in 5.8 almost repeats that of 5.6.

5.9 EXAMPLE: THE TENT MAP. Let  $X = [0, 1]$  and  $T : X \rightarrow X$  be defined by

$$T(x) = \begin{cases} 2x & \text{if } x \leq 1/2 \\ 2 - 2x & \text{if } x > 1/2 \end{cases}$$

This is the **tent map**, see Fig. 4a. Note that  $T$  is continuous and its graph is a “tent” with a sharp tip. Since  $T$  is a “two-to-one” map and  $|T'(x)| \equiv 2$ , it preserves the Lebesgue measure by 5.8(a).

5.10 EXAMPLE: A QUADRATIC MAP. Let  $X = [0, 1]$  and  $T : X \rightarrow X$  be defined by

$$T(x) = 4x(1 - x)$$

Note that  $T$  is continuous and even smooth, and its graph is a “tent” with a curved top, see Fig. 4b. Note that  $T$  is a “two-to-one” map and  $|T'(x)| = |4 - 8x|$ . It does not preserve the Lebesgue measure by 5.8(a).

5.11 EXERCISE. Show that the quadratic map in 5.10 preserves the measure  $\mu$  with density

$$f(x) = \frac{1}{\pi \sqrt{x(1-x)}}$$

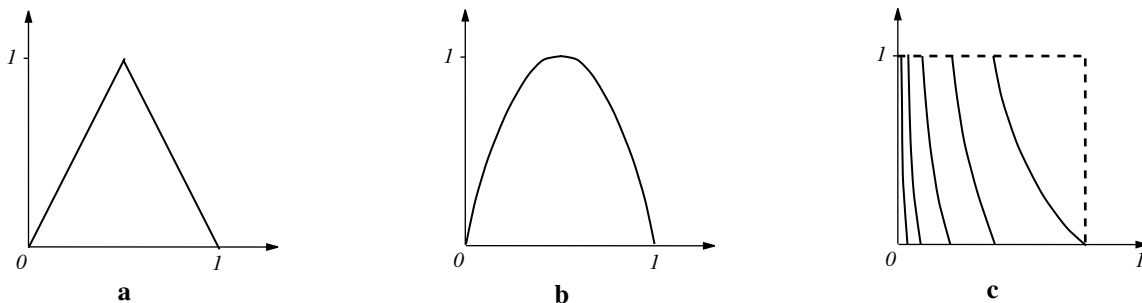


Figure 4: The tent map (a), the quadratic map (b) and the Gauss map (c).

Here  $1/\pi$  is just the normalizing factor introduced to make  $\mu$  a probability measure.

**5.12 EXAMPLE: THE GAUSS MAP.** Let  $X = [0, 1]$  and  $T : X \rightarrow X$  be defined by  $T(x) = 1/x \pmod{1}$  for  $x \neq 0$  and  $T(0) = 0$ . That is,  $T(x)$  is the fractional part of  $1/x$ . The graph of this function has infinitely many branches, see Fig. 4c. Its discontinuity points are  $x = 1/n$ ,  $n \in \mathbb{N}$ , and  $x = 0$ . Every point  $x < 1$  has infinitely many preimages.

**5.13 EXERCISE.** Show that the Gauss map in 5.12 preserves the measure  $\mu$  with density

$$f(x) = \frac{1}{\ln 2} \cdot \frac{1}{1+x}$$

Here  $1/\ln 2$  is just the normalizing factor introduced to make  $\mu$  a probability measure. This density was found by K. F. Gauss in the Nineteenth Century.

**5.14 REMARK.** The density of a  $T$ -invariant measure is a solution of the functional equation (5). Generally, it is very hard (if at all possible) to solve it. The solutions given in Exercises 5.11 and 5.13 for two particular maps have been found basically “by accident” or “by trial and error”. There are no general algorithms to solve functional equations.

**5.15 DEFINITION.** An map  $T : I \rightarrow I$  of an interval  $I \subset \mathbb{R}$  is said to have an **absolutely continuous invariant measure** (**a.c.i.m.** for short) if there is a  $T$ -invariant measure on  $I$  with a density.

We have seen that the doubling map, the tent map, the quadratic map, and the Gauss map all have a.c.i.m.

**5.16 REMARK.** The existence or uniqueness of an a.c.i.m. is not guaranteed. For all of the above interval maps the a.i.c.m. is indeed unique, and we will prove that in Section 8.

By far, we have seen maps of various kinds: some were one-to-one (e.g., circle rotations and the baker's transformation) and some others were two-to-one (the doubling map and the tent map) or even infinitely-many-to-one (the Gauss map). If a map  $T : X \rightarrow X$  is a bijection, one can use  $T^{-1}$  as well.

**5.17 DEFINITION.** A bijective map  $T : X \rightarrow X$  is called an **automorphism** if both  $T$  and  $T^{-1}$  are measurable maps. (Examples: circle rotations and the baker's map.)

On the contrary, if  $T$  is not an automorphism, then  $T$  is called an **endomorphism**.

**5.18 EXERCISE.** If an automorphism  $T$  preserves a measure  $\mu$ , then its inverse  $T^{-1}$  also preserves  $\mu$ .

## 6 Recurrence

Here we turn back to the general properties of measure-preserving transformations. In this section,  $T : X \rightarrow X$  always denotes a measurable map with a  $T$ -invariant measure  $\mu$ .

6.1 DEFINITION. Let  $A \subset X$  be a measurable set and  $x \in A$ . Denote by

$$\tau_A(x) = \min\{n \in \mathbb{N} : T^n(x) \in A\}$$

the **first return time** when the point  $x$  comes back to the set  $A$ . If such a time does not exist (i.e., if  $x$  never returns to  $A$ ) then we set  $\tau_A(x) = \infty$ .

6.2 THEOREM (POINCARÉ RECURRENCE THEOREM). If  $\mu(A) > 0$ , then almost every point  $x \in A$  does return to  $A$ , i.e.

$$\mu(\{x \in A : \tau_A(x) = \infty\}) = 0$$

*Proof.* See Walters, p. 26 or Pollicott, p. 9.

6.3 REMARK. Poincaré Theorem 6.2 is false if the measure  $\mu$  is infinite. For example, let  $X = \mathbb{R}$ ,  $\mu$  be the Lebesgue measure and  $T(x) = x + 1$ . Then the set  $A = (0, 1)$  has no returning points at all.

6.4 COROLLARY. If  $\mu(A) > 0$ , then almost every point  $x \in A$  returns to  $A$  infinitely many times, i.e. for almost every  $x \in A$  there is a sequence  $0 < n_1 < n_2 < \dots$  of natural numbers such that  $T^{n_i}(x) \in A$  for each  $i$ .

6.5 REMARK. Let  $\tilde{A} \subset A$  denote the set of points that return to  $A$  infinitely many times. Note that

$$\tilde{A} = A \cap \left( \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} T^{-m}(A) \right)$$

Since  $\tilde{A}$  is obtained by a countable number of set operations, it is a measurable set, i.e.  $\tilde{A} \in \mathcal{B}$ .

6.6 DEFINITION. Let  $A \subset X$  and  $\mu(A) > 0$ . The map  $T_A : A \rightarrow A$  defined by

$$T_A(x) = T^{\tau_A(x)}(x)$$

is called the **first return map** (or **Poincaré map**) on  $A$ . It is actually defined on the set  $\{x \in A : \tau_A(x) < \infty\}$ , but that set coincides with  $A$  up to a set of measure zero.

It is more convenient to restrict  $T_A$  to the set  $\tilde{A}$ . Note that  $T_A(\tilde{A}) \subset \tilde{A}$ , i.e.  $\tilde{A}$  is invariant under  $T_A$ . We now can consider  $T_A : \tilde{A} \rightarrow \tilde{A}$  as a new transformation, induced by  $T$  and  $A$ .



Note that  $\mu(\tilde{A}) = \mu(A)$  and  $\mu(A \Delta \tilde{A}) = 0$  by 6.4, so one commonly identifies  $A$  with  $\tilde{A}$  by neglecting a set of zero measure. Still, for the sake of clarity we will keep working with  $\tilde{A}$ .

6.7 DEFINITION. Let  $\mathcal{B}_{\tilde{A}}$  be the  $\sigma$ -algebra induced on  $\tilde{A}$ , i.e.

$$\mathcal{B}_{\tilde{A}} = \{B \in \mathcal{B} : B \subset \tilde{A}\}$$

Define a probability measure  $\mu_{\tilde{A}}$  on  $(\tilde{A}, \mathcal{B}_{\tilde{A}})$  by

$$\mu_{\tilde{A}}(B) = \mu(B)/\mu(\tilde{A}) \quad \text{for all } B \in \mathcal{B}_{\tilde{A}}$$

The measure  $\mu_{\tilde{A}}$  is called the **conditional measure**.

6.8 THEOREM. Let  $A \subset X$  and  $\mu(A) > 0$ . Then the first return map  $T_A : \tilde{A} \rightarrow \tilde{A}$  preserves the conditional measure  $\mu_{\tilde{A}}$ .

*Proof.* Let  $B \in \mathcal{B}_{\tilde{A}}$ . Define a sequence of sets  $B_n \subset \tilde{A}$  by

$$B_n = \{x \in \tilde{A} : \tau_A(x) = n \text{ \& } T^n(x) \in B\}$$

Then  $T_A^{-1}(B) = \cup_{n \geq 1} B_n$ . Next, the sets  $B_n$  are pairwise disjoint, because  $\tau_A$  takes different values on different  $B_n$ 's. Lastly, we need to show that  $\mu(B) = \sum_n \mu(B_n)$ . For each  $n \geq 1$ , let

$$C_n = \{x \in T^{-n}(B) : T^i x \notin A \text{ } \forall i = 0, 1, \dots, n-1\}$$

Verify by direct inspection that  $T^{-1}(C_n) = C_{n+1} \cup B_{n+1}$  and  $C_{n+1} \cap B_{n+1} = \emptyset$ . Therefore,

$$\mu(C_n) = \mu(C_{n+1}) + \mu(B_{n+1})$$

Also,  $T^{-1}(B) = C_1 \cup B_1$  and  $C_1 \cap B_1 = \emptyset$ , hence  $\mu(B) = \mu(C_1) + \mu(B_1)$ . Adding all these equations for measures together proves the theorem.

Right? Not quite. We also need to show that  $\lim_{n \rightarrow \infty} \mu(C_n) = 0$ . Do this as an exercise. Hint: show that the sets  $C_n$  are pairwise disjoint.  $\square$

6.9 EXERCISE. Let  $f : X \rightarrow \mathbb{R}^+$  be a *positive* measurable function, i.e.  $f(x) > 0$  for all (or almost all)  $x \in X$ . Show that

$$\sum_{n=0}^{\infty} f(T^n(x)) = \infty$$

for almost every point  $x \in X$ .

Find an example of a strictly positive function  $f > 0$  so that the above series converges for some points  $x \in X$ . (Hint: you can use the map 2.26e).

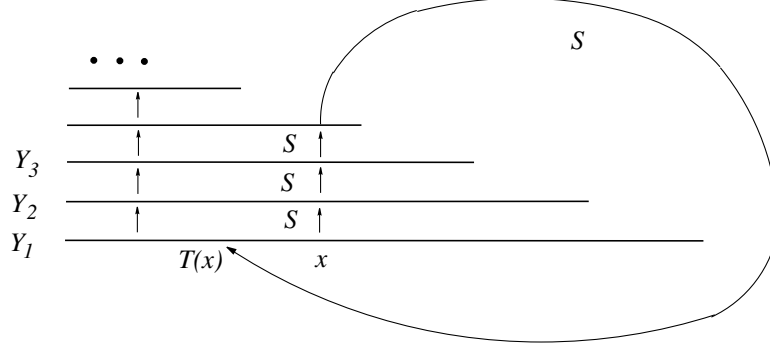


Figure 5: The tower map  $S$ . A point  $x$  with  $\tau(x) = 4$  is shown.

The following construction is in a sense opposite to that of the first return map. Before, we shrunk the space  $X$  to a smaller set  $A$ . Now we will construct a “larger” space to replace  $X$ .

**6.10 DEFINITION.** Let  $T : X \rightarrow X$  be a transformation preserving a measure  $\mu$  and  $\tau : X \rightarrow \mathbb{N}$  a measurable positive integral-valued function on  $X$  such that

$$D = \int_X \tau d\mu < \infty$$

Let

$$Y = \{(x, k) : x \in X, k = 1, 2, \dots, \tau(x)\}$$

This is called a **tower**. It is naturally partitioned into levels (“floors”)  $Y = \cup_{n \geq 1} Y_n$  where

$$Y_n = \{(x, k) \in Y : k = n\}$$

Let  $\varphi : Y \rightarrow X$  be a natural projection  $\varphi(x, n) = x$ . Define a map  $S : Y \rightarrow Y$  by

$$S(x, k) = \begin{cases} (x, k+1) & \text{if } k < \tau(x) \\ (T(x), 1) & \text{if } k = \tau(x) \end{cases}$$

Note that  $S$  moves each point straight up the tower until it reaches the top level (“the ceiling”), then  $S$  takes it down to the level zero and at that time applies the “old” map  $T$ , see Fig. 5. The function  $\tau(x)$  is called the **ceiling function**.

Define a measure  $\mu'$  on each level  $Y_n$ ,  $n \geq 1$ , by

$$\mu'(B) = \mu(\varphi(B)) \quad B \subset Y_n$$

(this also defines a  $\sigma$ -algebra on  $Y_n$ ). Then we obtain a measure  $\mu'$  on  $Y$ . It is finite and  $\mu'(Y) = D$ , then the measure  $\tilde{\mu} = \mu'/D$  is a probability measure on  $Y$ . The map  $S$  preserves the measure  $\tilde{\mu}$ , which can be verified directly.

## 7 Ergodicity

7.1 DEFINITION (NOT A GOOD ONE!). A set  $B \subset X$  is said to be  $T$ -invariant if  $T(B) \subset B$ .

7.2 REMARK. The above definition is standard in some mathematical courses, but not so convenient for dynamical systems. First of all,  $T(B)$  may not be measurable, even if  $B$  is. Second, when  $B$  is invariant, then  $B^c = X \setminus B$  may not be. These considerations motivate the following:

7.3 DEFINITION A measurable set  $B \subset X$  is **(fully)  $T$ -invariant** if  $T^{-1}(B) = B$ .

Note that now when  $B$  is invariant, then so is  $B^c$ , i.e.  $T^{-1}(B^c) = B^c$ . The following exercise shows, though, that Definition 7.1 is essentially equivalent to 7.3:

7.4 EXERCISE. Let  $T(B) \subset B$  for a set  $B \in \mathcal{B}$ . Consider the “forward limit”  $B_+ = \bigcap_{n \geq 0} T^n(B)$  and the “backward limit”  $B_- = \bigcup_{n \geq 0} T^{-n}(B)$ .

- (a) Show that the set  $B_-$  is measurable and (fully) invariant, i.e.  $T^{-1}(B_-) = B_-$ . Is  $B_+$  also fully invariant? (Prove or give a counterexample.)
- (b) Show that if  $\mu$  is a  $T$ -invariant measure, then  $\mu(B \Delta B_-) = 0$ . Also, if  $T^n(B) \in \mathcal{B}$  for all  $n \geq 1$ , then prove  $\mu(B \Delta B_+) = 0$ .

This shows that for any invariant set  $B$  in the sense of 7.1 there is a fully invariant set  $B_-$  that coincides with  $B$  up to a null set.

In dynamical systems, we work with measures. Sets of zero measure are treated as negligible and often ignored.

7.5 DEFINITIONS. Let  $\mu \in \mathcal{M}$  be a probability measure. We call  $A$  a **null set** if  $\mu(A) = 0$ . We say that  $A$  and  $B$  **coincide (mod 0)** if  $\mu(A \Delta B) = 0$ . In this case we write  $A = B \pmod{0}$ .

7.6 REMARK. The relation  $A = B \pmod{0}$  is an equivalence relation. In particular, to check that  $A = B \pmod{0}$  and  $B = C \pmod{0}$  implies  $A = C \pmod{0}$ , one can use a simple formula  $A \Delta C \subset (A \Delta B) \cup (B \Delta C)$ .

7.7 DEFINITION. Let  $\mu$  be an invariant measure. We say that a set  $B \subset X$  is **invariant (mod 0)** if  $B = T^{-1}(B) \pmod{0}$ , i.e.  $\mu(B \Delta T^{-1}B) = 0$ .

Note: in this case  $B = T^{-n}B \pmod{0}$  for all  $n \geq 1$ . This easily follows from 7.6.

7.8 DEFINITION. Let  $\mu$  be a probability measure. If  $f, g : X \rightarrow \mathbb{R}$  are two measurable functions, then we say that  $f$  and  $g$  are  $\mu$ -**equivalent**, and write  $f = g \pmod{0}$ , if  $\mu\{x : f(x) \neq g(x)\} = 0$ .

7.9 EXERCISE. Let  $B \in \mathcal{B}$ . Consider the set

$$B_\infty = \bigcap_{n=0}^{\infty} \bigcup_{m=n}^{\infty} T^{-m}(B)$$

consisting of points whose orbits visit  $B$  infinitely many times.

- (a) Show that the set  $B_\infty$  is measurable and (fully) invariant, i.e.  $T^{-1}(B_\infty) = B_\infty$ .
- (b) Let  $\mu$  be a  $T$ -invariant measure. Suppose that  $B = T^{-1}(B) \pmod{0}$ , i.e.  $B$  is invariant  $\pmod{0}$ . Prove that  $B = B_\infty \pmod{0}$ .

This shows that for any  $\pmod{0}$  invariant set  $B$  there is a fully invariant set  $B_\infty$  that coincides  $\pmod{0}$  with  $B$ .

7.10 EXAMPLES.

- (a) If  $X$  is a finite set and  $T : X \rightarrow X$  is a permutation (see 2.26f), then every set whose elements form a cycle is fully invariant.
- (b) If  $X = [0, 1]$  and  $T(x) = x^2$ , see 2.26e, then the sets  $B_1 = \{0\}$ ,  $B_2 = (0, 1)$  and  $B_3 = \{1\}$  are fully invariant.

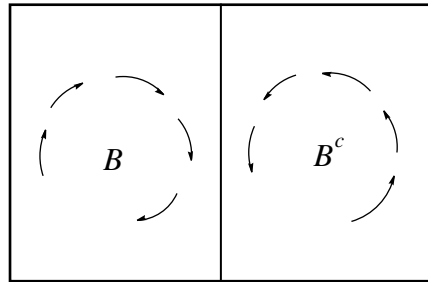


Figure 6: The decomposition of  $X$  into two invariant subsets.

7.11 REMARK. Let  $B \subset X$  be a (fully) invariant set. Then  $T(B) \subset B$  and  $T(B^c) \subset B^c$ . So, the orbits that originate in  $B$  never visit  $B^c$  and the orbits originating in  $B^c$  never visit  $B$ . The space  $X$  naturally decomposes into two “noninteracting” parts:  $X_1 = B$  and  $X_2 = B^c$ . We can restrict  $T$  to  $X_1$  and  $X_2$  and consider two measurable maps  $T_i = T : X_i \rightarrow X_i$  with  $i = 1, 2$ , separately.

Whenever  $\mu_1$  is a  $T_1$ -invariant measure on  $X_1$  and  $\mu_2$  is a  $T_2$ -invariant measure on  $X_2$ , then any weighted sum  $p\mu_1 + (1-p)\mu_2$  for  $0 \leq p \leq 1$  will be a  $T$ -invariant measure on  $X$ .

Also, any  $T$ -invariant measure  $\mu$  on  $X$  such that  $\mu(B) > 0$  and  $\mu(B^c) > 0$  is a weighted sum of the conditional measures  $\mu_B$  and  $\mu_{B^c}$  on  $B$  and  $B^c$ , which are invariant under  $T_1$  and  $T_2$ , respectively. In this case one can reduce the study of the map  $T : X \rightarrow X$  to the study of two “smaller” maps  $T_1 : X_1 \rightarrow X_1$  and  $T_2 : X_2 \rightarrow X_2$ .

**7.12 DEFINITION.** A  $T$ -invariant measure  $\mu$  is said to be **ergodic** if there is no  $T$ -invariant subset  $B \subset X$  such that  $\mu(B) > 0$  and  $\mu(B^c) > 0$ . (Equivalently: there is no  $T$ -invariant set  $B$  such that  $0 < \mu(B) < 1$ .)

We also say that  $T$  is **ergodic** (with respect to the measure  $\mu$ ). The set of ergodic measures is denoted by  $\mathcal{M}_{\text{erg}} \subset \mathcal{M}_{\text{inv}}$ .

### 7.13 EXAMPLES.

- (a) If  $X$  is a finite set and  $T : X \rightarrow X$  is a permutation (see 2.26f), then a  $T$ -invariant measure is ergodic iff it is concentrated on one cycle.
- (b) If  $X = [0, 1]$  and  $T(x) = x^2$ , see 2.26e, then the only ergodic measures are  $\delta_0$  and  $\delta_1$ . Note: a  $T$ -invariant set  $(0, 1)$  cannot carry any  $T$ -invariant measure, see 2.26c.

We see that ergodic measures are, in a sense, “extreme” invariant measures.

Recall that the collection of  $T$ -invariant measures,  $\mathcal{M}_{\text{inv}}$  is a convex subset of  $\mathcal{M}(X)$ , see 4.1. For any convex set  $R$  a point  $x \in R$  is called an **extremal point** of  $R$  if there is no pair of points  $x_1, x_2 \in R$ ,  $x_1 \neq x_2$ , such that  $x = px_1 + (1 - p)x_2$  for some  $0 < p < 1$ .



Figure 7: Extremal points of convex sets: the entire boundary of an oval and just four vertices of a rectangle.

**7.14 PROPOSITION.** Ergodic measures are precisely extremal points of the set  $\mathcal{M}_{\text{inv}}$ .

*Proof.* Let  $\mu$  be a nonergodic measure. Let  $B$  be an invariant set such that  $0 < \mu(B) < 1$ . Then  $\mu$  is a weighted sum of the conditional measures  $\mu_B$  and  $\mu_{B^c}$ , see Remark 7.11, i.e.  $\mu$  is not an extremal point of  $\mathcal{M}$ .

Conversely, let  $\mu$  be another measure that satisfies  $\mu = p\mu_1 + (1 - p)\mu_2$  for some distinct  $T$ -invariant measures  $\mu_1 \neq \mu_2$  and  $0 < p < 1$ . Consider the signed measure  $\nu = \mu_1 - \mu_2$ .

By Hahn decomposition theorem, there is a set  $B \in \mathcal{B}$  such that  $\nu = \nu_+ - \nu_-$ , where  $\nu_+$  and  $\nu_-$  are positive measures concentrated on  $B$  and  $B^c$ , respectively. Moreover,  $\nu_+(A) = \nu(A \cap B)$  and  $\nu_-(A) = -\nu(A \cap B^c)$  for all  $A \in \mathcal{B}$ .

Since  $\mu_1$  and  $\mu_2$  are  $T$ -invariant, then so is  $\nu$ . Therefore,  $\nu(B \setminus T^{-1}B) = \nu((T^{-1}B) \setminus B)$ . On the other hand,  $\nu(B \setminus T^{-1}B) \geq 0$  and  $\nu((T^{-1}B) \setminus B) \leq 0$ . Hence,  $\nu(B \setminus T^{-1}B) = \nu((T^{-1}B) \setminus B) = 0$ , i.e.  $\bar{\nu}(B \Delta T^{-1}B) = 0$ , where  $\bar{\nu} = \nu_+ + \nu_-$  is the total variation of  $\nu$ . Next,  $\nu_+$  is an invariant measure since for any  $A \subset X$

$$\nu_+(T^{-1}A) = \nu((T^{-1}A) \cap B) = \nu(T^{-1}(A \cap B)) = \nu(A \cap B) = \nu_+(A)$$

Similarly,  $\nu_-$  is invariant, hence  $\bar{\nu}$  is invariant as well. Therefore, for the  $T$ -invariant set  $B_\infty$  we have  $\bar{\nu}(B \Delta B_\infty) = 0$ , see Exercise 7.9. Hence,  $\nu(B_\infty) > 0$  and  $\nu(B_\infty^c) < 0$ . Thus,  $\mu(B_\infty) > 0$  and  $\mu(B_\infty^c) > 0$ , so the measure  $\mu$  is not ergodic.  $\square$

We see that a nonergodic measure  $\mu$  can be represented by a weighted sum of two other invariant measures. If those measures are not ergodic either, this decomposition can be continued, and ultimately we can represent  $\mu$  by ergodic measures only:

**7.15 THEOREM (ERGODIC DECOMPOSITION).** Let  $X$  be a topological space. Given any invariant measure  $\mu \in \mathcal{M}_{\text{inv}}$  there exists a probability measure  $\rho_\mu$  on the space  $\mathcal{M}_{\text{inv}}$  such that

- (a)  $\rho_\mu(\mathcal{M}_{\text{erg}}) = 1$ ;
- (b) for any  $f \in L^1_\mu(X)$

$$\int_X f d\mu = \int_{\mathcal{M}_{\text{erg}}} \left( \int_X f d\nu \right) d\rho_\mu(\nu)$$

(i.e. the invariant measure  $\mu$  is an affine combination, weighted by  $\rho_\mu$ , of ergodic measures  $\nu \in \mathcal{M}_{\text{erg}}$ ).

We assume this statement without proof. A proof is outlined in Pollicott, pp. 18-19. The ergodic decomposition theorem 7.15 is conceptually important, but practically it is hardly useful (see Remark 7.21, however).

**7.16 COROLLARY.** Let  $X$  be a compact metrisable topological space and  $T : X \rightarrow X$  a continuous transformation. Then  $\mathcal{M}_{\text{inv}}$  is closed and  $\mathcal{M}_{\text{erg}} \neq \emptyset$ , i.e. there exists at least one ergodic measure.

**7.17 DEFINITION.** Two measures  $\mu_1, \mu_2 \in \mathcal{M}(X)$  are said to be **mutually singular** (denoted  $\mu_1 \perp \mu_2$ ) if there is a set  $B \in \mathcal{B}$  such that  $\mu_1(B) = 1$  and  $\mu_2(B) = 0$ .

**7.18 PROPOSITION.** If  $\mu_1, \mu_2 \in \mathcal{M}_{\text{erg}}$  are two ergodic measures, then they either coincide ( $\mu_1 = \mu_2$ ) or are mutually singular ( $\mu_1 \perp \mu_2$ ).

*Proof.* Consider the signed measure  $\nu = \mu_1 - \mu_2$ . Arguing as in the proof of Proposition 7.14 gives  $\mu_1(B_\infty) > 0$  and  $\mu_2(B_\infty^c) > 0$ . Now use the ergodicity of  $\mu_1$  and  $\mu_2$ .

**7.19 PROPOSITION.** Let  $\mu \in \mathcal{M}_{\text{inv}}$  be an invariant measure. Suppose the measure  $\rho_\mu$  in 7.15 is concentrated on a finite or countable set of ergodic measures  $\nu_1, \nu_2, \dots$ . Since all of them are mutually singular by 7.18, there is a partition  $X = \cup_{n \geq 1} X_n$ ,  $X_i \neq X_j$  for  $i \neq j$ , such that for each  $n \geq 1$  the measure  $\nu_n$  is concentrated on  $X_n$ , i.e.  $\nu_n(X_n) = 1$ . In this case  $\nu_n = \mu_{X_n}$  is the conditional measure induced by  $\mu$  on  $X_n$ .

**7.20 REMARK.** The above proposition can be generalized to arbitrary invariant measures  $\mu$ . That is, the space  $X$  can be decomposed into smaller  $T$ -invariant subsets on each of which the conditional measure induced by  $\mu$  is ergodic. However, when those subsets have zero  $\mu$ -measure, the definition of conditional measures is quite involved and we do not give it here, see Pollicott, p. 18, and Walters, p. 9.

**7.21 REMARK.** Let  $A \subset X$  be a subset such that  $\nu(A) = 1$  for every ergodic measure  $\nu$ . Then  $\mu(A) = 1$  for every invariant measure  $\mu$ . Indeed, we can set  $f = \chi_A$ , the indicator of the set  $A$ , and apply Theorem 7.15.

Invariance of sets (defined by 7.3) is just as important as invariance of functions:

**7.22 DEFINITION.** A function  $f \in L^0(X)$  is **invariant** if  $U_T f = f$ , i.e.  $f(T(x)) = f(x)$  for all  $x \in X$ . In this case  $f$  is constant on every orbit  $\{T^n x\}$ .

Given an invariant measure  $\mu$ , a function  $f$  is **almost everywhere invariant** if  $f(T(x)) = f(x)$  for a.e.  $x \in X$ . This precisely means  $\mu(\{x : f(T(x)) \neq f(x)\}) = 0$ .

Note: if the function  $f$  is invariant a.e., then  $f$  is constant on the orbit of almost every point  $x \in X$ .

**7.23 LEMMA (CHARACTERIZING ERGODIC MEASURES).** An invariant measure  $\mu$  is ergodic iff any invariant (alternatively: any almost everywhere invariant) function  $f$  is constant a.e.

Note: it is enough to restrict this criterion to functions  $f \in L_\mu^1(X)$  or  $f \in L_\mu^2(X)$  or even  $f \in L^\infty(X)$ .

*Proof.* See Walters, p. 28, or Pollicott, p. 10.

**7.24 EXERCISE.** Let  $\mu$  be an ergodic measure and  $\mu(A) > 0$ . Show that  $\mu(A_\infty) = 1$ , i.e. almost every point  $x \in X$  visits the set  $A$  infinitely many times.

**7.25 COROLLARY.** Let  $X$  be a topological space with a countable basis. Let  $\mu$  be an ergodic measure such that  $\mu(U) > 0$  for every nonempty open set  $U$  (this is quite common

in physics). Then the orbit of almost every point  $x \in X$  visits every open set infinitely many times. In particular, the orbit of almost every point is dense in  $X$ .

The following two exercises are very similar. Do only one of them (your choice):

**7.26 EXERCISE.** Let a map  $T : X \rightarrow X$  be ergodic with respect to a measure  $\mu$ . Let  $\mu(A) > 0$ . Show that the first return map  $T_A : A \rightarrow \tilde{A}$  constructed in 6.6 is ergodic with respect to the conditional measure  $\mu_{\tilde{A}}$  (defined in 6.7).

**7.27 EXERCISE.** Let a map  $T : X \rightarrow X$  be ergodic with respect to a measure  $\mu$ . Let  $\tau : X \rightarrow \mathbb{N}$  be a function with a finite integral. Show that the map  $S$  on the tower  $Y$  defined in 6.10 is ergodic with respect to the measure  $\tilde{\mu}$  (also defined in 6.10).

An important concept in dynamical systems is that of isomorphism.

**7.28 DEFINITION.** For  $i = 1, 2$ , let  $T_i : X_i \rightarrow X_i$  be a transformation preserving a probability measure  $\mu_i$ . We say that  $T_1$  and  $T_2$  are **isomorphic** if for each  $i = 1, 2$  there is a  $T_i$ -invariant set  $B_i \subset X_i$  of full measure, i.e.  $T(B_i) \subset B_i$  and  $\mu_i(B_i) = 1$ , and a bijection  $\varphi : B_1 \rightarrow B_2$  such that

- (i)  $\varphi$  preserves measures, i.e. for every measurable set  $A \subset B_1$  the set  $\varphi(A) \subset B_2$  is measurable and  $\mu_1(A) = \mu_2(\varphi(A))$  (and vice versa);
- (ii)  $\varphi$  preserves dynamics, i.e.  $\varphi \circ T_1 = T_2 \circ \varphi$  on  $B_1$ .

We call  $\varphi$  an **isomorphism** and write  $T_1 \simeq T_2$ . An isomorphism means that two dynamical systems,  $(X_1, T_1, \mu_1)$  and  $(X_2, T_2, \mu_2)$  are equivalent, up to sets of zero measure (which we neglect).

**7.29 REMARKS.** Isomorphism is an equivalence relation. Also, if  $T_1 \simeq T_2$ , then  $T_1^n \simeq T_2^n$  for all  $n \geq 1$ .

**7.30 EXERCISE.** Assume that  $T_1 \simeq T_2$ . Prove that  $T_1$  is ergodic if and only if so is  $T_2$ .



## 8 Examples of Ergodic Maps

Here we show that many of the maps we have discussed so far are ergodic. Let us start with the circle rotation  $T(x) = x + a \pmod{1}$ , where  $x \in X = [0, 1)$ .

8.1 CLAIM. If  $a \in \mathbb{Q}$ , i.e.  $a$  is a rational number, then every point  $x \in X$  is periodic.

*Proof.* Let  $a = p/q$ . Then  $T^n(x) = x + np/q \pmod{1}$  for all  $n \geq 1$ , hence  $T^q(x) = x$ .  
□

Note that if  $p$  and  $q$  are relatively prime, then  $q$  is the minimal period of every point  $x \in X$ .

8.2 CLAIM. If  $a$  is irrational, then for every  $x \in X$  the trajectory  $\{T^n x\}$ ,  $n \geq 0$ , is dense in  $X$ .

*Proof.* First, if  $a$  is irrational, then no point  $x$  can be periodic. Indeed, if  $x$  is periodic with period  $n$ , then  $x = T^n(x) = x + na \pmod{1}$ , hence  $na \in \mathbb{Z}$ , so  $a$  is a rational number.

Therefore, all points  $\{T^n x\}$ ,  $n \geq 1$ , are distinct. Hence for any  $\varepsilon > 0$  there are two of those points, say  $T^m(x)$  and  $T^{m+k}(x)$ , which are  $\varepsilon$ -close to each other. Put  $y = T^m(x)$ , then  $T^k(y) = T^{m+k}(x)$ , so  $\text{dist}(y, T^k y) \leq \varepsilon$ . Since  $T$  preserves distances, we have

$$\text{dist}(y, T^k y) = \text{dist}(T^k y, T^{2k} y) = \text{dist}(T^{2k} y, T^{3k} y) = \dots$$

Since  $T$  also preserves orientation, then the sequence  $y, T^k y, T^{2k} y, T^{3k} y, \dots$  moves in one direction on the circle, and eventually will go around it and make an  $\varepsilon$ -dense subset of  $X$ . Since  $\varepsilon > 0$  is arbitrary, we get our claim. □

8.3 CLAIM. If  $a$  is rational, then  $T$  is not ergodic with respect to the Lebesgue measure  $m$  on  $X$ .

*Proof.* Let  $a = p/q$ . For any  $\varepsilon > 0$  the set

$$B_\varepsilon = \cup_{i=0}^{q-1} (i/q, i/q + \varepsilon)$$

is invariant. Its measure is  $m(B_\varepsilon) = \varepsilon q > 0$ . □

8.4 CLAIM. If  $a$  is irrational, then  $T$  is ergodic with respect to the Lebesgue measure  $m$  on  $X$ .

*Proof.* We will need one fact from real analysis. Let  $B \subset \mathbb{R}$  be a Borel set. A point  $x \in \mathbb{R}$  is called a **Lebesgue density point** (or just a **density point**) of the set  $B$  if

$$\lim_{\varepsilon \rightarrow 0} \frac{m(B \cap [x - \varepsilon, x + \varepsilon])}{2\varepsilon} = 1$$

The fact is that  $m$ -almost every point  $x \in B$  is a density point.

Now, assume that  $T$  is not ergodic and  $B \subset X$  is a  $T$ -invariant set, such that  $0 < m(B) < 1$ . Let  $x$  be a density point for  $B$  and  $y$  a density point for  $B^c$  (one exists since  $m(B^c) > 0$ ). Find such a small  $\varepsilon > 0$  that  $m(B \cap [x - \varepsilon, x + \varepsilon]) > 1.9\varepsilon$  and  $m(B^c \cap [y - \varepsilon, y + \varepsilon]) > 1.9\varepsilon$ . Now, since the orbit of  $x$  is dense in  $X$ , there is an  $n \geq 1$  such that  $\text{dist}(T^n x, y) < 0.1\varepsilon$ . Because the set  $B$  is  $T$ -invariant, we have  $m(B \cap [T^n x - \varepsilon, T^n x + \varepsilon]) > 1.9\varepsilon$ . This easily leads to a contradiction.  $\square$

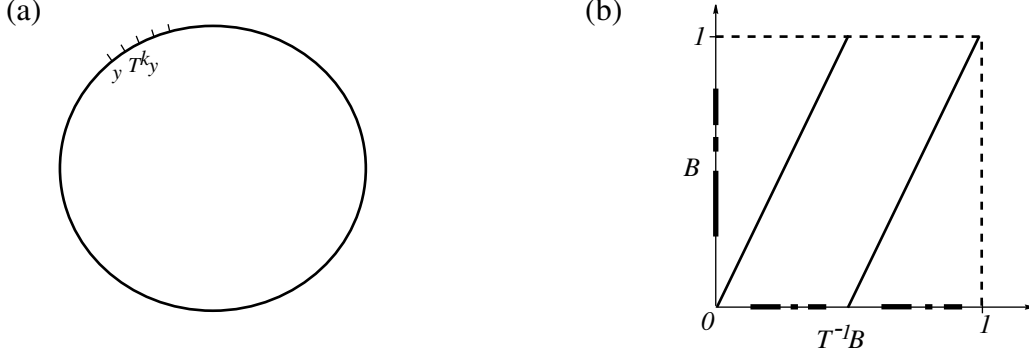


Figure 8: The sequence  $y, T^k y, T^{2k} y, \dots$  on the circle (a), and a set  $B$  with its preimage  $T^{-1}B$ , the latter consisting of two identical parts (b).

Next we turn to the doubling map  $T(x) = 2x \pmod{1}$ , where  $x \in X = [0, 1)$ .

**8.5 CLAIM.** The doubling map is ergodic with respect to the Lebesgue measure  $m$ .

*Proof.* Let  $B \subset X$  be a Borel set. Note that  $T^{-1}B$  consists of two identical copies of  $B$ , each is twice as small as the original  $B$ , one lies on the interval  $[0, 1/2)$  and the other on the interval  $[1/2, 1)$ . Hence,

$$m(T^{-1}B \cap [0, 1/2)) = m(T^{-1}B \cap [1/2, 1)) = m(B)/2$$

Applying the same argument to  $T^{-n}B$  we get the following. Let  $J_{i,n} = [i/2^n, (i+1)/2^n)$  for  $i = 0, 1, \dots, 2^n - 1$  (we call these **binary intervals**). Then

$$m(T^{-n}B \cap J_{i,n}) = m(B)/2^n = m(B) m(J_{i,n}) \quad (6)$$

for all  $i = 0, 1, \dots, 2^n - 1$ .

It is easy to see that the collection of all binary intervals  $\{J_{i,n}\}$ ,  $n \geq 1$ ,  $0 \leq i \leq 2^n - 1$ , generates the Borel  $\sigma$ -algebra on  $X$ . Finite unions of binary intervals make an algebra.

Let  $A \subset X$  be a Borel set. By the approximation theorem 1.19 for any  $\varepsilon > 0$  the set  $A$  can be approximated by a finite union of binary intervals, i.e. there is  $A_0 = \cup_j J_{i_j,n}$  so

that  $m(A \Delta A_0) < \varepsilon$ . Here  $n$  can be made the same for all  $j$ . Certainly,  $n$  depends on  $\varepsilon$ , so let us call it  $n_\varepsilon$ .

Now, (6) implies

$$m(T^{-n}B \cap A_0) = m(B) m(A_0)$$

for all  $n \geq n_\varepsilon$ . Since  $A_0$  approximates  $A$ , it is easy to derive that

$$|m(T^{-n}B \cap A) - m(B) m(A)| < 2\varepsilon$$

for all  $n \geq n_\varepsilon$ . Hence,

$$m(A \cap T^{-n}B) \rightarrow m(A) m(B) \quad \text{as } n \rightarrow \infty \quad (7)$$

Note: this is true for *any* (!) Borel sets  $A, B \subset X$ .

Now, if  $B$  is a  $T$ -invariant set, then  $T^{-n}B = B$  and (setting  $A = B$ ) we get  $m(B) = [m(B)]^2$ . This is only possible if  $m(B) = 0$  or  $m(B) = 1$ .  $\square$

**8.6 REMARK.** The same argument, without changes, applies to the tent map. Hence, the tent map is ergodic with respect to the Lebesgue measure.

Next, we take the quadratic map  $T(x) = 4x(1 - x)$  for  $x \in [0, 1]$ . This one has an absolutely continuous invariant measure (a.c.i.m.) with density given in 5.11. It is hard, in this case, to show the ergodicity directly, but there is a helpful trick. This map is shown on Fig. 4b and looks very much like the tent map shown on Fig. 4a (topologically, they are equivalent). Maybe they are isomorphic?

**8.7 CLAIM.** The tent map and the quadratic map are isomorphic.

*Proof.* Let  $T_1(x) = 2x$  for  $x \leq 1/2$  and  $T_1(x) = 2 - 2x$  for  $x > 1/2$  be the tent map. Let  $T_2(y) = 4y(1 - y)$  be the quadratic map. The isomorphism is established by the function

$$y = \varphi(x) = \frac{1 - \cos \pi x}{2}$$

It is a bijection of the unit interval  $[0, 1]$  onto itself. We need to verify the preservation of measures 7.28(i) and dynamics 7.28(ii).

We first check 7.28(i). Let  $x \in (0, 1)$  and  $y = \varphi(x)$ . Take a small interval  $(x, x + dx)$  and let  $y + dy = \varphi(x + dx)$ . The preservation of measures means

$$dx = f(y) dy + o(dy)$$

where  $f(y)$  is the density function given in 5.11. Dividing by  $dx$  and taking the limit  $dx \rightarrow 0$  gives

$$f(\varphi(x)) \varphi'(x) = 1$$

So, we need to verify this identity, which can be done by direct substitution.

Next, we show how to check 7.28(ii). Let  $x < 1/2$ . Then

$$\varphi(T_1(x)) = \varphi(2x) = \frac{1 - \cos 2\pi x}{2} = \sin^2 \pi x$$

On the other hand,

$$T_2(\varphi(x)) = 4\varphi(x)(1 - \varphi(x)) = 4 \times \frac{1 - \cos \pi x}{2} \times \frac{1 + \cos \pi x}{2} = \sin^2 \pi x$$

so we get  $\varphi \circ T_1 = T_2 \circ \varphi$ . The case  $x > 1/2$  is similar.  $\square$

**8.8 COROLLARY.** The quadratic map  $T(x) = 4x(1 - x)$  is ergodic with respect to the a.c.i.m. with density given in 5.11.

Lastly, we turn to the baker's transformation of the unit square.

**8.9 CLAIM.** The baker's map is ergodic with respect to the Lebesgue measure.

*Proof.* After you grasp the proof of 8.5, this should be pretty clear. Our main tool will be binary rectangles, rather than binary intervals. A **binary rectangle** is

$$R_{i,j,m,n} = \{(x, y) : i/2^m \leq x < (i+1)/2^m, j/2^n \leq y < (j+1)/2^n\}$$

It is easy to see that the collection of all binary rectangles generates the Borel  $\sigma$ -algebra on  $X$ . Also, finite unions of binary rectangles make an algebra.

Consider an arbitrary binary rectangle  $R_{i,j,m,n}$ . Note that  $T(R_{i,j,m,n}) = R_{s,t,m-1,n+1}$  for some  $s, t$ . Therefore,  $T^m(R_{i,j,m,n}) = R_{u,v,0,n+m}$  for some  $u, v$ . This last set is a rectangle, which stretches across  $X$  all the way in the  $x$  direction (from  $x = 0$  to  $x = 1$ ). Similarly,  $T^{-n}(R_{i,j,m,n}) = R_{e,f,m+n,0}$  is a rectangle stretching across  $X$  all the way in the  $y$  direction (from  $y = 0$  to  $y = 1$ ).

Now, let  $A$  and  $B$  be two Borel subsets of  $X$ . By the approximation theorem 1.19, for any  $\varepsilon > 0$  there are sets  $A_0$  and  $B_0$ , each being a finite union of some binary rectangles, such that  $m(A \Delta A_0) < \varepsilon$  and  $m(B \Delta B_0) < \varepsilon$ .

Let  $A_0 = \cup_{p,q} R_{i_p,j_q,m,n}$ , where  $m$  and  $n$  can be made the same for all  $(p, q)$ . Then  $T^m(A_0)$  is a union of binary rectangles stretched across  $X$  all the way in the  $x$  direction. The same is obviously true for  $T^k(A_0)$  whenever  $k \geq m$ .

Now let  $B_0 = \cup_{r,s} R_{i_r,j_s,m,n}$ , where  $m$  and  $n$  can be made the same for all  $(r, s)$  (and the same as above). Then  $T^{-n}(B_0)$  is a union of binary rectangles stretched across  $X$  all the way in the  $y$  direction. The same is obviously true for  $T^{-\ell}(B_0)$  whenever  $\ell \geq n$ .

The above observations imply

$$m(T^k A_0 \cap T^{-\ell} B_0) = m(T^k A_0) m(T^{-\ell} B_0)$$

for all  $k \geq m$  and  $\ell \geq n$ . Since the Lebesgue measure  $m$  is  $T$ -invariant and  $T$  is an automorphism,

$$m(A_0 \cap T^{-N} B_0) = m(A_0) m(B_0)$$

for all  $N \geq m + n$ .

Next, because  $A_0$  approximates  $A$  and  $B_0$  approximates  $B$ , it is easy to derive that

$$|m(A \cap T^{-N}B) - m(A)m(B)| < 4\epsilon$$

for all  $N \geq m + n$ . Hence,

$$m(A \cap T^{-N}B) \rightarrow m(A)m(B) \quad \text{as } N \rightarrow \infty \quad (8)$$

Note: this is true for *any* (!) Borel sets  $A, B \subset X$ .

Now, if  $B$  is a  $T$ -invariant set, then  $T^{-N}B = B$  and (setting  $A = B$ ) we get  $m(B) = [m(B)]^2$ . This is only possible if  $m(B) = 0$  or  $m(B) = 1$ .  $\square$

Finally, we discuss the uniqueness of the absolutely continuous invariant measures (a.c.i.m.'s) constructed in Section 5.

**8.10 THEOREM.** Let  $T : X \rightarrow X$  be a map of an interval  $X \subset \mathbb{R}$  that has an ergodic a.c.i.m.  $\mu$  with a positive density  $f(x) > 0$ . Then that a.c.i.m. is unique.

*Proof.* The assumption  $f(x) > 0$  implies that  $\mu$  is equivalent to the Lebesgue measure  $m$ , i.e.  $\mu(B) = 0$  iff  $m(B) = 0$ . Since  $\mu$  is ergodic, then for each  $T$ -invariant set  $B$  we have  $m(B) = 0$  or  $m(B^c) = 0$ .

If there were another a.c.i.m.  $\nu$  with density  $g(x)$ , then for any  $T$ -invariant set  $B$  we would have either  $\nu(B) = \int_B g dm = 0$  or  $\nu(B^c) = \int_{B^c} g dm = 0$ . Hence, the measure with density  $g$  would be ergodic, too. On the other hand, distinct ergodic measures are mutually singular by 7.18, a contradiction.  $\square$

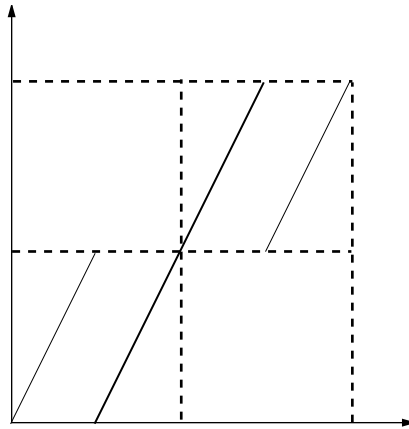


Figure 9: The function in Exercise 8.13.

8.11 REMARK. The last theorem (with the same proof) extends to maps on a unit square or, more generally, any domain of  $\mathbb{R}^d$  for any  $d \geq 1$ .

8.12 COROLLARY. The doubling map, the tent map, the quadratic map and the baker's transformation have unique a.c.i.m.'s.

8.13 EXERCISE. Let  $T : [0, 1) \rightarrow [0, 1)$  be defined by

$$T(x) = \begin{cases} 2x & \text{if } x < 1/4 \\ 2x - 1/2 & \text{if } 1/4 \leq x < 3/4 \\ 2x - 1 & \text{if } 3/4 < x \end{cases}$$

Find two distinct a.c.i.m.'s on  $X = [0, 1)$ . Are both ergodic? Can you find an ergodic a.c.i.m.? How many? Can you find an a.c.i.m. with a strictly positive density?

8.14 "EXERCISE". Let  $T : X \rightarrow X$  be a map of a unit interval  $X \subset \mathbb{R}$  that has an ergodic a.c.i.m.  $\mu$  with a positive density  $f(x) > 0$ . Assume, additionally, that

$$0 < C_1 \leq f(x) \leq C_2 < \infty \quad \forall x \in X$$

with some positive constants  $C_1 < C_2$ . Consider the sequence of measures

$$\nu_n = (m + Tm + \cdots + T^{n-1}m)/n$$

where  $m$  is the Lebesgue measure on  $X$ . Show that  $\nu_n$  weakly converges to  $\mu$ .

Note: this exercise requires substantial work, it is not recommended as a homework problem. It might be a large project for a student.

Hints: show that for every  $k \geq 1$  the measure  $T^k m$  is absolutely continuous and its density  $g_k(x)$  is bounded by  $C_1/C_2 \leq g_k \leq C_2/C_1$ . Then show that for every  $n \geq 1$  the measure  $\nu_n$  has the same property. Then show that every limit point of the sequence  $\nu_n$  has the same property. Next, show that every limit point of the sequence  $\nu_n$  is an invariant measure: to do that, prove that if a subsequence  $\nu_{n_k}$  weakly converges to a measure  $\nu$ , then  $\nu_{n_k}(A) \rightarrow \nu(A)$  for *every* Borel set  $A \subset X$  (this can be done via approximating  $A$  by finite unions of intervals  $A_0$  such that  $m(A \Delta A_0) < \varepsilon$ ). Lastly, use Theorem 8.10.

Note: it is not necessarily true that the sequence  $T^k m$  converges to any measure as  $k \rightarrow \infty$ .

## 9 Ergodic Theorem

This section presents the single most important result in the theory of dynamical systems. It is the ergodic theorem, the main fact in our course.

We again assume that  $T : X \rightarrow X$  is a transformation preserving a measure  $\mu$ .

**9.1 DEFINITION.** Let  $f \in L^0(X)$  be a measurable function. Then for every  $x \in X$  and  $n \geq 1$

$$S_n(x) = f(x) + f(Tx) + \cdots + f(T^{n-1}x)$$

is called a **partial sum** or **ergodic sum**. It is obtained from the time series  $\{f(T^i x)\}$ ,  $i \geq 0$ . The value  $S_n(x)/n$  is the **(partial) time average** of the function  $f$  at the point  $x$ . The limit

$$f_+(x) = \lim_{n \rightarrow +\infty} \frac{1}{n} S_n(x) \quad (9)$$

(if it exists) is called the **(asymptotic) time average** of the function  $f$  along the orbit of  $x$ .

**9.2 THEOREM (BIRKHOFF-KHINCHIN ERGODIC THEOREM).** Let  $f$  be integrable, i.e.  $f \in L^1_\mu(X)$ . Then

- (a) For almost every point  $x \in X$  the limit  $f_+(x)$  defined by (9) does exist.
- (b) The function  $f_+(x)$  is  $T$ -invariant. Moreover, if  $f_+(x)$  exists, then  $f_+(T^n x)$  exists for all  $n$  and  $f_+(T^n x) = f_+(x)$ .
- (c)  $f_+$  is integrable and

$$\int_X f_+ d\mu = \int_X f d\mu$$

- (d) If  $\mu$  is ergodic, then  $f_+(x)$  is constant almost everywhere and

$$f_+(x) = \int_X f d\mu \quad \text{for a.e. } x \in X$$

We postpone the proof for a short while. First, we make some remarks and derive some corollaries.

**9.3 REMARK.** The integral  $\int_X f d\mu$  is called the **space average** of the function  $f$ . The part (d) of the ergodic theorem asserts that

If  $T$  is ergodic, then the time averages are equal to the space average (almost everywhere)

In probability theory, the identity between time and space averages is known as the **strong law of large numbers**. It is quite common in statistical physics to replace time averages with space averages, this practice goes back to L. Boltzmann and others in the Nineteenth Century. This is how the idea of ergodicity was born.

9.4 COROLLARY. Let  $T : X \rightarrow X$  be an automorphism. Then the “past” time average

$$f_-(x) = \lim_{n \rightarrow \infty} \frac{f(x) + f(T^{-1}x) + \cdots + f(T^{-(n-1)}x)}{n} \quad (10)$$

exists almost everywhere, and  $f_+(x) = f_-(x)$  for a.e.  $x \in X$ .

*Proof.* The existence follows from Ergodic Theorem 9.2 applied to  $T^{-1}$ , it also implies that  $\int f_- d\mu = \int f d\mu$ . To show that  $f_+ = f_-$ , consider the  $T$ -invariant set  $A = \{x : f_+(x) > f_-(x)\}$ . If  $\mu(A) > 0$ , then apply Ergodic Theorem to the restriction of  $T$  to  $A$  preserving the conditional measure  $\mu_A$ . This yields  $\int_A f_+ d\mu = \int_A f d\mu = \int_A f_- d\mu$ , a contradiction. Similarly one can show that the set  $\{x : f_+(x) < f_-(x)\}$  has zero measure.  $\square$

9.5 DEFINITION. Let  $A \subset X$ . The limit

$$r_A(x) = \lim_{n \rightarrow +\infty} \frac{\#\{0 \leq i \leq n-1 : T^i(x) \in A\}}{n} \quad (11)$$

(if one exists) is called the **(asymptotic) frequency** of visits of the point  $x$  to the set  $A$ .

9.6 COROLLARY. For every set  $A$  the limit  $r_A(x)$  defined by (11) exists for almost every point  $x \in X$ . If  $\mu$  is ergodic, then  $r_A(x) = \mu(A)$  for a.e.  $x \in X$ .

Hence, if  $\mu$  is ergodic, then the orbit of a.e. point  $x$  spends time in the set  $A$  proportional to its measure  $\mu(A)$ . In this sense, the ergodic measure  $\mu$  describes the asymptotic distribution of almost every orbit  $\{T^n(x)\}$ ,  $n \geq 0$  in the space  $X$ .

9.7 PROOF OF ERGODIC THEOREM 9.2. This is done in three major steps. Step 1 consists in proving a lemma:

9.8 LEMMA (MAXIMAL ERGODIC THEOREM). For every  $N \geq 0$ , define a function

$$F_N = \max\{S_0, S_1, \dots, S_N\}$$

where we set  $S_0 = 0$ . Let  $A_N = \{x : F_N(x) > 0\}$  and  $A = \cup_{N \geq 1} A_N$ . Then

$$\int_{A_N} f d\mu \geq 0 \quad \text{and} \quad \int_A f d\mu \geq 0 \quad (12)$$



*Proof.* See Walters, pp. 37-38. We just sketch the proof here. First we prove that  $\int_{A_N} f d\mu \geq 0$ . For  $0 \leq n \leq N$  we have  $F_N \geq S_n$ , hence  $F_N \circ T \geq S_n \circ T$ , and hence  $F_N \circ T + f \geq S_{n+1}$ . Therefore, for all  $x \in A_N$  we have

$$F_N(T(x)) + f(x) \geq \max_{1 \leq n \leq N} \{S_n(x)\} = F_N(x)$$

(since  $S_0(x) = 0$  and  $F_N(x) > 0$ ). Thus,  $f(x) \geq F_N(x) - F_N(T(x))$  for  $x \in A_N$ . Note also that  $F_N(y) = 0$  and  $F_N(T(y)) \geq 0$  for all  $y \in A_N^c$ . Hence,

$$\int_{A_N} f d\mu \geq \int_{A_N} F_N d\mu - \int_{A_N} F_N \circ T d\mu \geq \int_X F_N d\mu - \int_X F_N \circ T d\mu = 0$$

the last equation is due to the invariance of  $\mu$ .

Next, since  $F_N \leq F_{N+1}$ , then  $A_N \subset A_{N+1}$  for all  $N$ . Now

$$\int_A f d\mu = \lim_{N \rightarrow \infty} \int_{A_N} f d\mu \geq 0 \quad \square$$

Step 2 consists in using Lemma 9.8 to prove the clause (a) of Theorem 9.2. See Walters, p. 38. We sketch the argument here. Let

$$\bar{f}(x) = \limsup_{n \rightarrow +\infty} \frac{1}{n} S_n(x)$$

and

$$\underline{f}(x) = \liminf_{n \rightarrow +\infty} \frac{1}{n} S_n(x)$$

It is enough to show that  $\bar{f} = \underline{f}$  a.e. If this is not the case, then there are real numbers  $\alpha > \beta$  such that the set

$$E = E_{\alpha, \beta} = \{x : \bar{f}(x) > \alpha \text{ and } \underline{f}(x) < \beta\}$$

has positive measure, i.e.  $\mu(E) > 0$ . Note that the functions  $\bar{f}$  and  $\underline{f}$  are invariant, hence  $E$  is a fully  $T$ -invariant set.

Consider a function  $g = (f - \alpha) \chi_E$ , where  $\chi_E$  is the indicator of  $E$ . Then for all  $x \in E$  we have

$$\sup_{N \geq 1} (g(x) + g(Tx) + \cdots + g(T^{N-1}x)) > 0$$

and  $g \equiv 0$  on  $E^c$ . Applying (12) to the function  $g$  we get  $\int_E g d\mu \geq 0$ , hence

$$\int_E f d\mu \geq \alpha \mu(E) \tag{13}$$

Similarly, we can show that

$$\int_E f d\mu \leq \beta \mu(E) \tag{14}$$

(by applying the previous argument to the function  $-f$ ). But then (13) and (14) imply  $\mu(E) = 0$ .

Step 3 consists of deriving the clauses (b), (c) and (d) from (a). See Walters, p. 39. The clause (b) is trivial. The clause (d) follows by 7.23. The integrability of  $f_+$  in the clause (c) follows from Fatou's lemma, which in fact gives

$$\|f_+\|_1 \leq \|f\|_1 \quad (15)$$

It remains to prove the integral identity in (c). Our argument is different from that of Walters. First, observe that  $\int_X S_n/n d\mu = \int_X f d\mu$  for all  $n$ , because  $\mu$  is invariant. If  $f$  is bounded, then  $\|S_n/n\|_\infty \leq \|f\|_\infty$ , hence the sequence  $S_n/n$  is uniformly bounded. In this case the integral identity in (c) follows from the dominated convergence theorem.

For an arbitrary  $f \in L^1_\mu(X)$ , we take  $\varepsilon > 0$  and approximate  $f$  with a bounded function  $\phi$  so that  $\|f - \phi\|_1 < \varepsilon$ . Since  $\phi$  is bounded, we have  $\int \phi d\mu = \int \phi_+ d\mu$ . Lastly,

$$\|f_+ - \phi_+\|_1 = \|(f - \phi)_+\|_1 \leq \|f - \phi\|_1 < \varepsilon$$

where we applied (15) to the function  $f - \phi$ . This implies  $|\int f d\mu - \int f_+ d\mu| < 2\varepsilon$ . Ergodic Theorem 9.2 is now proved.  $\square$

**9.9 COROLLARY ( $L^p$  ERGODIC THEOREM OF VON NEUMANN).** Let  $1 \leq p < \infty$ . If  $f \in L^p_\mu(X)$ , then  $\|S_n/n - f_+\|_p \rightarrow 0$  as  $n \rightarrow \infty$ .

*Proof.* See Walters, p. 36. By the way, it is quite similar to our proof of the clause (c) of Theorem 9.2.  $\square$

**9.10 EXERCISE.** Let  $A \subset X$  and  $\mu(A) > 0$ . Prove that  $r_A(x) > 0$  (defined in 9.5) for almost every point  $x \in A$ . [Hint: consider the set  $B = \{x \in A : r_A(x) = 0\}$ .]

**9.11 EXAMPLE.** There is an interesting application of the ergodic theorem to number theory. Recall Example 2.27, which involves the map  $T(x) = 10x \pmod{1}$  on the unit interval  $X = [0, 1)$ . This map preserves the Lebesgue measure  $m$ , as shown in 2.27. One can also show that  $m$  is ergodic by using the same argument as in 8.5, we omit details.

Consider the set  $A_r = [r/10, (r+1)/10)$  for some  $r = 0, 1, \dots, 9$ . For  $x \in X$ , the inclusion  $T^n(x) \in A_r$  means that the  $n$ -th digit in the decimal representation of  $x$  is  $r$ , see 2.27. For  $n \geq 1$ , let  $K_r(n, x)$  be the number of occurrences of the digit  $r$  among the first  $n$  digits of the decimal representation of  $x$ . This is exactly  $\#\{0 \leq i \leq n-1 : T^i(x) \in A_r\}$ . Corollary 9.6 now implies that

$$\lim_{n \rightarrow \infty} K_r(n, x)/n = m(A_r) = 0.1$$

for almost every  $x \in X$ .

The above fact is known in number theory. A number  $x \in [0, 1)$  is called **normal** if for every  $r = 0, 1, \dots, 9$  the asymptotic frequency of occurrences of the digit  $r$  in the decimal representation of  $x$  is exactly 0.1. The fact, which we just proved, that almost every point  $x \in X$  is normal is known as **Borel Theorem on Normal Numbers**.

Next we derive some further consequences of Ergodic Theorem 9.2 along the lines of 9.5 and 9.6. We need to assume that  $X$  is a compact metrisable topological space, but the map  $T$  does not have to be continuous. For  $x \in X$ , consider the sequence of uniform atomic measures (recall Definition 3.13)

$$\mu_x^{(n)} = (\delta_x + \delta_{Tx} + \dots + \delta_{T^{n-1}x})/n \quad (16)$$

As  $n \rightarrow \infty$ , the measure  $\mu_x^{(n)}$  may converge to a probability measure  $\mu \in \mathcal{M}(X)$  in the weak\* topology.

**9.12 DEFINITION.** A point  $x \in X$  is said to be  **$\mu$ -generic** for a measure  $\mu \in \mathcal{M}(X)$  if the sequence  $\mu_x^{(n)}$  defined by (16) weakly converges to  $\mu$  as  $n \rightarrow \infty$ . Equivalently,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} F(T^i(x)) = \int_X F d\mu \quad \forall F \in C(X)$$

i.e. the *time averages* are equal to the *space averages* for all continuous functions.

If a point  $x$  is  $\mu$ -generic, then the trajectory  $\{T^n x\}$  of  $x$  is distributed in the space  $X$  according to the measure  $\mu$  (one can say that  $x$  is “attracted” by the measure  $\mu$ ).

**9.13 DEFINITION.** Let  $T : X \rightarrow X$  be a map and  $\mu \in \mathcal{M}(X)$ . The set

$$B_\mu = \{x : x \text{ is } \mu\text{-generic}\}$$

is called the **basin of attraction** of the measure  $\mu$ .

Note: if  $\mu \neq \nu$  are two distinct measures, then  $B_\mu \cap B_\nu = \emptyset$ .

**9.14 PROPOSITION.** If  $\mu$  is an ergodic measure, then  $\mu$ -almost every point  $x \in X$  is  $\mu$ -generic, i.e.  $\mu(B_\mu) = 1$ .

*Proof.* Let  $\mathcal{J}$  be a countable basis in the topology on  $X$  and  $\mathcal{A}(\mathcal{J}) = \{A_k\}$  a countable algebra of  $X$  generated by  $\mathcal{J}$ , see 1.17. For every  $k$  we have  $\mu_x^{(n)}(A_k) \rightarrow \mu(A_k)$  as  $n \rightarrow \infty$  a.e. by 9.6, i.e. we have this convergence for all  $x \in X_k$  with  $\mu(X_k) = 1$ . Let  $X_\infty = \bigcap_k X_k$ . Obviously,  $\mu(X_\infty) = 1$ . Since every open set  $U \subset X$  is a union of some disjoint elements of  $\mathcal{A}(\mathcal{J})$ , one can easily derive that  $\liminf_n \mu_x^{(n)}(U) \geq \mu(U)$  for every  $x \in X_\infty$ . Therefore,  $\mu_x^{(n)}$  weakly converges to  $\mu$  by 3.11(iii).  $\square$

**9.15 APPLICATIONS IN PHYSICS.** Suppose a map  $T : X \rightarrow X$  models a physical process. In this case, usually,  $X$  is a compact topological space with some natural coordinates on it (examples: a compact domain in  $\mathbb{R}^d$ , a sphere, a torus, etc.). The coordinates allow us to define the Lebesgue measure  $m$  on  $X$ . It measures area or volume in  $X$ , depending on the dimension of  $X$ . Let  $m$  be normalized so that  $m(X) = 1$ .

A typical physical experiment (or a numerical test done with the aid of a computer) consists of choosing a point  $x \in X$  *at random* and experimentally following (or numerically generating) its trajectory  $x, T(x), \dots, T^{n-1}(x)$  until some large time  $n$ . The points  $\{x, T(x), \dots, T^{n-1}(x)\}$  represent the measure  $\mu_x^{(n)}$  defined by (16).

Proposition 9.14 shows that if  $\mu$  is an ergodic  $T$ -invariant measure and the point  $x \in X$  is typical with respect to  $\mu$ , then the measure  $\mu^{(n)}$  weakly converges to  $\mu$  as  $n \rightarrow \infty$ , i.e. the measure  $\mu$  describes the distribution of typical orbits in the space  $X$ .

However, in practice one may NOT want (or may NOT be able) to choose a point  $x$  typical with respect to some ergodic measure  $\mu$ . Why should those points be physically interesting? Physicists may not even have any ergodic measure at hands! What they want is to choose a point  $x$  typical with respect to the Lebesgue measure  $m$  on  $X$ . It is a fundamental principle in statistical physics that only such points are physically relevant (or experimentally observable). Such points are also easy to generate by computer programs (using so called random number generators). This motivates the following definition:

**9.16 DEFINITION.** Let  $X$  be a compact space with natural coordinates and a (normalized) Lebesgue measure  $m$ . A  $T$ -invariant measure  $\mu$  is said to be **physically observable** if  $m(B_\mu) > 0$ . Such measures are also referred to as **Sinai-Bowen-Ruelle (SRB) measures** in the modern theory of dynamical systems.

A measure  $\mu$  is physically observable if there is a chance to “observe”  $\mu$  by following a trajectory chosen at random with respect to the Lebesgue measure  $m$ , i.e. observe  $\mu$  in a physical experiment or by a computer simulation.

**9.17 PROPOSITION.** Let  $T : X \rightarrow X$  be the irrational circle rotation, or the doubling map, or the tent map, or the quadratic map, or the baker’s transformation, and  $\mu$  the absolutely continuous invariant measure on  $X$ . Then  $\mu$  is physically observable. Moreover,  $m(B_\mu) = 1$ , i.e. the a.c.i.m.  $\mu$  is the *only* physically observable measure.

*Proof.* Since  $\mu$  is ergodic, then  $\mu(B_\mu) = 1$  by 9.14. Also,  $\mu$  is equivalent to the Lebesgue measure  $m$ , hence  $m(B_\mu) = 1$ .  $\square$

**9.18 EXERCISE.** Let  $T(x) = x^2$  for  $x \in X = [0, 1]$ . Which invariant measures are physically observable? Recall Exercise 2.26(e).

**9.19 EXERCISE (OPTIONAL; IT IS RATHER TRICKY).** Let  $T(x) = x + a \pmod{1}$  for

$X = [0, 1)$  be a circle rotation with a rational  $a \in \mathbb{Q}$ . Are there any physically observable measures?

## 10 Symbolic Dynamics

Recall that in Example 2.28 we represented every point  $x \in X = [0, 1]$  by a sequence of binary digits  $i_0 i_1 i_2 \dots$  and the doubling map  $T(x) = 2x \pmod{1}$  then corresponded to the left shift of the sequence (with the removal of the first digit  $i_0$ ).

In Example 5.5 we represented every point  $(x, y) \in X = [0, 1] \times [0, 1]$  of the unit square by a double infinite sequence of binary digits  $\{\omega_n\}_{n=-\infty}^{\infty}$  and then the baker's map  $T : X \rightarrow X$  corresponded to the left shift of the entire sequence.

Here we generalize these examples.

**10.1 DEFINITIONS.** Let  $S$  be a finite set of  $r \geq 2$  elements. We call  $S$  an **alphabet** and elements  $s \in S$  **letters**. We label the letters by  $1, 2, \dots, r$ , i.e. we assume  $S = \{1, \dots, r\}$ . Let  $\Omega_+ = \Omega_{+,r} = S^{\mathbb{Z}_+}$  denote the space of infinite sequences of letters. So a point  $\underline{\omega} \in \Omega_+$  is a sequence  $\underline{\omega} = \{\omega_n\}_{n=0}^{\infty}$  with  $\omega_n \in S$  for each  $n \geq 0$ . Also, let  $\Omega = \Omega_r = S^{\mathbb{Z}}$  be the space of double infinite sequences of elements of  $S$ , i.e.  $\Omega$  consists of sequences  $\underline{\omega} = \{\omega_n\}_{n=-\infty}^{\infty}$  with  $\omega_n \in S$  for each  $n \in \mathbb{Z}$ . We call  $\Omega_+$  and  $\Omega$  **symbolic spaces**. (For brevity, we suppress the index  $r$ .)

We equip the finite set  $S$  with the discrete topology (i.e., all its subsets are open). We consider the **product topology** on  $\Omega_+$  and  $\Omega$ . The product topology is generated by cylinders, see below. Cylinders make a countable basis in the product topology and play the same important role as intervals in  $\mathbb{R}$ .

Let  $0 \leq m \leq n < \infty$  and  $\omega'_i \in S$  for  $m \leq i \leq n$ . A **cylinder** is the set

$$C_{m,n}(\omega'_m, \dots, \omega'_n) = \{\underline{\omega} \in \Omega_+ : \omega_i = \omega'_i \text{ for all } m \leq i \leq n\}$$

In other words, the cylinder  $C_{m,n}(\omega'_m, \dots, \omega'_n)$  consists of all sequences whose “coordinates” from  $m$  to  $n$  are fixed (equal to the given letters  $\omega'_m, \dots, \omega'_n$ ). Similarly, a cylinder  $C_{m,n} \subset \Omega$  can be defined for any  $-\infty < m \leq n < \infty$ .

**10.2 EXERCISE.** Let  $C_{m,n}$  and  $C_{m',n'}$  be two cylinders.

- (a) Show that  $C_{m,n} \cap C_{m',n'}$  is a cylinder or a finite union of disjoint cylinders. Describe it.
- (b) Assume that  $C_{m,n} \subset C_{m',n'}$  (note that this implies  $m \leq m'$  and  $n' \leq n$ ). Show that  $C_{m',n'}$  is a finite union of “smaller” cylinders, one of which is  $C_{m,n}$ .
- (c) Show that the complement  $C_{m,n}^c$  to a cylinder is a finite union of cylinders.

Note: It follows from (a)-(b) that if  $A$  is a finite union of cylinders, then  $A$  is a finite union of some *disjoint* cylinders.

**10.3 COROLLARY.** Finite disjoint unions of cylinders make an algebra. This algebra generates the Borel  $\sigma$ -algebra on  $\Omega$  and  $\Omega_+$ .

#### 10.4 REMARKS.

- (a) In the product topology, the convergence  $\underline{\omega}^{(k)} \rightarrow \underline{\omega}$  is equivalent to the following: for any  $n \geq 1$  there is a  $k_n \geq 1$  such that for all  $k \geq k_n$

$$\omega_i^{(k)} = \omega_i \quad \text{for all } |i| \leq n$$

i.e. the variable sequence  $\underline{\omega}^{(k)}$  stabilizes as  $k \rightarrow \infty$ .

- (b) The product topology is metrisable. The corresponding metric on  $\Omega$  can be defined by

$$\text{dist}(\underline{\omega}, \underline{\omega}') = \sum_{n=-\infty}^{\infty} \frac{1 - \delta_{\omega_n \omega'_n}}{2^{|n|}}$$

(in the space  $\Omega_+$ , we only need to sum over  $n \geq 0$ ).

- (c) In the product topology, both spaces  $\Omega$  and  $\Omega_+$  are compact and totally disconnected.

**10.5 DEFINITION.** Let  $\mu_0$  be a probability measure on the finite set  $S$ . It is characterized by  $r$  numbers  $p_i = \mu_0(\{i\})$  such that  $p_i \geq 0$  and  $\sum_{i=1}^r p_i = 1$ .

Then  $\mu_0$  induces the **product measure**  $\mu$  on  $\Omega$  (and on  $\Omega_+$ ). For any cylinder  $C_{m,n}(\omega'_m, \dots, \omega'_n)$  its measure is given by

$$\mu(C_{m,n}(\omega'_m, \dots, \omega'_n)) = \prod_{i=m}^n p_{\omega'_i} \quad (17)$$

The measure  $\mu$  is also called the **Bernoulli measure** on  $\Omega$  (resp., on  $\Omega_+$ ).

#### 10.6 REMARKS.

- (a) The existence and uniqueness of the measure  $\mu$  satisfying (17) follows from Extension Theorem 1.21. The proof involves the verification of countable additivity, which is quite tedious.
- (b) The measure  $\mu$  is nonatomic (has no atoms), unless  $p_i = 1$  for some  $i$ , in which case  $\mu$  is concentrated on one sequence  $\{\omega_n\}$ , for which  $\omega_n = i \forall n$ .
- (c) The product measure  $\mu$  makes the coordinates  $\omega_n$ ,  $n \in \mathbb{Z}$ , independent random variables, in terms of probability theory. This explains the name *Bernoulli*.

**10.7 DEFINITION.** The **left shift** map  $\sigma$  can be defined on the spaces  $\Omega$  and  $\Omega_+$ . For every  $\underline{\omega} \in \Omega$  we define  $\underline{\omega}' = \sigma(\underline{\omega})$  by  $\omega'_i = \omega_{i+1}$  for all  $i \in \mathbb{Z}$ . For every  $\underline{\omega} \in \Omega_+$  we define  $\underline{\omega}' = \sigma(\underline{\omega})$  by  $\omega'_i = \omega_{i+1}$  for all  $i \geq 0$ .

10.8 EXERCISE. Show that the left shift map  $\sigma$  is “onto” and continuous. On the space  $\Omega$  it is a homeomorphism, and on  $\Omega_+$  it is an  $r$ -to-one map.

10.9 EXERCISE. Show that the left shift  $\sigma$  preserves the Bernoulli measure  $\mu$  defined by 10.5. Hint: take a cylinder  $C_{m,n}$  and describe its image  $\sigma(C_{m,n})$ . Then use the algebra made by finite disjoint unions of cylinders.

10.10 DEFINITION. The symbolic space  $\Omega$  (or  $\Omega_+$ ) with a Bernoulli measure  $\mu$  defined by 10.5 and the left shift map  $\sigma$  is called a **Bernoulli system** (or a **Bernoulli shift**). We denote it by  $B_r(p_1, \dots, p_r)$  (resp.,  $B_{+,r}(p_1, \dots, p_r)$ ). Note: its only parameters are  $r$  and  $p_1, \dots, p_r$ .

10.11 PROPOSITION. The doubling map  $T(x) = 2x \pmod{1}$  with the Lebesgue measure is isomorphic to the (one-sided) Bernoulli shift  $B_{+,2}(1/2, 1/2)$ .

*Proof.* Let  $B' \subset [0, 1)$  be the set of binary rational numbers, i.e.

$$B' = \{k/2^n : n \geq 0, 0 \leq k < 2^n\}$$

Clearly,  $B'$  is a countable set, so  $m(B') = 0$ . Let  $r = 2$  and  $\Omega'_+ \subset \Omega_+$  be the set of eventually constant sequences, i.e.

$$\Omega'_+ = \{\underline{\omega} : \omega_i = \omega_{i+1} \ \forall i > i_0\}$$

Clearly,  $\Omega'_+$  is countable set, so  $\mu(\Omega'_+) = 0$ .

Now, for each  $x \in [0, 1) \setminus B'$  the binary representation  $x = 0.i_0i_1i_2\dots$  constructed in 2.28 is unique. We define a sequence  $\underline{\omega} = \phi(x) \in \Omega_+$  by  $\omega_n = i_n + 1$  for all  $n \geq 0$ . This defines a bijection between  $[0, 1) \setminus B'$  and  $\Omega_+ \setminus \Omega'_+$ . One can check by direct inspection that  $\phi$  preserves the measure and dynamics, i.e. it is an isomorphism.  $\square$

10.12 PROPOSITION. The baker’s map defined by 5.1 with the Lebesgue measure  $m$  is isomorphic to the (two-sided) Bernoulli shift  $B_2(1/2, 1/2)$ .

*Proof.* Similar to the previous one.

10.13 EXERCISE. Let  $C$  and  $C'$  be two cylinders and  $\mu$  a Bernoulli measure. Show that there is an  $n_0 \geq 0$  such that

$$\mu(C \cap \sigma^{-n}(C')) = \mu(C) \mu(C')$$

for all  $n \geq n_0$ . Note: this applies to both  $\Omega$  and  $\Omega_+$ .

10.14 THEOREM. Every Bernoulli shift is ergodic.



*Proof.* Let  $A$  and  $B$  be two Borel subsets of  $\Omega$  (or  $\Omega_+$ ). By the approximation theorem 1.19, for any  $\varepsilon > 0$  there are sets  $A_0$  and  $B_0$ , each being a finite disjoint union of some cylinders, such that  $m(A \Delta A_0) < \varepsilon$  and  $m(B \Delta B_0) < \varepsilon$ .

The result of Exercise 10.13 implies that there is an  $n_0 \geq 0$  such that

$$\mu(A_0 \cap \sigma^{-n}(B_0)) = \mu(A_0) \mu(B_0)$$

for all  $n \geq n_0$ . Since  $A_0$  approximates  $A$  and  $B_0$  approximates  $B$ , it is easy to derive that

$$|\mu(A \cap \sigma^{-n}(B)) - \mu(A) \mu(B)| < 4\varepsilon$$

for all  $n \geq n_0$ . Hence,

$$\mu(A \cap \sigma^{-n}(B)) \rightarrow \mu(A) \mu(B) \quad \text{as } n \rightarrow \infty \quad (18)$$

Now, if  $B$  is a  $\sigma$ -invariant set, then  $\sigma^{-n}(B) = B$  and (setting  $A = B$ ) we get  $\mu(B) = [\mu(B)]^2$ . This is only possible if  $\mu(B) = 0$  or  $\mu(B) = 1$ .  $\square$

Whenever we establish an isomorphism between a given dynamical system  $(X, T, \mu)$  and a symbolic system  $(\Omega, \sigma, \nu)$  or  $(\Omega_+, \sigma, \nu)$  with some  $\sigma$ -invariant measure  $\nu$ , we call this a **symbolic representation** of  $(X, T, \mu)$ . We now outline a standard method of constructing symbolic representations.

**10.15 DEFINITION.** Let  $X = X_1 \cup \dots \cup X_r$  be a finite partition of  $X$  into disjoint parts,  $X_i \cap X_j = \emptyset$  for  $i \neq j$ . Let  $T : X \rightarrow X$  be a map. For every point  $x \in X$  its **itinerary** is a sequence defined by

$$\{\omega_n\}_{n=0}^{\infty} : \quad T^n(x) \in X_{\omega_n} \quad \forall n \geq 0$$

If the map  $T$  is a bijection, i.e.  $T^{-1} : X \rightarrow X$  is also defined, then the **full itinerary** of a point  $x \in X$  is a double infinite sequence defined by

$$\{\omega_n\}_{n=-\infty}^{\infty} : \quad T^n(x) \in X_{\omega_n} \quad \forall n \in \mathbb{Z}$$

**10.16 DEFINITION.** A partition  $X = X_1 \cup \dots \cup X_r$  is called a **generating partition** if distinct points have distinct itineraries. Equivalently, for any  $x \neq y$  there is an  $n$  such that  $T^n(x) \in X_i$  and  $T^n(y) \in X_j$  with some  $i \neq j$ .

**10.17 CONSTRUCTION OF A SYMBOLIC REPRESENTATION.** Let  $T : X \rightarrow X$  be a map and  $X = X_1 \cup \dots \cup X_r$  a generating partition. Let  $\phi : X \rightarrow \Omega_+$  (or  $\phi : X \rightarrow \Omega$ , if  $T$  is an automorphism) be the map that takes every point  $x \in X$  to its itinerary constructed in 10.15. This map is injective for any generating partition.

Let  $\Omega_X = \phi(X)$  be the image of  $X$ . Then  $\Omega_X$  is  $\sigma$ -invariant, i.e.  $\sigma(\Omega_X) \subset \Omega_X$ . Moreover,  $\phi \circ T = \sigma \circ \phi$ . If  $T$  has an invariant measure  $\mu$  on  $X$ , one can define a measure  $\nu$  on  $\Omega_X$  by  $\nu(B) = \mu(\phi^{-1}(B))$ . Then the dynamical systems  $(X, T, \mu)$  and  $(\Omega_X, \sigma, \nu)$  will be isomorphic.

This is a general principle for the construction of a symbolic representation.

10.18 REMARK. In the above symbolic representation of  $T : X \rightarrow X$ , any cylinder  $C_{m,n}(\omega_m, \dots, \omega_n) \subset \Omega_X$  corresponds to the set

$$X_{m,n}(\omega_m, \dots, \omega_n) = \bigcap_{k=m}^n T^{-k} X_{\omega_k}$$

that is,  $\phi^{-1}(C_{m,n}(\omega_m, \dots, \omega_n)) = X_{m,n}(\omega_m, \dots, \omega_n)$ .

10.19 REMARKS. The symbolic representation of the doubling map corresponds to the partition  $X_1 = [0, 0.5)$  and  $X_2 = [0.5, 1)$ . The symbolic representation of the baker's map corresponds to the partition of the square  $X$  by the line  $x = 0.5$ .

## 11 Mixing

In this section, we again assume that  $T : X \rightarrow X$  is a transformation preserving a measure  $\mu$ .

11.1 PROPOSITION.  $T$  is ergodic iff for every two functions  $f, g \in L^2_\mu(X)$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_X f(T^i x) g(x) d\mu = \int_X f d\mu \cdot \int_X g d\mu \quad (19)$$

*Proof.* If  $T$  is not ergodic, then there is an invariant set  $A$  with  $0 < \mu(A) < 1$ . The choice of  $f = g = \chi_A$  shows that (19) fails. Let  $T$  be ergodic. By Ergodic Theorem,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \int_X f d\mu \quad \text{a.e.}$$

Multiplying by  $g(x)$  and integrating over  $X$  should give the result. The exact argument is similar to the proof of (c) in Ergodic Theorem. First, for bounded functions  $f$  and  $g$  we can use the dominated convergence theorem. Then, for arbitrary functions  $f, g \in L^2_\mu(X)$ , we approximate them by bounded functions (in the  $L^2$  metric) and use standard integral estimates (including the Schwarz inequality) to obtain the result. The (boring) details are left out.  $\square$

11.2 REMARKS.

- (a) We assume in 11.1 that  $f, g \in L^2_\mu(X)$  (not just  $f, g \in L^1_\mu(X)$ ) in order to ensure the existence of the integrals on the left hand side of (19).
- (b) Physicists use a more convenient notation:  $\int_X f d\mu = \langle f \rangle_\mu$  (or just  $\langle f \rangle$ ). Then (19) can be rewritten as

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \langle (f \circ T^i) g \rangle = \langle f \rangle \langle g \rangle$$

11.3 PROPOSITION.  $T$  is ergodic iff for every two measurable sets  $A, B \subset X$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i} A \cap B) = \mu(A) \mu(B) \quad (20)$$

*Proof.* If  $T$  is not ergodic, let  $A$  be an invariant set with  $0 < \mu(A) < 1$  and choose  $B = A$ , then (20) fails. If  $T$  is ergodic, apply Proposition 11.1 with  $f = \chi_A$  and  $g = \chi_B$ .  $\square$

11.4 REMARK. In the proofs of Claims 8.5 and 8.9 and Theorem 10.14, we have obtained a stronger convergence than (20):

$$\lim_{n \rightarrow \infty} \mu(T^{-n} A \cap B) = \mu(A) \mu(B)$$

for all pairs of measurable sets  $A, B \subset X$ . This motivates the following definition:

11.5 DEFINITION. We say that a map  $T : X \rightarrow X$  is **mixing** (with respect to  $\mu$ ) if for all measurable subsets  $A, B \subset X$

$$\lim_{n \rightarrow \infty} \mu(T^{-n}A \cap B) = \mu(A) \mu(B) \quad (21)$$

We also say that the measure  $\mu$  is mixing if this holds.

11.6 COROLLARY. Mixing transformations are ergodic.

11.7 EXERCISE. Assume that two maps  $T_1 : X_1 \rightarrow X_1$  and  $T_2 : X_2 \rightarrow X_2$  are isomorphic, i.e.  $T_1 \simeq T_2$ . Prove that  $T_1$  is mixing if and only if  $T_2$  is.

11.8 PROPOSITION. The doubling map, the tent map, the quadratic map and the baker's map are mixing with respect to their absolutely continuous invariant measures.

*Proof.* This follows from our proofs of Claims 8.5 and 8.9 (formulas (7) and (8)), Remark 8.6 and Exercise 11.7.  $\square$

11.9 PROPOSITION. Every Bernoulli shift is mixing.

*Proof.* See the proof of Theorem 10.14 (formula (18)).  $\square$

11.10 CLAIM. Circle rotations are never mixing (with respect to the Lebesgue measure).

*Proof.* Let  $T(x) = x + a \pmod{1}$  be a circle rotation. If  $a \in \mathbb{Z}$ , then  $T$  is an identity and the claim is trivial. Otherwise  $T^{-1}(x) \neq x$ . Take a small interval (an arc)  $A' = (x - \varepsilon, x + \varepsilon)$  such that  $T^{-1}(A') \cap A' = \emptyset$ . Let  $A = (x - \varepsilon/2, x + \varepsilon/2)$  and  $B = A$ . If (21) holds, then  $T^{-n}A \cap A \neq \emptyset$  for all large enough  $n$ . On the other hand, if  $T^{-n}A \cap A \neq \emptyset$  for some  $n$ , then  $T^{-(n+1)}A \subset T^{-1}A'$  is disjoint from  $A$ , a contradiction.  $\square$

Note: if an automorphism  $T$  is mixing, then so is  $T^{-1}$ . Hence, the mixing of an automorphism  $T$  can be defined by

$$\lim_{n \rightarrow \infty} \mu(T^n A \cap B) = \mu(A) \mu(B)$$

This is so because  $\mu(T^n A \cap B) = \mu(A \cap T^{-n}B)$ .

11.11 PROPOSITION.  $T$  is mixing iff for every two functions  $f, g \in L^2_\mu(X)$

$$\lim_{n \rightarrow \infty} \langle (f \circ T^n) g \rangle = \langle f \rangle \langle g \rangle \quad (22)$$

*Proof.* If  $T$  is not mixing, we take two sets  $A$  and  $B$  on which (21) fails and set  $f = \chi_A$  and  $g = \chi_B$ , then (22) will fail as well. Let  $T$  be mixing. Then (22) follows from (21) directly, for the characteristic functions  $f = \chi_A$  and  $g = \chi_B$  of any pair of sets  $A, B \subset X$ . By taking linear combinations of characteristic functions, we obtain (22) for any pair of simple functions. Then we approximate arbitrary functions  $f, g \in L^2_\mu(X)$  by simple functions (in the  $L^2$  metric) and use standard integral estimates (including the Schwarz inequality) to obtain the result. Once again, the boring details are left out.  $\square$

11.12 REMARK. Because  $\mu(A) = \mu(T^{-n}A)$ , the characteristic equation of the mixing property (21) can be rewritten as

$$\lim_{n \rightarrow \infty} \left| \mu(T^{-n}A \cap B) - \mu(T^{-n}A)\mu(B) \right| = 0 \quad (23)$$

In probability theory, the events  $A$  and  $B$  that satisfy  $P(A \cap B) = P(A)P(B)$  are called **independent**. The mixing condition (23) in fact says that  $T^{-n}A$  and  $B$  become *asymptotically independent* as  $n \rightarrow \infty$ .

In other words, the events  $x \in B$  and  $x \in T^{-n}A$  are almost independent for large  $n$ . Note that  $x \in T^{-n}A$  is equivalent to  $T^n(x) \in A$ . Now we can say that the event  $x \in B$  (a condition on the initial point  $x$ ) becomes independent of the event  $T^n(x) \in A$  (a condition on its image at time  $n$ ) as  $n \rightarrow \infty$ . Or shortly, the **distant future will become independent of the present** as time goes on. Alternatively, thinking of  $T^n(x)$  as a *present* point and  $x$  as initial, or *past* point, we can say that the **present becomes independent of the remote past**.

Another way to look at mixing is this. Given a set  $A \subset X$ , the condition (23) must hold for every set  $B$ , as  $n \rightarrow \infty$ . Then  $T^{-n}A$  must be a rather weird set that tends to overlap with every subset  $B \subset X$  as  $n \rightarrow \infty$ . One may expect that  $T^{-n}A$  “spreads out” and behaves like an octopus that penetrates with its tentacles every tiny corner of the space  $X$ , see Fig. 10.

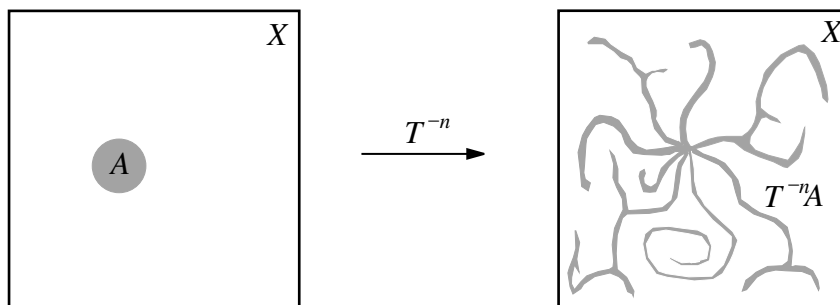


Figure 10: A set  $A$  and its preimage  $T^{-n}A$  under a mixing map.

11.13 REMARK. In physics and ergodic theory, the quantity

$$C(f, g) = \langle f g \rangle - \langle f \rangle \langle g \rangle$$

is called the **correlation**<sup>7</sup> between  $f$  and  $g$ . The quantity

$$C_n(f, g) = \langle (f \circ T^n) g \rangle - \langle f \rangle \langle g \rangle$$

is called the correlation between  $f$  and  $g$  at time  $n$ . Now the mixing property (22) is equivalent to the convergence of correlations to zero, as time goes on. In this case, physicists say that the **correlations decay**. For physical theories, it is very important to know just how fast the correlations decay for particular functions  $f$  and  $g$ . The rate (or speed) of the decay of correlations characterizes various physics models.

**11.14 DEFINITION.** We say that a map  $T$  is **weakly mixing** (with respect to  $\mu$ ) if for all measurable subsets  $A, B \subset X$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0$$

We can also say that the measure  $\mu$  is weakly mixing if this holds.

**11.15 PROPOSITION.** If  $T$  is mixing, then it is weakly mixing. If  $T$  is weakly mixing, then it is ergodic.

Note: the circle rotation  $T(x) = x + a \pmod{1}$  is not even weakly mixing.

**11.16 DEFINITION.** Let  $k \geq 2$ . We say that a map  $T : X \rightarrow X$  is  **$k$ -mixing**, or **mixing of multiplicity  $k$**  (with respect to  $\mu$ ), if for any measurable subsets  $A_1, A_2, \dots, A_k \subset X$  and  $0 \leq n_1 < n_2 < \dots < n_k$

$$\mu(T^{-n_1}A_1 \cap T^{-n_2}A_2 \cap \dots \cap T^{-n_k}A_k) \rightarrow \mu(A_1)\mu(A_2) \cdots \mu(A_k) \quad (24)$$

whenever

$$\min\{n_2 - n_1, n_3 - n_2, \dots, n_k - n_{k-1}\} \rightarrow \infty$$

We also say that the measure  $\mu$  is  $k$ -mixing if this holds.

The meaning of  $k$ -mixing is the asymptotic independence of the sets taken at  $k$  different moments of time when the time intervals between them grow.

**11.17 PROPOSITION.** Mixing is equivalent to 2-mixing. If  $T$  is  $k$ -mixing, then it is  $(k - 1)$ -mixing for every  $k \geq 3$ .

---

<sup>7</sup>In probability theory, the quantity  $C(f, g)$  is called the **covariance**, while the ratio  $\rho_{f, g} = C(f, g) / \sqrt{C(f, f)C(g, g)}$  is called the correlation. We will use here the physics/ergodic terminology and call  $C(f, g)$  the correlation.

*Proof.* Just set  $A_k = X$  in the above definition.  $\square$

11.18 REMARK. The weak mixing and  $k$ -mixing properties are invariant under isomorphisms. This generalizes Exercises 7.30 and 11.7.

11.19 PROPOSITION. Every Bernoulli shift is  $k$ -mixing for all  $k \geq 2$ .

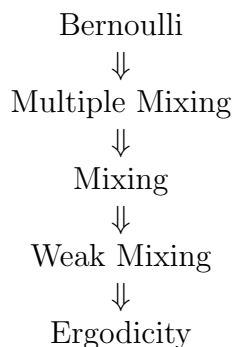
*Proof.* As in 10.14, we can approximate arbitrary sets  $A_1, \dots, A_k$  with finite unions of cylinders. Thus, it is enough to prove (24) for cylinders only. This generalizes Exercise 10.13, we omit details.  $\square$

11.20 DEFINITION. A map  $T : X \rightarrow X$  preserving a measure  $\mu$  is said to be **Bernoulli** if it is isomorphic to a Bernoulli shift. In this case we also call  $\mu$  a Bernoulli measure on  $X$ .

11.21 PROPOSITION. Any Bernoulli map is  $k$ -mixing for all  $k \geq 2$ .

*Proof.* This follows from 11.18 and 11.19.

11.22 HIERARCHY OF ERGODIC PROPERTIES. As we see, the following hierarchy of properties (listed from strongest to weakest) takes places:



Every map that has a higher property, also has the lower properties. Bernoulli is regarded as the highest degree of ergodicity (or, to some extent, the highest degree of chaos).

## 12 Linear Toral Automorphisms

Rich classes of dynamical systems can be constructed by using linear transformations on a torus.

**12.1 DEFINITION.** An  $d$ -dimensional torus  $\text{Tor}^d$  (for short, a  $d$ -torus) is a unit cube in  $\mathbb{R}^d$

$$K_d = \{\mathbf{x} = (x_1, \dots, x_d) : 0 \leq x_i \leq 1 \quad i = 1, \dots, d\}$$

whose opposite faces are identified, i.e. we assume  $x_i + 1 = x_i$  for all  $i = 1, \dots, d$ . Alternatively,  $\text{Tor}^d$  can be defined as the factor space  $\text{Tor}^d = \mathbb{R}^d / \mathbb{Z}^d$ .

Note: the 1-torus  $\text{Tor}^1$  is just the unit circle introduced in 2.1.

Note: the 2-torus  $\text{Tor}^2$  is a square with identified opposite sides. One is used to visualize  $\text{Tor}^2$  as the surface of a doughnut.

**12.2 DEFINITION.** Let  $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$ . A **translation** of the  $d$ -torus  $\text{Tor}^d$  is

$$T_{\mathbf{a}}(\mathbf{x}) = \mathbf{x} + \mathbf{a} \quad (\text{mod } 1)$$

i.e.  $T_{\mathbf{a}}(x_1, \dots, x_d) = (x_1 + a_1, \dots, x_d + a_d)$ , and each  $x_i + a_i$  is taken modulo 1.

Note: if  $d = 1$ , the translation  $T_{\mathbf{a}}$  is just the rotation of the unit circle  $\text{Tor}^1$  through the angle  $a_1$ .

**12.3 REMARK.** Every translation  $T_{\mathbf{a}}$  is a diffeomorphism of the torus  $\text{Tor}^d$ . Moreover,  $T_{\mathbf{a}}$  is a linear map and an isometry (i.e., it preserves distances between points and angles between vectors). Also, for each  $n \in \mathbb{Z}$  the  $n$ th iteration of  $T_{\mathbf{a}}$  is  $T_{\mathbf{a}}^n = T_{n\mathbf{a}}$ .

**12.4 PROPOSITION.**  $T_{\mathbf{a}}$  preserves the Lebesgue measure  $m$  on the torus  $\text{Tor}^d$ .

*Proof.* The derivative  $DT_{\mathbf{a}}$  is the identity matrix, hence  $\det DT_{\mathbf{a}} = 1$ . Now the invariance of the Lebesgue measure follows as in Exercise 5.6(a).  $\square$

**12.5 THEOREM.** The following are equivalent:

- (a)  $T_{\mathbf{a}}$  is ergodic with respect to the Lebesgue measure;
- (b)  $\forall \mathbf{x} \in \text{Tor}^d$  the trajectory  $\{T_{\mathbf{a}}^n(x)\}_{n=0}^{\infty}$  is dense in  $\text{Tor}^d$ ;
- (c) the numbers  $a_1, \dots, a_d$  are rationally independent of 1, i.e.

$$m_0 + m_1 a_1 + m_2 a_2 + \dots + m_d a_d \neq 0 \tag{25}$$

for any integers  $m_0, m_1, \dots, m_d \in \mathbb{Z}$  unless  $m_0 = m_1 = \dots = m_d = 0$ .



*Proof.* We start with the implication (a) $\Rightarrow$ (b). By 7.25, if  $T_{\mathbf{a}}$  is ergodic, then the trajectory of almost every point  $\mathbf{x} \in \text{Tor}^d$  is dense. On the other hand, if the trajectory of at least one point  $\mathbf{x}_0 \in \text{Tor}^d$  is dense, then the trajectory of every point  $\mathbf{x} \in \text{Tor}^d$  is dense. Indeed,

$$T_{\mathbf{a}}^n(\mathbf{x}) = \mathbf{x} + n\mathbf{a} = \mathbf{x}_0 + n\mathbf{a} + (\mathbf{x} - \mathbf{x}_0) = T_{\mathbf{a}}^n(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)$$

(all calculations here are done modulo 1). So, to find a subsequence  $T_{\mathbf{a}}^{n_k}(\mathbf{x})$  converging to a given point  $\mathbf{c} \in \text{Tor}^d$ , it is enough to find a subsequence  $T_{\mathbf{a}}^{n_k}(\mathbf{x}_0)$  converging to the point  $\mathbf{c} - \mathbf{x} + \mathbf{x}_0 \pmod{1}$ .

Next we prove the implication (b) $\Rightarrow$ (c). Suppose (c) fails, i.e. a nontrivial relation (25) exists. Consider the trajectory of the point  $\mathbf{x}_0 = \mathbf{0}$ , which consists of the points  $\mathbf{x}_n = T_{\mathbf{a}}^n(\mathbf{0}) = n\mathbf{a} \pmod{1}$ . Their coordinates must satisfy the relation

$$\sum_{i=1}^d m_i x_{n,i} = n \sum_{i=1}^d m_i a_i = -nm_0 = 0 \pmod{1}$$

We now show that the sequence  $\{\mathbf{x}_n\}$  cannot be dense in  $\text{Tor}^d$ . Let  $M = \max\{|m_1|, \dots, |m_d|\}$  and consider a small cube  $K' \subset \text{Tor}^d$  defined by  $K' = \{\mathbf{x} : 0 \leq x_i \leq 1/(2dM) \text{ for } i = 1, \dots, d\}$ . For every point  $\mathbf{x} \in K'$  the relation  $\sum_i m_i x_i = 0 \pmod{1}$  is equivalent to  $\sum_i m_i x_i = 0$  (without being taken modulo 1). This is just a hyperplane in  $\mathbb{R}^d$ , which cannot be dense in any open set, in particular in  $K'$ .

Lastly, we prove the implication (c) $\Rightarrow$ (a). The proof uses Fourier analysis. Let  $T_{\mathbf{a}}$  be not ergodic and  $B \subset \text{Tor}^d$  an invariant set with measure  $0 < m(B) < 1$ . The function  $f = \chi_B - m(B)$  is  $T_{\mathbf{a}}$ -invariant and bounded, so it belongs in  $L_m^2(\text{Tor}^d)$ . Consider functions

$$\phi_{\mathbf{m}}(\mathbf{x}) = e^{2\pi i \langle \mathbf{m}, \mathbf{x} \rangle} = \cos 2\pi \langle \mathbf{m}, \mathbf{x} \rangle + i \sin 2\pi \langle \mathbf{m}, \mathbf{x} \rangle$$

where  $\mathbf{m} \in \mathbb{Z}^d$  and  $\langle \mathbf{m}, \mathbf{x} \rangle = m_1 x_1 + \dots + m_d x_d$  is the scalar product in  $\mathbb{R}^d$ . These functions are periodic with period 1 in each coordinate, so that they are well defined on  $\text{Tor}^d$ . They make an orthonormal basis (the Fourier basis) in the space  $L_m^2(\text{Tor}^d)$  of complex-valued functions on  $\text{Tor}^d$ . This means that if

$$\int_{\text{Tor}^d} f(\mathbf{x}) \overline{\phi_{\mathbf{m}}(\mathbf{x})} dm = 0 \quad \forall \mathbf{m} \in \mathbb{Z}^d \quad (26)$$

then  $f(\mathbf{x}) = 0$  almost everywhere. By using change of variable, we now see that for any  $T_{\mathbf{a}}$ -invariant function  $f(\mathbf{x})$

$$\begin{aligned} \int_{\text{Tor}^d} f(\mathbf{x}) \overline{\phi_{\mathbf{m}}(\mathbf{x})} dm &= \int_{\text{Tor}^d} f(\mathbf{x} + \mathbf{a}) \overline{\phi_{\mathbf{m}}(\mathbf{x})} dm \\ &= \int_{\text{Tor}^d} f(\mathbf{y}) \overline{\phi_{\mathbf{m}}(\mathbf{y} - \mathbf{a})} dm \\ &= e^{2\pi i \langle \mathbf{m}, \mathbf{a} \rangle} \int_{\text{Tor}^d} f(\mathbf{y}) \overline{\phi_{\mathbf{m}}(\mathbf{y})} dm \end{aligned}$$

By the assumption (c), we have  $e^{2\pi i \langle \mathbf{m}, \mathbf{a} \rangle} \neq 1$  for all  $\mathbf{m} \in \mathbb{Z}^d$  except  $\mathbf{m} = \mathbf{0}$ . Therefore, we get (26) for all  $\mathbf{m} \neq \mathbf{0}$ . For  $\mathbf{m} = \mathbf{0}$  we have  $\phi_{\mathbf{m}}(\mathbf{x}) \equiv 1$ , hence (26) holds as well for our function  $f(\mathbf{x}) = \chi_B - m(B)$ . Thus,  $f(\mathbf{x}) = 0$  almost everywhere, a contradiction.  $\square$

Note: the translation  $T_{\mathbf{a}}$  is not mixing, not even weakly mixing (just like circle rotations).

Next, we restrict ourselves to the 2-torus  $\text{Tor}^2 = \{(x, y) : 0 \leq x, y < 1\}$  (just for the sake of simplicity). Let  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  be a matrix with integral entries  $a, b, c, d \in \mathbb{Z}$ . Assume that

$$\det A = ad - bc = \pm 1 \quad (27)$$

**12.6 DEFINITION.** Any matrix  $A$  with the above properties defines a **linear toral automorphism**  $T_A : \text{Tor}^2 \rightarrow \text{Tor}^2$  by

$$T_A(\mathbf{x}) = A\mathbf{x} \pmod{1}$$

i.e.  $T_A(x, y) = (ax + by \pmod{1}, cx + dy \pmod{1})$ .

Note:  $T_A$  is well defined on  $\text{Tor}^2$  whenever the entries of  $A$  are integers.

**12.7 EXERCISE.** Show that  $T_A$  is one-to-one. Hint: if  $T_A(\mathbf{x}) = T_A(\mathbf{y})$ , then  $T_A(\mathbf{z}) = \mathbf{0} \pmod{1}$  for  $\mathbf{z} = \mathbf{x} - \mathbf{y} \pmod{1}$ . Next consider the system  $az_1 + bz_2 = m_1$  and  $cz_1 + dz_2 = m_2$  for some  $m_1, m_2 \in \mathbb{Z}$  and show that its only solution  $(z_1, z_2)$  is a pair of integers  $z_1, z_2$ .

**12.8 REMARK.** The map  $T_A$  is a linear diffeomorphism of the torus  $\text{Tor}^2$ . Moreover, for every  $n \in \mathbb{Z}$  its  $n$ th iterate  $T_A^n$  is  $T_A^n = T_{A^n}$ . In particular,  $T_A^{-1} = T_{A^{-1}}$  (see also Walters, pp. 14–16, for more details). The point  $\mathbf{0} = (0, 0)$  is always a fixed point:  $T_A(\mathbf{0}) = \mathbf{0}$ .

**12.9 EXERCISE.** Show that every point  $(x, y) \in \text{Tor}^2$  with rational coordinates  $x, y \in \mathbb{Q}$  is periodic.

**12.10 PROPOSITION.**  $T_A$  preserves the Lebesgue measure  $m$  on the torus  $\text{Tor}^2$ .

*Proof.* The derivative  $DT_A$  is the matrix  $A$  itself, hence  $\det DT_A = \pm 1$ . Now the invariance of the Lebesgue measure follows as in Exercise 5.6(a).

**12.11 EXAMPLE.** Let  $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ . The line  $x = y$  consists entirely of fixed points, while all the other points are periodic with period 2, because  $T_A^2 = \text{id}$ . The map  $T_A$  “flips” the torus across its main diagonal  $x = y$ . There are no dense orbits, hence the map is not ergodic. Note that the eigenvalues of  $A$  are  $\lambda = \pm 1$ .

12.12 EXAMPLE. Let  $A = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$  with any  $m \in \mathbb{Z}$ . The line  $y = 0$  consists entirely of fixed points. Every line  $y = \text{const}$  is invariant under  $T_A$ . No dense orbits exist, so the map  $T_A$  is not ergodic. Note that  $A$  has one eigenvalue  $\lambda = 1$  of multiplicity 2.

Note: if we picture  $\text{Tor}^2$  as a cylinder with the base circle  $\{(x, y) : 0 \leq x < 1, y = 0\}$  and the vertical coordinate  $y$ , then  $T_A$  rotates every horizontal section  $y = \text{const}$  by the angle  $my$ , since it acts by  $x \mapsto x + my \pmod{1}$ . The higher the section, the larger the angle of rotation, so that the whole cylinder is twisted upward (unscrewed). Such maps are called **twist maps**.

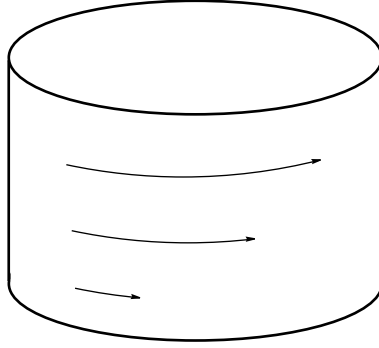


Figure 11: A twist map of a cylinder.

12.13 REMARK. Examples 12.11 and 12.12 illustrate what ergodic components of a map may look like. In 12.11, each ergodic component of  $T_A$  is either a periodic orbit of period 2 or a fixed point (the latter can be ignored since they make a null set). In 12.12, most of the ergodic components are sections  $y = c$  (precisely, such are all sections with irrational  $c$ ). The sections  $y = c$  for rational  $c$  can be further decomposed into periodic orbits or ignored altogether, since their total measure is zero.

12.14 EXAMPLE. Let  $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ . In this case, the only fixed point is  $\mathbf{0}$ , and the description of the map  $T_A$  is not so simple. Linear algebra will help us, though. Note that the eigenvalues of  $A$  are

$$\lambda_{1,2} = \frac{1 \pm \sqrt{5}}{2}$$

and the corresponding eigenvectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are orthogonal. Let  $L_1$  and  $L_2$  be the perpendicular lines on  $\text{Tor}^2$  spanned by the vectors  $v_1$  and  $v_2$ , respectively. Both lines are invariant under  $T_A$ . Since  $\lambda_1 > 1$ , the line  $L_1$  is expanded (“stretched out”) by a factor of  $\lambda_1$  under  $T_A$ . On the other hand,  $|\lambda_2| < 1$ , so the other line  $L_2$  is compressed (contracted) by a factor of  $|\lambda_2|$  under  $T_A$  (and it is flipped over, because  $\lambda_2 < 0$ ). Locally,

near the fixed point  $\mathbf{0}$ , the action of  $T_A$  is shown on Fig. 12, it looks like a “saddle”. The orbit of any point near  $\mathbf{0}$  lies on a hyperbola (or a pair of hyperbolas). In differential equations such fixed points are referred to as hyperbolic.

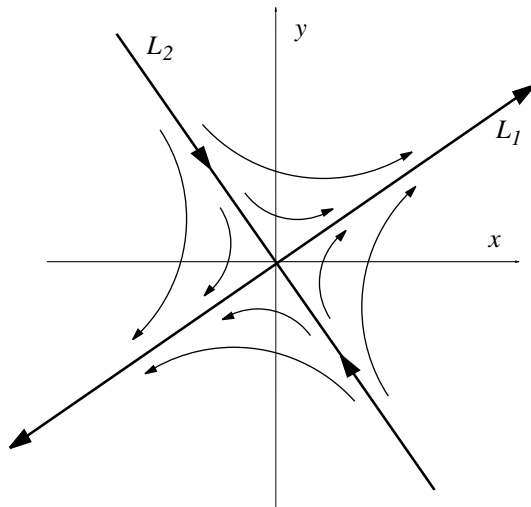


Figure 12: A fixed point of a hyperbolic toral automorphism.

**12.15 DEFINITION.** A linear total automorphism  $T_A$  is **hyperbolic** if the eigenvalues of  $A$  are real numbers different from  $\pm 1$ .

**12.16 EXERCISE.** Show that for any hyperbolic toral automorphism both eigenvalues are irrational, the absolute value of one of them is greater than 1, and that of the other eigenvalue is less than 1.

Let  $\lambda = \min\{|\lambda_1|, |\lambda_2|\}$ . Note that  $\lambda^{-1} = \max\{|\lambda_1|, |\lambda_2|\}$ . Note also that the inverse matrix  $A^{-1}$  has eigenvalues  $\lambda_1^{-1}$  and  $\lambda_2^{-1}$  and the same eigenvectors as  $A$  does. So,  $A^{-1}$  contracts  $L_1$  by a factor of  $\lambda$  and expands  $L_2$  by a factor of  $\lambda^{-1}$ .

**12.17 DEFINITION.** The line  $L_1$  spanned by the eigenvector  $v_1$  corresponding to the larger (in absolute value) eigenvalue of  $A$  is called the **unstable manifold**. It is expanded (“stretched out”) under  $T_A$ . The line  $L_2$  spanned by the eigenvector  $v_2$  corresponding to the other, smaller eigenvalue of  $A$  is called the **stable manifold**. Note that both lines extend infinitely long, they wrap around the torus infinitely many times.

**12.18 EXERCISE.** Show that for any hyperbolic toral automorphism the lines  $L_1$  and  $L_2$  are dense on the torus  $\text{Tor}^2$ . Hint: verify that the equation of the line  $L_i$  for  $i = 1, 2$  is  $y = \gamma_i x$  where  $\gamma_i = (\lambda_i - a)/b$  is an irrational number by 12.16 (assume that  $b \neq 0$ ).

for simplicity). For any real number  $\alpha$  the points  $(n\alpha, n\alpha\gamma_i) \pmod{1}$  for  $n = 0, 1, 2, \dots$  belong in  $L_i$ . Now use Theorem 12.5 to show that for some  $\alpha \neq 0$  these points make a dense set. (Note: your  $\alpha$  should be chosen carefully, so that (c) will be satisfied!)

Note: with a little extra effort one can show that for any  $\varepsilon > 0$  there is a  $d > 0$  such that every segment of length  $d$  on the line  $L_1$  intersects every disk of radius  $\varepsilon > 0$  on the torus  $\text{Tor}^2$ .

**12.19 RECTANGULAR PARTITIONS.** Further analysis of the map  $T_A$  involves symbolic dynamics. According to 10.17, one needs to start with a generating partition. Here we partition the torus  $\text{Tor}^2$  into rectangles with sides parallel to the stable and unstable lines. Figure 13(a) shows the partition of the torus into three rectangles for Example 12.14. The sides of the rectangles are made by pieces of the lines  $L_1$  and  $L_2$ . Fig. 13(b) shows the images of those three rectangles under  $T_A$ , respectively. Note that each rectangle is stretched by  $T_A$  in the direction of  $L_1$  (the unstable direction) and compressed in the direction of  $L_2$  (the stable direction), but it retains its rectangular shape.

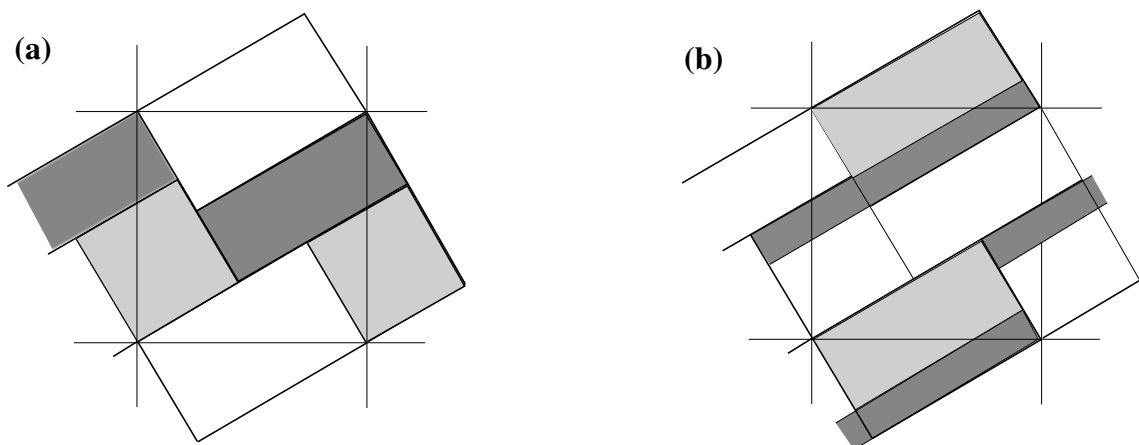


Figure 13: A Markov partition of  $\text{Tor}^2$  for Example 12.14.

**12.20 PROPER INTERSECTION.** Denote the rectangles by  $R_1$  (white),  $R_2$  (light grey) and  $R_3$  (dark grey). Let's closely examine the intersections  $T_A(R_i) \cap R_j$  for each pair  $i, j$ . If it is not empty, then it is a subrectangle in  $R_j$ , which stretches completely across  $R_j$  in the unstable direction. Also, it is a subrectangle in  $T_A(R_i)$ , which stretches completely across  $T_A(R_i)$  in the stable direction. In other words,  $T_A(R_i)$  intersects  $R_j$  **properly** (transversely), as illustrated in Fig. 14.

**12.21 DEFINITION.** A partition of  $\text{Tor}^2$  into rectangles  $\{R_i\}_{i=1}^r$  with sides parallel to  $L_1$  and  $L_2$  is called a **Markov partition** if all intersections  $T_A(R_i) \cap R_j$ ,  $1 \leq i, j \leq r$ , with

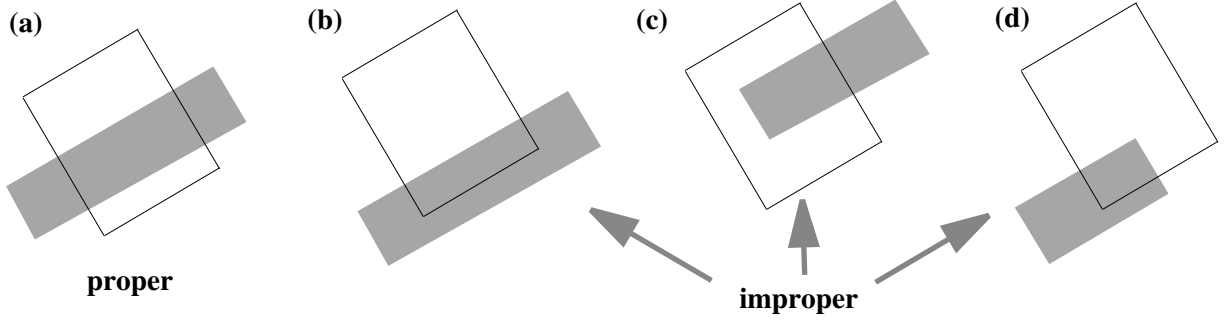


Figure 14: A proper intersection (a) of  $T_A(R_i)$  (grey) and  $R_j$  (white). The figures (b), (c), and (d) illustrate improper intersections.

nonempty interior are connected and proper.

A partition shown in Fig. 13(a) is then a Markov partition.

**12.22 LEMMA.** Any Markov partition is generating.

*Proof.* If not, then some distinct points  $x \neq y$  have the same itinerary, i.e.  $x, y \in \cap_{k=-\infty}^{\infty} T_A^{-k} R_{i_k}$  for some sequence  $\{i_k\}_{k=-\infty}^{\infty}$ . However, the diameter of the set  $\cap_{k=-n}^n T_A^{-k} R_{i_k}$  is  $O(\lambda^n)$ , which converges to zero as  $n \rightarrow \infty$ , a contradiction.  $\square$

Note: it is essential for this proof that the intersections  $T(R_i) \cap R_j$  are connected, without this assumption Lemma 12.22 may fail.

Recall that a generating partition  $\text{Tor}^2 = R_1 \cup \dots \cup R_r$  into  $r$  disjoint subsets gives rise to a symbolic representation of an automorphism  $T_A : \text{Tor}^2 \rightarrow \text{Tor}^2$  by a shift  $\sigma : \Omega_r \rightarrow \Omega_r$  on a symbolic space with  $r$  symbols as defined by 10.15–10.17. By 10.18, every cylinder  $C_{m,n}(i_m, \dots, i_n)$  corresponds to the intersection

$$R_{m,n}(i_m, \dots, i_n) = \cap_{k=m}^n T_A^{-k}(R_{i_k})$$

that is,  $\phi^{-1}(C_{m,n}(i_m, \dots, i_n)) = R_{m,n}(i_m, \dots, i_n)$ .

We now study the Lebesgue measure  $m$  and the induced measure  $\mu$  on  $\Omega_r$ .

For each rectangle  $R_i$ , denote by  $s_i$  and  $u_i$  its sides parallel to the stable direction ( $L_2$ ) and the unstable direction ( $L_1$ ), respectively. Then  $m(R_i) = s_i u_i$ . Due to the properness of intersections, if  $T_A R_i \cap R_j$  has nonempty interior, then  $m(T_A R_i \cap R_j) = \lambda s_i u_j$ , where  $\lambda = \min\{|\lambda_1|, |\lambda_2|\}$ .

12.23 LEMMA. (a) For any integers  $m < n$  and any  $i_m, i_{m+1}, \dots, i_n$  the set  $R_{m,n}(i_m, \dots, i_n)$  either has empty interior (and then zero measure) or has measure

$$m(R_{m,n}) = \lambda^{n-m} s_{i_m} u_{i_n}$$

(b) The set  $R_{m,n}(i_m, \dots, i_n)$  has nonempty interior if and only if  $T_A(R_{i_k}) \cap R_{i_{k+1}}$  has nonempty interior for every  $k = m, \dots, n-1$ .

*Proof* is a simple geometric inspection.  $\square$

12.24 DEFINITION. The **matrix of transition probabilities** is an  $r \times r$  matrix  $\Pi = (\pi_{ij})$  with entries

$$\pi_{ij} = \frac{m(T_A R_i \cap R_j)}{m(R_i)} = \begin{cases} \lambda u_j / u_i & \text{if } \text{int}(T_A R_i \cap R_j) \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

The **stationary probability vector** is the row vector  $\mathbf{p} = (p_1, \dots, p_r)$  with components

$$p_i = m(R_i) = s_i u_i$$

Note:  $\mathbf{p}\Pi = \mathbf{p}$ , i.e.  $\mathbf{p}$  is a left eigenvector for the matrix  $\Pi$  with eigenvalue 1. In other words, it remains invariant (stationary) under the right multiplication by  $\Pi$ .

Note: for each  $i$  we have  $\sum_j \pi_{ij} = 1$ . Matrices with nonnegative entries whose rows sum up to one are called **stochastic matrices**.

Note:  $\pi_{ij}$  is the fraction of  $R_i$  that is mapped into  $R_j$  by  $T_A$ . We interpret  $\pi_{ij}$  as the probability for a point starting in  $R_i$  to move into  $R_j$ .

12.25 LEMMA. For any integers  $m < n$  and any  $i_m, \dots, i_n$  the intersection  $R_{m,n}(i_m, \dots, i_n)$  has measure

$$m(R_{m,n}(i_m, \dots, i_n)) = p_{i_m} \pi_{i_m i_{m+1}} \cdots \pi_{i_{n-1} i_n}$$

12.26 DEFINITION. A measure  $\mu$  on  $\Omega_r$  is called a **Markov measure** with a transition probability matrix  $\Pi = (\pi_{ij})$  of size  $r \times r$  and a stationary probability vector  $\mathbf{p} = (p_1, \dots, p_r)$  if for every cylinder  $C_{m,n}(i_m, \dots, i_n)$  its measure is

$$\mu(C_{m,n}(i_m, \dots, i_n)) = p_{i_m} \pi_{i_m i_{m+1}} \cdots \pi_{i_{n-1} i_n}$$

This model is known in probability theory as a *Markov chain*. This analogy explains the term *Markov partition*.

12.27 COROLLARY. The Lebesgue measure  $m$  on  $\text{Tor}^2$  corresponds to the Markov measure  $\mu$  on  $\Omega_r$ . The dynamical systems  $(\text{Tor}^2, T_A, m)$  and  $(\Omega_r, \sigma, \mu)$  are isomorphic.

Now the study of the hyperbolic toral automorphism  $T_A$  reduces to the study of the shift  $\sigma$  on  $\Omega_r$  with the Markov measure  $\mu$ . Surprisingly, this reduction makes things a lot simpler. A little piece of probability theory will help us.

**12.28 NOTATION.** For  $k \geq 1$ , denote by  $\pi_{ij}^{(k)}$  the elements of the matrix  $\Pi^k$  (the  $k$ th power of  $\Pi$ ). Note that  $\mathbf{p}\Pi^k = \mathbf{p}$  for all  $k \geq 1$ .

**12.29 EXERCISE.** Show that

$$\pi_{ij}^{(k)} = \frac{m(T_A^k R_i \cap R_j)}{m(R_i)}$$

**12.30 EXERCISE.** Let  $C = C_{m,n}(i_m, \dots, i_n)$  and  $C' = C_{m',n'}(i_{m'}, \dots, i_{n'})$  be two cylinders such that  $n < m'$ . Let  $t = m' - n$ . Show that

$$\mu(C \cap C') = \mu(C) \mu(C') \pi_{i_n i_{m'}}^{(t)} / p_{i_{m'}}$$

Hint: do that first for  $t = 1$  and then use induction on  $t$ . A helpful formula is  $C = \cup_{i=1}^r C_{m,n+1}(i_m, \dots, i_n, i)$ .

**12.31 LEMMA.** There is an  $s \geq 1$  such that the matrix  $\Pi^s$  has all positive entries, i.e.  $\pi_{ij}^{(s)} > 0$  for all  $1 \leq i, j \leq r$ .

*Proof.* For large  $s > 1$ , the set  $T_A^s(R_i)$  is a very long narrow rectangle, one long side of which lies on the line  $L_1$ . Then by Exercise 12.18 (and the remark after it),  $T_A^s(R_i)$  intersects every rectangle  $R_j$  of the Markov partition. Now 12.29 completes the proof.  $\square$

**12.32 THEOREM (LIMIT THEOREM FOR MARKOV CHAINS).** If  $\Pi^s$  has all positive entries for some  $s > 0$ , then for all  $1 \leq i, j \leq r$  we have  $\pi_{ij}^{(t)} \rightarrow p_j$  as  $t \rightarrow \infty$ .

*Proof.* We outline the argument. Fix  $1 \leq j \leq r$  and let  $n, t \geq 1$ . Since  $\Pi^{n+t} = \Pi^n \Pi^t$ ,

$$\pi_{ij}^{(n+t)} = \sum_{k=1}^r \pi_{ik}^{(n)} \pi_{kj}^{(t)}$$

Let  $\delta_n = \min_{i,k} \pi_{ik}^{(n)} \geq 0$ . Note that  $\max_{i,k} \pi_{ik}^{(n)} \leq 1 - \delta_n$ . Now let

$$m_t = \min_k \pi_{kj}^{(t)} \quad \text{and} \quad M_t = \max_k \pi_{kj}^{(t)}$$

(remember  $j$  is fixed). The following estimate is rather elementary:

$$(1 - \delta_n)m_t + \delta_n M_t \leq \pi_{ij}^{(n+t)} \leq \delta_n m_t + (1 - \delta_n)M_t \quad (28)$$



Hence,  $M_{t+n} \leq M_t$  and  $m_{t+n} \geq m_t$ . Next we show that  $M_t - m_t \rightarrow 0$  as  $t \rightarrow \infty$ . Let  $t = ms + n$  with  $0 \leq n < s$ . Then  $M_t - m_t \leq M_{ms} - m_{ms}$ . Now, put  $n = s$  and  $t = (m-1)s$  in (28) and get

$$M_{ms} - m_{ms} \leq (1 - 2\delta_s) (M_{(m-1)s} - m_{(m-1)s})$$

Since  $\delta_s > 0$ , we have  $M_{ms} - m_{ms} < C\lambda^{m-1}$  where  $C = M_s - m_s$  and  $\lambda = 1 - 2\delta_s < 1$ .

Therefore,  $\pi_{ij}^{(t)} \rightarrow q_j$  as  $t \rightarrow \infty$  with some  $q_j \geq 0$  (independent of  $i$ ). This implies  $\mathbf{e}_i \Pi^t \rightarrow \mathbf{q}$  as  $t \rightarrow \infty$  for each  $i = 1, \dots, r$ , where  $\mathbf{e}_i$  is the  $i$ th canonical basis row-vector and  $\mathbf{q} = (q_1, \dots, q_r)$ . By linearity,  $\mathbf{p} \Pi^t \rightarrow \mathbf{q}$  as  $t \rightarrow \infty$ , hence  $\mathbf{q} = \mathbf{p}$ .  $\square$

Two important facts follow from the above proof. First, the stationary vector  $\mathbf{p}$  is unique. Second, the convergence  $\pi_{ij}^{(t)} \rightarrow p_j$  is exponentially fast (in  $t$ ).

**12.33 COROLLARY.** For any two cylinders  $C, C' \subset \Omega_r$  we have

$$\lim_{n \rightarrow \infty} \mu(C \cap \sigma^{-n}(C')) = \mu(C) \mu(C')$$

*Proof.* This follows from 12.30, 12.31 and 12.32.

**12.34 PROPOSITION.** The shift  $\sigma : \Omega_r \rightarrow \Omega_r$  with the Markov measure  $\mu$  is mixing, and so is the original hyperbolic toral automorphism  $T_A$ .

*Proof.* This follows from 12.33 and standard approximation arguments used in the proof of 10.14.

**12.35 REMARK.** The hyperbolic toral automorphism  $T_A$  is actually Bernoulli, but the proof of this fact is rather sophisticated, we omit it.

## 13 Lyapunov Exponents

Recall that the baker's transformation expands (stretches) the unit square  $X$  horizontally and contracts it (or compresses) vertically. The hyperbolic toral automorphism  $T_A$  expands  $\text{Tor}^2$  along  $L_1$  and contracts it along  $L_2$ . Such expansion and contraction happen to be quite a common phenomena for smooth maps and, in particular, for many physical models. To study these phenomena, we need a piece of linear algebra.

**13.1 FACTS (FROM LINEAR ALGEBRA).** Let  $A$  be a  $d \times d$  matrix (with real entries), and assume that  $\det A \neq 0$ . For a nonzero vector  $\mathbf{u} \in \mathbb{R}^d$ , consider the sequence  $\{A^n \mathbf{u}\}$ ,  $n \in \mathbb{Z}$ . We would like to see if  $A^n$  expands or contracts the vector  $\mathbf{u}$  as  $n \rightarrow \infty$  and  $n \rightarrow -\infty$ .

Let  $\{\lambda_j\}$ ,  $1 \leq j \leq q$  ( $q \leq n$ ), be all distinct roots (real and complex) of the characteristic polynomial  $P_A(\lambda) = \det(\lambda I - A)$ . We arrange them so that  $\lambda_1, \dots, \lambda_r$  are all the distinct real roots and  $\lambda_{r+1}, \bar{\lambda}_{r+1}, \dots, \lambda_s, \bar{\lambda}_s$  are all the distinct conjugate pairs of complex roots. Denote by  $m_j$ ,  $1 \leq j \leq s$ , the respective multiplicities of the roots. Jordan theorem (the real canonical form) says that the roots  $\lambda_j$  (the eigenvalues of  $A$ ) are associated to  $A$ -invariant generalized eigenspaces  $E_j$ ,  $1 \leq j \leq s$ , whose respective dimensions equal  $\dim E_j = m_j$  for  $1 \leq j \leq r$  and  $\dim E_j = 2m_j$  for  $r < j \leq s$  (note: in the latter case the space  $E_j$  is associated to the pair  $\lambda_j, \bar{\lambda}_j$ , rather than to a single root  $\lambda_j$ ). Moreover,  $\mathbb{R}^d = \bigoplus_{j=1}^s E_j$ . We recall that  $\det A \neq 0$ , hence  $\lambda_j \neq 0$  for all  $j$ .

Now, if  $\mathbf{u} \in E_j$  is an eigenvector (such a vector only exists for  $j \leq r$ ), then  $A^n \mathbf{u} = \lambda_j^n \mathbf{u}$ , hence

$$\ln \|A^n \mathbf{u}\| = n \ln |\lambda_j| + \ln \|\mathbf{u}\|$$

for all  $n \in \mathbb{Z}$ . While this is not true for *any* vector  $\mathbf{u} \in E_j$ , it is still true that

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \ln \|A^n \mathbf{u}\| = \ln |\lambda_j| \quad \forall \mathbf{u} \in E_j, \quad \mathbf{u} \neq \mathbf{0} \quad (29)$$

This shows that when  $|\lambda_j| > 1$ , the vector  $A^n \mathbf{u}$  grows exponentially fast as  $n \rightarrow \infty$  and shrinks exponentially fast as  $n \rightarrow -\infty$ . If  $|\lambda_j| < 1$ , then it is vice versa. If  $|\lambda_j| = 1$ , then there is no exponential growth or contraction, but there might be a slow (subexponential) growth or contraction of the vector  $A^n \mathbf{u}$ , see an example in 13.5.

**13.2 EXERCISE.** A complete proof of the equation (29) might be quite lengthy and tedious. But you can at least verify it in the simple case  $\dim E_j = 2$ . There are two principal subcases here. If  $\lambda_j$  is a real root of multiplicity 2, then  $A$  restricted to  $E_j$  is given by a Jordan matrix  $J = \begin{pmatrix} \lambda_j & 1 \\ 0 & \lambda_j \end{pmatrix}$  in some basis. Verify that  $J^n = \begin{pmatrix} \lambda_j^n & n\lambda_j^{n-1} \\ 0 & \lambda_j^n \end{pmatrix}$  for all  $n \in \mathbb{Z}$  and then derive (29). If  $\lambda_j = a + bi$  is a complex root and  $b \neq 0$ , then the corresponding Jordan canonical block is  $J = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ . Verify that  $J^n =$

$|\lambda_j|^n \begin{pmatrix} \cos n\varphi & \sin n\varphi \\ -\sin n\varphi & \cos n\varphi \end{pmatrix}$  for some  $\varphi \in [0, 2\pi)$  and all  $n \in \mathbb{Z}$ , and then derive (29).

For an extra credit, try to prove (29) in any dimension.

**13.3 DEFINITION.** The numbers  $\chi_j = \ln |\lambda_j|$  that appear in (29) are called the **characteristic exponents** or the **Lyapunov exponents** of the matrix  $A$ .

Note: some distinct eigenvalues  $\lambda_i \neq \lambda_j$  may correspond to the same Lyapunov exponent, this happens whenever  $|\lambda_i| = |\lambda_j|$ . In this case each nonzero vector  $\mathbf{u} \in E_i \oplus E_j$  satisfies (29).

**13.4 PROPOSITION (LYAPUNOV DECOMPOSITION).** Every nonsingular real matrix  $A$  has distinct Lyapunov exponents  $\chi_1 > \chi_2 > \cdots > \chi_m$  (with  $m \leq s$ ) and there is a decomposition  $\mathbb{R}^d = \mathcal{E}_1 \oplus \cdots \oplus \mathcal{E}_m$  such that  $A(\mathcal{E}_j) = \mathcal{E}_j$  and

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \ln \|A^n \mathbf{u}\| = \chi_j \quad \forall \mathbf{u} \in \mathcal{E}_j, \quad \mathbf{u} \neq \mathbf{0} \quad (30)$$

The number  $\dim \mathcal{E}_j$  is called the **multiplicity** of the Lyapunov exponent  $\chi_j$ . We also call  $\mathcal{E}_j$  the **characteristic spaces** for  $A$ .

**13.5 REMARK.** We have

$$\sum_j \chi_j \cdot \dim \mathcal{E}_j = \ln |\det A|$$

because  $\det A$  equals the product of all the eigenvalues of  $A$  (counting multiplicity).

**13.6 EXAMPLES.** For the matrices  $A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  and  $A_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  (these are taken from Examples 12.11 and 12.12, respectively) all Lyapunov exponents are zero. The same is true for the matrix  $A_3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ , whose eigenvalues are  $\pm i$ . Note that  $A_2^n$  does expand and contract vectors, but very slowly (at most linearly in  $n$ ).

On the other hand, for any matrix  $A$  defining a hyperbolic toral automorphism  $T_A$  (see 12.15) one Lyapunov exponent is positive,  $\chi_1 = \ln \lambda^{-1} > 0$ , and the other is negative,  $\chi_2 = \ln \lambda < 0$ . Note that  $\chi_1 + \chi_2 = 0$ , because  $\det A = \pm 1$ .

**13.7 DEFINITION.** A matrix  $A$  and the corresponding linear map  $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$  are called **hyperbolic** if none of the eigenvalues of  $A$  (real or complex) lie on the unit circle  $|z| = 1$ .

Equivalently, we have the following principle:.

$A$  is hyperbolic iff all the Lyapunov exponents of  $A$  are different from zero

13.8 DEFINITION. The  $A$ -invariant subspaces

$$\mathcal{E}^s = \oplus_{\chi_j < 0} \mathcal{E}_j, \quad \mathcal{E}^u = \oplus_{\chi_j > 0} \mathcal{E}_j \quad \text{and} \quad \mathcal{E}^c = \mathcal{E}_j|_{\chi_j=0}$$

are called **stable**, **unstable**, and **neutral** (or **central**) subspaces, respectively. Note that

$$\mathbb{R}^d = \mathcal{E}^s \oplus \mathcal{E}^u \oplus \mathcal{E}^c$$

If the matrix  $A$  is hyperbolic, then  $\mathcal{E}^c = \{\mathbf{0}\}$ , hence  $\mathcal{E}^c$  can be omitted from the above decomposition.

The equation (29) easily implies

$$\limsup_{n \rightarrow \pm\infty} \frac{1}{n} \ln \|A^n \mathbf{u}\| < 0 \quad \forall \mathbf{u} \in \mathcal{E}^s, \quad \mathbf{u} \neq \mathbf{0}$$

and

$$\liminf_{n \rightarrow \pm\infty} \frac{1}{n} \ln \|A^n \mathbf{u}\| > 0 \quad \forall \mathbf{u} \in \mathcal{E}^u, \quad \mathbf{u} \neq \mathbf{0}$$

Let the matrix  $A$  have at least one nonzero Lyapunov exponent  $\chi_i \neq 0$ . Denote  $\chi = \min\{|\chi_i| : \chi_i \neq 0\}$  and  $\lambda = e^{-\chi}$ . Note that  $\chi > 0$  and  $\lambda < 1$ .

13.9 PROPOSITION. For any  $\varepsilon > 0$  there is a  $K > 0$  such that for all  $n \geq 0$

$$\|A^n \mathbf{u}\| \leq K(\lambda + \varepsilon)^n \|\mathbf{u}\| \quad \text{and} \quad \|A^{-n} \mathbf{u}\| \geq K^{-1}(\lambda + \varepsilon)^{-n} \|\mathbf{u}\| \quad \forall \mathbf{u} \in \mathcal{E}^s$$

and

$$\|A^n \mathbf{u}\| \geq K^{-1}(\lambda + \varepsilon)^{-n} \|\mathbf{u}\| \quad \text{and} \quad \|A^{-n} \mathbf{u}\| \leq K(\lambda + \varepsilon)^n \|\mathbf{u}\| \quad \forall \mathbf{u} \in \mathcal{E}^u$$

*Proof.* It is enough to prove the above bounds for unit vectors only. For every unit vector  $\mathbf{u}$  there is a  $K = K(\varepsilon, \mathbf{u})$  such that all these bounds hold, but  $K$  may depend on  $\mathbf{u}$ . Then we pick an orthonormal basis  $e_1, \dots, e_k$  in  $\mathcal{E}^s$  (resp.,  $\mathcal{E}^u$ ), ensure the above bounds with the same constant  $K(\varepsilon)$  for all the vectors  $e_1, \dots, e_k$ . Then we use the triangle inequality to derive the proposition for all unit vectors  $\mathbf{u}$  in  $\mathcal{E}^s$  and  $\mathcal{E}^u$ .  $\square$

Thus, vectors  $\mathbf{u} \in \mathcal{E}^u$  grow exponentially fast under  $A^n$  as  $n \rightarrow \infty$  and shrink exponentially fast as  $n \rightarrow -\infty$ . For vectors  $\mathbf{u} \in \mathcal{E}^s$ , it is exactly the opposite. Now what happens to other vectors in  $\mathbf{u} \in \mathbb{R}^d$ ?

13.10 COROLLARY. For any vector  $\mathbf{u} \notin \mathcal{E}^u \cup \mathcal{E}^s$  and any  $\varepsilon > 0$  there is a  $K > 0$  such that for all  $n \in \mathbb{Z}$

$$\|A^n \mathbf{u}\| \geq K(\lambda + \varepsilon)^{-|n|} \|\mathbf{u}\|$$

that is, the vector  $\mathbf{u}$  grows under  $A^n$  exponentially fast in both time directions: as  $n \rightarrow +\infty$  and as  $n \rightarrow -\infty$ .

Next, we extend the above results to nonlinear maps.

13.11 DEFINITION. Let  $U \subset \mathbb{R}^d$  be an open set and  $T : U \rightarrow \mathbb{R}^d$  a smooth one-to-one map with a fixed point  $x$ , i.e.  $T(x) = x$ . Then the matrix  $A = D_x T$  acts on tangent vectors  $\mathbf{u} \in \mathcal{T}_x \mathbb{R}^d$ , and the tangent space  $\mathcal{T}_x \mathbb{R}^d$  can be naturally identified with  $\mathbb{R}^d$ . Note that  $D_x T^n = (D_x T)^n = A^n$  by the chain rule.

Assume that  $\det A \neq 0$ . The Lyapunov exponents of the matrix  $A$  are called the **Lyapunov exponents** of the map  $T$  at the point  $x$ . The corresponding subspaces  $\mathcal{E}^s, \mathcal{E}^u, \mathcal{E}^c \subset \mathcal{T}_x \mathbb{R}^d$  are called the **stable**, **unstable**, and **neutral** (or **central**) subspaces, respectively, for the map  $T$  at the point  $x$ .

Note: the subspaces  $\mathcal{E}^s, \mathcal{E}^u, \mathcal{E}^c$  are invariant under  $D_x T$  but not necessarily under the map  $T$  itself. On the other hand,  $D_x T$  is a linear approximation to the map  $T$  at the point  $x$ . This allows us to obtain the following theorem, whose proof we omit.

13.12 THEOREM (HADAMARD-PERRON). There exist two submanifolds  $W^s \subset U$  and  $W^u \subset U$  such that

- (a)  $W^s \cap W^u = \{x\}$ ;
- (b) the spaces  $\mathcal{E}^s$  and  $\mathcal{E}^u$  are tangent to  $W^s$  and  $W^u$ , respectively, at the point  $x$ ;
- (c)  $T(W^s) \subset W^s$  and  $T^{-1}(W^u) \subset W^u$ ;
- (d)  $T^n(y) \rightarrow x$  for every  $y \in W^s$  and  $T^{-n}(y) \rightarrow x$  for every  $y \in W^u$ , as  $n \rightarrow \infty$ .

We omit the proof, but remark that the manifold  $W^u$  is constructed as a limit of  $(T^n \mathcal{E}^u) \cap V(x)$ , as  $n \rightarrow \infty$ , where  $V(x)$  is a sufficiently small neighborhood of  $x$ . The existence of this limit is proved by the contraction mapping principle. Similarly,  $W^s$  is constructed as a limit of  $(T^{-n} \mathcal{E}^s) \cap V(x)$ , as  $n \rightarrow \infty$ .

Since  $A = D_x T$  is a linear part of the map  $T$  at  $x$ , it is easy to obtain the following corollary to 13.9 and 13.12:

13.13 COROLLARY. For any  $\varepsilon > 0$  there is a neighborhood  $V(x)$  of the point  $x$  and a  $K > 0$  such that for all  $n \geq 0$

$$\text{dist}(T^n y, x) \leq K(\lambda + \varepsilon)^n \cdot \text{dist}(y, x) \quad \forall y \in W^s \cap V(x)$$

and

$$\text{dist}(T^{-n}y, x) \leq K(\lambda + \varepsilon)^n \cdot \text{dist}(y, x) \quad \forall y \in W^u \cap V(x)$$

Locally, near the point  $x$ , the map  $T$  acts as shown on Fig. 15.

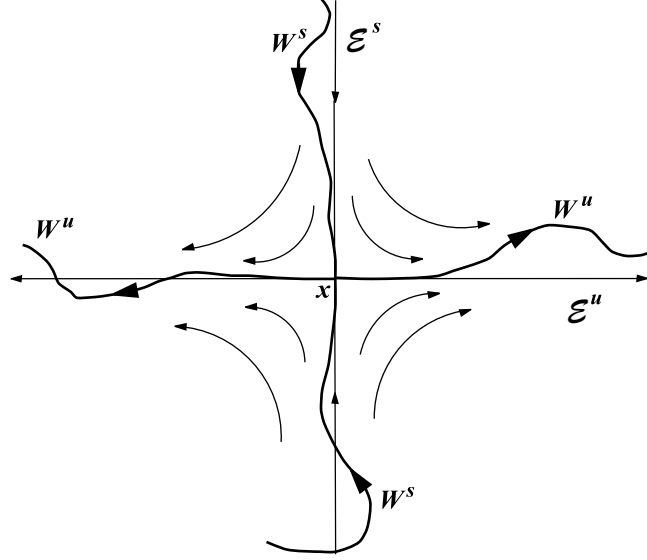


Figure 15: The action of  $T$  near a hyperbolic fixed point  $x$ .

**13.14 DEFINITION.**  $W^s$  and  $W^u$  are called the **stable** and **unstable manifolds**, respectively, for the map  $T$  at the point  $x$ . The map  $T$  is called **hyperbolic** at a fixed point  $x$  (and then  $x$  is called a **hyperbolic fixed point** for  $T$ ) if  $\dim \mathcal{E}^c = 0$ . In this case  $\dim W^s + \dim W^u = d$ .

**13.15 DEFINITION.** A hyperbolic point  $x$  is called a **source** (a **repeller**) if  $\dim \mathcal{E}^s = 0$  (hence  $\mathcal{E}^u$  coincides with  $\mathcal{T}_x \mathbb{R}^d$ ). It is called a **sink** (an **attractor**) if  $\dim \mathcal{E}^u = 0$  (hence  $\mathcal{E}^s$  coincides with  $\mathcal{T}_x \mathbb{R}^d$ ). It is called a **saddle** (a truly hyperbolic point) if both  $\mathcal{E}^s$  and  $\mathcal{E}^u$  are not trivial.

**13.16 REMARK.** Let  $x$  be a saddle point and  $y \neq x$  another point very close to  $x$ . If  $y \in W^u$ , then the trajectory  $T^n y$  moves away from  $x$  exponentially fast for  $n > 0$ , at least until  $T^n y$  leaves a certain neighborhood of  $x$ . If  $y \in W^s$ , then the trajectory  $T^n y$  moves away from  $x$  exponentially fast for  $n < 0$ . Now, if  $y \notin W^u \cup W^s$ , then the trajectory  $T^n y$  moves away from  $x$  exponentially fast for both  $n > 0$  and  $n < 0$ . This fact is known as the **separation principle**: nearby trajectories tend to separate exponentially fast, either in the future or in the past or (in most cases) both.

All the above definitions and results extend to any diffeomorphism  $T : U \rightarrow T(U) \subset M$  on an open subset  $U \subset M$  of a Riemannian manifold  $M$ , rather than  $U \subset \mathbb{R}^d$ . A

Riemannian structure in  $M$  is necessary for the norm  $\|\cdot\|$  to be well defined on  $M$ . Henceforth we assume that  $T$  is defined on an open subset of a Riemannian manifold  $M$ .

**13.17 REMARK.** All the above definitions and results easily apply to a periodic point  $x \in U$  rather than a fixed point. If  $T^p(x) = x$ , we can just consider  $T^p$  instead of  $T$ .

Note:  $D_x T^n = D_{T^{n-1}x} T \cdots D_{Tx} T \cdot D_x T$  by the chain rule.

Next, we turn to nonperiodic points. This is the most interesting and important part of the story.

**13.18 DEFINITION.** Let the map  $T^n$  be differentiable at a point  $x \in M$  for all  $n \in \mathbb{Z}$ . Assume that there are numbers  $\chi_1 > \cdots > \chi_m$  and the tangent space  $\mathcal{T}_x M$  is a direct sum of subspaces  $\mathcal{E}_1 \oplus \cdots \oplus \mathcal{E}_m$  such that if  $\mathbf{0} \neq \mathbf{u} \in \mathcal{E}_i$ , then

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \ln \|(D_x T^n) \mathbf{u}\| = \chi_i \quad (31)$$

Then the values  $\chi_i$  are called the **Lyapunov exponents** of the map  $T$  at the point  $x$ . The number  $\dim \mathcal{E}_i$  is called the **multiplicity** of the Lyapunov exponent  $\chi_i$ . The spaces  $\mathcal{E}_j$  are called **characteristic subspaces** at  $x$ .

The subspaces

$$\mathcal{E}^s = \oplus_{\chi_i < 0} \mathcal{E}_i, \quad \mathcal{E}^u = \oplus_{\chi_i > 0} \mathcal{E}_i \quad \text{and} \quad \mathcal{E}^c = \mathcal{E}_j|_{\chi_j=0}$$

are called **stable**, **unstable**, and **neutral** (or **central**) subspaces of  $\mathcal{T}_x M$ , respectively.

We note that the existence of the Lyapunov exponents  $\chi_i$  and the subspaces  $\mathcal{E}_i$  is not guaranteed for any point  $x \in U$ , as an example will show soon. We say that a point  $x$  **has all Lyapunov exponents** if  $\chi_i$  and  $\mathcal{E}_i$  exist.

**13.19 REMARK.** If a point  $x \in U$  has all Lyapunov exponents, then so do points  $T^n(x)$  for all  $n \in \mathbb{Z}$ . Moreover, the points  $T^n(x)$  have the same Lyapunov exponents (with the same multiplicity) as  $x$  does, and the characteristic subspaces are invariant along the trajectory of  $x$ :

$$(D_x T^n)(\mathcal{E}_i(x)) = \mathcal{E}_i(T^n x)$$

for all  $n \in \mathbb{Z}$  and each  $i$ .

In particular, observe that the Lyapunov exponents  $\chi_i$  and their multiplicities  $\dim \mathcal{E}_i$  are  $T$ -invariant functions.

**13.20 EXAMPLE.** Let  $T_A : \text{Tor}^2 \rightarrow \text{Tor}^2$  be a hyperbolic toral automorphism. Then all Lyapunov exponents exist everywhere on  $\text{Tor}^2$ , and they are  $\chi_1 = \ln \lambda^{-1} > 0$  and

$\chi_2 = \ln \lambda < 0$ . The corresponding subspaces  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are parallel to the lines  $L_1$  and  $L_2$ , respectively.

**13.21 EXAMPLE.** Let  $T : X \rightarrow X$  be the baker's transformation of the unit square  $X$ . Let  $X' \subset X$  be the set of points where  $T^n$  is differentiable for all  $n \in \mathbb{Z}$ . Then for every  $x \in X'$  all Lyapunov exponents exist, and they are  $\chi_1 = \ln 2 > 0$  and  $\chi_2 = -\ln 2 < 0$ . The corresponding subspaces  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are parallel to the  $x$  axis and  $y$  axis, respectively. Note that  $T^{\pm n}$  fails to be differentiable on the lines  $x = k/2^n$  and  $y = m/2^n$ , with  $k, m = 0, 1, \dots, 2^n$ . Hence  $m(X') = 1$ , i.e. all Lyapunov exponents exist almost everywhere.

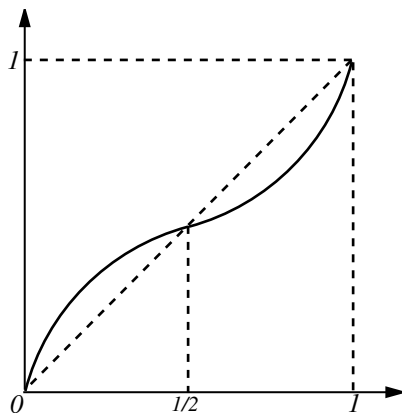


Figure 16: The map in Example 13.22.

**13.22 EXAMPLE.** Let  $T : X \rightarrow X$  be a diffeomorphism of the unit circle  $X = S^1$  given by  $T(x) = x + \frac{1}{3\pi} \sin 2\pi x$ , where  $0 \leq x < 1$  is the cyclic coordinate on  $X$ . We have two fixed points here,  $x_0 = 0$  and  $x_1 = 1/2$ . Lyapunov exponents exist at both fixed points:  $\chi(x_0) = \ln |T'(x_0)| = \ln(5/3) > 0$  and  $\chi(x_1) = \ln |T'(x_1)| = \ln(1/3) < 0$ . Since  $\lambda(x_0) > 0$ , the point  $x_0$  is unstable (a **repeller**). Likewise,  $x_1$  is a stable point (an **attractor**). For any point  $x \in (0, 1/2)$  we have  $T^n(x) \rightarrow x_1$  and  $T^{-n}(x) \rightarrow x_0$  as  $n \rightarrow \infty$ . Hence, by the chain rule, for any nonzero tangent vector  $\mathbf{u} \in \mathcal{T}_x(X)$  we have

$$\lim_{n \rightarrow -\infty} \frac{1}{n} \ln \|(T^n)'_x \mathbf{u}\| = \ln(5/3) \neq \ln(1/3) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|(T^n)'_x \mathbf{u}\|$$

This shows that the limit in (31) does not exist. The same conclusion holds for any  $x \in (1/2, 1)$ .

**13.23 REMARK.** In the last example, Lyapunov exponents only exist at two fixed points,  $x_0$  and  $x_1$ , and nowhere else on  $X$ . Hence, Lyapunov exponents seem to be very scarce on  $X$  in any “reasonable” sense: topologically, with respect to the usual Lebesgue measure,



by simple count. But none of these considerations are relevant in the dynamical sense, where only invariant measures count. The Lebesgue measure is not invariant under  $T$ , so it does not characterize this map. It is not hard to see that any invariant measure in this example is  $p\delta_{x_0} + (1-p)\delta_{x_1}$  with some  $0 \leq p \leq 1$ , compare this to Exercise 2.26(e). Hence, with respect to any invariant measure, Lyapunov exponents do exist almost everywhere!

It is remarkable that the above fact is very general, and this is the content of the Oseledec<sup>8</sup> multiplicative ergodic theorem, which we call shortly Oseledec's theorem.

**13.24 THEOREM (OSELEDEC).** Assume that  $M$  is a compact manifold and  $T : M \rightarrow M$  is a  $C^1$  diffeomorphism preserving a Borel probability measure  $\mu$ . Then there exists a  $T$ -invariant set  $N \subset M$ ,  $\mu(N) = 1$ , such that for every point  $x \in N$  all Lyapunov exponents exist.

First of all, it is enough to prove the theorem for ergodic measures  $\mu$ , because then we can apply Ergodic Decomposition Theorem 7.15 to any nonergodic invariant measure (see Remark 7.21). So, we assume from now on that  $\mu$  is ergodic.

We will only prove this theorem for surfaces, i.e. for  $\dim M = 2$ . The proof will be done in two major steps. The first one consists in proving the following theorem.

**13.25 THEOREM (UPPER LYAPUNOV EXPONENT).** Under the above assumptions, there is a  $\chi_+ \in \mathbb{R}$  such that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \ln \|D_x T^n\| = \chi_+$$

for almost every point  $x \in M$ . Here  $\|D_x T^n\| = \sup_{\|\mathbf{u}\|=1} \|D_x T^n \mathbf{u}\|$  is the norm of the matrix  $D_x T^n$ .

*Proof.* By the chain rule,

$$\|D_x T^{n+m}\| \leq \|D_x T^n\| \|D_{T^n x} T^m\|$$

Let  $F_n(x) = \ln \|D_x T^n\|$ , then

$$F_{n+m}(x) \leq F_n(x) + F_m(T^n x) \tag{32}$$

This condition is referred to as the **subadditivity** of the sequence of functions  $\{F_n\}$ . Now 13.25 follows from the next general statement:

**13.26 THEOREM (SUBADDITIVE ERGODIC THEOREM).** Let  $T : X \rightarrow X$  be a transformation preserving an ergodic measure  $\mu$ , and  $\{F_n\} \in L^1_\mu(X)$ ,  $n \geq 1$ , a sequence of

---

<sup>8</sup>The name is pronounced *Oseledets*.

integrable functions on  $X$  such that (32) holds for almost every  $x \in X$  and all  $n, m \geq 1$ . Then there is a  $\chi \in \mathbb{R} \cup \{-\infty\}$  such that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} F_n(x) = \chi$$

for almost every point  $x \in X$ .

**13.27 REMARKS.** In Theorem 13.25,  $F_1(x) = \ln \|D_x T\|$  is a continuous function on  $M$ , because  $T$  is  $C^1$ . Hence there is an upper bound

$$F_{\max} := \sup_{x \in M} F_1(x) < \infty \quad (33)$$

By iterating the subadditivity condition (32) we obtain for all  $x \in M$

$$\frac{1}{n} F_n(x) \leq \frac{1}{n} \sum_{i=0}^{n-1} F_1(T^i x) \leq F_{\max} \quad (34)$$

Also, by the chain rule  $(D_{T^n x} T^{-n})(D_x T^n) = I$  (the identity matrix), hence

$$1 = \|I\| \leq \|D_x T^n\| \|D_{T^n x} T^{-n}\| \quad (35)$$

therefore

$$\frac{1}{n} F_n(x) = \frac{1}{n} \ln \|D_x T^n\| \geq -\frac{1}{n} \ln \|D_{T^n x} T^{-n}\| \geq -\ln \max_{x \in M} \|D_x T^{-1}\| =: F_{\min} > -\infty \quad (36)$$

As a result,  $\chi_+$  in 13.25 is finite and  $\chi_+ \in [F_{\min}, F_{\max}]$ .

The proof of Subadditive Ergodic Theorem is given in Pollicott, pp. 37–40. We sketch the principal steps here, assuming the bounds (33) and (36) for simplicity.

**13.28 LEMMA (ON SUBADDITIVE SEQUENCES).** If  $\{a_n\}$ ,  $n \geq 1$ , is a sequence of real numbers and  $a_{n+m} \leq a_n + a_m$  for all  $n, m \geq 1$ , then  $a_n/n$  converges as  $n \rightarrow \infty$  and

$$\lim_{n \rightarrow \infty} a_n/n = \inf_{n \geq 1} \{a_n/n\}$$

The proof of Lemma 13.28 is elementary and we leave it as an exercise. We now continue the proof of Subadditive Ergodic Theorem 13.26.

The subadditivity condition (32) and the invariance of  $\mu$  imply  $\int F_{n+m} d\mu \leq \int F_n d\mu + \int F_m d\mu$ . Hence, by 13.28, there is a limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \int_X F_n d\mu = \chi := \inf \left\{ \frac{1}{n} \int_X F_n d\mu \right\}$$

Next, let

$$F_+(x) = \limsup_{n \rightarrow \infty} \frac{1}{n} F_n(x) \quad \text{and} \quad F_-(x) = \liminf_{n \rightarrow \infty} \frac{1}{n} F_n(x)$$

By subadditivity,

$$\frac{1}{n}F_n(x) \leq \frac{1}{n}F_1(x) + \frac{1}{n}F_{n-1}(Tx)$$

Letting  $n \rightarrow \infty$  gives  $F_+(x) \leq F_+(Tx)$  as well as  $F_-(x) \leq F_-(Tx)$ .

**13.29 LEMMA.** If an integrable function  $F : X \rightarrow \mathbb{R}$  has the property  $F(x) \leq F(Tx)$  for almost every  $x \in X$ , then  $F(x) = F(Tx)$  almost everywhere, i.e.  $F$  is  $T$ -invariant.

*Proof.* Let  $\Delta(x) = F(Tx) - F(x)$ . Then  $\Delta(x) \geq 0$  a.e. and  $\int \Delta d\mu = 0$ , hence  $\Delta(x) = 0$  a.e.

Thus, the functions  $F_-(x)$  and  $F_+(x)$  are  $T$ -invariant and therefore constant almost everywhere (by the ergodicity of  $\mu$ ). Denote their values by  $\bar{F}_-$  and  $\bar{F}_+$ , respectively.

We now prove that  $\bar{F}_- = \bar{F}_+$ . Fix a small  $\varepsilon > 0$  and define

$$n_-(x) = \min \left\{ n \geq 1 : \frac{1}{n}F_n(x) < \bar{F}_- + \varepsilon \right\}$$

Choose  $n_\varepsilon \geq 1$  so large that  $\mu\{x : n_-(x) > n_\varepsilon\} < \varepsilon$ . Denote  $A_\varepsilon = \{x : n_-(x) > n_\varepsilon\}$ . Define

$$n(x) = \begin{cases} n_-(x) & \text{if } n_-(x) \leq n_\varepsilon \\ 1 & \text{otherwise} \end{cases}$$

Fix a point  $x \in X$ . For  $N \geq 1$  let

$$k_N = \#\{0 \leq i \leq N-1 : T^i x \in A_\varepsilon\}$$

Define a sequence  $0 = n_0 < n_1 < \dots$  inductively by  $n_{i+1} = n_i + n(T^{n_i}x)$  for all  $i \geq 0$ . Let  $j = \max\{i : n_i \leq N\}$ , and note that  $N - n_j < n_\varepsilon$ . By using the subadditivity condition (32) and our definitions

$$F_N(x) \leq \sum_{i=0}^{j-1} F_{n_{i+1}-n_i}(T^{n_i}x) + F_{N-n_j}(T^{n_j}x)$$

Note that if  $T^{n_i}x \in A_\varepsilon$ , then  $n_{i+1} - n_i = 1$  and  $F_1(T^{n_i}x) \leq F_{\max}$ . Otherwise  $F_{n_{i+1}-n_i}(T^{n_i}x) \leq (n_{i+1} - n_i)(\bar{F}_- + \varepsilon)$ . Also,  $F_{N-n_j}(T^{n_j}x) \leq (N - n_j)F_{\max} \leq n_\varepsilon F_{\max}$ . Hence,

$$F_N(x) \leq (k_N + n_\varepsilon)F_{\max} + (N - k_N - n_\varepsilon)(\bar{F}_- + \varepsilon)$$

By the ergodic theorem,  $k_N/N \rightarrow \mu(A_\varepsilon) < \varepsilon$  as  $N \rightarrow \infty$  for almost every  $x \in X$ , hence  $F_+(x) \leq \bar{F}_- + 2\varepsilon$ . Since  $\varepsilon > 0$  is arbitrary,  $\bar{F}_+ = \bar{F}_-$ .

Thus, for a.e.  $x \in M$

$$\lim_{n \rightarrow \infty} \frac{1}{n}F_n(x) = \bar{F}_+ = \bar{F}_-$$

Finally, by (34) and (36), we have  $F_{\min} \leq \frac{1}{n}F_n(x) \leq F_{\max}$ , and then by Dominated Convergence Theorem  $\bar{F}_+ = \bar{F}_- = \chi$ .

This proves Theorem 13.26, and thus 13.25.  $\square$

This completes the first major step in the proof of 13.24. We turn to the second one starting with some extensions of Theorem 13.25.

### 13.30 COROLLARY (UPPER LYAPUNOV EXPONENTS WITH SHIFT).

(a) We have

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \ln \|D_{T^{-n}x} T^n\| = \chi_+$$

for almost every point  $x \in M$ .

(b) There is a  $\chi_- \in \mathbb{R}$  such that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \ln \|D_x T^{-n}\| = \lim_{n \rightarrow +\infty} \frac{1}{n} \ln \|D_{T^n x} T^{-n}\| = \chi_-$$

for almost every  $x \in M$ .

(c) Lastly,  $\chi_+ + \chi_- \geq 0$ .

*Proof.* Consider the function  $G_n(x) = \ln \|D_{T^{-n}x} T^n\|$ . By the chain rule  $G_{n+m}(x) \leq G_n(x) + G_m(T^{-n}x)$ , so that  $\{G_n(x)\}$  is a subadditive sequence of functions in the sense of (32), but now for the map  $T^{-1}$ , which preserves the same measure  $\mu$ . By Theorem 13.26 there is a  $\chi'$  such that  $\lim_{n \rightarrow \infty} \frac{1}{n} G_n(x) = \chi'$  almost everywhere. Using the notation of the proof of 13.25 we observe that  $G_n(x) = F_n(T^{-n}x)$ , hence  $\int F_n d\mu = \int G_n d\mu$  for every  $n \geq 1$ . Therefore,

$$\chi_+ = \inf \left\{ \frac{1}{n} \int_X F_n d\mu \right\} = \inf \left\{ \frac{1}{n} \int_X G_n d\mu \right\} = \chi'$$

so the claim (a) is proved. To prove (b), we just apply 13.25 and 13.30(a) to the map  $T^{-1}$ . Lastly, according to the chain rule, cf. (35),

$$\frac{1}{n} \ln \|D_x T^n\| + \frac{1}{n} \ln \|D_{T^n x} T^{-n}\| \geq \frac{1}{n} \ln \|D_x T^0\| = 0$$

Taking the limit as  $n \rightarrow +\infty$  yields (c).  $\square$

13.31 COROLLARY. For almost every  $x \in M$  there is a sequence  $n_k \rightarrow \infty$  as  $k \rightarrow \infty$  such that

$$\lim_{k \rightarrow \infty} \frac{1}{2n_k} \ln \|D_{T^{-n_k}x} T^{2n_k}\| = \chi_+$$

A similar statement also holds for  $\chi_-$ .

*Proof.* By Theorem 13.25, for every  $k \geq 1$  and almost every  $y \in M$ , there is an  $m_k(y)$  such that for all  $m \geq m_k(y)$  we have

$$\left| \frac{1}{m} \ln \|D_y T^m\| - \chi_+ \right| < \frac{1}{k}$$

Clearly, there is an  $M_k \geq 1$  such that  $\mu\{y : m_k(y) < M_k\} > 0$ . Fix any such  $M_k$  and let  $A_k = \{y : m_k(y) < M_k\}$ . By Exercise 7.24, the trajectory of almost every point  $x \in M$  visits  $A_k$  infinitely many times (now both in the future and the past, since  $T$  is an automorphism). Hence, for a.e.  $x \in M$  there is an  $n_k > M_k$  such that  $T^{-n_k}x \in A_k$ . This implies

$$\left| \frac{1}{2n_k} \ln \|D_{T^{-n_k}x} T^{2n_k}\| - \chi_+ \right| < \frac{1}{k}$$

This proves Corollary 13.31.  $\square$ .

We now outline the proof of Oseledec's theorem 13.24 referring to Pollicott, pp. 31–36, for some more details.

**13.32 LEMMA (FROM LINEAR ALGEBRA).** If  $B$  is a  $d \times d$  matrix, then there exists a symmetric positive semidefinite matrix  $A = A^T$  such that

- (a)  $A^2 = B^T B$ ;
- (b)  $\|A\mathbf{u}\| = \|B\mathbf{u}\|$  and  $\langle A\mathbf{u}, A\mathbf{v} \rangle = \langle B\mathbf{u}, B\mathbf{v} \rangle$  for all vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ .
- (c) There are orthogonal unit vectors  $e_1, \dots, e_d$  such that  $Be_1, \dots, Be_d$  are also orthogonal vectors.

If  $B$  is not singular, then  $A$  is positive definite.

*Proof.* Since  $B^T B$  is a symmetric positive semidefinite matrix, then it has real eigenvalues  $\beta_i \geq 0$  and corresponding orthogonal unit eigenvectors  $e_1, \dots, e_d$ . Therefore,  $B^T B = QDQ^T$ , where  $D = \text{diag}(\beta_1, \dots, \beta_d)$  and  $Q$  is an orthogonal matrix (whose columns are the vectors  $e_1, \dots, e_d$ ). Let  $D^{1/2} = \text{diag}(\beta_1^{1/2}, \dots, \beta_d^{1/2})$ , then

$$B^T B = QD^{1/2}D^{1/2}Q^T = QD^{1/2}Q^T QD^{1/2}Q^T$$

and we can set  $A = QD^{1/2}Q^T$ . The claim (a) and the fact  $A^T = A$  are obvious. To prove (b), just note that  $\langle B\mathbf{u}, B\mathbf{v} \rangle = \langle B^T B\mathbf{u}, \mathbf{v} \rangle = \langle A^2\mathbf{u}, \mathbf{v} \rangle = \langle A\mathbf{u}, A\mathbf{v} \rangle$ . Now, observe that  $\{e_1, \dots, e_d\}$  are eigenvectors of  $A$  with corresponding eigenvalues  $\lambda_1 = \beta_1^{1/2}, \dots, \lambda_d = \beta_d^{1/2}$ . Hence, by (b) we have  $\langle Be_i, Be_j \rangle = \langle Ae_i, Ae_j \rangle = \lambda_i \lambda_j \langle e_i, e_j \rangle$ . If  $i \neq j$ , then  $Be_i$  and  $Be_j$  must be orthogonal<sup>9</sup>, which proves (c). Setting  $i = j$ , we obtain another useful

---

<sup>9</sup>In linear algebra, the vectors  $e_1, \dots, e_d$  are called the **singular vectors** and  $\lambda_1, \dots, \lambda_d$  the **singular values** of the matrix  $B$ . This is the content of Singular Value Decomposition (SVD) theorem. It is easy to see that if  $B$  is invertible, then the singular values of  $B^{-1}$  are  $\lambda_1^{-1}, \dots, \lambda_d^{-1}$  and its singular vectors are the normalized vectors  $Be_1, \dots, Be_d$ .

identity:  $\|Be_i\| = \lambda_i$ .  $\square$

Now, fix a point  $x \in M$ . For each  $n \in \mathbb{Z}$  denote  $B_n(x) = D_x T^n$  and let  $A_n(x)$  be the corresponding symmetric positive definite matrix defined in 13.32. Let  $\lambda_1^{(n)}(x) \geq \lambda_2^{(n)}(x)$  be the eigenvalues of  $A_n(x)$  and  $E_1^{(n)}(x)$ ,  $E_2^{(n)}(x)$  the corresponding eigenspaces spanned by unit vectors  $e_1^{(n)}(x)$  and  $e_2^{(n)}(x)$ , respectively. Note that  $\lambda_i^{(n)}(x) > 0$  and  $E_1^{(n)}(x)$  is orthogonal to  $E_2^{(n)}(x)$ , i.e.  $e_1^{(n)} \perp e_2^{(n)}$ . We can choose  $e_j^{(0)} \in E_j^{(0)}$  arbitrarily and then adjust the orientation of the other vectors  $e_j^{(n)}$ ,  $n \neq 0$ , so that  $\langle e_j^{(n)}, e_j^{(n-1)} \rangle \geq 0$  for all  $n \in \mathbb{Z}$  and for each  $j = 1, 2$ . Observe that  $\|B_n(x)\| = \|A_n(x)\| = \lambda_1^{(n)}(x)$ .

By 13.25, for almost every  $x \in M$  there exists a limit

$$\chi_1 := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|B_n(x)\| = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \lambda_1^{(n)}(x)$$

Also, by Corollary 13.30(b), there exists a limit

$$\chi_2 := - \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|B_{-n}(x)\| = - \lim_{n \rightarrow \infty} \frac{1}{n} \ln \lambda_1^{(-n)}(x)$$

Note that  $\chi_1 = \chi_+$  and  $\chi_2 = -\chi_-$ , hence by 13.30(c) we have  $\chi_1 \geq \chi_2$ . These are the characteristic Lyapunov exponents in the statement of Oseledec's theorem 13.24.

If  $\chi_1 > \chi_2$ , then we need also to construct the characteristic subspaces  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , which we do next (assuming  $\chi_1 > \chi_2$ ).

**Claim 1.** The following limits exist:

$$e_j^\pm = \lim_{n \rightarrow \infty} e_j^{(\pm n)} \quad \text{for } j = 1, 2 \quad (37)$$

This follows from certain elementary geometric estimates, we refer to Pollicott, pp. 33-34. It is shown there that the convergence in (37) is exponentially fast, i.e.

$$\|e_j^{(\pm n)} - e_j^\pm\| \leq C e^{-\delta n} \quad (38)$$

for some constants  $C, \delta > 0$  and all  $n \geq 0$ ,  $j = 1, 2$ .

Now we define

$$\mathcal{E}_1 = \text{span}(e_2^-) \quad \text{and} \quad \mathcal{E}_2 = \text{span}(e_2^+)$$

We note that while the Lyapunov exponents  $\chi_j$ ,  $j = 1, 2$ , are defined by the dominant eigenvalues  $\lambda_1^{(\pm n)}$ , the characteristic subspaces  $\mathcal{E}_j$  are spanned by the limit eigenvectors  $\lim_{n \rightarrow \infty} e_2^{(\pm n)}$  corresponding to the other (smaller) eigenvalues  $\lambda_2^{(\pm n)}$ , which is an interesting fact.

Fig. 17 illustrates our construction. There we assumed that  $\chi_1 > 0$  and  $\chi_2 < 0$ , so that we would have exponential growth and contraction of vectors.

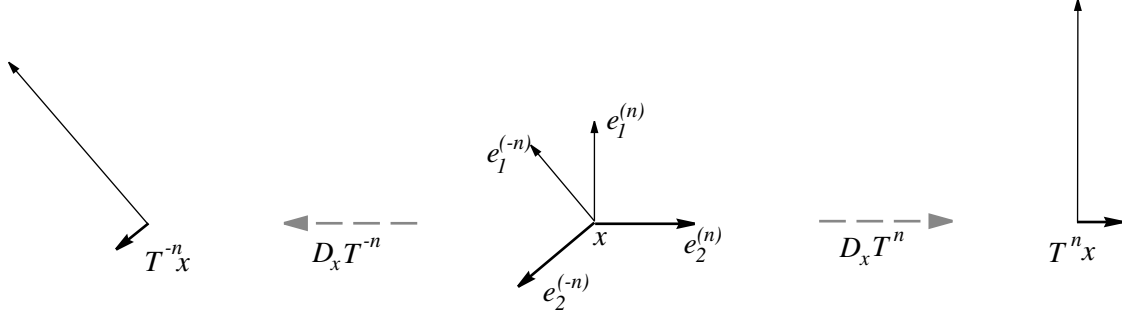


Figure 17: Two orthonormal bases at a point  $x$ , one for  $D_x T^n$  and the other for  $D_x T^{-n}$ . The contracting vectors  $e_2^{(\pm n)}$  are shown bold – they give rise to  $\mathcal{E}_1$  and  $\mathcal{E}_2$ .

**Claim 2.** The vector  $e_1^+$  grows under  $D_x T^n$  exponentially with the rate  $\chi_1$ :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \|D_x T^n(e_1^+)\| = \chi_1$$

Similarly,  $\lim_{n \rightarrow \infty} \frac{1}{n} \ln \|D_x T^{-n}(e_1^-)\| = -\chi_2$ .

This follows from certain elementary geometric estimates, we refer to Pollicott, p. 35.

**Claim 3.** The families of subspaces  $\mathcal{E}_1(x)$  and  $\mathcal{E}_2(x)$  are  $D_x T$ -invariant. Equivalently,  $D_x T(e_2^\pm(x)) = \beta^\pm(x) e_2^\pm(Tx)$  for some scalar functions  $\beta^\pm(x)$ .

Indeed, consider an arbitrary nonzero vector  $\mathbf{u}$  and let  $\mathbf{u} = c_1^+ e_1^+ + c_2^+ e_2^+$  be its decomposition in the basis  $\{e_1^+, e_2^+\}$ , and  $\mathbf{u} = c_1^{(n)} e_1^{(n)} + c_2^{(n)} e_2^{(n)}$  be its decomposition in the basis  $\{e_1^{(n)}, e_2^{(n)}\}$  for all  $n \geq 0$ . Of course,  $c_j^{(n)} \rightarrow c_j^+$  as  $n \rightarrow \infty$ , for  $j = 1, 2$ , according to Claim 1. If  $c_1^+ \neq 0$ , then  $|c_1^{(n)}| \geq |c_1^+|/2 > 0$  for all large enough  $n$ , hence  $D_x T^n(\mathbf{u})$  will grow exponentially at the rate  $\chi_1$  as  $n \rightarrow \infty$  by Claim 2, precisely

$$c_1^+ \neq 0 \implies \lim_{n \rightarrow \infty} \frac{1}{n} \ln \|D_x T^n(\mathbf{u})\| = \chi_1 \quad (39)$$

Now, when  $c_1^+ = 0$ , then  $|c_1^{(n)}| < \text{const } e^{-\delta n}$  by (38). By a little more detailed estimation, one obtains

$$c_1^+ = 0 \implies \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \|D_x T^n(\mathbf{u})\| \leq \max\{\chi_1 - \delta, \chi_2\} \quad (40)$$

Since the limits in (39) and (40) are different, Claim 3 is proved. Note: the families of subspaces spanned by the limit dominant eigenvectors,  $e_1^+(x)$  and  $e_1^-(x)$ , are generally not (!) invariant under  $D_x T$ .

**Claim 4.** The spaces  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are distinct, i.e. the vectors  $e_2^+$  and  $e_2^-$  are not parallel.

Indeed, suppose  $\mathcal{E}_1 = \mathcal{E}_2$ . Let  $\mathbf{v} \in \mathcal{T}_{T^{-n}x}M$  be an arbitrary unit vector at the point  $T^{-n}x$ . Let  $\mathbf{v} = c_1 e_1^{(n)}(T^{-n}x) + c_2 e_2^{(n)}(T^{-n}x)$  be its decomposition in the basis  $\{e_1^{(n)}(T^{-n}x), e_2^{(n)}(T^{-n}x)\}$ . Then

$$\mathbf{u} := D_{T^{-n}x}T^n(\mathbf{v}) = \frac{c_1}{\lambda_1^{(-n)}(x)}e_1^{(-n)}(x) + \frac{c_2}{\lambda_2^{(-n)}(x)}e_2^{(-n)}(x)$$

Recall that  $\frac{1}{n} \ln \lambda_1^{(-n)}(x) \rightarrow -\chi_2$  and  $\frac{1}{n} \ln \lambda_2^{(-n)}(x) \rightarrow -\chi_1$  as  $n \rightarrow \infty$ . Our assumption  $\mathcal{E}_1 = \mathcal{E}_2$  and (38) imply  $\mathbf{u} = c'_1 e_1^{(n)} + c'_2 e_2^{(n)}$  with some  $c'_1, c'_2$  satisfying

$$\limsup_{n \rightarrow \infty} \sup_{\mathbf{v}} \frac{1}{n} \ln |c'_1| \leq \max\{\chi_2, \chi_1 - \delta\}$$

and

$$\limsup_{n \rightarrow \infty} \sup_{\mathbf{v}} \frac{1}{n} \ln |c'_2| \leq \chi_1$$

Since  $e_1^{(n)}$  and  $e_2^{(n)}$  are eigenvectors for  $A_n(x)$ , we see that

$$D_x T^n \mathbf{u} = D_{T^{-n}x} T^{2n}(\mathbf{v}) = c'_1 \lambda_1^{(n)} e_1^{(-n)}(T^n x) + c'_2 \lambda_2^{(n)} e_2^{(-n)}(T^n x)$$

Therefore,

$$\limsup_{n \rightarrow \infty} \frac{1}{2n} \ln \|D_{T^{-n}x} T^{2n}\| \leq \max\left\{\frac{\chi_1 + \chi_2}{2}, \chi_1 - \frac{\delta}{2}\right\} < \chi_1$$

which contradicts Corollary 13.31. Claim 4 is proved.

We are now ready to show that for every nonzero vector  $\mathbf{u} \in \mathcal{E}_1$

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \ln \|D_x T^n(\mathbf{u})\| = \chi_1 \quad (41)$$

as required by the definition of Lyapunov exponents 13.18. For  $n \rightarrow +\infty$ , this follows from (39) (note: by Claim 4 the vector  $\mathbf{u}$  cannot be parallel to  $e_2^+$ , hence it has a nonzero projection onto  $e_1^+$ ). Now, recall that by Claim 3 we have  $\|D_x T(\mathbf{u})\| = |\beta^-(x)| \|\mathbf{u}\|$  and  $\|D_x T^{-1}(\mathbf{u})\| = |1/\beta^-(T^{-1}x)| \|\mathbf{u}\|$ . Let  $G(x) = \ln |\beta^-(x)|$ , then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \|D_x T^n(\mathbf{u})\| = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} G(T^i x) = G_+(x)$$

in the notation of Section 9, and we just have seen that  $G_+(x) = \chi_1$  almost everywhere. On the other hand,

$$\lim_{n \rightarrow -\infty} \frac{1}{n} \ln \|D_x T^n(\mathbf{u})\| = \lim_{n \rightarrow -\infty} \frac{1}{|n|} \sum_{i=1}^{|n|} G(T^{-i}x) = G_-(x)$$



and we have  $G_-(x) = G_+(x)$  almost everywhere by 9.4. This completes the proof of (41). The corresponding limit for vectors  $\mathbf{u} \in \mathcal{E}_2$  is computed similarly. Hence we proved Oseledec Theorem in the main case  $\chi_1 > \chi_2$ .

If  $\chi_1 = \chi_2$ , then the spaces  $\mathcal{E}_j$ ,  $j = 1, 2$ , need not be constructed, and the argument is much simpler, we refer to Pollicott, p. 36, and leave the details as an exercise.  $\square$

Lyapunov exponents are the key tool in the study of smooth dynamics. Interestingly, they exist even for maps that are only differentiable almost everywhere, rather than everywhere. The following extension of Oseledec's theorem takes care of such maps:

**13.33 THEOREM (OSELEDEC THEOREM, EXTENDED).** Let  $U \subset M$  be an open subset of a manifold  $M$ . Let  $T : U \rightarrow M$  be a  $C^1$  diffeomorphism of  $U$  onto  $T(U)$  preserving a probability measure  $\mu$  on  $M$ . Assume that

$$\int_M \ln^+ \|D_x T\| d\mu < \infty \quad \text{and} \quad \int_M \ln^+ \|D_x T^{-1}\| d\mu < \infty \quad (42)$$

where  $\ln^+ a = \max\{0, \ln a\}$ . Then there exists a  $T$ -invariant subset  $N \subset M$ ,  $\mu(N) = 1$ , such that for every point  $x \in N$  all Lyapunov exponents exist.

Note: the assumption (42) is necessary to ensure the integrability of functions  $F_n(x)$  in Subadditive Ergodic Theorem 13.26.

Oseledec Theorem contains a clause that we have omitted so far. We state it now without proof. It deals with angles between the characteristic subspaces  $\mathcal{E}_j$ . Even though these subspaces are not necessarily orthogonal to each other, the angle between them cannot become too small, in the following sense.

**13.34 ADDENDUM.** Let  $x \in M$  be a point where all Lyapunov exponents  $\chi_1 > \dots > \chi_m$  exist. For any subset  $S \subset \{1, \dots, m\}$  let  $\gamma_S(x)$  denote the angle between the spaces  $\oplus_{j \in S} \mathcal{E}_j$  and  $\oplus_{j \notin S} \mathcal{E}_j$  in  $\mathcal{T}_x M$ . Then for almost every  $x \in M$

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \ln \gamma_S(T^n x) = 0$$

i.e. the angles between the characteristic subspaces can only approach zero very slowly (more slowly than any exponential function).

**13.35 THEOREM (LYAPUNOV EXPONENTS VERSUS VOLUME).** Let  $J_n(x) = |\det D_x T^n|$  be the Jacobian of the map  $T^n$  at  $x$  (this is the factor by which  $T^n$  changes volume in an infinitesimal neighborhood of  $x \in M$ ). Then for almost every  $x$

$$\lim_{n \rightarrow \pm\infty} \frac{1}{n} \ln J_n(x) = \sum_{j=1}^m \chi_j \dim \mathcal{E}_j$$

i.e. the asymptotic rate of change of volume equals the sum of all Lyapunov exponents (counting multiplicity).

*Proof.* For simplicity, assume that  $\dim M = 2$  and  $\chi_1 > \chi_2$ . Pick any nonzero vectors  $e_1 \in \mathcal{E}_1$  and  $e_2 \in \mathcal{E}_2$ . The area of the parallelogram spanned by  $e_1, e_2$  equals  $\|e_1\| \|e_2\| \sin \gamma(x)$ , where  $\gamma(x)$  is the angle between  $\mathcal{E}_1$  and  $\mathcal{E}_2$  in  $\mathcal{T}_x M$ . Therefore,

$$J_n(x) = \frac{\|D_x T^n e_1\| \|D_x T^n e_2\| \sin \gamma(T^n x)}{\|e_1\| \|e_2\| \sin \gamma(x)}$$

Taking the logarithm, dividing by  $n$  and letting  $n \rightarrow \pm\infty$  proves the theorem. Note that the term containing  $\sin \gamma(T^n x)$  is eliminated by 13.34. In the general case, when  $\dim M > 2$ , some elementary but boring geometric estimates must be involved, so we do not elaborate.  $\square$

**13.36 COROLLARY.** Assume that  $\ln J_1(x) \in L^1_\mu(M)$ . Then

$$\int_M \ln J_1(x) d\mu = \sum_{j=1}^m \chi_j \dim \mathcal{E}_j$$

i.e. the average one-step rate of change of volume equals the sum of all Lyapunov exponents (counting multiplicity).

*Proof.* By the chain rule,

$$J_n(x) = J_1(x) \cdots J_1(T^i x)$$

Taking the logarithm, dividing by  $n$  and letting  $n \rightarrow \pm\infty$  proves the theorem in view of 13.35 and Ergodic Theorem 9.2.

**13.37 COROLLARY.** Let the invariant measure  $\mu$  be absolutely continuous with density  $f(x)$  with respect to the Lebesgue measure (volume). Assume that  $\ln f(x) \in L^1_\mu(M)$ . Then

$$\sum_{j=1}^m \chi_j \dim \mathcal{E}_j = 0$$

i.e. the sum of all Lyapunov exponents vanishes.

*Proof.* Note that  $J_1(x) = f(x)/f(Tx)$ , hence

$$\int_M \ln J_1(x) d\mu = \int_M \ln f(x) d\mu - \int_M \ln f(Tx) d\mu = 0$$

by the invariance of  $\mu$ .  $\square$

## 14 Dynamical Systems with Continuous Time

So far we have studied measurable maps  $T : X \rightarrow X$  and their iterations  $T^n$ ,  $n \in \mathbb{Z}$ , interpreting  $n$  as time. Our time variable  $n$  was restricted to integers only. Naturally, in real life and in physical sciences, time is a continuous variable that can take all real values.

If we allow the time variable  $t$  take all real values, then for any initial point  $x \in X$ , its image  $S^t x$  must be defined for all  $t \in \mathbb{R}$ . Hence, we need to define a one-parameter family of transformations  $S^t : X \rightarrow X$ , where  $t \in \mathbb{R}$ . Note that  $S^{t+s}x = S^t(S^s x)$  for every  $x \in X$ , so that  $S^{t+s} = S^t \circ S^s$ . Hence the transformations  $S^t$  make a *group* with respect to composition. In particular,  $S^0 = \text{id}$ , the identity map.

There are two ways to look at such a system  $S^t : X \rightarrow X$ . One way is to fix some  $t$ 's and consider the corresponding maps  $S^t$  on  $X$ , hence the entire system is viewed as an uncountable collection of maps of  $X$  into itself, which are related to each other by the group rule  $S^{t+s} = S^t \circ S^s$ . The other way is to fix some  $x$ 's and consider the corresponding trajectories (orbits)  $S^t(x)$ ,  $-\infty < t < \infty$ . Often these orbits are continuous (or even smooth) curves in  $X$ . Note that orbits of different points either coincide or are disjoint, hence they foliate  $X$ .

**14.1 DEFINITION.** Let  $X$  be a measurable space. A dynamical system with **continuous time** (also called a **flow**) is a one-parameter family  $\{S^t\}$  of measurable maps on  $X$  such that

- (a)  $S^{t+s} = S^t \circ S^s$  (the group property);
- (b) for every measurable function  $F : X \rightarrow \mathbb{R}$  the function  $G(x, t) = F(S^t x)$  is measurable on the product space<sup>10</sup>  $X \times \mathbb{R}$ .

The condition (b) ensures that the orbits  $S^t x$  of individual points  $x \in X$  are not “too wild”, but “nice enough”.

Commonly, flows are defined by differential equations as explained next.

**14.2 MAIN CLASS OF FLOWS.** Let  $M$  be a smooth Riemannian manifold and  $\mathbf{v}(x)$  a smooth vector field<sup>11</sup> on  $M$ . Then the differential equation

$$\dot{x}(t) = \mathbf{v}(x(t)) \tag{43}$$

has a smooth and unique solution  $x(t)$  passing through every point  $x \in M$ .

---

<sup>10</sup>The  $\sigma$ -algebra on the product space  $X \times \mathbb{R}$  is generated by sets  $A \times B$ , where  $A \subset X$  is a measurable set and  $B \subset \mathbb{R}$  is a Borel set.

<sup>11</sup>This is a function assigning a tangent vector  $\mathbf{v}(x) \in \mathcal{T}_x(M)$  to every point  $x \in M$ , which smoothly depends on  $x$ .

Now, for any  $x \in M$  and  $t \in \mathbb{R}$  we define  $S^t(x) = x(t)$ , where  $x(t)$  is the solution of the above differential equation with initial condition  $x(0) = x$ .

One can interpret  $\mathbf{v}(x)$  as a **velocity** vector field on  $M$ .

**14.3 REMARK.** The solution  $x(t)$  of the equation (43) with initial condition  $x(0) = x$  always exists for small enough  $t$ . But it may fail to exist globally, i.e. for large  $t \in \mathbb{R}$ .

One can remedy this situation by requiring that  $M$  be a complete manifold and  $\mathbf{v}(x)$  be a bounded vector field. This is certainly true whenever  $M$  is a smooth compact manifold, for example. We are not going into this technical discussion, and will always assume that the solutions of (43) exist globally, so that  $S^t$  is defined for all  $t \in \mathbb{R}$ .

#### 14.4 EXAMPLES.

- (a) Let  $X = \mathbb{R}^d$  and  $\mathbf{v} \in \mathbb{R}^d$  a fixed vector. Then the differential equation  $\dot{x} = \mathbf{v}$  defines a flow  $S^t$ , which is simply a parallel translation of  $\mathbb{R}^d$  along the vector  $\mathbf{v}$  with constant speed;
- (b) Let  $X = S^1$  be a unit circle and  $a \in \mathbb{R}$  a fixed number. Then the equation  $\dot{x} = a$  defines a flow on  $S^1$ , which is a continuous rotation (revolution) of  $S^1$  at constant speed.

Many concepts and results that we have developed for maps, easily extend to flows.

**14.5 DEFINITION.** A flow  $S^t : X \rightarrow X$  **preserves a measure**  $\mu$  if  $\mu(S^t(A)) = \mu(A)$  for all measurable subsets  $A \subset X$  and all  $t \in \mathbb{R}$ .

Note: both flows in Examples 14.4 preserve the Lebesgue measure. In the case (a) that measure is infinite, though, so we must say that  $S^t$  has no invariant measures.

**14.6 DEFINITION.** A measurable set  $A \subset X$  is invariant under the flow  $S^t$  if  $S^t(A) = A$  for all  $t \in \mathbb{R}$ . A function  $f : X \rightarrow \mathbb{R}$  is invariant under the flow  $S^t$  if  $f(S^t x) = f(x)$  for all  $t \in \mathbb{R}$  and all  $x \in X$  (a function is almost everywhere invariant if the above holds for a.e.  $x \in X$ ).

Note: in Example 14.4(b) the only invariant sets are  $X = S^1$  and  $\emptyset$ . All invariant functions are constant.

**14.7 DEFINITION.** A flow  $S^t : X \rightarrow X$  is **ergodic** with respect to an invariant measure  $\mu$  (we also say that a measure  $\mu$  is ergodic) if any  $S^t$ -invariant set  $A \subset X$  has measure 0 or 1. Equivalently:  $S^t$  is ergodic if any a.e. invariant function is a.e. constant. (Compare this to 7.12 and 7.23.)

14.8 REMARK The flow in Example 14.4(b) is ergodic. It is interesting to note that for some  $t \neq 0$  the map  $S^t$  will be an identity map, so it cannot be ergodic. Hence, the ergodicity of the flow  $S^t$  does not imply the ergodicity of any particular map  $S^t$ .

The map  $S^t$  (for a given  $t \in \mathbb{R}$ ) is called the **time  $t$  map**.

14.9 DEFINITION. For a function  $f : X \rightarrow \mathbb{R}$  its (forward and backward) time averages  $f_{\pm}(x)$  are defined by

$$f_{\pm}(x) = \lim_{T \rightarrow \pm\infty} \frac{1}{T} \int_0^T f(S^t(x)) dt$$

(Compare this to 9.1 and 9.4.)

14.10 THEOREM (BIRKHOFF-KHINCHIN ERGODIC THEOREM FOR FLOWS). Let  $S^t : X \rightarrow X$  preserve a measure  $\mu$  and  $f : X \rightarrow \mathbb{R}$  be an integrable function, i.e.  $f \in L^1_{\mu}(X)$ . Then

- (a) For almost every point  $x \in X$  the functions  $f_{\pm}(x)$  defined above exist and coincide, i.e.  $f_+(x) = f_-(x)$  a.e.
- (b) The function  $f_{\pm}(x)$  is  $S^t$ -invariant. Moreover, if  $f_{\pm}(x)$  exists, then  $f_{\pm}(S^t x)$  exists for all  $t$  and  $f_{\pm}(S^t x) = f_{\pm}(x)$ .
- (c)  $f_{\pm}$  is integrable and

$$\int_X f_{\pm} d\mu = \int_X f d\mu$$

- (d) If  $\mu$  is ergodic, then  $f_{\pm}(x)$  is constant almost everywhere and

$$f_{\pm}(x) = \int_X f d\mu \quad \text{for a.e. } x \in X$$

14.11 DEFINITION. A flow  $S^t : X \rightarrow X$  is **mixing** with respect to an invariant measure  $\mu$  (we also say that a measure  $\mu$  is mixing) if for any sets  $A, B \subset X$  we have

$$\lim_{t \rightarrow \pm\infty} \mu(A \cap S^t(B)) = \mu(A) \mu(B)$$

Equivalently:  $S^t$  is mixing if for any functions  $f, g \in L^2_{\mu}(X)$  we have

$$\lim_{t \rightarrow \pm\infty} \langle (f \circ S^t) g \rangle = \langle f \rangle \langle g \rangle$$

(Compare this to 11.5 and 11.11.)

14.12 REMARK. If a flow  $S^t$  is mixing, then every map  $S^t$ ,  $t \neq 0$ , is also mixing. (Compare this to 14.8.) On the other hand, if  $S^t$  is mixing for some  $t \neq 0$ , this does not imply

that the flow  $S^t$  is mixing. (Compare this to 14.13.)

**14.13 DEFINITION.** A flow  $S^t$  is called a **Bernoulli flow** if the map  $S^t$  is Bernoulli for every  $t \in \mathbb{R}$ .

It is interesting (but we do not give a proof) that if  $S^t$  is Bernoulli for some  $t \in \mathbb{R}$ , then  $S^t$  is Bernoulli for every  $t \in \mathbb{R}$ .

**14.14 EXAMPLE (LINEAR FLOW ON A TORUS).** Let  $\text{Tor}^d$  be a  $d$ -dimensional torus, cf. 12.1, and  $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$  a fixed vector. Define a flow  $S_{\mathbf{a}}^t$  on  $\text{Tor}^d$  by

$$S_{\mathbf{a}}^t(\mathbf{x}) = \mathbf{x} + t\mathbf{a} \pmod{1}$$

i.e.  $S_{\mathbf{a}}^t(x_1, \dots, x_d) = (x_1 + ta_1, \dots, x_d + ta_d)$ , and each  $x_i + td_i$  is taken modulo 1. (Compare this to 12.2.)

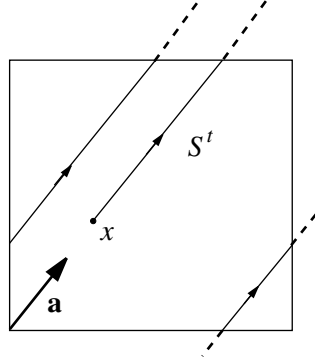


Figure 18: Linear flow on the torus.

Alternatively, one can define the linear flow on  $\text{Tor}^d$  by differential equations  $\dot{x}(t) = \mathbf{a}$  in the spirit of 14.2. Note that every coordinate  $x_i$  changes periodically with period  $1/a_i$  (hence, with frequency is  $a_i$ ). The flow  $S^t$  is also called a **conditionally periodic flow**, and the numbers  $a_i$  are called **frequencies**. It is easy to verify, that  $S_{\mathbf{a}}^t$  preserves the Lebesgue measure  $m$  on  $\text{Tor}^d$ , see also 14.18 below.

**14.15 THEOREM.** The following are equivalent:

- (a)  $S_{\mathbf{a}}^t$  is ergodic;
- (b)  $\forall \mathbf{x} \in \text{Tor}^d$  the orbit  $\{S_{\mathbf{a}}^t(x)\}$ ,  $t > 0$ , is dense in  $\text{Tor}^d$ ;
- (c) the numbers  $a_1, \dots, a_d$  are rationally independent, i.e.

$$m_1 a_1 + m_2 a_2 + \dots + m_d a_d \neq 0 \tag{44}$$

for any integers  $m_1, \dots, m_d$  unless  $m_1 = \dots = m_d = 0$ .

The proof of this theorem very much repeats that of 12.5. The differences are minor and can be easily worked out.

14.16 REMARK. Let  $d = 2$  in the above theorem. Then either  $S_{\mathbf{a}}^t$  is ergodic, or  $a_1/a_2$  is a rational number. In the latter case there is a  $T > 0$  such that  $S_{\mathbf{a}}^T(x) = x$  for every point  $x \in \text{Tor}^2$ . Hence, we arrive at an alternative: on a 2-D torus all the orbits are either periodic or dense.

14.17 REMARK. The flow  $S_{\mathbf{a}}^t$  is never mixing. In fact, all maps  $S_{\mathbf{a}}^t$ ,  $t \in \mathbb{R}$ , are isometries, i.e. they preserve distances and angles on  $\text{Tor}^d$ , so this flow is a very regular (as opposed to chaotic) dynamical system, despite its ergodicity for some  $\mathbf{a} \in \mathbb{R}^d$ .

There is a simple criterion for the invariance of the Lebesgue measure under a smooth flow:

14.18 LEMMA. Let  $X \subset \mathbb{R}^d$  be an open domain and  $S^t$  a flow defined by a differential equation, see 14.2. Then  $S^t$  preserves the Lebesgue measure  $m$  on  $X$  iff the divergence of the vector field  $\mathbf{v}$  vanishes:

$$\text{div } \mathbf{v} = \frac{\partial v_1}{\partial x_1} + \cdots + \frac{\partial v_d}{\partial x_d} = 0$$

at every point  $x \in X$ . (Compare this to 5.6(a).)

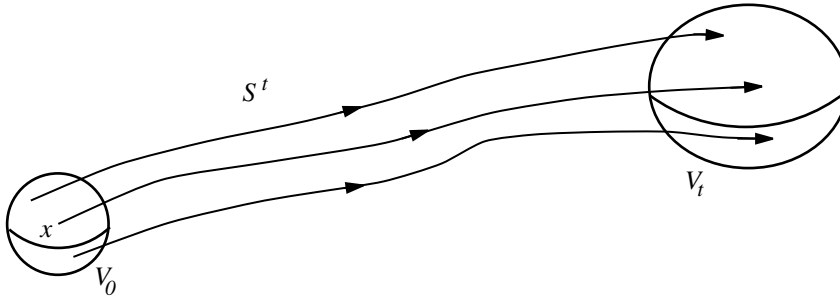


Figure 19: Change of volume by the flow  $S^t$ .

*Proof.* It is a well known fact in vector analysis that  $\text{div } \mathbf{v}$  is the rate of change of volume by the vector field  $\mathbf{v}$ . More precisely, let  $x(s)$ ,  $0 \leq s \leq t$ , be a segment of an orbit of the flow. Take an infinitesimal ball  $B_0$  of volume  $V_0$  around the point  $x(0)$  and denote by  $V(t)$  the volume of the image  $S^t(B_0)$ , see Fig. 19. Then

$$\frac{V_t}{V_0} = \exp \left[ \int_0^t \text{div } \mathbf{v}(x(s)) ds \right]$$

This proves the lemma.  $\square$

The following fact extends 14.18 to more general flows and measures.

**14.19 THEOREM (LIOUVILLE).** Let  $X \subset \mathbb{R}^d$  be an open domain and  $S^t$  a flow defined by a differential equation, see 14.2. Then  $S^t$  preserves a measure  $\mu$  with a continuous density  $f(x)$  on  $X$  iff the divergence of the vector field  $f\mathbf{v}$  vanishes:

$$\operatorname{div}(f\mathbf{v}) = \frac{\partial(fv_1)}{\partial x_1} + \cdots + \frac{\partial(fv_d)}{\partial x_d} = 0$$

at every point  $x \in X$ .

*Proof.* In terms of the previous lemma, the invariance of the measure with density  $f(x)$  means that

$$f(x(t)) V_t = f(x(0)) V_0$$

hence

$$f(x(t)) \exp \left[ \int_0^t \operatorname{div} \mathbf{v}(x(s)) ds \right] = f(x(0))$$

Differentiating with respect to  $t$  and cancelling the exponential factor yields

$$\langle \operatorname{grad} f, \mathbf{v} \rangle + f \operatorname{div} \mathbf{v} = 0$$

which is equivalent to  $\operatorname{div}(f\mathbf{v}) = 0$ .  $\square$

**14.20 EXAMPLE.** In classical mechanics, the motion of objects is governed by Newtonian equations constructed in the following way. Consider a system of  $n$  objects (particles) with masses  $m_1, \dots, m_n$ . Denote their coordinates by  $q_1, \dots, q_n$  and momenta by  $p_1, \dots, p_n$  (for simplicity, let all the particles move on a line, so that  $q_i$ 's and  $p_i$ 's are one-dimensional variables). The Newtonian equations are

$$\dot{q}_i = p_i/m_i \quad \text{and} \quad \dot{p}_i = F_i \tag{45}$$

for  $1 \leq i \leq n$ , where  $F_i$  is the force acting on the  $i$ -th particle. Classical forces (gravitational, etc.) only depend on the particle positions, not velocities, i.e.  $F_i = F_i(q_1, \dots, q_n)$ . For example, the gravitational force is obtained by

$$F_i(q_1, \dots, q_n) = -\frac{\partial}{\partial q_i} U(q_1, \dots, q_n)$$

where  $U(q_1, \dots, q_n)$  is the potential energy of the system.

The differential equations (45) define a flow on the  $2n$ -dimensional space with coordinates  $(q_1, \dots, q_n, p_1, \dots, p_n)$ , which is called the **phase space** of the mechanical system.



14.21 REMARK. The resulting flow on the phase space preserves the Lebesgue measure by Lemma 14.8 (indeed, all the partial derivatives involved in 14.8 vanish). This is a fundamental fact in classical mechanics. It is also called Liouville Theorem.

14.22 HAMILTONIAN SYSTEMS. More generally, the laws of classical mechanics can be expressed by a system of differential equations

$$\dot{q}_i = \frac{\partial}{\partial p_i} \mathcal{H} \quad \dot{p}_i = -\frac{\partial}{\partial q_i} \mathcal{H} \quad (46)$$

where the function  $\mathcal{H}(q_1, \dots, q_n, p_1, \dots, p_n)$  is the so called **Hamiltonian**. In the above example with gravitational forces

$$\mathcal{H}(q_1, \dots, q_n, p_1, \dots, p_n) = \sum_{i=1}^n \frac{p_i^2}{2m_i} + U(q_1, \dots, q_n)$$

This is a classical expression for the total energy of the system: it is the sum of the kinetic energy and potential energy.

Lastly, note that a Hamiltonian flow always preserves the Hamiltonian itself, i.e.  $\mathcal{H}$  is an invariant function under the flow. This follows from (46) by

$$\frac{d}{dt} \mathcal{H}(q_1(t), \dots, q_n(t), p_1(t), \dots, p_n(t)) = \sum_{i=1}^n \dot{q}_i \times \frac{\partial}{\partial q_i} \mathcal{H} + \sum_{i=1}^n \dot{p}_i \times \frac{\partial}{\partial p_i} \mathcal{H} = 0$$

In physics this fact is called the **conservation of energy**.