# Dynamical Systems, MA 797

## 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) $\{B_i\}_{i=1}^{\infty} \in \mathcal{B} \implies \cup_{i=1}^{\infty} B_i \in \mathcal{B}$

(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 $m : \mathcal{B} \to \mathbb{R} \cup \{+\infty\}$ 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 $i \neq j \implies m\left(\cup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} m(B_i)$

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 $m : \mathcal{B} \to \mathbb{R} \cup \{+\infty\}$ 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. $\{B_i\}_{i=1}^{\infty} \in \mathcal{B} \implies \cap_{i=1}^{\infty} B_i \in \mathcal{B}$ [to prove this, just use the formula $\cap_{i=1}^{\infty} B_i = \left(\cup_{i=1}^{\infty} B_i^c\right)^c$ ]. 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 $f : X \to \mathbb{R}$ is an integrable nonnegative function. Then

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

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, $X = \{1, 2, \ldots\}$. Then any measure $m$ on $(X, 2^X)$ is determined by the numbers $p_i = m(\{i\})$, $i \in X$, because

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

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, $X = \{1, 2, \ldots, n\}$. 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 $0 \le p \le 1$. 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[1] 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).

[1]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 $m_0 : \mathcal{A} \to \mathbb{R}$ 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 $m_0 \left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m_0(A_i)$.

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 $m_0 : \mathcal{A} \to \mathbb{R}$ 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 $T : X \to X$ 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 $0 \le x < 1$. 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 $T : X \to X$ 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**.

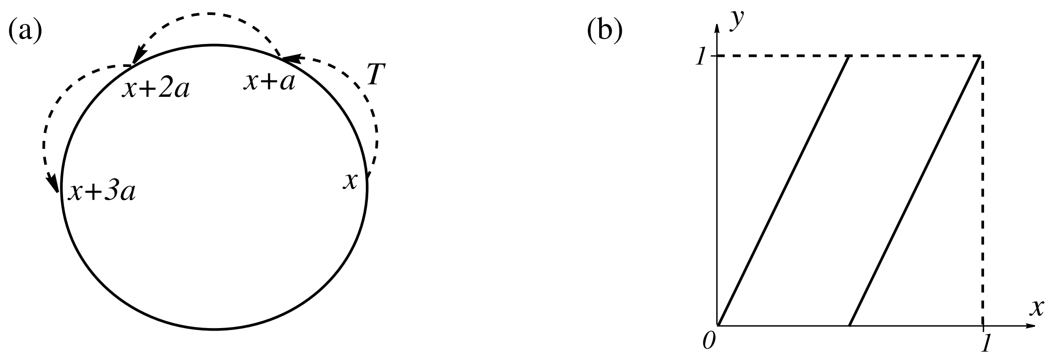


Figure 1: The circle rotation (a) and the doubling map (b).

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

$$T^n = \underbrace{T \circ T \circ \cdots \circ T}_{n}$$

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

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

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**[2]. We think of $x_0$ as the **initial point** and of $x_n = T^n(x_0)$

[2]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 $T : X \to X$ 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 $T : X \to X$ 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 $T : X \to X$ 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 $T : X \to X$ 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 $T : X \to X$ 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 $T : X \to X$ a measurable transformation. Then, for every $n \geq 1$, the function $\mu_n : \mathcal{B} \to \mathbb{R}$ given by

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

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

*Proof*: This easily follows from 2.8. □

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 $T : X \to X$ induces a map $T : \mathcal{M} \to \mathcal{M}$. For every $\mu \in \mathcal{M}$ the measure $T\mu$ is defined by

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

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 $T : X \to X$ 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 $T : X \to X$ 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. □

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 $x_0 = x$, $x_1 = T(x), \ldots, x_{n-1} = T^{n-1}(x)$. 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 $(\delta_{x_0} + \delta_{x_1} + \cdots + \delta_{x_{n-1}})/n$ 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 $T : X \to X$ has many invariant measures. If $T : X \to X$ 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 $T : X \to X$ defined by $T(x) = x + 1$. Show that $T$ has no invariant measures.

(b) Let $X = \mathbb{R}$ and $T : X \to X$ defined by $T(x) = x + a$ with a constant $a \neq 0$. Show that $T$ has no invariant measures[3].

(c) Let $X = (0, 1)$ and $T : X \to X$ defined by $T(x) = x^2$. Show that $T$ has no invariant measures.

(d) Let $X = [0, 1]$ and $T : X \to X$ 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 $T : X \to X$ defined by $T(x) = x^2$. Find all $T$-invariant measures.

(f) Let $X = \{1, 2, \ldots, m\}$ be a finite set, and $T : X \to X$ a permutation (i.e., a bijection of $X$). Describe all $T$-invariant measures.

[3] 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 $x = 0.i_0 i_1 i_2 \dots$ be the infinite decimal fraction representing $x$. Then $10x = i_0.i_1 i_2 \dots$, hence $T(x) = 0.i_1 i_2 i_3 \dots$. It is just as easy to see that $T^n(x) = 0.i_n i_{n+1} i_{n+2} \dots$ 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 $x = 0.i_0 i_1 i_2 \dots$ 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 $x = (0.i_0 i_1 i_2 \dots)_2$ where $i_n$, $n \geq 0$, are binary digits, i.e. zeroes and ones. Then $2x = (i_0.i_1 i_2 \dots)_2$, hence $T(x) = (0.i_1 i_2 i_3 \dots)_2$. It is again easy to see that $T^n(x) = (0.i_n i_{n+1} i_{n+2} \dots)_2$ for all $n \geq 1$.

2.29 Remark. We note how $T$ acts on the sequence of digits $i_0 i_1 i_2 \dots$ 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 $f : X \to \mathbb{R}$ is denoted by $C(X)$. (Very rarely, we will need to consider complex-valued continuous functions $f : X \to \mathbb{C}$, 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 $J_\mu : C(X) \to \mathbb{R}$ defined by

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

3.2 FACT. For each probability measure $\mu \in \mathcal{M}(X)$ the map $J_\mu : C(X) \to \mathbb{R}$ 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 $J : C(X) \to \mathbb{R}$ 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 $n \to \infty$

$$\mu_n \to \mu \qquad \Longleftrightarrow \qquad \int_X f\, d\mu_n \to \int_X f\, d\mu \quad \forall f \in C(X)$$

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

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

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[4] 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 $N \ge 1$, let $x_i = i/N$ for $1 \le i \le N$. Consider the measure

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

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

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

[4] A countable dense subset of $C(X)$ exists whenever $X$ is a metrisable compact Hausdorff space.

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

$$\int_X f \, d\mu^{(N)} \to \int_X f \, dm \qquad \forall f \in C(X)$$

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

3.11 Remarks.
(i) The map $X \to \mathcal{M}(X)$ given by $x \mapsto \delta_x$ is continuous in the weak* topology.
(ii) The convergence $\mu_n \to \mu$ 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 $\mu_n \to \mu$ 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 $\mu_n \to \mu$ 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 $x_1, \ldots, x_N$.

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 $T$.

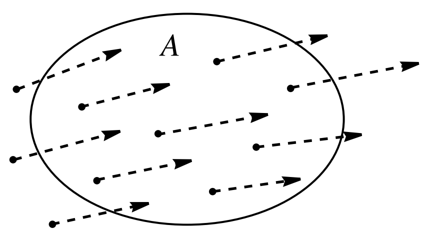


Figure 2: Three points leave the set $A$ and three new points come in.

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, $(X, B)$ is a measurable space and $T : X \to X$ 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 $0 \le p \le 1$). 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 $T : X \to X$ a continuous transformation, then $\mathcal{M}_{\text{inv}} \ne \emptyset$, 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 $T : \mathcal{M} \to \mathcal{M}$ 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 Shauder (or Tykhonov-Shauder) 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 $T : \mathcal{M} \to \mathcal{M}$ and the Shauder 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 $f : X \to \mathbb{R}$ (or, more generally, $f : X \to \mathbb{C}$) in the study of invariant measures.

4.5 Notation. We denote by $L^0(X)$ the set of all measurable functions $f : X \to \mathbb{R}$. 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[5] defined by

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

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

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

(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

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

(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 $T : X \to X$ 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 $T : X \to X$ induces a map $U_T : L^0(X) :\to L^0(X)$ defined by

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

4.11 Simple properties.

[5] 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 $L^p_\mu(X)$ into itself and preserves the norm $||\cdot||_p$, i.e. $||U_T f||_p = ||f||_p$. It also preserves the scalar product in $L^2_\mu(X)$, 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 $f : X \to \mathbb{R}$ and a point $x \in X$, the sequence

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

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

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

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

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

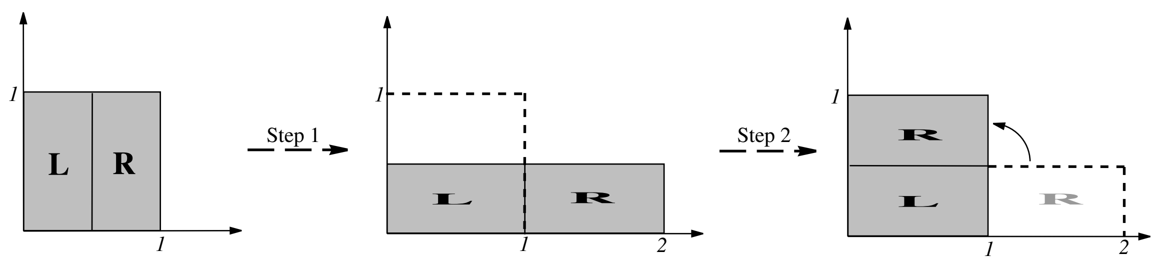


Figure 3: The baker's map constructed in two steps.

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, $T^{-1} : X \to X$, satisfies similar equations:

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

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[6] $X$. Then the baker's map $T : X \to X$ is measurable and preserves $m$.

[6]Let us adopt a convention: whenever $X \subset \mathbb{R}^d$, $d \ge 2$, 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. □

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 $\varepsilon \to 0$, 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 $x = 0.i_0 i_1 i_2 \dots$ be an infinite representation of $x$ in the binary system, cf. 2.28. Let $y = 0.j_1 j_2 j_3 \dots$ be the binary representation of $y$. Then the image $(x', y') = T(x, y)$ has representation

$$x' = 0.i_1 i_2 i_3 \dots \qquad \text{and} \qquad y' = 0.i_0 j_1 j_2 \dots$$

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 $j_1 j_2 j_3 \dots$ and append it to the sequence $i_0 i_1 i_2 \dots$ on the left:

$$\dots j_3 j_2 j_1 i_0 i_1 i_2 \dots \tag{2}$$

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 $T : X \to X$ 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 $T : X \to X$ be a **piecewise monotonic map**, i.e. there are points $0 = a_0 < a_1 < a_2 < \cdots < a_{k-1} < a_k = 1$ such that $T$ is strictly monotonic and differentiable on each of the $k$ intervals $(a_i, a_{i+1})$, $0 \le i \le k - 1$.

(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

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

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 $T : X \to X$ be defined by

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

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 $T : X \to X$ 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

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

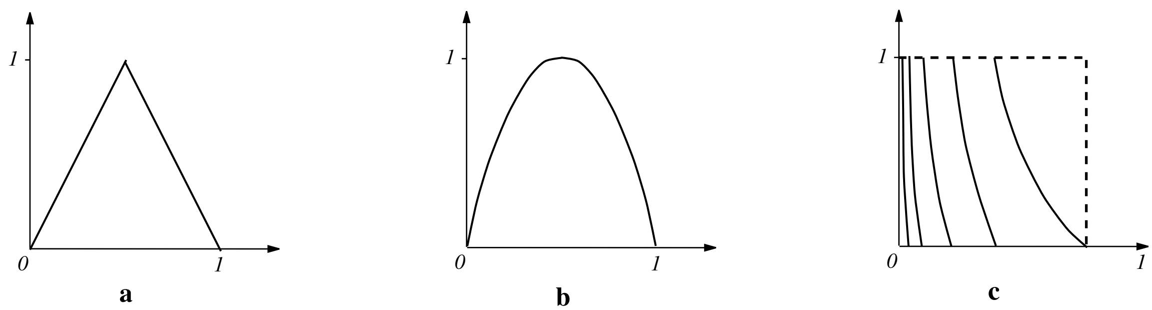


Figure 4: The tent map (a), the quadratic map (b) and the Gauss map (c).

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 $T : X \to X$ 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 $T : I \to I$ 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 $T : X \to X$ is a bijection, one can use $T^{-1}$ as well.

5.17 DEFINITION. A bijective map $T : X \to X$ 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, $T : X \to X$ 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 $0 < n_1 < n_2 < \cdots$ 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

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

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 $T_A : A \to A$ 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 $T_A : \tilde{A} \to \tilde{A}$ 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

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

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 $T_A : \tilde{A} \to \tilde{A}$ 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

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

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

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

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 $\lim_{n \to \infty} \mu(C_n) = 0$. Do this as an exercise. Hint: show that the sets $C_n$ are pairwise disjoint. $\square$

6.9 Exercise. Let $f : X \to \mathbb{R}^+$ 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).

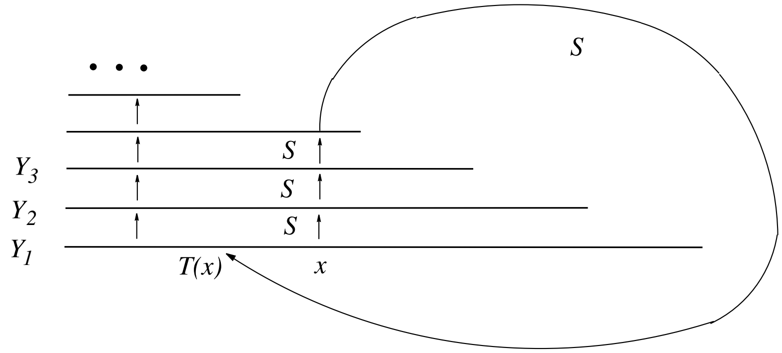


Figure 5: The tower map $S$. A point $x$ with $\tau(x) = 4$ is shown.

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 $T : X \to X$ be a transformation preserving a measure $\mu$ and $\tau : X \to \mathbb{N}$ a measurable positive integral-valued function on $X$ such that

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

Let

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

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

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

Let $\varphi : Y \to X$ be a natural projection $\varphi(x, n) = x$. Define a map $S : Y \to Y$ 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 $Y_n$, $n \ge 1$, by

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

(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" $B_+ = \cap_{n\geq 0} T^n(B)$ and the "backward limit" $B_- = \cup_{n\geq 0} T^{-n}(B)$.

(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 $A = B$ (mod 0).

7.6 Remark. The relation $A = B$ (mod 0) is an equivalence relation. In particular, to check that $A = B$ (mod 0) and $B = C$ (mod 0) implies $A = C$ (mod 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 $B = T^{-1}(B)$ (mod 0), i.e. $\mu(B\Delta T^{-1}B) = 0$.

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

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

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

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

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 $B = T^{-1}(B)$ (mod 0), i.e. $B$ is invariant (mod 0). Prove that $B = B_\infty$ (mod 0).

This shows that for any (mod 0) invariant set $B$ there is a fully invariant set $B_\infty$ that coincides (mod 0) with $B$.

7.10 Examples.

(a) If $X$ is a finite set and $T : X \to X$ 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.

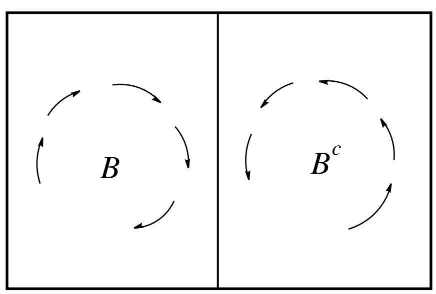


Figure 6: The decomposition of $X$ into two invariant subsets.

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 $T_i = T : X_i \to X_i$ 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 $T : X \to X$ to the study of two "smaller" maps $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$.

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 $T : X \to X$ 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 $T : X \to X$ 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 $\mu \in \mathcal{M}_{\rm inv}$ 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 $f \in L^1_\mu(X)$ or $f \in L^2_\mu(X)$ 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 $T : X \to X$ be ergodic with respect to a measure $\mu$. Let $\mu(A) > 0$. Show that the first return map $T_A : \tilde{A} \to \tilde{A}$ 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 $T : X \to X$ be ergodic with respect to a measure $\mu$. Let $\tau : X \to \mathbb{N}$ 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 $T_i : X_i \to X_i$ 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 $\varphi : B_1 \to B_2$ 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 $\mathrm{dist}(y, T^k y) \leq \varepsilon$. Since $T$ preserves distances, we have

$$\mathrm{dist}(y, T^k y) = \mathrm{dist}(T^k y, T^{2k} y) = \mathrm{dist}(T^{2k} y, T^{3k} y) = \cdots$$

Since $T$ also preserves orientation, then the sequence $y, T^k y, T^{2k} y, T^{3k} y, \ldots$ 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

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

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 $\mathrm{dist}(T^n x, y) < 0.1\,\varepsilon$. 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. □

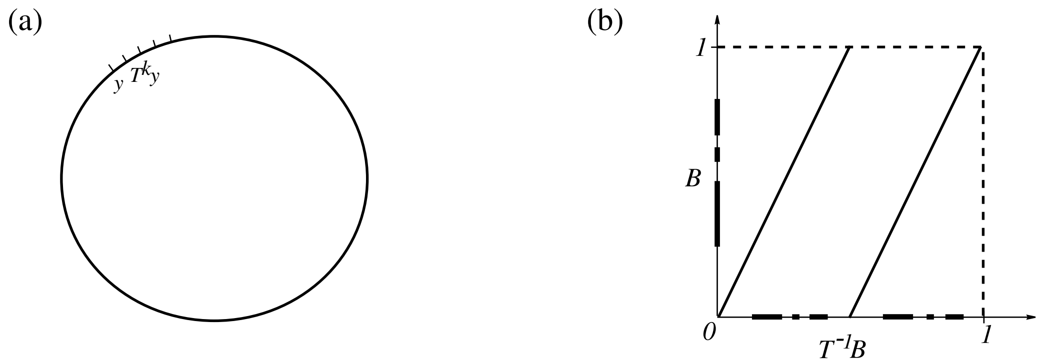


Figure 8: The sequence $y, T^k y, T^{2k} y, \ldots$ on the circle (a), and a set $B$ with its preimage $T^{-1}B$, the latter consisting of two identical parts (b).

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 $i = 0, 1, \ldots, 2^n - 1$ (we call these **binary intervals**). Then

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

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

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,

$$m(A \cap T^{-n}B) \to m(A)\, m(B) \qquad \text{as } \; n \to \infty \tag{7}$$

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 $m(B) = 1$. □

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 $dx \to 0$ 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. □

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

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

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 $k \ge m$.

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 $\ell \ge n$.

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 $k \ge m$ and $\ell \ge n$. 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

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

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

$$m(A \cap T^{-N} B) \to m(A)\, m(B) \qquad \text{as} \quad N \to \infty \tag{8}$$

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 $m(B) = 1$. □

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

8.10 THEOREM. Let $T : X \to X$ 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. □

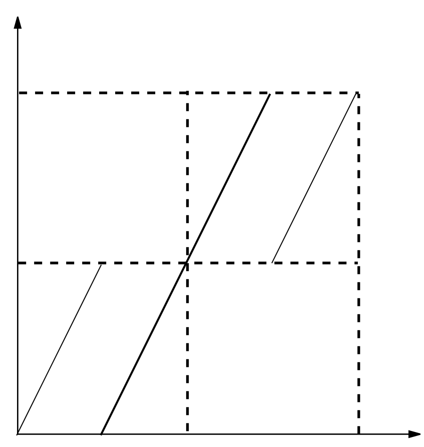

Figure 9: The function in Exercise 8.13.

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 $T : [0, 1) \to [0, 1)$ 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 $T : X \to X$ 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

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

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 $\nu_{n_k}(A) \to \nu(A)$ 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 $k \to \infty$.

# 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 $T : X \to X$ 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

$$f_+(x) = \lim_{n\to+\infty} \frac{1}{n}\, S_n(x) \tag{9}$$

(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

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

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 $T : X \to X$ be an automorphism. Then the "past" time average

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

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

$$r_A(x) = \lim_{n\to+\infty} \frac{\#\{0 \le i \le n-1 : \; T^i(x) \in A\}}{n} \tag{11}$$

(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 $\{T^n(x)\}$, $n \ge 0$ 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 $N \ge 0$, 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 $A = \cup_{N\ge1} A_N$. Then

$$\int_{A_N} f \, d\mu \ge 0 \quad \text{and} \quad \int_A f \, d\mu \ge 0 \tag{12}$$

*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

$$\int_A f\,d\mu = \lim_{N\to\infty} \int_{A_N} f\,d\mu \geq 0 \qquad \square$$

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

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

and

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

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

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

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

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

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

$$\|f_+\|_1 \le \|f\|_1 \tag{15}$$

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 $\|S_n/n\|_\infty \le \|f\|_\infty$, 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,

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

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. □

9.9 Corollary ($L^p$ Ergodic Theorem of Von Neumann). Let $1 \le p < \infty$. If $f \in L^p_\mu(X)$, then $\|S_n/n - f_+\|_p \to 0$ as $n \to \infty$.

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

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 $r = 0, 1, \ldots, 9$. 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 $n \ge 1$, 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 $\#\{0 \le i \le n-1 : \; T^i(x) \in A_r\}$. Corollary 9.6 now implies that

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

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 $r = 0, 1, \ldots, 9$ 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)

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

As $n \to \infty$, 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 $n \to \infty$. Equivalently,

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

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 $T : X \to X$ 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 $\mu_x^{(n)}(A_k) \to \mu(A_k)$ as $n \to \infty$ a.e. by 9.6, i.e. we have this convergence for all $x \in X_k$ with $\mu(X_k) = 1$. Let $X_\infty = \cap_k X_k$. 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 $T : X \to X$ 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 $n \to \infty$, 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 $T : X \to X$ 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?