Chapter 1

The course in a nutshell through an elementary example

This course provides an introduction into chaotic dynamical systems, highlighting topological as well as probabilistic aspects. The focus is primarily on examples that can be explicitly understood analytically. This is not representative of dynamical systems in general, but chosen here in order to optimize the clarity of exposition of fundamental concepts and ideas.

For instance, an elementary but surprisingly rich example is the mapping E_{10} : $[0, 1) \to [0, 1)$ given by

$$E_{10}(x) = 10x \bmod 1. \tag{1.1}$$

The map E_{10} represents a continuous map of the unit circle $S^1 \simeq \mathbb{R}/\mathbb{Z} \simeq [0, 1)$ with the geometric interpretation that it stretches the unit circle uniformly to a circle of ten times its length and then wraps it back (tenfold) around the original unit circle. E_{10} is an example of a so-called *uniformly expanding circle map*. But for practical purposes, we can consider it as a map of the half open interval $[0, 1)$, equipped with the usual Euclidean (arclength) metric

The action of E_{10} on a point $x \in [0, 1)$ is easiest understood if we express x in its decimal expansion $x = 0.x_0x_1x_2x_3x_4\dots$ with $x_i \in \{0, \dots, 9\}$ for all $i \in \mathbb{N}$, with the understanding that we identify 0.99999 with 0.0000 . . .. Namely

$$E_{10}(0.x_0x_1x_2x_3x_4\dots) = 0.x_1x_2x_3x_4\dots \tag{1.2}$$

Chaos

The observation (1.2) does not only provide ease of calculation, but also profound insights about how orbits move around in $[0, 1)$. Namely, the first n digits $x_0 \dots x_{n-1}$ indicate the location of x with reference to a uniform partition of $[0, 1)$ by intervals of size 10^{-n}:

$$x = 0.x_0 \dots x_{n-1} \dots \quad \Leftrightarrow \quad x \in \left[\sum_{i=0}^{n-1} x_i 10^{-i-1}, 10^{-n} + \sum_{i=0}^{n-1} x_i 10^{-i-1}\right].$$

This representation provides us with a lot of information. For instance:

- E_{10} is *topologically transitive*: there exists $x \in [0, 1)$ such that the orbit
$$O^+_{E_{10}}(x) := \{x, E_{10}(x), E^2_{10}(x), E^3_{10}(x), \ldots\}$$
with
$$E^n_{10} := \underbrace{E_{10} \circ \ldots \circ E_{10}}_{n \text{ times}}(x),$$
is dense in $[0, 1)$, in the sense that it intersects every open subset of $[0, 1)$.
- The set of periodic orbits of E_{10} forms a dense subset of $[0, 1)$.
- E_{10} is fundamentally unpredictable as its orbits display an inherent *sensitive dependence on initial conditions*: the first n decimals $x_0 \ldots x_{n-1}$ of an initial condition x do not provide any information about where $E^n_{10}(x)$ lies within $[0, 1)$, as $E^n_{10}(x) = 0.x_n x_{n+1} \ldots$ could lie anywhere in $[0, 1)$.

Devaney [4] coined dynamical systems with these three properties to be *chaotic*.

Symbolic Dynamics

The observation of how E_{10} acts on points in their decimal representation, provides an abstraction of the dynamics of E_{10} as a *left shift* on a space Σ^+_{10} of semi-infinite sequences of integers from the set $\{0, \ldots, 9\}$:

$$\sigma(x_0 x_1 x_2 x_3 x_4 \ldots) = x_1 x_2 x_3 x_4 \ldots \tag{1.3}$$

More precisely, σ and E_{10} are formally related as follows:

$$E_{10} \circ h = h \circ \sigma, \tag{1.4}$$

where $h : \Sigma^+_{10} \to [0, 1)$ is defined as

$$h(x_0 x_1 \ldots) = \sum_{i=0}^{\infty} x_i 10^{-i-1},$$

unless all decimals are equal to 9 in which case $h(9999 \ldots) = 0$.

It is important to note that h is not invertible, since not every number has a unique decimal expansion. For instance $0.01 = h(0100000 \ldots) = h(0099999 \ldots)$. If h would have been invertible, (1.4) would have represented a transformation of coordinates, or *conjugacy*. The relation (1.4) with h non-invertible is called a *semi-conjugacy* between E_{10} and σ and that E_{10} is a *factor* of σ.

Because of this relationship, it is natural to study the properties of E_{10} through the conceptually simpler shift dynamical system $\sigma : \Sigma^+_{10} \to \Sigma^+_{10}$. We in fact already informally did so when observing the relation (1.2). It only still remains to equip Σ^+_{10}

with a suitable metric $d^{\Sigma^+} : \Sigma_{10}^+ \times \Sigma_{10}^+ \to \mathbb{R}$,

$$d^{\Sigma^+}(x,y) := \sum_{i=0}^{\infty} \frac{\delta(x_i,y_i)}{3^i}, \text{ where } \delta(x_i,y_i) = \begin{cases} 0 \text{ if } x_i = y_i, \\ 1 \text{ if } x_i \neq y_i. \end{cases} \tag{1.5}$$

Note that this metric is such that deviations between decimals of two sequences at lower index weigh exponentially more heavily than deviations of decimals at higher index. In turns out that with this metric h is continuous, so that (1.4) constitutes a *topological semi-conjugacy*.

The overarching general philosophy we explore in this course is to reduce a potentially complicated dynamical system like E_{10}, via a topological (semi-)conjugacy, to the *symbolic dynamics* of a shift σ. Indeed, in this case, the shift σ can be readily demonstrated to be chaotic: it is transitive and it has dense periodic orbits and sensitive dependence on initial conditions. It is then seen to follow from the topological semi-conjugacy that E_{10} has these same properties.

Ergodic theory

When dynamics is very complicated from the topological point of view, it is natural to address questions about average properties rather than about individual orbits. A classical question in this context is whether avergaes of observables along orbits exist and if they depend on the initial condition of the orbit.

For instance, in the example E_{10} we may ask whether the relative frequency of visiting a subset $A \subset [0,1)$ exists. This is a special case of the general question posed above, with the observable being the characteristic function

$$\chi_A(x) := \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ if } x \notin A. \end{cases}$$

The relative frequency $F(A)$ of visiting A is precisely the average of χ_A along an orbit $O^+_{E_{10}}(x)$. The relative frequency with which the orbit $O^+_{E_{10}}(x)$ has visited A of up to time n is

$$F(A)(n,x) := \frac{1}{n}\sum_{i=0}^{n-1} \chi_A(E_{10}^i(x)).$$

So the question is whether $\lim_{n\to\infty} F(A)(n,x)$ exists and if so, whether it depends on x.

The answer to this question is intricate and addressed by the field of *Ergodic Theory*. A main ingredient of this theory is the identification of so-called *invariant probability measures*. In the context of our example, such measures are naturally defined on the measurable space that is obtained by equipping $[0,1)$ with its so-called *Borel σ-algebra* $\mathcal{B}([0,1))$ (which is a set of subsets of $[0,1)$ containing all countable unions and intersections of open and closed intervals). Probability measures $\mu : \mathcal{B}([0,1)) \to \mathbb{R}$ are non-negative functions such that $\mu([0,1)) = 1$ and $\mu(A \cup B) = \mu(A) + \mu(B)$ if $A \cap B = \emptyset$. Such a measure μ is called E_{10}-*invariant* if $\mu(E_{10}^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}([0,1))$, where $E_{10}^{-1}(A) := \{x \in [0,1) \mid E_{10}(x) \in A\}$. In addition, it is called *ergodic* if $\mu(A) \in \{0,1\}$ for every $A \in \mathcal{B}([0,1))$ satisfying $E_{10}^{-1}(A) = A$.

It turns out that the Lebesgue measure, defined for all $A \in \mathcal{B}([0,1))$ as

$$\lambda(A) = \int_A dx,$$

is an ergodic invariant probability measure for E_{10}.

It then follows from *Birkhoff's Ergodic Theorem* that for λ-almost all initial conditions x,[1] we have

$$\lim_{n\to\infty} F(A)(n,x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(E_{10}^i(x)) = \int_0^1 \chi_A(x) d\lambda(x) = \int_A dx. \quad (1.6)$$

In particular, this implies that the relative frequency in which Lebesgue typical orbits visit an interval $I \subset [0,1)$ is equal to the length $|I|$ of this interval.

Interestingly, the existence of this ergodic invariant measure can be constructively derived using the same symbolic dynamics we developed to understand the topological dynamics. Namely, it turns out that the transformation h at the basis of the semi-conjugacy between E_{10} and the shift σ is measurable with respect to the relevant Borel σ-algebras on $[0,1)$ and Σ_{10}^+, which implies a correspondence of ergodic invariant measures of E_{10} and σ: if μ is an ergodic invariant measure for σ, then $h_*\mu := \mu \circ h^{-1}$ is an ergodic invariant measure for E_{10}. There is a large set of natural ergodic invariant measures for σ on Σ_{10}^+. One of these ergodic invariant measures μ is the uniform Bernoulli measure, defined by its values on cylinder sets as

$$\mu(C_{x_0\ldots x_{n-1}}) = 10^{-n}.$$

It turns out that the Lebesgue measure $\lambda = h_*\mu$ and that in fact no other ergodic invariant measure can exist with a support that has positive Lebesgue measure.

While Birkhoff's Ergodic Theorem establishes the fact that the relative frequency exists and is constant for Lebesgue almost all initial conditions, on a set of initial conditions that has zero Lebesgue measure, such averages may exist and take different values, or may even not exist at all. Such exceptional initial conditions can rather easily be found. For instance, when $x \in \mathbb{Q}$, the corresponding orbit $O_{E_{10}}^+(x)$ is periodic and the relative frequency to visit a subinterval will typically not be equal to the length of this interval, but instead be equal to the proportion of points of this periodic orbit that lie inside this interval. The symbolic dynamics can be used to find examples of initial conditions for which relative frequencies do not exist for all measurable sets A. For example, if the symbolic sequence for the initial condition is

[1] This means that the set of initial conditions for which (1.6) does not hold, has Lebesgue measure zero.

such that is alternatingly dominated by 0 and 9, eg $x = 0.090099990^8 9^{16} \ldots$, then the relative frequency for the forward orbit of x to visit the interval $[0.9, 1)$ does not converge.

In this course, we will lie the foundations of the theory of topological and probabilistic properties of dynamical systems with a focus on one-dimensional maps that admit symbolic dynamics, as in the example E_{10} discussed in this introduction.

Chapter 2
Topological dynamics

2.1 Continuous maps and their orbits

In these notes, we consider dynamical systems with discrete time, represented by the iteration of a continuous map

$$f : X \to X,$$

where X is a compact metric space.[1] The metric space X is the *state space*, where each $x \in X$ represents a state of the system. The metric $d^X : X \times X \to \mathbb{R}$ provides a measure of distance between two different states in X. The map f represents the evolution of the system in one time-step.

We adopt the following notation for the nth iterate of f:

$$f^n := \underbrace{f \circ \ldots \circ f}_{n \text{ times}}.$$

Thus, if f represents the time-one map of a system, then f^n represents the time-n map.

The evolution in time of an initial condition $x \in X$ is represented by its forward orbit

$$O_f^+(x) := \{x, f(x), f^2(x), f^3(x), \ldots\}.$$

Some types of orbits have special names. For instance, if $O_f^+(x) = \{x\}$, then $f(x) = x$ and x is called a *fixed point* of f. Likewise, $O_f^+(x)$ is called a *periodic orbit* if $f^p(x) = x$ for some positive integer p and $O_f^+(x) = \{x, f(x), \ldots, f^{p-1}(x)\}$. The least p such that $f^p(x) = x$ is called the *period* of x and of its orbit. Fixed points are special cases of periodic orbits, namely those with period equal to one.

There is a surprising potential richness in the set of periodic orbits that dynamical systems may possess. A celebrated result that illustrates this, is the following

[1] Compactness is assumed to facilitate the exposition. Most concepts and results have straightforward generalisations to the case when X is not compact.

corollary of *Sharkovskii's Theorem* [8], which was also - much later - independently obtained in a seminal paper by Li and Yorke [6].

Theorem 2.1 (For interval maps, period 3 implies all periods.) *Let $f : I \to I$ be a continuous map of the interval with a periodic orbit of period 3. Then f has periodic orbits of any period.*

Proof Let $a < b < c$ be the three points of the three-periodic orbit in I. We discuss the case that $f(a) = b$ in detail. The remaining case when $f(a) = c$ can be treated analogously with changed roles of $[a, b]$ and $[b, c]$, cf. Exercise 2.4.

We thus have $f(a) = b$, $f(b) = c$ and $f(c) = a$. This implies $b, c \in f([a, b])$ and $a, c \in f([b, c])$. Since closed intervals are mapped onto closed inervals, this yields

$$[b, c] \subset f([a, b]) \text{ and } [a, c] \subset f([b, c]) . \tag{2.1}$$

We first establish the existence of a fixed point. This follows from the observation that by (2.1), $[b, c] \subset [a, c] \subset f([b, c])$, which implies that f has a fixed point in $[b, c]$ by the following elementary lemma.

Lemma 2.1 *Consider a closed interval $I = [a, b]$ and a continuous function $f : I \to \mathbb{R}$. Suppose that $f(I) \subset I$ or $I \subset f(I)$. Then f has a fixed point.*

Proof Define the continuous function $g : I \to \mathbb{R}$ by $g(x) := f(x) - x$. Suppose first that $f(I) \subset I$. Then $f(a), f(b) \in [a, b]$ and

$$g(a) = f(a) - a \geq 0 \geq f(b) - b = g(b).$$

The latter implies by the intermediate value theorem that $g(x) = 0$ for some $x \in [a, b]$, so that $f(x) = x$.

If $I \subset f(I)$ there exist $x_1, x_2 \in [a, b]$ with $f(x_1) \leq a$ and $f(x_2) \geq b$. This implies that

$$g(x_1) = f(x_1) - x_1 \leq a - x_1 \leq 0 \leq b - x_2 \leq f(x_2) - x_2 = g(x_2),$$

which implies the existence of a fixed point of f in $[x_1, x_2]$. □

We proceed with the proof of the theorem to show that for all natural numbers $n \geq 2$, there exists a fixed point of f^n in the interval $[b, c]$. This follows, by Lemma 2.1, from the existence of a closed interval $I_n \subset [b, c]$ with $I_n \subset f^n(I_n)$. To demonstrate the latter, we use the following result.

Lemma 2.2 *Let $I \subset \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ a continuous function. Then for every closed interval $K \subset f(I)$, there exists a closed interval $I_K \subset I$ with $K = f(I_K)$.*

Proof Let $K := [a, b] \subset f(I)$, then there exists $c, d \in I$ with $f(c) = a$ and $f(d) = b$. Let $c \leq d$ and define $r := \sup\{x \in [c, d] : f(x) \leq a\}$ and $s := \inf\{x \in [r, d] : f(x) \geq b\}$. Then for all $x \in [r, s]$, we have $a \leq f(x) \leq b$, and this means that $f([r,s]) \subset [a,b]$. Conversely, for all $y \in [a,b]$ there exists $x \in [r,s]$ with $y = f(x)$. so that $f([r,s]) = [a,b]$. The remaining case $c > d$ follows analogously (cf. Exercise 2.3). $\square$

Let $I_0 := [b,c]$, then $I_0 \subset [a,c] \overset{(2.1)}{\subset} f(I_0)$, and Lemma 2.2 establishes the existence of a closed interval $I_1 \subset I_0$ with $f(I_1) = I_0$. In turn, there is a closed interval $I_1 \subset I_0 = f(I_1)$, so that from Lemma 2.2 we obtain a closed interval $I_2 \subset I_1$ with $f(I_2) = I_1$. We may repeat this procedure to obtain a sequence of closed intervals

$$I_0 \supset I_1 \supset \cdots \supset I_{n-2} \text{ with } f(I_k) = I_{k-1}\ \forall k \in \{1,\ldots,n-2\}\,. \tag{2.2}$$

By induction, we thus obtain

$$f^k(I_k) = I_0 = [b,c],\ \forall k \in \{0,\ldots,n-2\}\,. \tag{2.3}$$

We now observe that $f^{n-1}(I_{n-2}) = f(f^{n-2}(I_{n-2})) \overset{(2.3)}{=} f([b,c]) \overset{(2.1)}{\supset} [a,c] \supset [a,b]$. Hence, by Lemma 2.2, there exists a closed interval $I_{n-1} \subset I_{n-2}$ such that

$$f^{n-1}(I_{n-1}) = [a,b]\,. \tag{2.4}$$

Similarly, $f^n(I_{n-1}) = f(f^{n-1}(I_{n-1})) \overset{(2.4)}{=} f([a,b]) \overset{(2.1)}{\supset} [b,c]$, so that there exists a closed interval $I_n \subset I_{n-1}$ that satisfies

$$f^n(I_n) = [b,c] \supset I_n\,.$$

Then, by Lemma 2.1, f^n has a fixed point $x_n \in I_n \subset I$.

In order to finalize the proof, it remains to be shown that x_n has period n. Since $I_n \subset I_k$ for all $k \in \{1,\ldots,n-1\}$, (2.3) implies

$$f^k(x_n) \in [b,c]\ \forall k \in \{0,\ldots,n-2\}\,. \tag{2.5}$$

But because of (2.4), it follows that

$$f^{n-1}(x_n) \in [a,b]\,. \tag{2.6}$$

We consider the cases $x_n = b$ and $x_n \in (b,c]$, separately:
If $x_n = b$, we assert that $n = 3$. Firstly, $n = 1$ and $n = 2$ are not possible, since b is not a fixed point of f or f^2 since $f(b) = c \neq b$ $f^2(b) = a \neq b$. Moreover, $x_n \neq b$ if $n \geq 4$, since if $x_n = b$, (2.5) would imply that $f^2(b) = f^2(x_n) \in [b,c]$, which would contradict the fact that $f^2(b) = a < b$.

Finally, if $x_n \in (b,c]$, because of (2.6), we have $f^{n-1}(x_n) \neq x_n$. This x_n cannot have period $n-1$. Moreover, if x_n has a period that is less than or equal to $n-2$, then (2.5) implies that its entire orbit $O_f^+(x_n)$ lies in $(b,c]$, which contradicts (2.6). Thus x_n has period n, which concludes the proof this theorem. $\square$

Theorem 2.1 is a special case of a more general result by Sharkovskii [8] that concerns the forced existence of periodic orbits for continuous maps of the interval. He introduced the following ordering of the natural numbers, known as *Sharkovskii ordering*:

$$3 \triangleright 5 \triangleright 7 \triangleright \cdots \triangleright 2 \cdot 3 \triangleright 2 \cdot 5 \triangleright 2 \cdot 7 \triangleright \cdots \triangleright 2^2 \cdot 3 \triangleright 2^2 \cdot 5 \triangleright 2^2 \cdot 7 \triangleright \dots$$
$$\dots \triangleright 2^m \cdot 3 \triangleright 2^m \cdot 5 \triangleright 2^m \cdot 7 \triangleright \cdots \triangleright 2^3 \triangleright 2^2 \triangleright 1\,.$$

Theorem 2.2 (Sharkovskii [8]) *Let $I \subset \mathbb{R}$ be an interval, and $f : I \to \mathbb{R}$ be continuous. If f has a periodic orbit of period n, then f has m-periodic points for all $n \triangleright m$.*

While the proof of this theorem uses in essence the same ingredients as the proof of Theorem 2.1, the necessary book-keeping is more involved. For a concise exposition, see for instance [3, §7.3].

Exercises[2]

2.1 Continuous maps $f : S^1 \to S^1$ of the circle to itself are in practice conveniently represented by a continuous *lift* to the real line, $F : \mathbb{R} \to \mathbb{R}$, with the property that after projection by "mod 1" to $[0, 1)$, it represents the circle map f, i.e.

$$f(x \bmod 1) = F(x) \bmod 1. \tag{2.7}$$

We use here the fact that $S^1 \simeq \mathbb{R}/\mathbb{Z}$, by virtue of which the circle is parametrizable by the half-open interval $[0, 1)$.

(a) Show that the lift F must satisfy the following property:

$$F(x + 1) = F(x) + k \text{ for some } k \in \mathbb{Z}.$$

k is known as the *degree* of the circle map f.

(b) Show that the lift F of f is not unique.

2.2 For a continuous map $f : X \to X$, let

$$P_n(f) := \#\{x \in X \mid f^n(x) = x\}, \tag{2.8}$$

denote the cardinality of the set of fixed points of f^n.

(a) Let $E_k : S^1 \to S^1$, with $k \in \mathbb{Z}$ and $|k| > 1$, be defined as $E_k(x) := kx \bmod 1$. Show that $P_n(E_k) = |k^n - 1|$.
(b) Let $f : S^1 \to S^1$ be a C^1 circle map with $|f'(x)| > 1$, $\forall x \in S^1$, of degree k. Show that $P_n(f) = |k^n - 1|$.

[2] Throughout this text, exercises are labeled * when challenging and ** when very challenging.

(c) Let $f_a : \mathbb{R} \to \mathbb{R}$ be the logistic map $f_a(x) := ax(1-x)$. Show that

(i) $P_n(f_a) \leq 2^n$ for all $a \in \mathbb{R}$.
(ii) $*P_n(f_a) = 2^n$ if $a \geq 4$.

2.3 Work out the details of the final part of the proof of Lemma 2.2, when $c > d$.

2.4 Work out the details of the proof of Theorem 2.1 when $f(a) = c$.

2.5 *Let $f : S^1 \to S^1$ be a C^1 circle map with $|f'(x)| > 1$, $\forall x \in S^1$, of degree $k \in \mathbb{Z}$. Show that such f has periodic orbits of all periods if $k > 1$. Show that this is not true for all $k < -1$.

2.6 **Consider the logistic map $f_r : [0,1] \to [0,1]$, defined as $f_r(x) = rx(1-x)$, with $0 < r \leq 4$. Determine for which values of r, f_r has periodic orbits of all periods.

2.2 ω-limit sets, invariant sets and attractors

A central objective of dynamical systems theory is to characterize the behaviour of orbits as time goes to infinity. In particular, one may wonder whether

$$\lim_{n\to\infty} f^n(x) \tag{2.9}$$

exists. It can be shown that if this limit exists and $\lim_{n\to\infty} f^n(x) = \tilde{x}$, then necessarily $\tilde{x}$ is a fixed point of f.

However, in many cases the limit (2.9) does not exist. In order to describe long term behaviour, it is useful to identify those points in X to which an infinite subsequence of $O_f^+(x)$ converges.

Definition 2.1 (ω-limit point) A point $\tilde{x} \in X$ is an ω-limit point of $x \in X$ for a continuous map $f : X \to X$ if there exists a strictly increasing sequence $\{n_k\}_{k\in\mathbb{N}}$ of positive integers such that $\lim_{k\to\infty} f^{n_k}(x) = \tilde{x}$.

The collection of all ω-limit points of a point $x \in X$ is called the ω-limit set of x.

Definition 2.2 (ω-limit set) The ω-limit set $\omega(x)$ of a point $x \in X$ for a continuous map $f : X \to X$ is the set of all ω-limit points of x.

For instance, if $O_f^+(x)$ is a periodic orbit, then $\omega(x) = O_f^+(x)$.

ω-limit sets are examples of so-called *invariant sets*[3] of f.

Definition 2.3 (Invariant set) Let $f : X \to X$ be a continuous map. We call $A \subset X$ *positively f-invariant* if $f(A) \subset A$ and *f-invariant* if $f(A) = A$.

[3] There are various definitions of invariant sets in the literature and the definition given here suits the exposition in this course.

Proposition 2.1 *Let $f : X \to X$ be a continuous map and $x \in X$. Then $\omega(x)$ is closed and f-invariant.*

Proof We first note that due to the assumed compactness of X in these notes, $\omega(x) \neq \emptyset$ since every sequence in a compact space has a converging subsequence.

$\omega(x)$ is closed since it is a countable intersection of closed sets (and thus closed), cf. Exercise 2.8.

For f-invariance, we assert that $f(\omega(x)) = \omega(f(x))$ and $\omega(f(x)) = \omega(x)$. The first equality is a direct consequence of the continuity of f, which implies that $f(\lim_{k\to\infty} f^{n_k}(x)) = \lim_{k\to\infty} f^{n_k}(f(x))$. The second equality follows from observing that if there exists a strictly increasing sequence of positive integers $\{n_k\}_{k\in\mathbb{N}_0}$ such that $\lim_{k\to\infty} f^{n_k}(x) = y$, then with $m_k := n_k - 1$, $\{m_k\}_{k\in\mathbb{N}}$ is such that $\lim_{k\to\infty} f^{m_k}(f(x)) = y$. □

ω-limit sets exist for every $x \in X$. An alternative approach to capturing long-term behaviour of dynamical systems, is by the identification of so-called *attractors*, which are objects that attract a neigbourhood of itself, rather than only individual orbits.[4]

Definition 2.4 Let $f : X \to X$ be a continuous map. Then a compact subset $A \subset X$ is called an *attractor* of f if there exists an open $U \subset X$ such that $f(\overline{U}) \subset U$ and $A = \cap_{i\in\mathbb{N}_0} f^i(U)$.

The set U in this definition is known as a *trapping region* for the attractor. The set of all points whose forward orbits converge to the attractor A is called its *basin of attraction* $B(A)$:

$$B(A) := \bigcup_{n\in\mathbb{N}_0} f^{-n}(U), \tag{2.10}$$

where

$$f^{-n}(U) := \{x \in X \mid f^n(x) \in U\}.$$

Example 2.1 (Contractions) The simplest dynamical systems with attractors are contractions on a complete metric space. By the *contraction mapping theorem*, such contractions have a unique fixed point and for all initial conditions, forward orbits of contractions converge exponentially fast to the unique fixed point.

More generally, attractors are invariant sets (Exercise 2.11) which are *asymptotically stable*, the definition of which involves the so-called semi-Hausdorff distance

$$\mathrm{dist}(x, A) := \inf_{\tilde{x}\in A} d^X(x, \tilde{x}). \tag{2.11}$$

Definition 2.5 An invariant set $A \subset X$ of a continuous map $f : X \to X$ is *asymtotically stable* if there exists an open neighbourhood U of A such that for every $x \in U$, $\lim_{n\to\infty} \mathrm{dist}(f^n(x), A) = 0$.

[4] There are various definitions of attractors in the literature and this relatively uncomplicated definition is chosen to suit the context of this course.

Proposition 2.2 *Attractors of continuous maps are* asymptotically stable.

Proof Recall that we consider the setting of a continuous map on a compact state space.

Let $V \subset X$ be open such that $A \subset V \subset \overline{U} \subset X$. We show that there exists $N \in \mathbb{N}_0$ such that $f^n(x) \in V$ for all $x \in U$ and all $n \geq N$.

Let $S_n := X \setminus f^n(\overline{U})$. Then

$$\begin{aligned} x \notin A \Leftrightarrow \exists n \geq 0 \text{ such that } x \notin f^n(U) \\ \Rightarrow \exists n \geq 0 \text{ such that } x \notin f^n(\overline{U}) \Leftrightarrow \exists n \geq 0 \text{ such that } x \in S_n. \end{aligned}$$

Hence, $X = A \cup \left(\bigcup_{n\geq 0} S_n\right) = V \cup \left(\bigcup_{n\geq 0} S_n\right)$ is an infinite open cover of X. Note that S_n is open since it is the complement of a closed set in a topological space. Since X is assumed to be compact, there exists a finite subcover, i.e. $\exists N \in \mathbb{N}_0$ such that

$$X = V \cup \left(\bigcup_{n=0}^{N} S_n\right). \tag{2.12}$$

We further note that since $f(\overline{U}) \subset U$ also $f(\overline{U}) \subset \overline{U}$ and hence

$$f^{n+1}(\overline{U}) \subset f^n(\overline{U}),\ \forall n \in \mathbb{N}_0. \tag{2.13}$$

In turn, this implies that

$$X \setminus f^{n+1}(\overline{U}) \supset X \setminus f^n(\overline{U}) \ \Leftrightarrow\ S_{n+1} \supset S_n$$

which, together with (2.12), leads to

$$X = V \cup S_N, \text{ for some } N \in \mathbb{N}.$$

Hence, $f^N(\overline{U}) \subset V$ and, in combination with (2.13), $f^n(\overline{U}) \subset V$, for all $n \geq N$. We note that orbits thus converge uniformly, in the sense that N does not depend on $x \in U$. □

Example 2.2 Let $M : [0, 1] \to [0, 1]$ be defined as

$$M(x) = \begin{cases} 3x & \text{if } x \in [0, \frac{1}{4}], \\ \frac{5}{4} - 2x & \text{if } x \in [\frac{1}{4}, \frac{1}{2}], \\ 2x - \frac{3}{4} & \text{if } x \in [\frac{1}{2}, \frac{3}{4}], \\ 3 - 3x & \text{if } x \in (\frac{3}{4}, 1]. \end{cases} \tag{2.14}$$

$[\frac{1}{4}, \frac{3}{4}]$ is an invariant set of M since

$$\inf_{x\in[\frac{1}{4},\frac{3}{4}]} M(x) = \frac{1}{4} \text{ and } \sup_{x\in[\frac{1}{4},\frac{3}{4}]} M(x) = \frac{3}{4}.$$

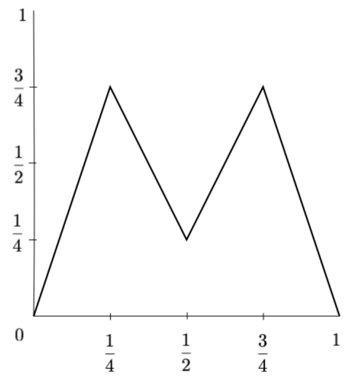

Fig. 2.1: Graph of the map M (2.14).

It turns out that $[\frac{1}{4}, \frac{3}{4}]$ is also an attractor of M. Namely, consider $U = (\frac{1}{9}, \frac{8}{9}) \supset [\frac{1}{4}, \frac{3}{4}]$. Then $f(\overline{U}) = [\frac{1}{4}, \frac{3}{4}] \subset U$ and $\cap_{i \in \mathbb{N}} f^i(U) = [\frac{1}{4}, \frac{3}{4}]$ is an attractor. The basin of this attractor, as defined in (2.10), is given by

$$B([\tfrac{1}{4}, \tfrac{3}{4}]) = (0, 1),$$

since $f^{-n}([\frac{1}{4}, \frac{3}{4}]) = [\frac{3^{-n}}{4}, 1 - \frac{3^{-n}}{4}]$ for all $n \in \mathbb{N}_0$.

Exercises

2.7 Let $f : X \to X$ be a continuous map on a metric space. Show that if $\lim_{n\to\infty} f^n(x) = \tilde{x}$, then $\tilde{x}$ is a fixed point of f.

2.8 Prove the following characterisation, alternative to Definition 2.2, of an ω-limit set of a continuous map on a metric space $f : X \to X$:

$$\omega(x) = \bigcap_{n \in \mathbb{N}_0} \overline{O_f^+(f^n(x))}.$$

2.9 Let $R_a : S^1 \to S^1$ denote the rigid translation on the unit circle: $R_a(x) = x + a \bmod 1$, where the unit circle is parametrized $S^1 \simeq [0, 1)$, as usual. R_a is the rigid rotation over the angle $2\pi a$. Determine the ω-limit sets for the map R_a. Note that the answer depends on a.

2.10 * * Prove that a continuous map on a compact metric space with no isolated points has a dense forward orbit if and only if it is topologically transitive. Recall that a point x in a metric space is called *isolated* if there exists $\varepsilon > 0$ such that $B_\varepsilon(x) = \{x\}$.

2.11 Show that attractors are invariant sets.

2.12 Consider a linear map L on $\mathbb{R}^n$, all eigenvalues of which have norm smaller than 1. Show that $\omega(x) = 0$ for all $x \in \mathbb{R}^n$ and that 0 is the only attractor of L. Which property of the eigenvalues of L would prevent global attraction to 0?

2.13 *Consider a C^2 map $f : X \to X$, where X is a smooth finite dimensional manifold (or $\mathbb{R}^n$), with fixed point $\tilde{x} \in X$ such that all eigenvalues of the derivative $Df(\tilde{x})$ have norm smaller than 1. Show that $\tilde{x}$ is an attractor.

2.14 Consider a C^2 map $f : X \to X$, where X is a smooth finite dimensional manifold (or $\mathbb{R}^n$), with periodic point $\tilde{x} \in X$. Formulate a sufficient condition for $O_f^+(\tilde{x})$ to be an attractor. You may use the result obtained in Exercise 2.13.

2.15 *Show that for a continuous map f on a compact metric space X, the following statements hold:

(a) $\omega(x)$ is the smallest subset of X satisfying $\lim_{n\to\infty} \mathrm{dist}(f^n(x), \omega(x)) = 0$.
(b) For all nonempty compact subsets $S \subsetneq \omega(x)$, $S \cap \overline{f(\omega(x)) \setminus S} \neq \emptyset$.
(c) $\omega(x)$ is finite if and only if it is a periodic orbit.

2.3 Chaos

In view of the observation that the dynamical behaviour of relatively simple systems, such as E_{10} in Chapter 1, can be very complicated, the term *chaos* was coined Li and Yorke in [6]. There are several manifestations of complicated behaviour, of which we here discuss a few important ones.

One of the most fundamental observations of complexity in dynamical systems, is a fundamental limitation to the accuracy of long term predictions, in the sense that arbitrarily small errors in the initial condition may imply a uniform prediction error in the long run. This is formalized as the property of *sensitive dependence on initial conditions* or, in short, *sensitive dependence*.

Definition 2.6 (Sensitive dependence) A continuous map $f : X \to X$ on a metric space X with metric d has *sensitive dependence* if there exists a *sensitivity constant* $\Delta > 0$ such that for all $x \in X$ and $\varepsilon > 0$, there exists $y \in X$ with $d^X(x, y) < \varepsilon$ and $n \in \mathbb{N}$ such that $d^X(f^n(x), f^n(y)) \geq \Delta$.

Example 2.3 (E_{10} has sensitive dependence.) From the shift action of E_{10} on points in terms of their decimal expansion, as detailed in Chapter 1, we observe that if two initial points have a different nth decimal digit, then after n iterates they are at least 10^{-2} apart.

The notion of *topological transitivity* relates to the possibility for a dynamical system to have orbits connecting all local regions of the state space.

Definition 2.7 (Topological transitivity) A continuous map $f : X \to X$ is *topologically transitive* if for any pair of open sets $U, V \subset X$ there exists $n \in \mathbb{N}_0$ such that $f^n(U) \cap V \neq \emptyset$.

Under some mild additional hypothesis on X, this is equivalent to f having a dense orbit, see Exercise 2.10.

Example 2.4 (E_{10} is topologically transitive) Every open subset of $[0, 1)$ contains a subinterval that is characterized by points having a given finite initial subsequence of decimals. It suffices to show that for each such interval $U = \{x \in [0, 1) \mid x = x_0 x_1 \ldots x_n \ldots\}$, we have $E_{10}(U) = [0, 1)$ which implies in particular that $E_{10}(U) \cap V \neq \emptyset$ for all open $V \subset [0, 1)$.

Devaney [4] proposed the following definition of *chaotic dynamics*.

Definition 2.8 (Chaotic dynamics) A continuous map $f : X \to X$ is *chaotic* if it has the following three properties:

1. the periodic points of f are dense in X,
2. f is topologically transitive,
3. f has sensitive dependence on initial conditions.

Example 2.5 (E_{10} is chaotic) We have seen in Chapter 1 that E_{10} has dense periodic orbits and in Examples 2.3 and 2.4 that it is also topologically transitive and has sensitive dependence.

It was later shown by Banks *et al.* [2] that the first two properties in Definition 2.8 imply the third if the entire state space does not consist of a single periodic orbit.

Theorem 2.3 ([2]) *A continuous map on a metric space is chaotic if it has dense periodic orbits and is topologically transitive, unless the metric space consists of a single periodic orbit.*

Proof We divide the proof into nine steps. We use the notation of Definition 2.8.

1. If X is a single periodic orbit of f, then f cannot have sensitive dependence.
2. If X is not a single periodic orbit of f, then density of periodic orbits implies the existence of at least two periodic orbits. Let $q, \tilde{q} \in X$ be elements of two different periodic orbits of f. We define

$$\Delta := \min\{d^X(f^n(q), f^m(\tilde{q})) \mid n, m \in \mathbb{N}\}/8$$

   and note that $\Delta > 0$. The aim of the proof is to show that Δ is a sensitivity constant for f.
3. Let $x \in X$, then by the triangle inequality the distance between x and one of these periodic orbits is at least 4Δ, i.e.

$$\max\{\min\{d^X(x, f^n(q)) \mid n \in \mathbb{N}\}, \min\{d^X(x, f^n(\tilde{q})) \mid n \in \mathbb{N}\}\} \geq 4\Delta.$$

   Without loss of generality, we let q be the periodic orbit at least distance 4Δ from x.

4. By density of periodic orbits, for any $\varepsilon \in (0, \Delta)$ there exists a periodic point $p \in B_\varepsilon(x)$. We denote the minimal period of p to be n.
5. Let
$$V := \bigcap_{i=0}^{n} f^{-i}(B_\Delta(f^i(q))$$
denote the set of points in X whose first n iterates shaddow the orbit of q up to distance Δ. By continuity of f, V is an open neighbourhood of q.
6. By transitivity of f there exists $k \in \mathbb{N}$ such that $f^k(B_\varepsilon(x)) \cap V \neq \emptyset$, or - equivalently - a $y \in B_\varepsilon(x)$ such that $f^k(y) \in V$.
7. In order to keep track of $O_f^+(y)$ relative to $O_f^+(p)$, we define
$$j := \left\lfloor \frac{k}{n} \right\rfloor + 1,$$
so that $\frac{k}{n} < j \leq \frac{k}{n} + 1$ and the length of the orbit of p fits $j-1$ times - but no more than $j-1$ times - in k:
$$k < nj \leq k + n \quad \Leftrightarrow \quad 0 < N - k \leq n,$$
where $N := nj$ so that $f^N(p) = p$.
8. We use the triangle inequality to find
$$d^X\left(x, f^{N-k}(q)\right) \leq d^X(x,p) + d^X\left(p, f^N(y)\right) + d^X\left(f^N(y), f^{N-k}(q)\right),$$
and use this to obtain
$$d^X(f^N(p), f^N(y)) \geq \underbrace{d^X\left(x, f^{N-k}(q)\right)}_{\geq 4\Delta} - \underbrace{d^X\left(f^{N-k}(q), f^{N-k}(y)\right)}_{<\Delta} - \underbrace{d^X(p,x)}_{<\Delta} \geq 2\Delta,$$
where the first and third estimates follow from the definitions in the third and fourth steps of this proof. The remaining second estimate uses the fifth and sixth step, noting that $0 < N - k \leq n$ and $f^N(y) = f^{N-k}(f^k(y))$, with $f^k(y) \in V$.
9. Consequently, another application of the triangle inequality on the conclusion of the last step yields
$$d^X\left(f^N(p), f^N(x)\right) + d^X\left(f^N(y), f^N(x)\right) \geq 2\Delta,$$
which in turn implies that $d^X\left(f^N(p), f^N(x)\right) \geq \Delta$ or $d^X\left(f^N(y), f^N(x)\right) \geq \Delta$ so that indeed Δ is a sensitivity constant. □

Another popular and useful characterisation of complicated dynamics is *topological mixing*.

Definition 2.9 (Topological mixing) A continuous map on a metric space $f : X \to X$ is *topologically mixing* if for any pair of non-empty open sets $U, V \subset X$ there exists $N \in \mathbb{N}$ such that for all $n > N$,

$$f^n(U) \cap V \neq \emptyset.$$

It follows immediately that topologically mixing implies topological transitivity, cf. Definition 2.7.

Example 2.6 (E_{10} is topologically mixing) Any open set $U \subset [0, 1)$ contains a closed subinterval $\tilde{U}$ of size 10^{-m} for some $m \in \mathbb{N}$. Since $E_{10}^n(\tilde{U}) = [0, 1)$ for all $n \geq m$. It then follows that E_{10} is topologically mixing.

Finally, we note that in practise, topological mixing implies sensitive dependence.

Theorem 2.4 (Topological mixing implies sensitive dependence) *Every topologically mixing continuous map, on a metric space that consists of more than one point, has sensitive dependence.*

Proof We use again the notation of Definition 2.8. Let $\Delta > 0$ be such that there exists a pair of points $x_1, x_2 \in X$ such that $d^X(x_1, x_2) > 4\Delta$. We show that Δ is a sensitivity constant.

Let $V_i := B_\Delta(x_i)$, $i = 1, 2$. Let furthermore $x \in X$ and U be a neighbourhood of x. By topological mixing, there exist $N_1, N_2 \in \mathbb{N}$ such that $f^n(U) \cap V_1 \neq \emptyset$ for all $n \geq N_1$ and $f^n(U) \cap V_2 \neq \emptyset$ for all $n \geq N_2$. Hence, for all $n \geq \max\{N_1, N_2\}$ there are points $y_1, y_2 \in U$ with $f^n(y_1) \in V_1$ and $f^n(y_2) \in V_2$. Hence, $d^X(f^n(y_1), f^n(y_2)) \geq 2\Delta$ so that by the triangle inequality $d^X(f^n(y_1), f^n(x)) \geq \Delta$ or $d^X(f^n(y_2), f^n(x)) \geq \Delta$. □

Exercises

2.16 Give an example of a continuous map of the circle $f : S^1 \to S^1$ with the following properties:

(a) f is transitive but has no sensitive dependence.
(b) f has dense periodic orbits but no sensitive dependence.
(c) **f has sensitive dependence but is not topologically mixing.

2.17 Show that for a topologically mixing map on a metric space X without isolated points, any number less than half its diameter is a sensitivity constant. Recall that the diameter of X is defined to be equal to $\sup_{x, \tilde{x} \in X} d^X(x, \tilde{x})$.

2.4 Topological entropy

Topological entropy is the exponential growth rate of the number of essentially different orbit segments of length n. It is a measure of the complexity of the orbit structure of a dynamical system. In particular, positive topological entropy is a considered a manifestation of *chaos*, alternative - in principle - to Devaney's definition, see Definition 2.8.

We recall our setting of a continuous map $f : X \to X$ on a compact metric space X with metric d^X. For each $n \in \mathbb{N}$, let

$$d_n^X(x, \tilde{x}) := \max_{0 \le k \le n-1} d^X(f^k(x), f^k(\tilde{x})) \tag{2.15}$$

be the maximum distance between the first n iterates of f of points $x, \tilde{x} \in X$. The following quantities count the number of orbit segments of length n that are indistinguishable at scale ε.

Definition 2.10 Let $\varepsilon > 0$.

- A subset $A \subset X$ is (n, ε)-*spanning* if for every $x \in X$ there is $\tilde{x} \in A$ such that $d_n^X(x, \tilde{x}) < \varepsilon$. Let $\operatorname{span}(n, \varepsilon, f)$ be the minimum cardinality of an (n, ε)-spanning set.
- A subset $A \subset X$ is (n, ε)-*separated* if any two distinct points in A are at least ε apart in the metric d_n^X. Let $\operatorname{sep}(n, \varepsilon, f)$ be the maximum cardinality of an (n, ε)-separated set.

We note that if X is compact, there are finite (n, ε)-spanning sets and any (n, ε)-separated set is finite, cf. Exercise 2.19.

The *topological entropy* of f, is defined as the exponential growth rate of $\operatorname{span}(n, \varepsilon, f)$ in n, in the limit $\varepsilon \to 0$.

Definition 2.11 The *topological entropy* of $f : X \to X$ is defined as

$$h_{\text{top}}(f) := \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \ln \operatorname{span}(n, \varepsilon, f).$$

It turns out that the corresponding exponential growth rate of $\operatorname{sep}(n, \varepsilon, f)$ is identical.

Theorem 2.5

$$h_{\text{top}}(f) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \ln \operatorname{sep}(n, \varepsilon, f).$$

Proof The proof relies on the assertion that for any $n \in \mathbb{N}$ and $\varepsilon > 0$, we have

$$\operatorname{sep}(n, 2\varepsilon, f) \le \operatorname{span}(n, \varepsilon, f) \le \operatorname{sep}(n, \varepsilon, f). \tag{2.16}$$

To prove the first inequality, let A be an (n, ε)-spanning set of minimal cardinality ($\operatorname{span}(n, \varepsilon, f) = \#(A)$). Note that A is (n, ε)-spanning if and only if

$$X = \bigcup_{y \in A} B_\varepsilon^n(y),$$

where $B^n_\varepsilon(x) := \{\tilde{x} \in X \mid d^X_n(x, \tilde{x}) < \varepsilon\}$ denotes the open ε-ball for the metric d^X_n.

Let A' be an $(n, 2\varepsilon)$-separated set. Then any two points $x_1, x_2 \in A'$ with $x_1 \neq x_2$ are not both contained in $B^n_\varepsilon(y)$ for any $y \in A$. Namely, suppose to the contrary that there exists a $y \in A$ such that $x_1, x_2 \in B^n_\varepsilon(y)$, then by the triangle inequality,

$$d^X_n(x_1, x_2) \le d^X_n(x_1, y) + d^X_n(y, x_2) < \varepsilon + \varepsilon = 2\varepsilon\,,$$

which contradicticts the fact that A' is $(n, 2\varepsilon)$-separated. This means that the number of elements in A' is bounded from above by the cardinality of A, which is $\operatorname{span}(n, \varepsilon, f)$. As the $(n, 2\varepsilon)$-separated A' set was chosen arbitrarily, the first inequality of (2.16) follows.

To prove the second inequality in (2.16), we show that if A is an (n, ε)-separated set of maximal cardinality ($\operatorname{sep}(n, \varepsilon, f) = \#(A)$) then A is also (n, ε)-spanning. Let $x \in X \setminus A$. Since A has maximal cardinality, the set $A \cup \{x\}$ is not (n, ε)-separated. This means that there exists a $y \in A$ such that $d^X_n(x, y) < \varepsilon$ (otherwise $A \cup \{x\}$ would be (n, ε)-separated). Hence, A is (n, ε)-spanning, which implies that $\operatorname{span}(n, \varepsilon, f) \le \#(A) = \operatorname{sep}(n, \varepsilon, f)$.

We now proceed to prove the theorem by observing that (2.16) implies that

$$\operatorname{sep}(n, 2\varepsilon, f) \le \operatorname{span}(n, \varepsilon, f) \le \operatorname{sep}(n, \varepsilon, f)\ \forall n \in \mathbb{N} \text{ and } \varepsilon > 0\,.$$

Consequently, for all $\varepsilon > 0$ we have

$$\limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{sep}(n, 2\varepsilon, f) \le \limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{span}(n, \varepsilon, f) \le \limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{sep}(n, \varepsilon, f)\,.$$

Taking the limit $\varepsilon \to 0$ leads to

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{sep}(n, 2\varepsilon, f) \le h_{\text{top}}(f) \le \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{sep}(n, \varepsilon, f)\,. \tag{2.17}$$

Since the leftmost and rightmost expressions in (2.17) coincide, all inequalities are in fact equalities, which concludes the proof. □

It turns out that the alternative definitions of topological entropy are useful in practice, to find upper and lower bounds in concrete examples.

Example 2.7 (Topological entropy of E_{10} on S^1.) Let A_n be a uniformly distributed set of points on the circle with nearest-neighbour spacing 10^{-n}. We claim that A_{n+k}, with $k \in \mathbb{N}$, is (n, ε)-spanning if $\varepsilon > 10^{-k-1}$. The distance of any $x \in S^1$ to A_{n+k} is smaller or equal than 10^{-n-k}. We note that the distance between any two points, closer to each other than 10^{-1}, grows by a factor 10 at one iteration of E_{10}. Hence, for any $x \in S^1$ there exists $y \in A_{n+k}$ (with $d^{S^1}(x, y) \le 10^{-n-k}$) such that with $0 \le m < n$,

$$d^{S^1}(E^m_{10}(x), E^m_{10}(y)) = 10^m d^{S^1}(x, y) \le 10^{-n-k+m} < \varepsilon.$$

Hence, $d^{S^1}_n(x, y) < \varepsilon$ so A_{n+k} is (n, ε)-spanning.

The fact that A_{n+k} is (n,ε)-spanning yields an upper bound for the topological entropy: since A_{n+k} has cardinality 10^{n+k}, with $k(\varepsilon)$ denoting a choice of k such that $\varepsilon > 10^{-1-k}$, we have

$$\limsup_{n\to\infty}\frac{1}{n}\ln\operatorname{span}(n,\varepsilon,E_{10}) \le \limsup_{n\to\infty}\frac{\ln(10^{n+k(\varepsilon)})}{n} = \ln 10, \quad \Rightarrow \quad h_{\text{top}}(E_{10}) \le \ln 10.$$

It turns out that A_{n+k} is also (n,ε)-separated for all $\varepsilon \le 10^{-k-1}$. Namely, let $x, y \in A_{n+k}$ with $x \ne y$, then $d^{S^1}(x,y) \ge 10^{-k-n}$. If $d^{S^1}(x,y) \ge 10^{-k-1}$ then obviously $d^{S^1}(x,y) \ge 10^{-k-1}$. If $d^{S^1}(x,y) < 10^{-k-1}$ then there exists $m \in \mathbb{N}_0$ such that $0 \le m < n-1$ and

$$10^{-k-n+m} \le d^{S^1}(x,y) < 10^{-k-n+m+1} \le 10^{-1-k},$$

so that

$$d^{S^1}(E_{10}^{n-1-m}(x), E_{10}^{n-1-m}(y)) = 10^{n-1-m} d^{S^1}(x,y) \ge 10^{-k-1}$$

and consequently $d_n^{S^1}(x,y) \ge \varepsilon$.

The fact that A_{n-1+k} is (n,ε)-separated implies that

$$\limsup_{n\to\infty}\frac{1}{n}\ln\operatorname{sep}(n,\varepsilon,E_{10}) \ge \limsup_{n\to\infty}\frac{\ln(10^{n-1+k(\varepsilon)})}{n} = \ln 10 \quad \Rightarrow \quad h_{\text{top}}(E_{10}) \ge \ln 10.$$

It this follows that

$$h_{\text{top}}(E_{10}) = \ln 10,$$

confirming that E_{10}, which was already shown to be chaotic in the sense of Devaney, also has positive topological entropy.

Exercises

2.18 Show that d_n^X, as defined in (2.15), is a metric on X, $d_n^X \ge d_{n-1}^X$ and $d_1^X = d^X$.

2.19 Show that compactness of X implies for a continuous map $f : X \to X$ that there exist finite (n,ε)-spanning sets and that any (n,ε)-separated set is finite.

2.20 Explain how the reasoning in Example 2.7, showing that $h_{\text{top}}(E_{10}) = \ln 10$, can be extended to establish that $h_{\text{top}}(E_k) = \ln k$ for all $k \in \mathbb{N}$ with $k \ge 2$, where $E_k : S^1 \to S^1$ is defined by $E_k(x) = kx \bmod 1$. What is $h_{\text{top}}(E_{-k})$ for $k \ge 2$?

2.21 Show that if $f : X \to X$ is an *isometry* of X, i.e. $d^X(f(x), f(\tilde{x})) = d^X(x,\tilde{x})$ for all $x, \tilde{x} \in X$, then $h_{\text{top}}(f) = 0$.

2.5 Topological conjugacy

Given two dynamical systems, defined by maps $f : X \to X$ and $g : Y \to Y$, it is natural to address the question whether f and g are essentially the same in the sense that there is a one-to-one correspondence between f and g, expressed by the existence of a bijection $h : X \to Y$ such that

$$h \circ f = g \circ h, \tag{2.18}$$

or, equivalently,

$$g = h \circ f \circ h^{-1}. \tag{2.19}$$

The equations (2.18) and (2.19) express an equivalence relation between f and g that we refer to as a *conjugacy*. It is readily checked that h maps (periodic) orbits of f to (periodic) orbits of g (of the same period).

If f and g are continuous, it is natural to consider conjugacies with the additional condition that h is a *homeomorphism*, which means that h and its inverse are also continuous. In this case we we refer to a *topological conjugacy*. It will follow that topologically conjugate continuous maps have very similar topological dynamics. In the case of a topological conjugacy it follows that X is homeomorphic to Y and it is indeed natural to consider topological conjugacies between maps that are defined on the same state space.

The characterisations of dynamics introduced in this chapter so far are all *topological*, in the sense that they are equal for topologically conjugate maps. We summarize this observation in the following propositions, the proofs of which are deferred to Exercise 2.22.

Proposition 2.3 *Let $f : X \to X$ and $g : Y \to Y$ be continuous and topologically conjugate. Then*

- *f is topologically transitive if and only if g is topologically transitive.*
- *f has dense periodic orbits if and only if g has dense periodic orbits.*
- *f is topologically mixing if and only if g is topologically mixing.*
- *f has chaotic dynamics if and only if g has chaotic dynamics.*

Proposition 2.4 (Topological entropy is a topological invariant.) *Let $f : X \to X$ and $g : Y \to Y$ be continuous and topologically conjugate. Then $h_{\mathrm{top}}(f) = h_{\mathrm{top}}(g)$.*

Topological conjugacies are an important tool in the study of topological dynamics. They can for instance be used to bring systems into a standard form (like by a continuous coordinate transformation), which may reveal topological dynamical properties that are more cumbersome to recognize otherwise.

Conjugacies that are not continuous may express an equivalence relation between dynamical systems with different (topological) dynamics, as the following example illustrates.

Example 2.8 Consider the surjective invertible map on a closed interval $f : [0, 1] \to [0, 1]$ with exactly 5 fixed points $0 < p_1 < p_2 < p_3 < 1$ where p_1 and p_3 are attracting (with basins of attraction $(0, p2)$ and $(p2, 1)$, repectively) and the remaining fixed points repelling. Let $h : [0, 1] \to [0, 1]$ be such that $h([0, p_1)) = (p_3, 1]$, $h((p_3, 1]) = [0, p_1)$ and $h|_{[p_1,p_3]} = \mathrm{id}$. Let $g := h \circ f \circ h^{-1}$. Then if $x_1 \in (0, p_1)$ and $x_2 \in (p_1, p_2)$ then $\lim_{n\to\infty} f^n(x_1) = \lim_{n\to\infty} f^n(x_2) = p_1$. However, $\lim_{n\to\infty} g^n(h(x_1)) = p_3 \neq \lim_{n\to\infty} g^n(h(x_2)) = f^n(x_2) = p_1$.

It turns out that the relation of *topological semi-conjugacy*, which is weaker than topological conjugacy, is of particular practical relevance.

Definition 2.12 (Semi-conjugacy) Let $f : X \to X$ and $g : Y \to Y$ be continuous. Then f is called an *extension* of g, and g a *factor* of f, if there exists a surjective map $h : X \to Y$ such that

$$h \circ f = g \circ h. \tag{2.20}$$

If h is moreover continuous, we speak of a *topological semi-conjugacy*.

We now show that the useful symbol representation of E_{10}'s action on decimal expansions, is related to a semi-conjugacy.

Example 2.9 (E_{10} is topologically semi-conjugate to a shift on 10 symbols.) We earlier observed that the understanding of the dynamics of E_{10} on S^1 is facilitated by realizing how E_{10} acts on the decimal expansion of points in $[0, 1)$, that parametrizes S^1. The decimal expansions $0.x_0x_1x_2\ldots$ with $x_i \in \{0, \ldots, 9\}$ of points in $[0, 1)$ are represented by the decimals which are (half infinite) sequences of 10 symbols. Let Σ_{10}^+ denote the set of all such symbol sequences. The interpretation of these symbol sequences as decimal expansions takes the form of the function $h : \Sigma_{10}^+ \to S^1 \simeq [0, 1)$

$$h(x_0x_1\ldots) := \sum_{i=0}^{\infty} \frac{x_i}{10^{i+1}} \bmod 1, \tag{2.21}$$

where the "mod 1" appears here only to guarantee that $h(\overline{9}) = 0$, using the notational convention that $\overline{a}$ represents a periodic repetition of a. In the decimal representation, the action of E_{10} is

$$E_{10}(0.x_0x_1x_2\ldots) = 0.x_1x_2\ldots,$$

thus inducing the shift $\sigma : \Sigma_{10}^+ \to \Sigma_{10}^+$ on the sequence representing the decimals, defined by

$$\sigma(x_0x_1x_2\ldots) = x_1x_2\ldots.$$

These observations can be formalized as the following semi-conjugacy

$$E_{10} \circ h = h \circ \sigma,$$

so that E_{10} is a factor of σ. We recall that h is surjective but not invertible since decimal expansions are not always unique, since sometimes two sequences have the same image under h, eg. $h(1\overline{0}) = h(0\overline{9})$.

The sequence space Σ_{10}^+ can now be equipped with a suitable metric that retains the property of the decimal expansions that the order of closeness is in first order represented by the position of the first differing digit, while on the other hand also facilitating the analysis of the shift dynamical system. We define $d^{\Sigma^+} : \Sigma_{10}^+ \times \Sigma_{10}^+ \to \mathbb{R}$ by[5]

$$d^{\Sigma^+}(x_0 x_1 \ldots, y_0 y_1 \ldots) := \sum_{i=0}^{\infty} \frac{\delta(x_i, y_i)}{3^i}, \text{ where } \delta(x_i, y_i) := \begin{cases} 0 \text{ if } x_i = y_i, \\ 1 \text{ if } x_i \neq y_i. \end{cases} \quad (2.22)$$

It is straightforward to check that d^{Σ^+} is a metric and it turns out that with this metric, Σ_{10}^+ is compact. If, moreover, S^1 is equipped with the standard (arclength) metric, then h is continuous.

While topological semi-conjugacies do not necessarily preserve all topological dynamical properties, they can have important implications.

Proposition 2.5 *Let $f : X \to X$ and $g : Y \to Y$ be topologically semi-conjugate, g being a factor of f, then:*

- *g has dense periodic orbits if f has dense periodic orbits.*
- *g is transitive if f is transitive.*
- *g is topologically mixing if f is topologically mixing.*
- *g is chaotic if f is chaotic, unless Y consists of a single periodic orbit of g.*

We defer the proof of this proposition to Exercise 2.25.

Theorem 2.6 *Let $f : X \to X$ be topologically semi-conjugated to $g : Y \to Y$, g being a factor of f, then $h_{\mathrm{top}}(g) \leq h_{\mathrm{top}}(f)$.*

Suppose that the semi-conjugacy is obtained by a map $h : X \to Y$ such that the cardinality of the set of preimages of h for any point $y \in Y$ is uniformly bounded, i.e. $\sup_{y \in Y} \#(h^{-1}(y)) \leq C$ for some $C \in \mathbb{N}$, then $h_{\mathrm{top}}(f) = h_{\mathrm{top}}(g)$.

Proof Due to compactness of X and Y, which is the implicit assumption throughout this text, the conjugacy h is uniformly continuous. That means that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x, \tilde{x} \in X$

$$d^X(x, \tilde{x}) < \delta \Longrightarrow d^Y(h(x), h(\tilde{x})) < \varepsilon\,.$$

This implies that for any $i \in \mathbb{N}$

$$d^X(f^i(x), f^i(\tilde{x})) < \delta \Longrightarrow d^Y(h(f^i(x)), h(f^i(\tilde{x}))) = d^Y(g^i(h(x)), g^i(h(\tilde{x}))) < \varepsilon\,,$$

so that

$$h(B_\delta^{n,X}(x)) \subset B_\varepsilon^{n,Y}(h(x)) \quad \text{for all } x \in X\,,$$

where $B_\varepsilon^{n,Z}$ denotes the open ball of radius ε in metric space Z with metric d_n^Z, cf. (2.15).

By the surjectivity of h, it follows that

$$\mathrm{span}(n, \delta, f) \geq \mathrm{span}(n, \varepsilon, g), \quad (2.23)$$

[5] This same metric is also used for Σ_k^+, with $k > 1$ integer. Hence we omit here the subscript "10".

which implies that

$$\limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{span}(n, \delta, f) \geq \limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{span}(n, \varepsilon, g),$$

so that by taking the limit $\varepsilon \to 0$ and $\delta \to 0$ (which is implied by $\varepsilon \to 0$) we obtain $h_{\mathrm{top}}(f) \geq h_{\mathrm{top}}(g)$.

Now, if $\sup_{y\in Y} \#(h^{-1}(y)) \leq C \in \mathbb{N}$, we note that by surjectivity of h^{-1}, h^{-1} maps any (ε, n)-spanning set in Y (for g) is mapped to a (δ, n)-spanning set in X (for f) with at most C times the cardinality. Hence, application of this observation to the minimal (ε, n)-spanning set in Y (for g) yields

$$C\operatorname{span}(n, \varepsilon, g) \geq \operatorname{span}(n, \delta, f) \overset{(2.23)}{\Rightarrow} C\operatorname{span}(n, \varepsilon, g) \geq \operatorname{span}(n, \delta, f) \geq \operatorname{span}(n, \varepsilon, g)$$

so that

$$\limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{span}(n, \varepsilon, g) \geq \limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{span}(n, \delta, f) \geq \limsup_{n\to\infty} \frac{1}{n} \ln \operatorname{span}(n, \varepsilon, g),$$

which implies by taking, as above, the limit $\varepsilon \to 0$ and $\delta \to 0$ (implied by $\varepsilon \to 0$) that $h_{\mathrm{top}}(f) = h_{\mathrm{top}}(g)$. □

We illustrate the use of these implications in the context of the circle map E_{10}. We show that the shift σ is topologically mixing and chaotic on Σ_{10}^+ and that its topological entropy is equal to ln 10. By the topological semi-conjugacy of E_{10} and σ, and the fact that $\#(h^{-1}(x)) \leq 2$ for all $x \in S^1$, it follows that E_{10} has the exact same attributes.

Example 2.10 We show that the dynamics of the shift map $\sigma : \Sigma_{10}^+ \to \Sigma_{10}^+$ is chaotic, topologically mixing and has topological entropy ln 10. At the basis of the argument lies the fact that the cylinder sets

$$C_{x_0\ldots x_{m-1}} := \{s_0 s_1 s_2 \ldots \in \Sigma_{10}^+ \mid x_i = s_i,\ i = 0, \ldots, m-1\} \tag{2.24}$$

are the open balls with radius 3^{-m} around any element from this cylinder set, cf. Exercise 2.24.

Topological mixing follows from the observation that every open set contains a cylinder set and for every cylinder set $C_{\tilde{x}_0\ldots\tilde{x}_{\tilde{m}-1}}$, $\tilde{m} > 1$, for any $n \geq m$ there exists $y \in C_{x_0\ldots x_{m-1}}$ such that $\sigma^n(y) \in C_{\tilde{x}_0\ldots\tilde{x}_{\tilde{m}-1}}$.

To show that the shift is chaotic, it suffices to show that periodic orbits are dense, since topological mixing already implies transitivity and sensitive dependence. Density of periodic orbits follows from the fact that every cylinder set $C_{x_0\ldots x_{n-1}}$ contains a periodic orbit, namely $\overline{x_0 \ldots x_{n-1}}$.

Finally, the topological entropy of the shift σ on Σ_{10}^+ can be determined exactly. While the cylinder set $C_{x_0\ldots x_{m-1}}$ is the open ball with radius 3^{-m+1} around any element from this cylinder set, $C_{x_0\ldots x_{m-1}\ldots x_{m-1+n-1}}$ is the open ball with radius 3^{-m+1} for the $d_n^{\Sigma^+}$ metric (2.15). Consequently, Σ_{10}^+ is covered entirely and by no fewer than exactly 10^{m+n-1} such open balls. Hence, $\mathrm{span}(n, 3^{-m+1}, \sigma) = 10^{m+n-1}$ and

$$h_{\mathrm{top}}(\sigma) = \lim_{m\to\infty} \lim_{n\to\infty} \frac{1}{n} \ln 10^{m+n-1} = \ln 10.$$

We finally note that our analysis of E_{10} readily extends to the maps $E_k : S^1 \to S^1$, $k \in \mathbb{Z}$ with $|k| \geq 2$, defined by

$$E_k(x) = kx \bmod 1, \tag{2.25}$$

equipping the space of semi-infinite sequences of k symbols, Σ_k^+, with the same metric d^{Σ^+} (2.22), with the conclusion that E_k is chaotic, topologically mixing and $h_{\mathrm{top}}(E_k) = \ln |k|$, cf. Exercise 2.26

Exercises

2.22 Prove Propositions 2.3 and 2.4.

2.23 Consider the discussion of circle maps and their lifts in Exercise 2.1. Show that continuous circle maps f are topologically semi-conjugate to their lifts F, in the sense that f is a factor of F.

2.24 Show that d^{Σ^+}, as defined in (2.22), is a metric and that the *cylinder sets* (2.24) are open balls of size 3^{-m+1} in this metric. Use the latter to show that h, as defined in (2.21), is continuous.

2.25 Prove Proposition 2.5.

2.26 Show that for all $k \geq 2$, E_k (2.25) is topologically conjugate to the shift map on Σ_k^+. Show that E_k is chaotic and topologically mixing and that $h_{\mathrm{top}}(E_k) = \ln k$.

Appendix: some preliminaries concerning metric spaces

In this appendix, we summarize some elementary notions and facts of metric spaces that we use in this chapter.

Definition 2.13 (Metric space) Let X be a set, then a function $d^X : X \times X \to \mathbb{R}$ is *metric* for X is it satisfies the following properties:

1. symmetry: $d^X(x, \tilde{x}) = d^X(\tilde{x}, x)$, $\forall\, x, \tilde{x} \in X$,

2. positivity: $d^X(x, \tilde{x}) = 0$ if and only if $x = \tilde{x}$, $\forall\, x, \tilde{x} \in X$,

3. triangle inequality: $d^X(x, \tilde{x}) + d^X(\tilde{x}, \hat{x}) \geq d^X(x, \hat{x}), \forall\, x, \tilde{x}, \hat{x} \in X$.

A set X endowed with a metric d^X is called a *metric space*.

A well-known example of a metric space is $\mathbb{R}^n$, endowed with the Euclidean metric

$$d^{\mathbb{R}^n}(x, \tilde{x}) := \sqrt{\sum_{i=1}^{n} (x_i - \tilde{x}_i)^2},$$

where the $x = (x_1, \ldots, x_n)$ and $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_n)$.

Definition 2.14 (Open ball) Let X be a metric space and $r > 0$, then the open r-ball around x is defined as

$$B_r(x) := \{\tilde{x} \in X \mid d^X(x, \tilde{x}) < r\}.$$

Definition 2.15 (Open set) Let X be a metric space then $U \subset X$ is called *open*, if for every $x \in U$, there exists $r > 0$ such that $U \supset B_r(x)$.

Open sets have the useful property that countable unions and countable intersections of open sets are open.

Definition 2.16 (Bounded set) Let X be metric space and $U \subset X$, then U is *bounded* if there exist $r > 0$ and $x \in X$ such that $U \subset B_r(x)$.

Definition 2.17 (Continuous function) Let X and Y be metric spaces, then a function $f : X \to Y$ is *continuous* if $\forall x \in X$ and $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $f(B_\delta(x)) \subset B_\varepsilon(f(x))$.

An alternative characterisation of continuity for a function $f : X \to Y$ is that for all open $A \subset Y$, its pre-image $f^{-1}(A) := \{x \in X \mid f(x) \in A\}$ is open.

Definition 2.18 (Closure) Let X be a metric space and $U \subset X$. Then the *closure* of U is defined as

$$\overline{U} := \{x \in X \mid \exists \{x_n \in U\}_{n \in \mathbb{N}} \text{ such that } \lim_{n \to \infty} x_n = x\}.$$

Definition 2.19 (Dense set) Let X be a metric space and $U \subset X$, then U is *dense* in X is for all $x \in X$ and all $r > 0$, $B_r(x) \cap U \neq \emptyset$.

Definition 2.20 (Closed set) A set $U \subset X$ is *closed* if $X \setminus U$ is open.

Countable unions and countable intersections of closed sets are also closed.

Definition 2.21 (Compact set) Let X be a metric space, then $U \subset X$ is *compact* if every sequence $\{x_n \in U\}_{n \in \mathbb{N}}$ has a converging subsequence with limit in U, i.e. $\forall \{x_n \in U\}_{n \in \mathbb{N}} \subset U$, $\exists$ sequence $\{n_m\}_{m=0}^{\infty} \subset \mathbb{N}$ with $\lim_{m \to \infty} n_m = \infty$ such that $\lim_{m \to \infty} x_{n_m} \in U$.

There is a useful practical characterisation of compact sets in Euclidean spaces:

Theorem 2.7 (Bolzano-Weierstrass) *The compact subsets of $\mathbb{R}^n$ endowed with the Euclidean metric are those subsets of $\mathbb{R}^n$ that are closed and bounded.*

In general, a metric space is compact if and only if it is *complete* (all Cauchy sequences in the metric space converge within the metric space) and *totally bounded* (for every $r > 0$, the metric space can be covered by finitely many open r-balls).

In this text, we use the following properties of compact metric spaces without proof:

- For any $\varepsilon > 0$, there exists a finite ε-cover of any compact metric space: for any compact metric space X and any $\varepsilon > 0$, there exists a finite $S \subset X$ such that $X \subset \{B_\varepsilon(x) \mid x \in S\}$.
- Any open cover of a compact metric space contains a finite subcover.
- Any compact metric space contains a countable dense subset.

We also use properties of continuous functions on a compact metric space:

- Any continuous function $f : X \to Y$ between a compact metric space X and another metric space Y is *uniformly continuous*: $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $\forall x \in X$, $f(B_\delta(x)) \subset B_\varepsilon(f(x))$.
- Any continuous function $f : X \to \mathbb{R}$ on a compact metric space X is bounded and attains its minimum and maximum in X, i.e. there exists $\tilde{x}, \hat{x} \in X$ such that $f(\tilde{x}) = \sup_{x \in X} f(x)$ and $f(\hat{x}) = \inf_{x \in X} f(x)$.

Chapter 3
Symbolic dynamics

In this chapter we develop the technique of *symbolic dynamics*, which establishes a comparison beween the dynamics of a system to a shift on a suitable space of symbolic sequences, generalizing the observations in Example 2.9 concerning the representation of E_{10} by a shift on 10 symbols.

3.1 Topological Markov partitions

The key to symbolic dynamics is the partitioning of the phase space (or an attractor) into a finite set of relevant labelling domains. We here confine the discussion to the one-dimensional setting of (finite sets of) closed intervals and the circle.

Definition 3.1 (Expanding map) A map $f : X \to X$ on a compact metric space X is *expanding* if there exist $\varepsilon > 0$ and $L > 1$ such that for all $x, \tilde{x} \in X$ with $d^X(x, \tilde{x}) < \varepsilon$, $d^X(f(x), f(\tilde{x})) \geq L d^X(x, \tilde{x})$.

It may happen that while f is not expanding, eventually, for some $n \in \mathbb{N}$, f^n is expanding. We call such f *topologically expanding*, cf. Exercise 3.1.

From here on we focus on the one-dimensional setting. In the remainder of this section, I denotes a one-dimensional compact set (in examples usually a closed interval, a finite set of closed intervals, or the circle).

Proposition 3.1 (Smooth expanding map) *A C^1 map $f : I \to I$, is expanding if and only if $|f'(x)| > 1$.*

Note that this definition of expanding map in the one-dimensional setting implies that I must be a circle. The proof of this proposition is deferred to Exercise 3.2. It should be noted that

Definition 3.2 (Finite topological partition) A finite set of pairwise disjoint open intervals $\mathcal{R} := \{R_0, \ldots, R_{k-1}\}$ is a finite *topological partition* of I if

$$I = \overline{R_0} \cup \ldots \cup \overline{R_{k-1}}, \tag{3.1}$$

The aim of symbolic dynamics is to obtain a representation of individual orbits by a recording of the partition intervals that are visited by the forward orbit. In particular, we aim to represent $O_f^+(x)$ be a sequence $i_n \in \{0, \dots, k-1\}$, $n \in \mathbb{N}_0$, such that

$$f^n(x) \in \overline{R_{i_n}}.$$

We note that as the intersections $\overline{R_i} \cap \overline{R_j}$ are not necessarily empty if $i \neq j$, there may be some ambiguity in this approach, but this turns out to be manageable.

Consider the refinement of the partition $\mathcal{R}$ of I to an induced partition $\mathcal{R}_1$ of I constructed as follows:

$$\mathcal{R}_1 = \{R_{ij} \mid i, j \in \{0, \dots, k-1\}\}, \text{ where } R_{ij} := R_i \cap f^{-1}(R_j).$$

This partition is such that if $x \in \overline{R_{ij}}$ then $x \in \overline{R_i}$ and $f(x) \in \overline{R_j}$. In a similar way we may define subsequent refinements as

$$\mathcal{R}_m = \{R_{i_0 \dots i_{m-1}} \mid i_0, \dots i_{m-1} \in \{0, \dots, k-1\}, \tag{3.2}$$

where

$$R_{i_0 \dots i_{m-1}} := \bigcap_{n=0}^{m-1} f^{-n}(R_{i_n}) \tag{3.3}$$

for any integer $m > 1$, so that if $x \in \overline{R_{i_0 \dots i_{m-1}}}$ then $f^n(x) \in \overline{R_{i_n}}$ for all $n \in \{1, \dots, m-1\}$.

We illustrate the above partition refinements for the circle map E_{10} on S^1.

Example 3.1 (Partition for symbolic dynamics of E_{10}.) Consider the topological partition of the circle S^1 into 10 equally sized adjacent open intervals R_i, $i = 0, \dots, 9$ so that

$$S^1 = \cup_{i=0}^9 \overline{R_i},$$

see Fig. 3.1. This partition coincides with the domains that define the first decimal in $[0, 1)$, starting at R_0, with $R_i := (i/10, (i+1)/10)$. As f^{-1} shrinks the length of any circle segment by a factor 10, one easily sees that for all $m > 1$

$$R_{i_0 \dots i_{m-1}} := \bigcap_{n=0}^{m-1} f^{-n}(R_{i_n}) = (\sum_{n=0}^{m-1} i_n \cdot 10^{-n-1}, \sum_{n=0}^{m-1} i_n \cdot 10^{-n-1} + 10^{-m}).$$

We observe that these intervals are non-empty for any choice of $i_0 \dots i_{m-1}$ and that

$$\lim_{m \to \infty} R_{i_0 \dots i_{m-1}} = \lim_{m \to \infty} \sum_{n=0}^{m-1} i_n \cdot 10^{-n-1} = 0.i_0 i_1 i_2 i_3 \dots \mod 1$$

confirming the fact that the symbolic sequence in Σ_{10}^+ corresponds to the decimal expansion of the limit point. Recall that the right-hand side expression corresponds to the function h in (2.21), that semi-conjugates E_{10} to the shift on Σ_{10}^+.

It is easy to extend these results to all the expanding circle maps E_k, $k \geq 2$.

Example 3.2 (Symbolic dynamics for expanding circle maps) We can further extend the analysis in the previous example also to general (nonlinear) expanding circle maps $f : S^1 \to S^1$ with positive degree. We first note that a circle map of positive degree k, cf. Exercise 2.1, can be expanding only if $k \geq 2$. Chosing without loss of generality a parametrisation of the circle $[0, 1)$ such that $f(0) = 0$, let $\{r_i \in [0, 1)\}_{i=1}^{k-1}$ denote the $k - 1$ other distinct points such that $f(r_i) = 0$. We now consider the partition of the circle induced by the points r_i, i.e. $\mathcal{R} = \{(0, r_1), \ldots, (r_{k-1}, 1)\}$. We now may proceed as before and define refined partitions $\mathcal{R}_n$ so that the intervals $\overline{R_{i_0 \ldots i_{n-1}}}$ converge to single points on the circle in the limit that $n \to \infty$. Defining $h : \Sigma_k^+ \to S^1$ accordingly as

$$h(i_0 i_1 i_2 \ldots) := \lim_{n\to\infty} \overline{\bigcap_{m=0}^{n-1} f^{-m}(R_{i_m})}, \tag{3.4}$$

one establishes the topological semi-conjugacy

$$h \circ \sigma = f \circ h, \tag{3.5}$$

in analogy to the case when $f = E_k$.

It is interesting to note that the above topological semi-conjugacies of E_k and any other expending circle map of degree k to the shift σ on Σ_k^+ can be used to prove the result that any two expanding circle maps of degree k are topologically conjugate, cf. Exercise 3.5.

The transformation h, as defined in (3.4), is key to establising symbolic dynamics, as it serves to semi-conjugate the symbolic (shift) dynamics to the map f under consideration.

To gain potential insight from the proposed symbolic description of orbits one would like to establish the fact that not more than one orbit can be associated to a certain symbolic representation. In particular, it is natural to ask whether if a

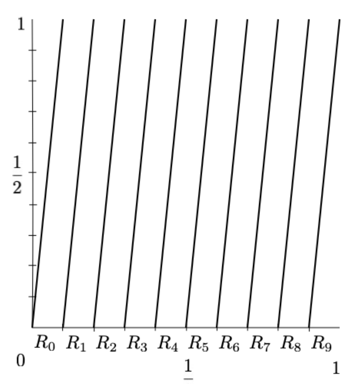

Fig. 3.1: Graph of E_{10} (3.6), with $S^1 \simeq [0, 1)$, and Markov partition $\{R_i\}_{i=0}^{9}$ of S^1 as discussed in Example 3.1.

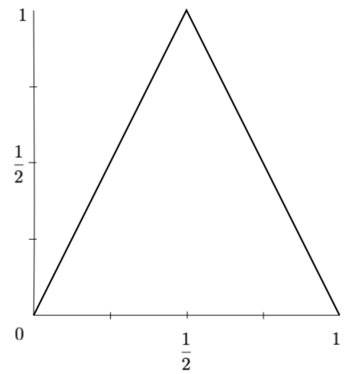

Fig. 3.2: Graph of the tent map (3.6).

semi-infinite sequence of k symbols $i_0 i_1 i_2 \ldots \in \Sigma_k^+$ is such that $R_{i_0 \ldots i_{m-1}} \neq \emptyset$ for all $m \in \mathbb{N}$, the sequence of intervals $\overline{R_{i_0 \ldots i_{m-1}}}$ converge to a single point in I, in the limit $m \to \infty$? This question can be answered in the affirmative if f is *piecewise expanding* on $\mathcal{R}$.

Definition 3.3 (Piecewise expanding one-dimensional map) A continuous map $f : I \to I$ is piecewise expanding if there exists a finite topological partition $\mathcal{R} = \{R_1, \ldots, R_k\}$ such that f is expanding on R_i for all $i \in \{1, \ldots, k\}$.

Piecewise expanding maps need not be expanding.

Example 3.3 (Tent map) Let $T : [0, 1] \to [0, 1]$ be defined as

$$T(x) = \begin{cases} 2x, & \text{if } x \in [0, 1/2], \\ 2 - 2x, & \text{if } x \in (1/2, 1]. \end{cases} \tag{3.6}$$

See Fig. 3.2 for a sketch of its graph. Then T is not expanding since in any neighbourhood of $1/2$ one can find two points that are mapped by T to the same target

$$T(1/2 - \varepsilon) = T(1/2 + \varepsilon) = 1 - 2\varepsilon,$$

violating the condition in Definition 3.1. However, T is piecewise expanding on the partition $\{(0, 1/2), (1/2, 1)\}$ by the result of Proposition 3.1 as $|T'(x)| = 2$ on both (open) partition elements.

Lemma 3.1 *Suppose $f : I \to I$ is piecewise expanding with respect to a partition $\mathcal{R}$ of I. Then, if $i_0 i_1 i_2 \ldots \in \Sigma_k^+$ is such that $R_{i_0 \ldots i_{n-1}} \neq \emptyset$ for all integer $n \geq 2$, we have*

$$\lim_{n \to \infty} \overline{R_{i_0 \ldots i_{n-1}}} \in I.$$

Proof By continuity of f, $R_{i_0 \ldots i_{n-1}}$ is an open interval inside $R_{i_0 \ldots i_{n-2}}$ for all $n \geq 2$. Also, by the fact that f is piecewise expanding, there exists $L > 1$ such that

$|R_{i_0\dots i_{n-1}}| \leq L^{-1}|R_{i_0\dots i_{n-2}}|$ for all $n \geq 2$ and consequently the lengths of the relevant intervals shrink exponentially fast by virtue of which the intervals converge to a single point in I. ◻

This lemma implies that in addition to the fact that every orbit of f having a symbolic representation, which follows by construction, every symbolic sequence represents the orbit of at most one initial point. In order to establish the desired topological semi-conjugacy between shift dynamics and f, it remains to determine which set of sequences $\Sigma_{\text{adm}} \subset \Sigma_k^+$ is *admissible*, in the sense that they represent orbits of f. In the case of orientation preserving expanding circle maps of degee $k > 1$ we have seen that $\Sigma_{\text{adm}} = \Sigma_k^+$ so that all sequences represent (initial conditions of) orbits. However, for piecewise expanding maps this need not necessarily be the case.

In any case, the above establishes the topological semi-conjugacy (3.5) with $\sigma : \Sigma_{\text{adm}} \to \Sigma_{\text{adm}}$ and $h : \Sigma_{\text{adm}} \to S^1$ surjective. However, the dynamics of the shift σ on Σ_{adm} depends on the admissible sequences contained in this space.

An important class of sequence spaces on which the dynamics of the shift can be comprehensively understood, are so-called *topological Markov chains*.

Definition 3.4 (Topological Markov chain.) A topological Markov chain $\Sigma_{k,A}^+$ with k symbols is a set of semi-infinite symbol sequences $i_0 i_1 i_2 \ldots \in \Sigma_k^+$ characterised by rules concerning the admissibility of consecutive symbols only. These rules are summarized in a $k \times k$ *connectivity matrix* A, where $A_{ij} = 1$ if symbol j is allowed to appear after symbol i, and $A_{ij} = 0$ otherwise. $\Sigma_{k,A}^+$ is a metric space, endowed with the metric $d^{\Sigma_k^+}$ (naturally induced from Σ_k^+).

The *Markov graph* associated with a topological Markov chain $\Sigma_{k,A}^+$ is a directed graph with vertices labelled 0 to $k-1$ and a directed edge from vertex i to vertex j, whenever $A_{ij} = 1$. The sequences in the topological Markov chain represent sequences of vertices on (semi-infinite) paths along directed edges in the Markov graph.

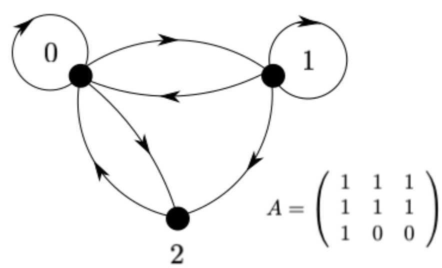

Fig. 3.3: Example of a Markov graph for a connectivity matrix A.

In order to avoid trivial redundancy, we always quietly that all k symbols appear in (some of the) semi-infinite sequences in $\Sigma_{k,A}^+$. This implies for the connectivity matrix A that it has no rows with only zeros, and for the associated Markov graph that every vertex has an outgoing directed edge.

Definition 3.5 (Finite Markov partition) Let $f : I \to I$ be piecewise expanding on a topological partition $\mathcal{R}$ of I. $\mathcal{R}$ is called a *finite Markov partition* of I if for all $i \in \{0, \ldots, k-1\}$ there exist $S_i \subset \{0, \ldots, k-1\}$ such that

$$\begin{aligned} f(R_i) &\supset R_j \quad\quad \text{if } j \in S_i, \\ f(R_i) \cap R_j &= \emptyset \text{ if } j \notin S_i. \end{aligned} \tag{3.7}$$

This definition of a Markov partition is tailored to the one-dimensional setting considered in this text. For a more general discussion, see for instance the expository paper by Adler [1].

Proposition 3.2 *Let $f : I \to I$ be piecewise expanding on a finite Markov partition $\mathcal{R}$. Then f is topologically semi-conjugate to the shift map on a topological Markov chain.*

Proof From the discussion above, it remains to be shown that $\Sigma_{\text{adm}} = \Sigma^+_{k,A}$ for some connectivity matrix A. Indeed, it is readily verified that for a partition satisfying (3.7), the set of admissible sequences is exactly given by a Markov chain with connectivity matrix A satisfying $A_{ij} = 1$ if and only if $j \in S_i$. □

If the connectivity matrix A of a topological Markov chain is such that $A_{ij} = 1$ for all $i, j \in \{0, \ldots, k-1\}$, then $\Sigma^+_k = \Sigma^+_{k,A}$ and we refer to the corresponding shift σ as the *full shift (on k symbols)*. The corresponding Markov graph is then fully connected. A shift on a finite topological Markov chain is an example of a so-called *subshift of finite type*.

Example 3.4 (Markov partitions for expanding circle maps.) The partitions proposed in Example 3.1 and Example 3.2 are Markov partitions since the closure of the image of any partition element is equal to the entire circle. The corresponding Markov graphs are fully connected and a topological semi-conjugacy to a full shift on two symbols is established.

Example 3.5 (Markov partition for the tent map.) The partition of the piecewise expanding tent map, introduced in Example 3.3, is also a Markov partition since in analogy to Example 3.4, the closure of the image of each partition element is equal to the entire state space. Hence, the tent map is a factor of the full shift on two symbols. As the cardinality of $h^{-1}(x)$ is at most two for each $x \in [0, 1]$, cf. Exercise 3.3, by Proposition 2.5 it then follows that the tent map is chaotic and topologically mixing.

The following example shows that Markov partitions of piecewise expanding maps need not give rise to full shifts.

Example 3.6 Consider the map of the interval $f : [0, 1] \to [0, 1]$ given by

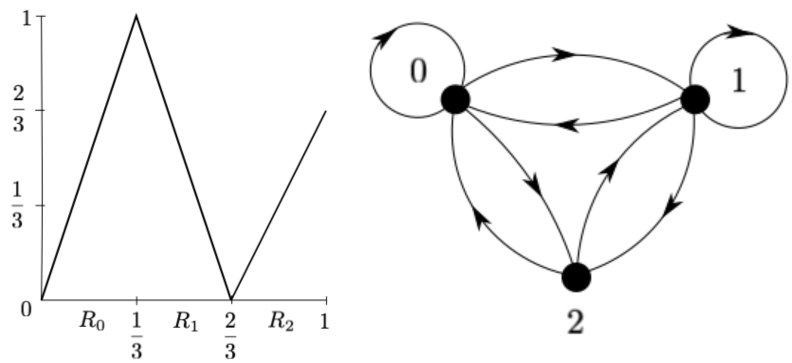

Fig. 3.4: Graph of the map (3.8) with partition (3.9) and the associated Markov graph for the connectivity matrix (3.10).

$$f(x) = \begin{cases} 3x, & \text{if } x \in \left[0, \frac{1}{3}\right], \\ 2 - 3x, & \text{if } x \in \left(\frac{1}{3}, \frac{2}{3}\right], \\ 2x - \frac{4}{3}, & \text{if } x \in \left(\frac{2}{3}, 1\right]. \end{cases} \tag{3.8}$$

This map is piecewise expanding with respect to the partition

$$\mathcal{R} = \left\{ \left(0, \tfrac{1}{3}\right), \left(\tfrac{1}{3}, \tfrac{2}{3}\right), \left(\tfrac{2}{3}, 1\right) \right\}. \tag{3.9}$$

This partition is also a Markov partition since

$$\overline{f\left(\left(0, \tfrac{1}{3}\right)\right)} = \overline{f\left(\left(\tfrac{1}{3}, \tfrac{2}{3}\right)\right)} = [0, 1] \text{ and } \overline{f\left(\left(\tfrac{2}{3}, 1\right)\right)} = \left[0, \tfrac{2}{3}\right] = \overline{\left(0, \tfrac{1}{3}\right) \cup \left(\tfrac{1}{3}, \tfrac{2}{3}\right)}.$$

It this follows that f is a factor of the shift on $\Sigma^+_{3,A}$ with

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \tag{3.10}$$

where $A_{22} = 0$ since orbits cannot move from $R_2 = \left(\frac{2}{3}, 1\right)$ to itself.

We consider the semi-conjugacy in a little more detail. In Fig. 3.5 we sketch the Markov partition $\mathcal{R}$ and its first two refinements $\mathcal{R}_2$ and $\mathcal{R}_3$, cf. (3.2) and (3.3). When $n \geq 2$, $|R_{i_0 \dots i_{n-1}}| = \left(\frac{1}{3}\right)^{n-m} \cdot \left(\frac{1}{2}\right)^m$ if $i_0 \dots i_{n-2}$ contains m times the symbol 2.

The semi-conjugacy $h \circ \sigma = f \circ h$ is established by $h : \Sigma^+_{3,A} \to [0, 1]$,

$$h(i_0 i_1 i_2 \dots) := \lim_{n \to \infty} \overline{\bigcap_{m=0}^{n-1} f^{-m}(R_{i_m})}, \tag{3.11}$$

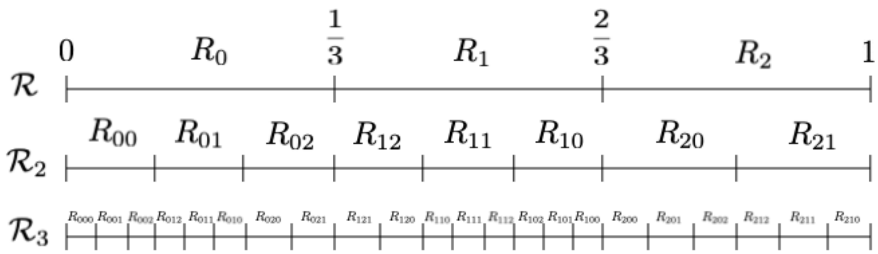

Fig. 3.5: The Markov partition (3.9) for f (3.8) and its refinements $\mathcal{R}_2$ and $\mathcal{R}_3$.

cf. (3.4). Due to the one-dimensionalilty of the subsequent refined partitions, it is clear that $\#(h^{-1}(x)) \leq 2$ for all $x \in [0,1]$, since each point can possibly be approached by converging labelling intervals from at most two directions. The points where ambiguity of coding arises are precisely the boundary points of the Markov partition and its refinements. These form a countable dense subset of $[0,1]$.

For example, we find that

$$h^{-1}(\tfrac{1}{3}) = \{021\overline{0}, 121\overline{0}\},\ h^{-1}(\tfrac{1}{9}) = \{0021\overline{0}, 0121\overline{0}\},\ h^{-1}(\tfrac{2}{3}) = \{1\overline{0}, 2\overline{0}\},$$

with the following corresponding trajectories under f:

$$O_f^+(\tfrac{1}{3}) = \{\tfrac{1}{3}, 1, \tfrac{2}{3}, 0, 0, 0, \ldots\},\ O_f^+(\tfrac{1}{9}) = \{\tfrac{1}{9}, \tfrac{1}{3}, \tfrac{2}{3}, 0, 0, 0, \ldots\}\ O_f^+(\tfrac{2}{3}) = \{\tfrac{2}{3}, 0, 0, 0, \ldots\}.$$

Finally, while a map may be piecewise expanding with respect to a given partition $\mathcal{R}$ that is not Markov, it should be noted that there may exist a Markov partition that is a refinement of $\mathcal{R}$, which is naturally also piecewise expanding, as the following example illustrates.

Example 3.7 Consider the map of the interval $f : [0,1] \to [0,1]$ given by

$$f(x) = \begin{cases} 3x, & \text{if } x \in \left[0, \frac{1}{3}\right], \\ 2 - 3x, & \text{if } x \in \left(\frac{1}{3}, \frac{2}{3}\right], \\ \frac{3}{2}x - 1, & \text{if } x \in \left(\frac{2}{3}, 1\right]. \end{cases} \tag{3.12}$$

This map is piecewise expanding with respect to the same partition (3.9) as in the previous example. However, this partition is not a Markov partition since $\overline{f\left(\frac{2}{3}, 1\right)} = \left[0, \frac{1}{2}\right]$ which is not the closure of adjacent partition elements from this partition. However, the refined partition

$$\mathcal{R} = \left\{\left(0, \tfrac{1}{3}\right), \left(\tfrac{1}{3}, \tfrac{1}{2}\right), \left(\tfrac{1}{2}, \tfrac{2}{3}\right), \left(\tfrac{2}{3}, 1\right)\right\} \tag{3.13}$$

is such that

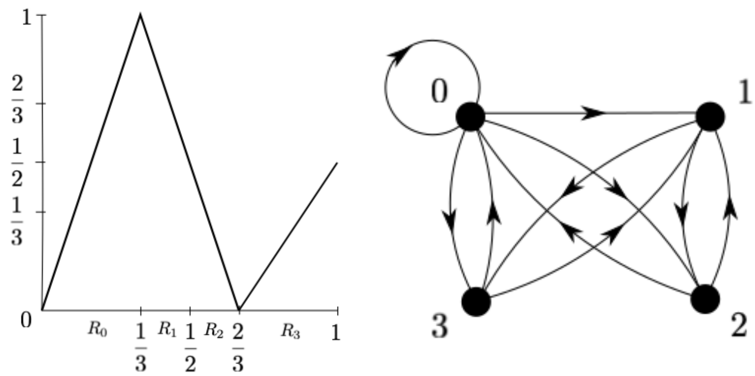

Fig. 3.6: Graph of the map (3.12), Markov partition (3.13) and Markov graph for the connectivity matrix (3.14).

$$\begin{aligned}
\overline{f\left(\left(0,\tfrac{1}{3}\right)\right)} &= [0,1] = \overline{\left(0,\tfrac{1}{3}\right) \cup \left(\tfrac{1}{3},\tfrac{1}{2}\right) \cup \left(\tfrac{1}{2},\tfrac{2}{3}\right) \cup \left(\tfrac{2}{3},1\right)},\\
\overline{f\left(\left(\tfrac{1}{3},\tfrac{1}{2}\right)\right)} &= \left[\tfrac{1}{2},1\right] = \overline{\left(\tfrac{1}{2},\tfrac{2}{3}\right) \cup \left(\tfrac{2}{3},1\right)},\\
\overline{f\left(\left(\tfrac{1}{2},\tfrac{2}{3}\right)\right)} &= \overline{f\left(\left(\tfrac{2}{3},1\right)\right)} = \left[0,\tfrac{1}{2}\right] = \overline{\left(0,\tfrac{1}{3}\right) \cup \left(\tfrac{1}{3},\tfrac{1}{2}\right)}
\end{aligned}$$

so that indeed $\mathcal{R}$ is a Markov partition. The associated Markov chain has connectivity matrix

$$A = \begin{pmatrix} 1&1&1&1\\ 0&0&1&1\\ 1&1&0&0\\ 1&1&0&0 \end{pmatrix}. \tag{3.14}$$

The concept of a Markov partition in Definition 3.5 can be extended to study so-called *non-escaping sets*. Let $f : I \to I$ be such that f is piecewise expanding on a compact subset $U \subset I$. Then one may ask to describe the set of points $N(U) \subset U$ that never escape U:

$$N(U) = \lim_{n\to\infty} \bigcap_{i=0}^{n-1} f^{-i}(U). \tag{3.15}$$

$N(U)$ is f-invariant and one may also want to describe the dynamics of f on $N(U)$. If f is piecewise expanding on U with respect to a finite partition $\mathcal{R} = \{R_0, \ldots, R_{k-1}\}$ of U, we call $\mathcal{R}$ a Markov partition for $N(U)$ if the partition satisfies the usual conditions (3.7).

Proposition 3.3 *Let $f : I \to I$ and f is piecewise expanding with respect to a finite Markov partition on $U \subset I$. Then the restriction of f to $N(U)$ (3.15) is topologically semi-conjugate to a shift on a topological Markov chain.*

Example 3.8 Consider the map f (3.8) from Example 3.6. Let $U = [0, \frac{2}{3}]$ and consider the partition $\mathcal{R} = \{(0, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3})\}$ of U. This is a Markov partition on U

since $\overline{f(R_0)} = \overline{f(R_1)} = U$. The refined Markov partitions $\mathcal{R}_n$ are such that for each $n > 1$ the closure of the union of its elements consists of 2^n disjoint intervals. These intervals converge uniformly to $N(U)$. $N(U)$ is a so-called *Cantor set* and the dynamics on $N(U)$ is topologically conjugate to the shift on Σ_2^+. The conjugacy arises due to the absence of ambiguity of the coding as the closures of the labelling intervals are always disjoint.

In the next section the dynamical properties of the shift map on finite Markov chains will be considered in more detail.

Exercises

3.1 Let X be a metric space with metric d^X and $f : X \to X$ be eventually expanding, i.e. f^n is expanding for some integer $n > 1$, but f not necessarily. Show that this implies that f is expanding on X endowed with the different metric $d^X_{n,L}$, which is defined as

$$d^X_{n,L}(x, \tilde{x}) := \sum_{m=0}^{n-1} L^{-m/n} d^X \left(f^m(x), f^m(\tilde{x})\right), \tag{3.16}$$

for all $x, \tilde{x} \in X$ and with $L > 1$ be the Lipschitz constant of f^n as in Definition 3.1. While different, the metrics d^X and d^X_n are *topologically equivalent* (they have the same open sets, and consequently also the same converging sequences). Hence, eventually expanding maps are also called *topologically expanding*.

3.2 Prove Proposition 3.1.

3.3 Show, as asserted in Example 3.5, that the tent map (3.6) is topologically semi-conjugate to a full shift on two symbols and show that h^{-1} is at most double-valued.

3.4 Show that for each value of $k \in \mathbb{Z}$, with $|k| \geq 2$, the expanding circle map E_k is topologically semi-conjugate to the shift on Σ_k^+. Are E_k and E_{-k} topologically conjugate to each other?

3.5 *Use the observation in Example 3.4, that all expanding circle maps are topologically semi-conjugate to a full shift, to prove that any two expanding maps of the circle are topologically conjugate if and only if they have the same degree.

3.2 Shift dynamics

This section is dedicated the dynamics of shift maps on topological Markov chains

$$\sigma_A : \Sigma^+_{k,A} \to \Sigma^+_{k,A}.$$

We identify necessary and sufficient conditions for chaotic and topologically mixing dynamics as well as an expression for the topological entropy.

We recall that we consider $\Sigma^+_{k,A}$ as a metric space endowed with metric d^{Σ^+}, cf. (2.22). With this metric, for each admissible sequence $i_0 \dots i_{m-1}$ for the connectivity matrix A, the cylinder set

$$C_{i_0 \dots i_{m-1}} := \{s_0 \dots s_{m-1} s_m \dots \in \Sigma^+_{k,A} \mid i_j = s_j,\ j = 0, \dots m-1\}$$

is non-empty and precisely the open ball of radius 3^{-m+1} around each point in this cylinder set, cf. Exercise 2.24.

It turns out that dynamical properties of a subshift of finite type are directly related to properties of the connectivity matrix A of the associated Markov chain.

The proof of the following elementary proposition is left as Exercise 3.6.

Proposition 3.4 *Consider the topological Markov chain $\Sigma^+_{k,A}$. Then*

(i) the number of distinct paths of length m on the associated Markov graph from vertex i to vertex j is given by $(A^m)_{ij}$,

(ii) the number of distinct paths in the Markov graph of length m that start and end at the same vertex is equal to $\mathrm{Tr}(A^m)$.

We note that (ii) in this proposition is equivalent the assertion that the number of fixed points of σ^m_A, denoted as $P_m(\sigma_A)$ in (2.8), is equal to $\mathrm{Tr}(A^m)$.

We define two useful and important notions.

Definition 3.6 (Irreducible topological Markov chain) A topological Markov chain $\Sigma^+_{k,A}$ is *irreducible* if its $k \times k$ connectivity matrix A is such that for all $i, j \in \{0, \dots, k-1\}$ there exists $m \in \mathbb{N}$ such that $(A^m)_{ij} \neq 0$.

A Markov chain is irreducible if and only if there exists a directed path from each vertex to each other vertex.

A slightly stronger notion than irreducibility is primitivy:

Definition 3.7 (Primitive topological Markov chain) A topological Markov chain $\Sigma^+_{k,A}$ is *primitive* if its $k \times k$ connectivity matrix A is such that there exists $m \in \mathbb{N}$ such that $(A^m)_{ij} \neq 0$ for all $i, j \in \{0, \dots, k-1\}$.

We also call the associated connectivity matrices of reducible and primitive Markov chains *irreducible* and *primitive*, respectively. It is clear from the definitions that primitivity implies irreducibility, but not vice-versa.

Example 3.9 The connectivity matrix

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \tag{3.17}$$

is irreducible but not primitive.

We proceed to examine the dynamical properties of the shift map on finite Markov chains.

Proposition 3.5 *The shift map on an irreducible topological Markov chain is transitive and has dense periodic orbits.*

Proof Density of periodic orbits follows from the observation that every non-empty cylinder set in $\Sigma^+_{k,A}$ contains a periodic orbit. Namely, consider $C_{i_0 \dots i_{n-1}} \subset \Sigma^+_{k,A}$ for any A-admissible sequence $i_0 \dots i_{n-1}$ $(n > 1)$ and let $m \in \mathbb{N}$ be such that $(A^m)_{i_{n-1} i_0} \neq 0$. This implies that there exists $s \in C_{i_0 \dots i_{n-1}}$ such that $s_{n-1+m} = i_0$. In turn this implies that there exists a periodic sequence $\overline{i_0 \dots s_{n-2+m}} \in C_{i_0 \dots i_{n-1}}$.

To establish transitivity it suffices to show that for any pair of non-empty cylinder sets $U = C_{i_0 \dots i_{n-1}}$ and $V = C_{j_0 \dots j_{p-1}}$, with $n, p > 1$, there exists $q \in \mathbb{N}$ such that $\sigma_A^q(U) \cap V \neq \emptyset$. Let $m \in \mathbb{N}$ be such that $(A^m)_{i_{n-1} j_0} \neq 0$. Then there exists $s \in U$ such that $s_{n-1+m} = j_0$, which in turn implies that there exists $s \in U$ such that $s_{n-1+m+\ell} = j_\ell$ for $\ell = 0, \dots, p-1$, thus confirming that $\sigma_A^{m+n-1}(U) \cap V \neq \emptyset$. □

In fact, irreducibility is necessary for transitivity.

Proposition 3.6 *The shift map on a topological Markov chain is transitive if and only if the topological Markov chain is irreducible.*

The proof of this result is deferred to Exercise 3.7.

It follows from Theorem 2.3 that the shift map on an irreducible topological Markov chain also has sensitive dependence, unless the topological Markov chain is uninteresting:

Corollary 3.1 *The shift map on a topological Markov chain is chaotic if and only if it is irreducible, unless the topological Markov chain consist of a single periodic sequence.*

Finally, we address the property of topological mixing. We note that irreducible topological Markov chains need not be topologically mixing.

Example 3.10 Consider the connectivity matrix A from Example 3.9. Then the corresponding topological Markov chain $\Sigma^+_{k,A}$ is not topologically mixing since for all $n \in \mathbb{N}$, $\sigma_A^{2n}(C_0 \cup C_1) = C_0 \cup C_1$ and $\sigma_A^{2n+1}(C_0 \cup C_1) = C_2 \cup C_3$. See Exercise 3.9 for an example of a piecewise expanding circle map that is topologically semi-conjugate to this topological Markov chain.

Proposition 3.7 *The shift map on a topological Markov chain is topologically mixing if and only if the topological Markov chain is primitive.*

Proof Primitivity of A means that there exists $\tilde{m} \in \mathbb{N}$ such that $(A^{\tilde{m}})_{ij} \neq 0$ for all $i, j \in \{0, \dots, k-1\}$. We note that this implies that also $(A^m)_{ij} \neq 0$ for all $m \geq \tilde{m}$. Let $m \in \mathbb{N}$ be such that $(A^m)_{i_{n-1} j_0} \neq 0$. Consider a pair of non-empty cylinder sets $U = C_{i_0 \dots i_{n-1}}$ and $V = C_{j_0 \dots j_{p-1}}$, with $n, p > 1$. Then there exists $s \in U$ such that $s_{n-1+m} = j_0$, which in turn implies that there exists $s \in U$ such that $s_{n-1+m+\ell} = j_\ell$ for $\ell = 0, \dots, p-1$, thus confirming that $\sigma_A^{m+n-1}(U) \cap V \neq \emptyset$. To establish topological mixing, it suffices to show that for any pair of cylinder sets U, V there exists $N \in \mathbb{N}$ such that $\sigma_A^q(U) \cap V \neq \emptyset$ for all $q \geq N$. With $\tilde{m}$ as defined above, the latter holds with $N = \tilde{m} + n - 1$.

If A is not primitive there exists $i, j \in \{0, \ldots, k-1\}$ such that for any $\tilde{m} \in \mathbb{N}$ there exists $m > \tilde{m}$ such that $(A^m)_{ij} = 0$. This implies that for any $\tilde{m} \in \mathbb{N}$ there exists $m > \tilde{m}$ such that $\sigma_A^m(C_i) \cap C_j = \emptyset$, contradicting topological mixing. □

Finally, it turns out that *topological entropy* of a topological Markov chain is directly computable from its connectivity matrix.

Theorem 3.1 *The shift σ_A on the topological Markov chain $\Sigma_{k,A}^+$ has* topological entropy

$$h_{\text{top}}(\sigma_A) = \ln r(A),$$

where $r(A)$ denotes the spectral radius of A, which is its largest real positive eigenvalue.[1]

Proof The argument is a straightforward generalisation of the one used in Example 2.10 to determine the topological entropy of the full shift on 10 symbols. Namely, it suffices to calculate the exponential growth rate of the minimal number of admissible balls of radius $3^{-(m-1)}$ that cover $\Sigma_{k,A}^+$ in the $d_n^{\Sigma_k^+}$ metric. The latter is equal to the minimal cardinality of the covering by $3^{-(m+n-2)}$-balls in the $d^{\Sigma_k^+}$ metric, which is precisely the number of admissible subsequences of length $(m+n-1)$ in $\Sigma_{k,A}^+$. By Proposition 3.4(i), this is equal to the sum of the entries of the matrix A^{m+n-1}. This sum turns out to be a matrix norm[2]

$$||A||_1 := \sum_{i,j=0}^{k-1} |A_{ij}|. \tag{3.18}$$

So,

$$h_{\text{top}}(\sigma) = \lim_{m\to\infty} \lim_{n\to\infty} \frac{1}{n} \ln ||A^{m+n-1}||_1 \tag{3.19}$$

In order to finalize the proof, we recall *Gelfand's formula*,[3] relating the exponential growth rate of the norm of powers of a matrix to its spectral radius:

$$\ln r(A) = \lim_{m\to\infty} \frac{1}{m} \ln ||A^m||,$$

for any matrix norm $||A||$. Applying Gelfand's formula to (3.19) yields

[1] We note that the characteristation of the spectral radius of A as its largest real eigenvalue relies on Perron-Frobenius theory, which establishes the existence of such an eigenvalue for non-negative matrices (such as the connectivity matrx A). This eigenvalue is also larger or equal than the norm of any other eigenvalue of A. For more details, see for instance C.D Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM (2000).

[2] This element-wise norm is known as the *vector 1-norm*.

[3] This result appears in many textbooks and we use it here without proof. It was first published in I. Gelfand, Normierte ringe, *Rech. Math. [Mat. Sbornik] N.S.* **9** (51) (1941), 3-24.

$$\begin{aligned} h_{\mathrm{top}}(\sigma_A) &= \lim_{m\to\infty}\lim_{n\to\infty}\frac{1}{n}\ln ||A^{m+n-1}||_1 \\ &= \lim_{m\to\infty}\lim_{n\to\infty}\frac{m+n-1}{n}\cdot\frac{1}{m+n-1}\ln ||A^{m+n-1}||_1 = \ln r(A), \end{aligned}$$

as asserted. □

We finally illustrate the use of symbolic dynamics to obtain results on an explicit example that was discussed before.

Example 3.11 We continue the discussion in Example 3.6, concerning the dynamics of the piecewise expanding map f (3.8). It was concluded that f is topologically semi-conjugate to the shift map on a topological Markov chain $\Sigma^+_{3,A}$ with A given by (3.10). This topological Markov chain is primitive since A^2 has no zero entries. This implies that the shift map σ_A is chaotic and topologicallly mixing and so is f, by Proposition 2.5. The topological entropy of the shift follows from Theorem 3.1,

$$h_{\mathrm{top}}(\sigma_A) = \ln r(A) = \ln(1+\sqrt{3}).$$

Moreover, since $\sup_{x\in[0,1]}\{\#(h^{-1}(x))\} \leq 2$, Theorem 2.6 implies that $h_{\mathrm{top}}(f) = h_{\mathrm{top}}(\sigma_A)$.

Exercises

3.6 Prove Proposition 3.4.

3.7 Prove Proposition 3.6.

3.8 Determine whether the map f (3.12) in Example 3.7 is chaotic and/or topologically mixing. Determine $h_{\mathrm{top}}(f)$.

3.9 Consider the sawtooth map $S : [0,1] \to [0,1]$ defined by[4]

$$S(x) = \begin{cases} 2x + \frac{1}{2}, & \text{if } x \in \left[0, \frac{1}{4}\right], \\ \frac{3}{2} - 2x, & \text{if } x \in \left(\frac{1}{4}, \frac{3}{4}\right], \\ 2x - \frac{3}{2}, & \text{if } x \in \left(\frac{3}{4}, 1\right], \end{cases} \tag{3.20}$$

see also Fig. 3.7. Show that S is chaotic but not topologically mixing. Determine $h_{\mathrm{top}}(S)$.

3.10 *Consider the map of the interval $f : [0,1] \to [0,1]$ given by

[4] The sawtooth map (3.20) may also be interpreted as a continuous circle map and serve as an answer to Exercise 2.16(c).

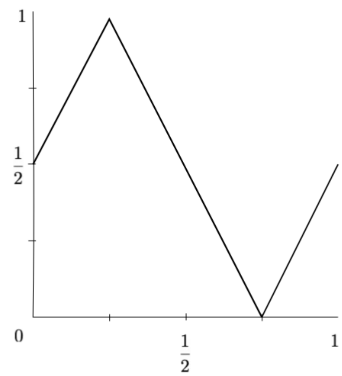

Fig. 3.7: Graph of the map S (3.20).

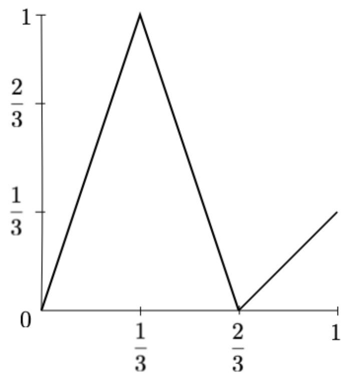

Fig. 3.8: Graph of the map f (3.21).

$$f(x) = \begin{cases} 3x, & \text{if } x \in \left[0, \frac{1}{3}\right], \\ 2 - 3x, & \text{if } x \in \left(\frac{1}{3}, \frac{2}{3}\right], \\ x - \frac{2}{3}, & \text{if } x \in \left(\frac{2}{3}, 1\right]. \end{cases} \tag{3.21}$$

The graph of f is sketched in Fig. 3.8. Show that f is chaotic and topologically mixing. Determine $h_{\mathrm{top}}(f)$.

3.11 *Explore the dynamics of the piecewise affine maps, the graphs of which are depicted in Fig. 3.9.

short version

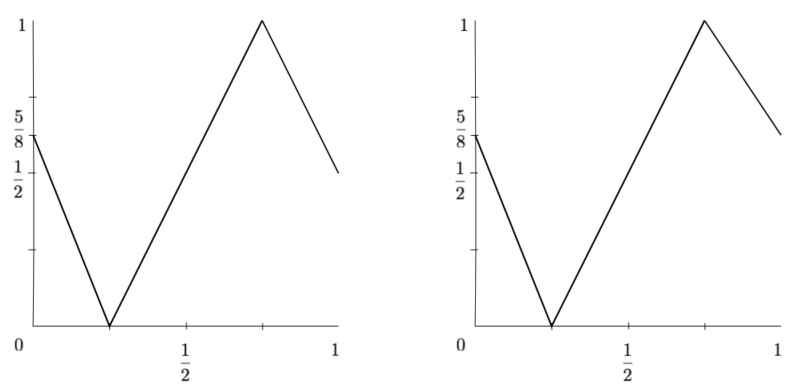

Fig. 3.9: Graphs of maps of [0, 1], for Exercise 3.11.

Chapter 4
Ergodic theory

Instead of considering trajectories of continuous maps $f : X \to X$ on a compact metric space X from the topological point of view, there is an alternative approach that consists of considering the induced action of f on Borel probability measures, by endowing X with its Borel σ-algebra $\mathscr{B}(X)$ - the smallest σ-algebra containing all open and closed subsets of X - on which these measures are defined.

For a concise discussion of essential preliminaries from measure theory that we employ in this chapter, see the Appendix to this chapter.

4.1 Invariant probability measures

Consider the set of probability measures $\mathcal{P}(X)$ on a measurable space $(X, \mathscr{F})$. Then any continuous map $f : X \to X$ induces an action $f_* : \mathcal{P}(X) \to \mathcal{P}(X)$ on the set $\mathcal{P}(X)$ of probability measures on $(X, \mathscr{F})$, defined by

$$f_*\mu(A) := \mu(f^{-1}(A)),\ \forall A \in \mathscr{F},$$

where $f^{-1}(A) := \{B \in \mathscr{F} \mid f(B) = A\}$.

The map f_* represents the transport of measure on $(X, \mathscr{F})$ induced by the map f: for all $A \in \mathscr{F}$, $f_*\mu(A)$ is the measure obtained by transferring the measure from $f^{-1}(A)$ to A. Note in this respect that $f^{-1}(A)$ is the largest subset of X that is mapped by f to A.

Definition 4.1 (Invariant probability measure[1].) $\mu \in \mathcal{P}(X)$ is an *f-invariant probability measure* if

$$\mu(A) = f_*\mu(A),\ \ \forall A \in \mathscr{F}, \tag{4.1}$$

As it is in general virtually impossible to directly establish invariance for important σ-algebras such as Borel σ-algebras, it is of practical importance that in order to

[1] The more general notion of invariant measure is analogous, allowing also for measures μ satisfying (4.2) that are not probability measures.

establish invariance, it in fact suffices to establish this for a much smaller semi-ring that generates the σ-algebra.

Proposition 4.1 *Let $(X, \mathscr{F})$ be a measure space and $\mathscr{S} \subset \mathscr{F}$ a semi-ring of subsets of X, generating the σ-algebra $\mathscr{F}$, i.e. $\sigma(\mathscr{S}) = \mathscr{F}$. Let $\mu \in \mathcal{P}(X)$ and $f : X \to X$ be μ-measurable, then*

$$\mu(A) = f_*\mu(A), \ \ \forall A \in \mathscr{S} \ \ \Leftrightarrow \ \ \mu(A) = f_*\mu(A), \ \ \forall A \in \mathscr{F}, \tag{4.2}$$

This proposition is related to Caratheodory's extension theorem, Theorem 4.7 in the Appendix to this chapter, which establishes the extension of a measure defined on a semi-ring to the σ-algebra that it generates. Its proof is beyond the scope of this course, but a detailed account can be found in [9].

Importantly, the existence of invariant measures is guaranteed by the following result.

Theorem 4.1 (Krylov-Bogoliubov) *Let X be a compact metric space and $f : X \to X$ be continuous. Then there exists an f-invariant Borel probability measure $\mu \in \mathcal{P}(X)$.*

Compactness of X is an essential assumption in this theorem. For instance, $f : (0, 1) \to (0, 1)$ defined by $f(x) = \frac{1}{2}x$, has no invariant probability measure.

A formal proof of Theorem 4.1 is beyond the scope of this course. The usual strategy of proof is to consider for any $\nu \in P(X)$ the sequence

$$\nu_n = \frac{1}{n}\sum_{i=0}^{n-1}(f^i)_*\nu.$$

Due to compactness of X, with a natural chosen topology, $\mathcal{P}(X)$ is also compact, so that $\{\nu_n\}_{n\in\mathbb{N}}$ contains a convergent subsequence, the limit of which is an invariant measure $\mu \in \mathcal{P}(X)$.

Example 4.1 We present some elementary examples of invariant probability measures.

- *Fixed point.* Let $f : X \to X$ and $f(x) = x$, then δ_x is an f-invariant measure.
- *Periodic orbit.* Let $f : X \to X$ and $f^p(x) = x$ for some $p \in \mathbb{N}$, then $\mu = \sum_{i=0}^{p-1} \delta_{f^i(x)}$ is an f-invariant measure.
- *Rigid rotation on the circle:* Let $f_a : S^1 \to S^1$ be a rigid rotation $f_a(x) = x + a \bmod 1$, with the usual parametrisation $S^1 \simeq [0, 1)$ and S^1 endowed with the standard (arclength) metric. Then the Lebesgue measure λ on S^1 is an f_a-invariant probability measure.
  If $a \in \mathbb{Q}$ then there are many other invariant measures, for instance on periodic orbits, but if $a \in \mathbb{R} \setminus \mathbb{Q}$ then it turns out that λ is the unique f_a-invariant Borel probability measure.
- *Piecewise affine full-branch interval maps.* Let $f : [0, 1] \to [0, 1]$ be a continuous piecewise-affine full-branch map, i.e. there exists a finite partition

$\mathcal{R} = \{I_i\}_{i=0}^{n-1}$ of I into open intervals such that $\cup_{i=0}^{n-1}\overline{I_i} = [0,1]$ and for all $i = 0, \ldots n-1$, $f_{I_i} : I_i \to (0,1)$ are affine and onto. Then the Lebesgue measure λ on $[0,1]$ is an f-invariant probability measure.

Let $(X, \mathscr{F}, \mu)$ be a measurable space and $f : X \to X$ measurable and μ-preserving, i.e. $f_*\mu = \mu$. Then $(X, \mathscr{F}, \mu, f)$ denotes a *measure-preserving dynamical system*. If $\mu(X) < \infty$, without loss of generality we may normally normalize μ so that μ is a *probability measure*, i.e. $\mu(X) = 1$. Hence any results stated for invariant probability measures directly extend also to systems with bounded invariant measures. In this course we mainly consider the canonical setting in which X is a (compact) metric space, $\mathscr{F} = \mathscr{B}(X)$, μ is a probability measure and f is continuous.

Exercises

4.1 Prove the assertions in Example 4.1

4.2 Poincaré recurrence

The phenomenon of *Poincaré recurrence* serves as a precursor to *ergodicity*.

Theorem 4.2 (Poincaré recurrence) *Let $(X, \mathscr{F}, \mu, f)$ be a probability measure-preserving dynamical system. Let $A \in \mathscr{F}$ with $\mu(A) > 0$, then for μ-almost every $x \in A$ there exist infinitely many $i \in \mathbb{N}$ such that $f^i(x) \in A$.*

Proof Let

$$B := \{x \in A \mid f^i(x) \notin A,\ \forall i \in \mathbb{N}\} = A \setminus \bigcup_{i\in\mathbb{N}} f^{-i}(A).$$

We note that $B \in \mathscr{F}$ by measurability of f, and likewise also $f^{-k}(B) \in \mathscr{F}$ for all $k \in \mathbb{N}$. Moreover, $(f^k)_*\mu(B) := \mu(f^{-k}(B)) = \mu(B)$ by f-invariance of μ.

We show that the pre-images $\{f^{-i}(B)\}_{i\in\mathbb{N}_0}$ are disjoint. Namely, suppose that there exist $m, n \in \mathbb{N}_0$ with $n > m$ such that $f^{-n}(B) \cap f^{-m}(B) \neq \emptyset$. Let $x \in f^{-n}(B) \cap f^{-m}(B)$, then

$$f^n(x) \in f^n(f^{-n}(B)\cap f^{-m}(B)) \subset f^n(f^{-n}(B))\cap f^n(f^{-m}(B)) = B\cap f^{n-m}(B) \neq \emptyset,$$

contradicting the definition of B.

Since $\mu(X) = 1$, it follows that $\sum_{k\in\mathbb{N}} \mu(f^{-k}(B)) < \infty$, but since $\mu(B) = \mu(f^{-k}(B))$ for all $k \in \mathbb{N}$ this implies that $\mu(f^{-k}(B)) = 0$ for all $k \in \mathbb{N}_0$.

To show that almost every point in A returns to A infinitely often, let for all $n \in \mathbb{N}_0$,

$$C_n := \{a \in A \mid f^n(x) \in A \text{ and } f^k(x) \notin A \text{ for all } k > n\}.$$

Note that $C_0 = B$. We assert that $\mu(C_n) = 0$ for all $n \in \mathbb{N}_0$. This implies that the set of points returning only a finite number of times to A,

$$C := \bigcup_{n \in \mathbb{N}_0} C_n \subset A$$

has zero μ-measure, i.e. $\mu(C) = 0$. Indeed, $f^n(C_n) \subset B$ for all $n \in \mathbb{N}$, so that $\mu(f^n(C_n)) \leq \mu(B) = 0$. Similarly, $C_n \subset f^{-n}(f^n(C_n)) \subset f^{-n}(B)$ for all $n \in \mathbb{N}$, which implies that $\mu(C_n) \leq \mu(f^{-n}(B)) = 0$. □

Example 4.2 Poincaré recurrence may not always reveal interesting new facts, for instance if applied to the invariant probability measure $\mu = \sum_{i=0}^{p-1} \delta_{f^i(x)}$ for periodic orbits, as discussed in Example 4.1.

Example 4.3 For an irrational rotation on the unit circle, as discussed in Example 4.1, invariance of Lebesgue measure ensures via Poincaré recurrence the accumulation of each forward orbit to its initial condition.

Poincaré-recurrence ensures that under a measure preserving transformation, almost every point of a subset A of positive measure will return to A. Namely, let $(X, \mathscr{F}, \mu, f)$ be a probability measure-preserving dynamical system and $A \in \mathscr{F}$ with $\mu(A) > 0$, then by Theorem 4.2, the integer

$$n_A(x) := \inf\{n \in \mathbb{N} \mid f^n(x) \in A\},$$

is well-defined for μ-almost all $x \in A$.

However, the Poincaré recurrence theorem does not provide any estimate of how long we would have to wait for returns to happen. One may expect that return times to sets of large measure are small, whereas return times to sets of small measure are large. Kac's lemma confirms that this intuition is indeed correct.

Lemma 4.1 (Kac's lemma[2]) *Let $(X, \mathscr{F}, \mu, f)$ be a probability measure-preserving dynamical system and $A \in \mathscr{F}$ with $\mu(A) > 0$. Let*

$$A^{c*} := \{x \in A^c \mid f^n(x) \notin A, \forall n \in \mathbb{N}\},$$

i.e. A^{c*} is the subset of points in X that never enter A. Then, n_A is μ-integrable and

$$\int_A n_A d\mu = 1 - \mu(A^{c*}) \tag{4.3}$$

Proof For all $n \in \mathbb{N}$, define

$$A_n := \{x \in A \mid f(x) \notin A, \dots, f^{n-1}(x) \notin A \text{ but } f^n(x) \in A\},$$
$$A_n^{c*} := \{x \in A^c \mid f(x) \notin A, \dots, f^{n-1}(x) \notin A \text{ but } f^n(x) \in A\}.$$

[2] M. Kac. On the notion of recurrence in discrete stochastic processes. *Bulletin of the American Mathematical Society* **53** (1947), 1002–1010.

This means that A_n is the set of point in A that return for the first time to A after precisely n iterates of f:

$$A_n = \{x \in A \mid n_A(x) = n\}.$$

$A_n^{c*} \subset A^c$ is the set of points outside A that enter A for the first time after n iterations. Measurability of A_n follows from its definition and correspondingly n_A is measurable since $n_A^{-1}(n) = A_n$ for all $n \in \mathbb{N}$.

Moreover, defining additionally[3]

$$A_0 := \{x \in A \mid f^n(x) \notin A,\ \forall n \in \mathbb{N}\},$$

and $A_0^{c*} := A^{c*}$, the sets $\{A_n, A_n^{c*}\}_{n\in\mathbb{N}_0}$ are pairwise disjoint and their union equals X. Hence, since $\mu(A_0) = 0$ by Theorem 4.2,

$$1 = \mu(X) = \sum_{n\in\mathbb{N}_0} [\mu(A_n) + \mu(A_n^{c*})] = \mu(A_0^{c*}) + \sum_{n\in\mathbb{N}} [\mu(A_n) + \mu(A_n^{c*})]. \tag{4.4}$$

Subsequently, observe that for all $n \in \mathbb{N}$

$$f^{-1}(A_n^{c*}) = A_{n+1}^{c*} \cup A_{n+1}.$$

Then, f-invariance of μ implies that

$$\mu(A_n^{c*}) = \mu(f^{-1}(A_n^{c*}) = \mu(A_{n+1}^{c*}) + \mu(A_{n+1}).$$

This implies that for all $m > n$,

$$\mu(A_n^{c*}) = \mu(A_m^{c*}) + \sum_{i=n+1}^{m} \mu(A_i). \tag{4.5}$$

Due to (4.4), it follows that $\lim_{m\to\infty} \mu(A_m^{c*}) = 0$ and in turn, from (4.5), that

$$\mu(A_n^{c*}) = \sum_{i=n+1}^{\infty} \mu(A_i). \tag{4.6}$$

Finally, substituting (4.6) back into (4.4), one obtains

$$1 - \mu(A_0^{c*}) = \sum_{n\in\mathbb{N}} \left(\sum_{i=n}^{\infty} \mu(A_i) \right) = \sum_{n\in\mathbb{N}} n\mu(A_n) = \int_A n_a d\mu,$$

which concludes the proof. □

---

[3] Note that the set B, defined in the proof of Theorem 4.2, is identical to A_0.

Exercises

4.2 Prove the assertion in Example 4.3.

4.3 Verify that Lemma 4.1 establishes the correct expected return times for periodic orbits.

4.4 Let $R_a : S^1 \to S^1$ denote the rigid translation on the unit circle: $R_a(x) = x + a \bmod 1$, as in Exercise 2.9. Examine the implications of Poincaré-recurrence and Kac's Lemma for R_a.

4.3 Birkhoff's ergodic theorem

Consider a probability measure-preserving dynamical system $(X, \mathscr{F}, \mu, f)$.

Let $\mathscr{G}$ denote the σ-subalgebra, $\mathscr{G} \subset \mathscr{F}$, consisting of f-invariant elements, i.e.[4]

$$\mathcal{G} := \{A \in \mathscr{F} \mid f^{-1}(A) = A\}.$$

Then, in this setting, the following result establishes an important relationship between the existence of time-averages of observables along orbits and (conditional) expectations with respect to the invariant measure.

Theorem 4.3 (Birkhoff's ergodic theorem) *Let $g : X \to \mathbb{R}$ be integrable, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x)) = \mathbb{E}[g|\mathscr{G}](x), \textit{ for } \mu\textit{-almost all } x \in X. \tag{4.7}$$

We defer the proof of this result to Section 4.4.

The left-hand-side of (4.7) represents the *time-average* of the *observable* g along $O_f^+(x)$. Birkhoff's Ergodic Theorem asserts that this time-average exists for μ-almost all $x \in X$ and is given by the conditional expectation of the observable g on the right-hand-side.

Given an initial condition $x \in X$ for the forward orbit $O_f^+(x)$, the conditional expectation on the right-hand side of (4.7) is uniquely determined by the smallest f-invariant measurable set in which this initial condition lies.

You may have observed that despite the appearance of the terminology "ergodic" in the title of Theorem 4.3, the statement makes no reference to ergodicity and we even have not given a definition of ergodicity, yet. *Ergodicity* is a property of an invariant probability measure.

Definition 4.2 (Ergodic probability measure) An f-invariant probability measure μ is called *ergodic* if for any f-invariant $A \in \mathscr{F}$, i.e. every $A \in \mathscr{G}$, $\mu(A) \in \{0, 1\}$.

[4] Please be aware that this notion of invariance is not identical to the notion of invariance introduced in the topological setting, cf. Definition 2.3.

The following corollary of Theorem 4.3 is better known. We assume the same setting.

Corollary 4.1 (Better known version of Birkhoff's Ergodic Theorem) *Let μ be* ergodic*, then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} g(f^i(x)) = \mathbb{E}[g] = \int_X g\,d\mu, \text{ for } \mu\text{-almost all } x \in X. \tag{4.8}$$

Proof $\mathbb{E}[g|\mathscr{G}] : X \to \mathbb{R}$ is $\mathscr{G}$-measurable, which means that for all $z \in \mathrm{Range}(\mathbb{E}[g|\mathscr{G}])$, $(\mathbb{E}[g|\mathscr{G}])^{-1}(z) \in \mathscr{G} \setminus \{\emptyset\}$. Let $A(x) := (\mathbb{E}[g|\mathscr{G}])^{-1}(\mathbb{E}[g|\mathscr{G}](x)) \in \mathscr{G}$, and note that $\mathbb{E}[g|\mathscr{G}]$ is constant on $A(x)$. Ergodicity implies that $\mu(A(x)) \in \{0, 1\}$, since $A(x) \in \mathscr{G}$. Now, let $x \in X$ be such that $\mu(A(x)) = 1$ (that such x exists, follows from surjectivity of $\mathbb{E}[g|\mathscr{G}]$ to its range). Then $\mathbb{E}[g|\mathscr{G}](x) = \int_{A(x)} \mathbb{E}[g|\mathscr{G}]d\mu = \int_X \mathbb{E}[g|\mathscr{G}]d\mu = \mathbb{E}[g]$. Hence $\mathbb{E}[g|\mathscr{G}](x) = \mathbb{E}[g]$ for μ-almost all $x \in X$. □

Birkhoff's ergodic theorem is often informally summarized as follows: in the presence of an ergodic invariant measure μ, the time-average of a (μ-integrable) observable along forward orbits, typically (for μ-almost every initial condition $x \in X$) converges to the (μ-)*space average* of this observable (i.e. its expectation with respect to μ). The left-hand side of (4.8) represents the time-average and the right-hand side the space average.

Ergodicity simplifies the expression for the expected return time for measurable sets as it turns out that in that case only a zero-measure set of points does not enter A, cf. Exercise 4.5.

Lemma 4.2 (Kac's lemma for ergodic invariant measures) *Let $(X, \mathscr{F}, \mu, f)$ be an ergodic probability measure preserving dynamical system and $A \in \mathscr{F}$ with $\mu(A) > 0$. Then, the expected return time for μ-almost all $x \in A$ to A, is given by*

$$\frac{\int_A n_A d\mu}{\mu(A)} = \frac{1}{\mu(A)}. \tag{4.9}$$

The following observations about ergodic invariant measures are useful:

Proposition 4.2 *Let $(X, \mathscr{F})$ be a measurable space and $f : X \to X$ measurable.*

(i) If μ_1 and μ_2 are ergodic f-invariant probability measures and $\mu_1 << \mu_2$, then $\mu_1 = \mu_2$.
(ii) If μ_1 and μ_2 are f-invariant probability measures such that $\mu_1 \neq \mu_2$ and $\mu = t\mu_1 + (1-t)\mu_2$ for some $t \in (0, 1)$, then μ is not ergodic.[5]
(iii) Let $\mu_1 \neq \mu_2$ be two ergodic f-invariant probability measures. Then μ_1 and μ_2 are mutually singular*, i.e. $\exists A \in \mathscr{F}$ such that $\mu_1(A) = 1$ and $\mu_2(A) = 0$.*

[5] In fact, the invariant probability measures form a convex set of which the extremal points are precisely ergodic.

Their proof is left as Exercise 4.6

We end this section with a historical footnote. Statistical mechanics, one of the scientific revolutions of the 20th century, is built on Boltzmann's *ergodic hypothesis*: the assumption that *Liouville measure* (phase space volume) is an ergodic invariant measure. Dynamical systems and ergodic theory research has been partly driven by the desire to prove the correctness of this hypothesis in certain idealized circumstances. However, extensive dynamical systems research has led to the observation that the Ergodic hypothesis in fact rarely holds exactly. Nevertheless, many physical systems appear to approximately satisfy this hypothesis, so that the adoption of this hypothesis was able to be used as the basis of a theory that has been able to provide new deep fundamental insights in many physical phenomena. In particular, statistical mechanics has enabled the understanding of macroscopic phenomena, hitherto predominantly covered by phenomenological theories like that of thermodynamics, on the basis fundamental microscopic assumptions (which are often experimentally accessible).

Exercises

4.5 Prove Lemma 4.2 by showing that in Lemma 4.1, $\mu(A^{c*}) = 0$ if μ is ergodic .

4.6 Prove Proposition 4.2.

4.7 Prove that the Lebesgue measure is an ergodic invariant measure for irrational rigid rotations of the circle. Hint: Recall the discussion in Example 4.1.

4.8 * Prove that the Lebesgue measure on S^1 is an ergodic invariant measure of E_k, $k \geq 2$, by using

Theorem (Lebesgue density theorem) *Let λ denote the Lebesgue measure on S^1 and $A \in \mathscr{B}(S^1)$ with $\lambda(A) > 0$. Then,*

$$\lim_{\varepsilon \to 0} \frac{\lambda(A \cap B_\varepsilon(x))}{\lambda(B_\varepsilon(x))} = 1, \quad \text{for } \lambda\text{-almost all } x \in A.$$

Proceed along the following steps:

(a) Let $P_n := \{[0, k^{-n}), [k^{-n}, 2k^{-n}), \ldots, [1 - k^{-n}, 1)\}$ be a partition of the circle in k^n intervals. Show that if $\lambda(A) > 0$,for all $\varepsilon > 0$, there exists $n \in \mathbb{N}$ and $J \in P_n$ such that

$$\frac{\lambda(A \cap J)}{\lambda(J)} \geq 1 - \varepsilon, \quad \text{or equivalently that} \quad \frac{\lambda(J \setminus A)}{\lambda(J)} \leq \varepsilon.$$

(b) Show that if $f^{-1}(A) = A$,

$$\frac{\lambda([0,1] \setminus A)}{\lambda([0,1])} \leq \varepsilon,$$

and, in turn, if $\lambda(A) > 0$, that $\lambda(A) = 1$.

4.9 Let it be given that the Lebesgue measure λ on S^1 is an ergodic invariant probability measure of the expanding circle map $f = E_{10}$ (1.1), cf. Exercise 4.8. Consider the observable $g = \mathbb{1}_A$ where $A := [0, \frac{1}{2})$ is half of the unit circle $S^1 \simeq [0, 1)$.

(a) Show that there exist $x \in S^1$ for which

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x)) \neq \int_{S^1} g d\lambda = \frac{1}{2}.$$

(b) Show that there exist $x \in S^1$ for which

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x)) \text{ does not exist.}$$

(c) Show that the sets points, satisfying the conditions specified in parts (a) and (b) above, are dense in S^1.

4.4 Proof of Birkhoff's ergodic theorem[6]

A first important ingredient in the proof of Birkhoff's ergodic theorem is the so-called *maximal ergodic theorem*, the proof of which relies on the *dominated convergence theorem* which we state here without proof.

Theorem 4.4 (Dominated convergence theorem) *Let $(X, \mathscr{F}, \mu)$ be a measure space, $f_n : X \to \mathbb{R}$ be measurable $\forall n \in \mathbb{N}$ and let $g : X \to \mathbb{R}$ be non-negative and μ-integrable. Suppose*

(i) $|f_n(s)| \leq g(s)$, $\forall s \in X$, $\forall n \in \mathbb{N}$,
(ii) $\lim_{n\to\infty} f_n(s) =: f(s)$, $\forall s \in X$,

then f is μ-integrable and $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.

Theorem 4.5 (Maximal ergodic theorem) *Let $(X, \mathcal{F}, \mu, f)$ be a measure-preserving dynamical system and $g : X \to \mathbb{R}$ be an integrable random variable. Let*

$$S_n(x) := \sum_{i=0}^{n-1} g(f^i(x)),$$
$$M_n^g(x) := \max_{i=1,\dots,n} S_i(x),$$
$$M^g(x) := \sup_{i\in\mathbb{N}} S_i(x),$$

[6] The details of this proof are not part of the examinable material for this course.

then

$$\int_{\{M^g>0\}} g d\mu \geq 0. \tag{4.10}$$

Proof Note that $g + (M^g_{n-1} \circ f)^+ = M^g_n$. Hence,

$$\begin{aligned}
\int_{\{M^g_n>0\}} g d\mu &= \int_{\{M^g_n>0\}} M^g_n d\mu - \int_{\{M^g>0\}} \left(M^g_{n-1} \circ f\right)^+ d\mu \\
&= \int_X \left(M^g_n\right)^+ d\mu - \left(\int_X \left(M^g_{n-1} \circ f\right)^+ d\mu - \int_{\{M^g\leq 0\}} \left(M^g_{n-1} \circ f\right)^+ d\mu\right) \\
&\geq \int_X \left(M^g_n\right)^+ d\mu - \int_X \left(M^g_{n-1} \circ f\right)^+ d\mu \\
&= \int_X \left(M^g_n\right)^+ d\mu - \int_X \left(M^g_{n-1}\right)^+ df_*\mu \\
&\geq \int_X \left(M^g_n\right)^+ d\mu - \int_X \left(M^g_n\right)^+ df_*\mu = 0
\end{aligned}$$

since $f_*\mu = \mu$. Finally,

$$\int_{\{M^g_n>0\}} g d\mu \geq 0,\ \forall n \in \mathbb{N} \quad \Rightarrow \int_{\{M^g>0\}} g d\mu \geq 0,$$

by application of Theorem 4.4, where $f_n := g \cdot \mathbb{1}_{\{M^g_n>0\}}$ and $f := g \cdot \mathbb{1}_{\{M^g>0\}}$ with $\lim_{n\to\infty} f_n = f$, pointwise. □

Proof of Theorem 4.3. We divide the proof into two parts: existence and evaluation of the limit.
Existence of the limit. Define

$$\overline{S} := \limsup_{n\to\infty} \frac{1}{n} S_n, \quad \underline{S} := \liminf_{n\to\infty} \frac{1}{n} S_n.$$

Let $\alpha, \beta \in \mathbb{R}$ and define

$$E_{\alpha,\beta} := \{x \in X \mid \underline{S} < \alpha \text{ and } \overline{S} > \beta\}.$$

It will be shown that $\mu(E_{\alpha,\beta}) = 0\ \forall \alpha, \beta \in \mathbb{Q}$, so that

$$\mu\left(\bigcup_{\alpha<\beta\in\mathbb{Q}} E_{\alpha,\beta})\right) = 0 \quad \Rightarrow \quad \overline{S} = \underline{S} \ \ \mu\text{-almost surely.}$$

We approve the assertion by contradiction: suppose $\mu(E_{\alpha,\beta}) > 0$ for some $\alpha, \beta \in \mathbb{R}$. Define the measure$\overline{\mu} : \overline{\mathscr{F}} \to \mathbb{R}_{\geq 0}$, where $\overline{\mathscr{F}} := \{A \in \mathscr{F} \mid A \subset E_{\alpha,\beta}\}$ and

$$\overline{\mu}(A) := \mu(A|E_{\alpha,\beta}) := \frac{\mu(A \cap E_{\alpha,\beta})}{\mu(E_{\alpha,\beta})}.$$

From the definition of $E_{\alpha,\beta}$ it follows that

$$E_{\alpha,\beta} \subset \{M^{g-\beta} > 0\} \cap \{M^{\alpha-g} > 0\}.$$

We apply Theorem 4.5 to $(E_{\alpha,\beta}, \overline{\mathscr{F}})$, to obtain

$$\int_{E_{\alpha,\beta}} (g-\beta) d\overline{\mu} \geq 0 \text{ and } \int_{E_{\alpha,\beta}} (\alpha - g) d\overline{\mu} \geq 0.$$

from which it follows (by addition of the inequalities) that

$$\int_{E_{\alpha,\beta}} (g-\beta) d\overline{\mu} \geq 0,$$

contradicting that $\alpha < \beta$. Hence, $\mu(E_{\alpha,\beta}) = 0$ for all $\alpha, \beta \in \mathbb{Q}$.
Evaluation of the limit. We aim to show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x)) - \mathbb{E}[g|\mathscr{G}](x) = 0.$$

which is equivalent to

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left(g(f^i(x)) - \mathbb{E}[g|\mathscr{G}]\right) \left(f^i(x)\right) = 0,$$

using f-invariance of $\mathbb{E}[g|\mathscr{G}]$, i.e. $\forall A \in \mathscr{G}$, if $x \in A$ then $f(x) \in A$. So without loss of generality, we can consider the observable $\tilde{g} := g - \mathbb{E}[g|\mathscr{G}]$ with the property that $\mathbb{E}[\tilde{g}|\mathscr{G}](x) = 0$ μ-almost surely.

The objective now is to show that $\underline{S} = \overline{S} = 0$ μ-almost surely. We first show that $\overline{S} \leq 0$. For any $\varepsilon > 0$, let $E_\varepsilon := \{x \in X \mid \overline{S} > \varepsilon\}$. We need to show that $\mu(E_\varepsilon) = 0$. Suppose $\mu(E_\varepsilon) > 0$, then $E_\varepsilon \subset \{M^g > 0\}$. Define the measure $\tilde{\mu}(A) := \mu(A|E_{\alpha,\beta})$ and apply Theorem 4.5 to get $\int_{E_\varepsilon} (g-\varepsilon) d\tilde{\mu} \geq 0$. This implies that $\int_{E_\varepsilon} (g-\varepsilon) d\mu \geq 0$ and

$$\varepsilon\mu(E_\varepsilon) \leq \int_{E_\varepsilon} g d\mu = \int_{E_\varepsilon} \mathbb{E}[g|\mathscr{G}] d\mu = 0,$$

using the fact that $E_\varepsilon \in \mathscr{G}$, which follows from the fact that $\overline{S}(f(x)) = \overline{S}(x)$, for all $x \in X$. But the latter implies that $\mu(E_\varepsilon) = 0$, contradicting the assumption.

The fact that $\underline{S} \geq 0$ follows from the same argument, replacing g by $-g$. □

4.5 Mixing

Ergodicity by itself does not necessarily reveal much detail about the underlying dynamics. For instance, the Lebesgue measure is ergodic for expanding circle maps

E_k, with $|k| > 1$, as well as for any irrational rigid rotation on the circle R_α, with $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, but the dynamics of expanding circle maps is very different from that of irrational rotations.

In order to distinguish dynamical behaviours in more detail, we discuss in this section the measure theoretic notion of *mixing*, not to be confused by the notion of topological mixing.[7]

We first present an alternative characterisation of ergodicity.

Proposition 4.3 *Let $(X, \mathscr{F}, \mu, f)$ be a probability measure preserving dynamical system. Then μ is ergodic if and only if*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(f^{-i}(A) \cap B) = \mu(A)\mu(B), \quad \forall A, B \in \mathscr{F}.$$

Proof Suppose μ is ergodic, then by Birkhoff's Ergodic Theorem, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_A(f^i(x)) = \mu(A) \text{ for } \mu\text{-almost all } x \in X .$$

Thus,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_A(f^i(x))\mathbb{1}_B(x) = \mu(A)\mathbb{1}_B(x) \text{ for } \mu\text{-almost all } x \in X,$$

and by application of the Dominated Convergence Theorem (Theorem 4.4) it follow that

$$\begin{aligned}
\mu(A)\mu(B) &= \int_X \mu(A)\mathbb{1}_B(x)d\mu(x) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_X \mathbb{1}_A(f^i(x))\mathbb{1}_B(x)d\mu(x) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(f^{-i}(A) \cap B) .
\end{aligned}$$

To prove the reverse implication, suppose that $A \in \mathscr{F}$ satisfies $f^{-1}(A) = A$. Then

$$\mu(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(f^{-i}(A) \cap A) = \mu(A)^2 .$$

This implies that $\mu(A) \in \{0, 1\}$. □

[7] Topological mixing neither implies, nor is implied by (weak) mixing: there are examples of systems that are weak mixing but not topologically mixing, and examples that are topologically mixing but not strong mixing.

We now introduce the notion of *mixing* for an ergodic invariant probability measure. While also a measure of dynamical complexity, it should not be confused with the separate notion of *topological mixing*.

Definition 4.3 (Mixing) Let $(X, \mathscr{F}, \mu, f)$ be a probability measure preserving dynamical system. Then μ is *mixing* if

$$\lim_{n\to\infty} \mu(A \cap f^{-n}(B)) = \mu(A)\mu(B), \ \forall A, B \in \mathscr{F}.$$

We note that if follows directly from Proposition 4.3 that mixing measures are ergodic.

As $f_*^n \mu(B) = \mu(f^{-n}(B)) = \mu(B)$, mixing means that

$$\lim_{n\to\infty} \frac{\mu(A \cap f^{-n}(B))}{\mu(f^{-n}(B))} = \mu(A)$$

for all $A, B \in \mathscr{F}$ with $\mu(B) > 0$. This means that A and $f^{-n}(B)$ are asymptotically, in the large n limit, *stochastically independent*.

The following notion of *weak mixing* is positioned between mixing and ergodicity.

Definition 4.4 (Weak mixing) Let $(X, \mathscr{F}, \mu, f)$ be a probability measure preserving dynamical system. Then μ is *weakly mixing* if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(f^{-i}(A) \cap B) - \mu(A)\mu(B)| = 0, \ \forall A, B \in \mathscr{F}.$$

It follows that mixing invariant probability measures are weakly mixing and that weakly mixing invariant probability measures are ergodic.

Example 4.4 The following facts are readily verified:

- Let $(X, \mathscr{F})$ be a measure space and $f : X \to X$ measurable. Let $f(x) = x$ for some $x \in X$. Then the f-invariant measure δ_x is mixing.
- Let $(X, \mathscr{F})$ be a measure space and $f : X \to X$ measurable. Let $p \in \mathbb{N}$ be least such that $f^p(x) = x$. with $p > 1$, for some $x \in X$. Then the ergodic invariant measure $\sum_{i=1}^{p} \delta_{f^i(x)}$ is not weakly mixing.
- Let $E_k : S^1 \to S^1$ be the uniformly expanding circle map, as defined in (2.25). Then the Lebesque measure on S^1 is E_k-invariant and mixing.

4.6 Metric entropy

4.7 Markov chains

We now develop the notion of a Markov chain with $k < \infty$ states, as an extension of the finite topological Markov chain introduced in Definition 3.4. The Markov chain is a stochastic process generating a (half-infinite) sequence in k symbols by providing for each possible symbol i, the probability P_{ij} that the next symbol is j. Importantly, the transition probability for "the next symbol in the sequence" only depends on the preceding symbol. The set of all relevant transition probabilities is represented by a transition matrix P, with the transition probabilities P_{ij}, as defined above, as elements. Such a matrix is called a *stochastic matrix*, with the defining property that

$$P_{ij} \geq 0 \text{ for all } i, j = 0, \dots, k-1 \text{ and } \textstyle\sum_{j=0}^{k-1} P_{ij} = 1 \text{ for all } i = 0, \dots, k-1.$$

The Markov chain provides a probabilistic description of forward trajectories along admissible paths of its associated Markov graph, whose connectivity matrix A is *compatible* with the transition matrix P in the following sense:

$$\begin{cases} A_{ij} = P_{ij} & \text{if } P_{ij} = 0, \\ A_{ij} = 1 & \text{if } P_{ij} > 0. \end{cases} \tag{4.11}$$

Moreover, if the Markov chain has k states, then the probability to find the symbol j, $m \geq 2$ locations after the symbol i, is given by

$$(P^m)_{ij} = \sum_{i_1,\dots,i_{m-1}=0}^{k-1} P_{ii_1} P_{i_1 i_2} \dots P_{i_{m-1} j}.$$

Note that $P^1 = P$ and that P^m is a stochastic matrix if P is.

The characterisation of connectivity matrices in topological Markov chains, as *(ir)reducible* and *primitive*, has a natural extension to stochastic transition matrices.

Definition 4.5 (Irreducible and primitive stochastic matrix) A $k \times k$ stochastic matrix P is called *irreducible* if $\forall i, j \in 0, \dots, k-1$, $\exists m \geq 1$ such that $(P^m)_{ij} > 0$, and *primitive* if $\exists m \geq 1$ such that $(P^m)_{ij} > 0$, $\forall i, j \in 0, \dots, k-1$.

The following properties of stochastic matrices are well-established. An element of $\mathbb{R}^k$ is called a *probability vector* if all its elements are non-negative and the sum of all its elements is equal to 1. If such a vector has no elements equal to zero, then it is called *positive*.

Proposition 4.4 *Let P be a $k \times k$ stochastic matrix, then*

(i) the largest real eigenvalue of P is equal to 1 and the spectral radius $r(P) = 1$,
(ii) there exists a probability vector v that is a left eigenvector of P for eigenvalue 1, i.e. $vP = v$,
(iii) if P is irreducible, then P has a unique left eigenvector v for eigenvalue 1 which is a positive probability vector.

We leave the proof of this proposition as Exercise 4.10.

We now show how any stochastic matrix P and (one of) its left probability eigenvector(s) for eigenvalue 1 constitute an invariant measure for the shift map σ_A on $\Sigma^+_{k,A}$, where A is to connectivity matrix associated to the stochastic matrix P as in (4.11).

Definition 4.6 (Markov measure) Let P be a $k \times k$ stochastic matrix, $v = (v_{i_0}, \ldots, v_{i_{n-1}})$ (one of) its left probability eigenvector(s) for eigenvalue 1 and A the associated connectivity matrix. Let $\mathscr{S}$ denote the semi-ring of cylinder sets $C_{i_0 \ldots i_{n-1}}$ of $\Sigma_{k,A}^+$ and the pre-measure $\mu_{v,P} : \mathscr{S} \to [0,1]$ be defined as

$$\mu_{v,P}(C_{i_0 \ldots i_{n-1}}) = v_{i_0} P_{i_0 i_1} \ldots P_{i_{n-2} i_{n-1}}.$$

Then the *Markov measure* $\mu_{v,P} : \mathscr{B}(\Sigma_{k,A}^+) \to [0,1]$ is the unique extension of this pre-measure to the Borel σ-algebra.

Note that the existence of the unique existence of the pre-measure on the semi-ring to the Borel σ-algebra, relies on Theorem 4.7. Markov measures corresponding to stochastic matrices P where P_{ij} only depends on j, are called *Bernoulli measures*, see also Example 4.11 in the Appendix to this chapter.

Theorem 4.6 (Invariance and ergodicity of Markov measures) *Markov measures $\mu_{v,P}$ are ergodic invariant probability measures for the shift map $\sigma_A : \Sigma_{k,A}^+ \to \Sigma_{k,A}^+$.*

The proof of this result is defered to Exercise 4.12 While the good news is that all ergodic invariant probability measures of the shift σ_A are Markov measures, we note that due to the freedom in chosing the stochastic matrix P, there exist an uncountably infinite number of ergodic measures for each and every (nontrivial) shift.

Exercises

4.10 Prove Proposition 4.4

4.11 Let P be an irreducible stochastic matrix and $v_1 P = v_1$. Show that for all probability vectors $v \in \mathbb{R}^k$

(a) $\lim_{n\to\infty} v \frac{1}{n} \sum_{i=0}^{n-1} P^i = v_1$,
(b) $\lim_{n\to\infty} v P^n = v_1$, if P is primitive.

4.12 Prove Theorem 4.6

4.13 * Consider two containers, one of which containing 100 balls, numbered 1 to 100, and the other bowl none. Suppose each second one draws i.i.d. a number between 1 and 100 (each with equal probability). Then the ball, whose number is drawn, is moved from whichever container it is, to the other container.[8]

One expects that this systems will settle into an equilibrium state in which there are roughly 50 balls in each container. There will of course continue to be fluctuations, but it would appear highly unlikely for the system to return to the initial state in which 100 balls are in the first container. Nevertheless, Poincaré's recurrence

[8] This model is due to Paul Ehrenfest, intended to model the behaviour of a gas distributed over 2 connected chambers.

theorem (Theorem 4.2) asserts that this situation will occur almost surely (although we may have to wait a long time for this to happen).

Use Kac's lemma (Lemma 4.2) to estimate how long it is expected to take for this system to return to its initial state. Compare this to the expected return time from the 50/50 equilibrium state to itself.

4.8 Measurable conjugacy

Birkhoff's ergodic theorem, Theorem 4.3, expresses the fact that $\mu_{v,P}$-almost surely, the time-average of an observable along a forward orbit of σ_A converges to the $\mu_{v,P}$-expectation of the observable. At the level of the shift map on the topological Markov chain, there appears to be no particular preference for one ergodic Markov measure above any other. In case a shift map is semi-conjugate (in the sense of extension) to a dynamical system of interest, this thus establishes the existence of many ergodic invariant measures of the latter. However, there may be specific types of such measures which are considered more important than others. For example, if one is interest in *Lebesgue-typical* phenomena, i.e. that hold for Lebesgue-almost all initial conditions, then it is natural to highlight ergodic invariant measures that are absolutely continuous with respect to the Lebesgue measure.

Definition 4.7 (Measurable conjugacy) Let $(X, \mathscr{F})$ and $(Y, \mathscr{G})$ be measurable spaces and $f : X \to X$ and $g : Y \to Y$ measurable. Then f and g are *measurably conjugate* if there exists a measurable bijection[9] $h : X \to Y$ such that

$$h \circ f = g \circ h. \tag{4.12}$$

Definition 4.8 (Measurable semi-conjugacy) Let $(X, \mathscr{F})$ and $(Y, \mathscr{G})$ be measurable spaces and $f : X \to X$ and $g : Y \to Y$ measurable. Then f and g are *measurably semi-conjugate* if there exists a measurable surjective $h : X \to Y$ satisfying (4.12).

Proposition 4.5 *Let f and g be measurably (semi-)conjugated by h, as in (4.12). Let the measure $\mu : \mathscr{F} \to [0, \infty]$ be f-invariant, then $h_*\mu : \mathscr{G} \to [0, \infty]$ is g-invariant. In addition, if μ is ergodic for f then $h_*\mu$ is ergodic for g.*

The proof of this proposition is left as Exercise 4.14.

This proposition is instrumental for establishing the existence of ergodic invariant measures if a dynamical system is topologically semi-conjugate to a shift map on a topological Markov chain.

We will now apply the above results to re-derive the fact that the Lebesgue measure is an ergodic invariant measure of the expanding circle map E_{10}.

Example 4.5 (Ergodicity of Lebesgue measure for E_{10} via symbolic dynamics.) E_{10} is topologically semi-conjugated to the full shift $\sigma : \Sigma_{10}^+ \to \Sigma_{10}^+$. Consider the

[9] So that both h and h^{-1} are measurable.

10×10 stochastic matrix P, with $P_{ij} = 10^{-1}$ for all $i, j = 0, \ldots, 9$. Then the unique probability left eigenvector for P is $v = 10^{-1}(1, 1, \ldots 1)$, which gives rise to the Markov (and Bernouilli) measure $\mu_{v,P}$ on $\mathscr{B}(\Sigma_{10}^+)$, characterised by its definition on cylinder sets: $\mu(C_{i_0 \ldots i_{n-1}}) = 10^{-n}$. Let $h : \Sigma_{10}^+ \to [0, 1]$ denote the transformation inducing the semi-conjugacy. By construction, $h_* \mu_{v,P}$ is defined by its value on the semi-ring of cylinder sets on Σ_{10}^+.[10] Ergodicity of λ follows from ergodicity of $\mu_{v,P}$, cf. Proposition 4.5.

For a more general approach towards identifying Markov measures inducing invariant measures that are absolutely continuous with respect to Lebesgue measure, it is useful to employ an alternative characterisation of absolute continuity.

Proposition 4.6 *Let μ and ν be two measures on a measurable space $(X, \mathscr{F})$. Then μ is absolutely continuous with respect to ν ($\mu << \nu$) if and only if $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $\forall A \in \mathscr{F}$, $\nu(A) < \delta \ \Rightarrow \ \mu(A) < \varepsilon$.*

Proof Recall that $\mu << \nu$ if and only if for all $A \in \mathscr{F}$, $\nu(A) = 0 \Rightarrow \mu(A) = 0$. We first show this implies that $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $\forall A \in \mathscr{F}$, $\nu(A) < \delta \ \Rightarrow \ \mu(A) < \varepsilon$. Namely, suppose the latter does not hold. Then for some $\varepsilon > 0$ and all $n \in \mathbb{N}$, there exists $E_n \in \mathscr{F}$, such that $\lambda(E_n) < 2^{-n}$ and $\mu(E_n) \geq \varepsilon$. Let

$$F_k = \bigcup_{n=k}^{\infty} E_n \quad \text{and} \quad F = \bigcap_{k=1}^{\infty} F_k.$$

Then, using the fact that $\sum_{n=k}^{\infty} 2^{-n} = 2^{1-k}$, we obtain

$$\lambda(F_k) < 2^{1-k} \quad \Rightarrow \quad \lambda(F) = \lim_{n\to\infty} \lambda(\cap_{k=1}^{n} F_k) = \lim_{n\to\infty} \lambda(F_n) = 0.$$

But also $\mu(F_k) \geq \varepsilon$ for all $k \in \mathbb{N} \Rightarrow \mu(F) \geq \varepsilon$ and hence μ is not absolutely continuous with respect to ν.

To establish the reverse implication, note that if for all $n \in \mathbb{N}$, there exists $\delta_n > 0$ such that $\nu(A) < \delta_n$ implies $\mu(A) < n^{-1}$ for all $A \in \mathscr{F}$, then $\nu(A) = 0$ implies that $\mu(A) = 0$. □

We now continue to address the identification of ergodic invariant measures of piecewise affine expanding maps of the interval $[0, 1]$. We use the following.

Proposition 4.7 *Let $f : [0, 1] \to [0, 1]$ be a piecewise affine expanding map, topologically semi-conjugate to the shift map on an irreducible topological Markov chain, $\sigma_A : \Sigma_{k,A}^+ \to \Sigma_{k,A}^+$. Then, a Markov measure, $\mu_{v,P}$, where A is compatible with P, induces an ergodic f-invariant Borel probability measure $h_* \mu_{v,P}$ on $[0, 1]$ that is absolutely continuous with respect to the Lebesgue measure λ, if there exists $K > 0$ such that for every cylinder set $C_{i_0 \ldots i_{n-1}} \subset \Sigma_{k,A}^+$,*

$$\lambda(h(C_{i_0 \ldots i_{n-1}})) \geq K \cdot \mu_{v,P}(C_{i_0 \ldots i_{n-1}}). \tag{4.13}$$

[10] It turns out that $h(\mathscr{S})$, where $\mathscr{S}$ denotes the semi-ring of cylinder sets, is a so-called *sufficient semi-ring* for $\mathscr{B}(S^1)$ [9].

Proof All of the essential ingredients have been discussed before and are by now canonical. The sufficiency of condition (4.13) implies the one in Proposition 4.6. □

It now remains to identify Markov measures that induce ergodic invariant measure on the factor that are absolutely continuous with respect to Lebesgue measure. Let $\mathcal{R} = \{R_0, \ldots, R_{k-1}\}$ denote the relevant Markov partition for a the piecewise affine expanding map. Let τ_i denote the (constant) derivative of f on R_i and A the transition matrix associated to the resulting topological Markov chain. Let

$$P_{ij} = \begin{cases} 0 & \text{if } A_{ij} = 0, \\ \frac{|R_j|}{|\tau_i| \cdot |R_i|} & \text{if } A_{ij} = 1. \end{cases}$$

Then

$$P_{i_0 i_1} \cdot \ldots \cdot P_{i_{n-2} i_{n-1}} = \frac{|R_{n-1}|}{|R_{i_0}|} \prod_{i=0}^{n-2} |\tau_i|^{-1}$$

and

$$\mu_{v,P}(C_{i_0 \ldots i_{n-1}}) := v_{i_0} \cdot P_{i_0 i_1} \cdot \ldots \cdot P_{i_{n-2} i_{n-1}} = \frac{v_{i_0} \cdot |R_{n-1}|}{|R_{i_0}|} \prod_{i=0}^{n-2} |\tau_i|^{-1}.$$

The affine action on each of the partitions provides the following explicit expression for the size of the refined partition elements:

$$|R_{ij}| = |R_i| \cdot \frac{|R_j|}{|\tau_i| \cdot |R_i|} = \frac{|R_j|}{|\tau_i|} \Rightarrow |R_{i_0 \ldots i_{n-1}}| = |R_{n-1}| \cdot \prod_{j=0}^{n-2} |\tau_{i_j}|^{-1},$$

so that

$$\lambda(R_{i_0 \ldots i_{n-1}}) = \frac{|R_{i_0}|}{v_{i_0}} \mu_{v,P}(C_{i_0 \ldots i_{n-1}}) = \frac{|R_{i_0}|}{v_{i_0}} h_* \mu_{v,P}(R_{i_0 \ldots i_{n-1}}). \tag{4.14}$$

Hence, the conditions of Proposition 4.7 are satisfied with

$$K = \min_{i \in \{0, \ldots, k-1\}} \left\{ \frac{|R_i|}{v_i} \right\}$$

in (4.13), and $h_* \mu_{v,P}$ is the (unique) ergodic f-invariant probability measure on $[0,1]$ that is absolutely continuous with respect to Lebesgue measure. Moreover, (4.14) shows that this measure has constant Lebesgue density on the elements (R_i) of the Markov partition, namely $v_i / |R_i|$ on R_i.

Example 4.6 Let $f : [0,1] \to [0,1]$ be defined as

$$f(x) = \begin{cases} 4x & \text{if } x \in [0, \frac{1}{4}], \\ \frac{3}{2} - 2x & \text{if } x \in [\frac{1}{4}, \frac{1}{2}], \\ 2x - \frac{1}{2} & \text{if } x \in [\frac{1}{2}, \frac{3}{4}], \\ 4 - 4x & \text{if } x \in (\frac{3}{4}, 1]. \end{cases} \tag{4.15}$$

Then $\mathcal{R} = \{(0, \frac{1}{4}), (\frac{1}{4}, \frac{1}{2}), (\frac{1}{2}, \frac{3}{4}), (\frac{3}{4}, 1)\}$ is a Markov partition and f is affine piecewise

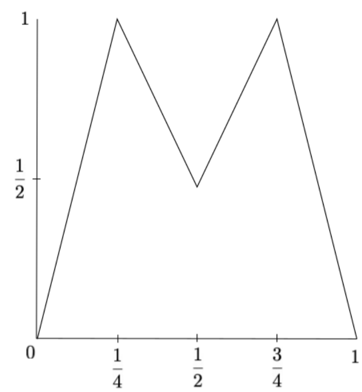

Fig. 4.1: Graph of the map f (4.15).

expanding on this partition. Following the above recipe, the relevant stochastic matrix P is found to be

$$P = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

P is irreducible with unique left probability eigenvector $v = \frac{1}{8}(1, 1, 3, 3)$. $h_*\mu_{v,P}$ is the unique ergodic f-invariant measure that is absolutely continuous with respect to the Lebesgue measure.

Exercises

4.14 Prove Proposition 4.5.

4.15 Verify the claims in Example 4.6 and find the density of the ergodic invariant measure $h_*\mu_{v,P}$.

4.16 Let $f_4(x) : [0, 1] \to [0, 1]$ be the logistic map at parameter value 4, $f_4(x) = 4x(1 - x)$ and T the tent map (3.6).

(a) Show that f_4 and T are topologically and measurably conjugate to each other by the transformation $h(x) = \sin^2(\frac{\pi}{2}x)$.
(b) Show $\mu : \mathscr{B}([0,1]) \to [0,1]$, defined for all $A \in \mathscr{B}([0,1])$ as

$$\mu(A) = \int_A \frac{1}{\pi\sqrt{x(1-x)}}dx,$$

is an ergodic invariant measure for f_4.
(c) Consider the observable $g_n(x) := |\frac{d}{dx}(f_4^n)(x)|$. Show that g_n grows exponentially (for Lebesgue-typical initial conditions x) and determine the growth rate. Interpret the answer.

4.17 Show that the Lebesgue measure is an ergodic invariant measure of the tent map (3.6), by demonstrating that the Lebesgue measure is the push-forward of a Markov measure of a topological Markov chain by the transformation that establishes a semi-conjugacy between the shift on this Markov chain and the tent map.

4.18 Let $f : [0,1] \to [0,1]$ be given by

$$f(x) = \begin{cases} 2 + \sqrt{2}(x-1) & \text{if } 0 \le x \le c, \\ \sqrt{2}(1-x) & \text{if } c < x \le 1, \end{cases} \tag{4.16}$$

where $c = \frac{1}{2}(2-\sqrt{2})$. The graph of f is sketched below:

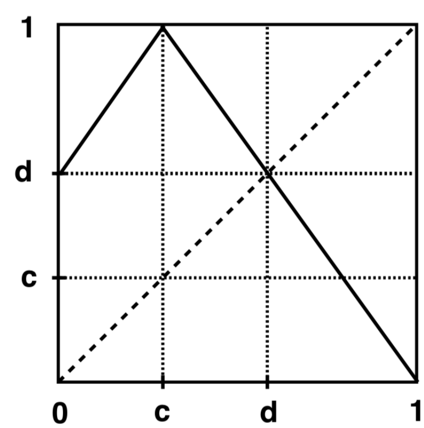

Fig. 4.2: Sketch of the graph of f (4.16). $d = 2c$.

Let the measure $\mu : \mathscr{B}([0,1]) \to [0,1]$ be defined as

$$\mu(A) := d^{-1} \int_A \left(\frac{\mathbb{1}_{[0,d]}(x)}{2} + \frac{\mathbb{1}_{(d,1]}(x)}{\sqrt{2}} \right) dx \tag{4.17}$$

with $d = 2 - \sqrt{2}$.

(a) Show that μ is a probability measure and sketch the graph of its density.
(b) Show that μ is an invariant measure for f.
(c) *Prove that μ is ergodic by showing that f is topologically semi-conjugate to a shift map σ_A on a finite topological Markov chain $\Sigma^+_{k,A}$ and identfiying the relevant ergodic Markov measure ν for this shift map so that $h_*\nu = \mu$, where $f \circ h = h \circ \sigma_A$.
(d) Show that ergodicity of μ implies that

$$\int_0^1 (\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f^i(x))dx = \frac{5}{4} - \frac{1}{\sqrt{2}}.$$

Appendix: some preliminaries concerning measure theory

In this appendix, we summarize some elementary notions and facts from measure theory that we use in this chapter. All of this, and more, can be found in more detail in standard references like [5, 8]

Measures, semi-rings and σ-algebras.

A *measurable space* is a set X with a σ-algebra $\mathscr{F}$ of subsets of X that will be measured.

Definition 4.9 (σ-algebra) Let X be a non-empty set and $\mathscr{F}$ be a collection of subsets of X, then $\mathscr{F}$ is a σ-algebra if

1. $\emptyset \in \mathscr{F}$,
2. $A \in \mathscr{F} \Rightarrow A^c := X \setminus A \in \mathscr{F}$,
3. for any countable collection $\{A_i\}_{i\in\mathbb{N}}$ of elements of $\mathscr{F}$, $\cup_{i\in\mathbb{N}} A_i \in \mathscr{F}$.

The elements of a σ-algebra of a measurable space are also referred to as *measurable sets*.

Let C be a set of subsets of X, then the σ-algebra *generated by* C is the smallest σ-algebr of X that contains C.

Definition 4.10 (Measure) Let $(X, \mathscr{F})$ be a measurable space, then a function $\mu : \mathscr{F} \to [0, \infty]$ is a *measure* if

1. $\mu(\emptyset) = 0$ (null empty set),
2. $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ for any countable (or finite) collection of pairwise disjoint sets $\{A_i\}_{k=1}^{\infty} \in \mathscr{F}$ (σ-additivity).

A *measure space* $(X, \mathscr{F}, \mu)$ is a measurable space $(X, \mathscr{F})$ with specific measure μ on it. If $\mu(X) = 1$, then the measure space $(X, \mathscr{F}, \mu)$ is also referred to as a *probability space*.

Example 4.7 (Dirac δ-measure) Let $(X, \mathscr{F})$ be a measurable space and $x \in X$. Then the Dirac δ-measure $\delta_x : \mathscr{F} \to \{0, 1\}$ is defined as

$$\delta_x(A) = \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ if } x \notin A. \end{cases}$$

The following elementary properties of measures are useful:

Proposition 4.8 *Let $(X, \mathscr{F}, \mu)$ be a measure space, then*

1. *Monotonicity: If $A, B \in \mathscr{F}$ and $A \subset B$ then $\mu(A) \leq \mu(B)$.*
2. *Subadditivity: If $\{A_i\}_{i\in\mathbb{N}} \subset \mathscr{F}$ then $\mu(\cup_{i\in\mathbb{N}} A_i) \leq \sum_{i\in\mathbb{N}} \mu(A_i)$.*
3. *Continuity from below: If $\{A_i\}_{i\in\mathbb{N}} \subset \mathscr{F}$ and $A_1 \subset A_2 \subset \ldots$, then $\mu(\cup_{i\in\mathbb{N}} A_i) = lim_{i\to\infty}\mu(A_i)$.*
4. *Continuity from above: If $\{A_i\}_{i\in\mathbb{N}} \subset \mathscr{F}$ and $A_1 \supset A_2 \supset \ldots$, and $\mu(A_1) < \infty$ then $\mu(\cap_{i\in\mathbb{N}} A_i) = lim_{i\to\infty}\mu(A_i)$.*

If a measure μ on a measurable space $(X, \mathscr{F})$ is such that $\mu(X) = 1$, then μ is called a *probability measure*.

If $(X, \mathscr{F})$ and $(Y, \mathscr{G})$ are measurable spaces, then $f : X \to Y$ is called *measurable* if for all $B \in \mathscr{G}$, the pre-image $f^{-1}(B) \in \mathscr{F}$. A measurable function $g : X \to \mathbb{R}$ is also called a *random variable*.

In general, σ-algebras may be very large and hard to deal with in practise. If a σ-algebra is generated by a *semi-ring* then the situation simplifies significantly.

Definition 4.11 (Semi-ring) Let X be a non-empty set and $\mathscr{S}$ be a collection of subsets of X, then $\mathscr{S}$ is a semi-ring if

1. $\emptyset \in \mathscr{S}$,
2. $\forall A, B \in \mathscr{S}$, $A \cap B \in S$,
3. $\forall A, B \in \mathscr{S}$ there exists $n \in \mathbb{N}$ and disjoint sets $C_1, \ldots C_n$ such that $A \setminus B = \cup_{i=1}^{n} C_i$.

A function $\mu : \mathscr{S} \to [0, \infty]$, satisfying conditions 1 and 2 of Definition 4.10 for all $A \in \mathscr{S}$, is called a *pre-measure* on the semi-ring $\mathscr{S}$.

Theorem 4.7 (Carathéodory's extension theorem) *Any pre-measure μ on a semi-ring S can be extended to a measure on $\sigma(\mathscr{S})$, the σ-algebra generated by $\mathscr{S}$.*

Moreover, if μ is σ*-finite* on $(X, \mathscr{S})$ – i.e. there exists $\{A_i\}_{i\in\mathbb{N}} \subset \mathscr{S}$ such that $\forall i \in \mathbb{N}$, $\mu(A_i) < \infty$, and $X = \cup_{i\in\mathbb{N}} A_i$ – then the extention of μ to a measure on $\sigma(\mathscr{S})$ is unique.

Example 4.8 (Lebesgue measure on $\mathbb{R}$) The set of "half-open" intervals in the real line, $\{(a, b] \mid a, b \in \mathbb{R},\ a \leq b\}$, forms a semi-ring and the function $\lambda((a, b])) := b - a$ is a pre-measure, which extends to the so-called *Lebesgue measure* on the σ-algebra $\mathscr{B}(\mathbb{R})$ generated by this semi-ring.

Example 4.9 (Absolutely continuous measure) Let $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a continuous function and let $\mu((a,b]) := \int_a^b f(y)dy$, then μ is a pre-measure on the semi-ring of half-open intervals introduced in Example 4.8 which extends to a unique measure on $\mathscr{B}(\mathbb{R})$. f is called the *Lebesgue density* of this measure. $\mu = \lambda$ if and only if $f = 1$.

$\mathscr{B}(\mathbb{R})$ is an example of a *Borel σ-algebra*: $\mathscr{B}(X)$ of a metric (or topological) space X is the smallest σ-algebra containing all open subsets of X.

Example 4.10 ($\mathscr{B}(\Sigma_k^+)$) The set of cylinder sets

$$\{C_{i_0 \dots i_{n-1}} \subset \Sigma_k^+ \mid n \in \mathbb{N},\ i_j \in \{0, \dots, k-1\}\ \forall j \in \{0, \dots, n-1\}\}$$

forms a semi-ring generating the Borel σ-algebra $\mathscr{B}(\Sigma_k^+)$ (as usual, considering the topology induced by the metric d^{Σ^+}).

Example 4.11 (Bernoulli measure) Consider the function μ on the semi-ring of cylinder sets of Σ_k^+, cf. Example 4.10, defined by $\mu_p(C_{i_0 \dots i_{n-1}}) := \Pi_{j=0}^{k-1} p_{i_j}$, where $p = (p_0, \dots, p_{k-1})$ with $p_j \geq 0\ \forall j \in \{0, \dots, k-1\}$ and $\sum_{j=0}^{k-1} p_j = 1$. Then μ_p extends uniquely to $\mathscr{B}(\Sigma_k^+)$.

If X is a topological (or metric) space, it is natural to consider the measurable space $(X, \mathscr{B}(X))$. Measures on $(X, \mathscr{B}(X))$ are referred to as *Borel measures*.

In this course, where we consider the dynamics of continuous maps on metric spaces, we always consider this setting. The following elementary result is very useful.

Lemma 4.3 *Let X, Y be topological (or metric) spaces and $h : X \to Y$ be continuous and surjective, then h is* Borel measurable*, i.e. $\forall A \in \mathscr{B}(Y)$, $h^{-1}(A) \in \mathscr{B}(X)$.*

Lebesgue integration

Measures are important for integration, in particular in settings where the standard Riemann integral does not exist.

Let $(X, \mathscr{F}, \mu)$ be a measure space. Given finitely many disjoint sets $A_1, \dots, A_n \in \mathscr{F}$ and positive real constants $c_1, \dots, c_n$, a function $h : \mathscr{F} \to \mathbb{R}_{\geq 0}$ of the form $h = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$ is called an *elementary function*, where the characteristic function $\mathbb{1}_A : X \to \{0, 1\}$

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ if } x \notin A. \end{cases}$$

The *Lebesgue integral* of such an elementary function is defined as

$$\int_X h d\mu = \sum_{i=1}^n c_i \mu(A_i).$$

This integral can be extended to measurable functions, i.e. functions $g : X \to \mathbb{R}$ where $g^{-1}(B) \in \mathscr{F}$ for all $B \in \mathscr{B}(\mathbb{R})$.

First, we define the Lebesgue integral for non-negative measurable functions $g : X \to \mathbb{R}_{\geq 0}$ as

$$\int_X g d\mu := \sup\{\int_X h d\mu \mid h \text{ is elementary and } h \leq g \},$$

where $h \leq g$ means that $h(x) \leq g(x)$ for all $x \in X$. This integral is well defined, as follows from the following lemma.

Lemma 4.4 *Let $g : X \to \mathbb{R}_{\geq 0}$ be measurable, then h_n converges pointwise to g and $\int_X g d\mu = \lim_{n\to\infty} \int_X h_n d\mu$ where*

$$h_n = \sum_{i=0}^{2^n-1} \frac{i}{2^n} \mathbb{1}_{\{\frac{i}{2^n} \leq g \leq \frac{i+1}{2^n}\}} + n\mathbb{1}_{\{n \leq g\}}.$$

Finally to integrate general measurable functions $g : X \to \mathbb{R}$ we write $g = g^+ - g^-$ where

$$g^+(x) := \max\{g(x), 0\} \text{ and } g^-(x) := \max\{-g(x), 0\},$$

where e note that g^+ and g^- are both non-negative measurable functions.

We call a measurable function $g : X \to \mathbb{R}$ *(Lebesgue μ-)integrable*, if $\int_X g^+ d\mu < \infty$ and $\int_X g^- d\mu < \infty$. For integrable g, the Lebesgue integral is defined as

$$\int_X g d\mu = \int_X g^+ d\mu - \int_X g^- d\mu.$$

We finally list some useful properties of integrable functions.

Proposition 4.9 *Let $(X, \mathscr{F}, \mu)$ be a measure space and $g, h : X \to \mathbb{R}$ be integrable. Then*

1. *$\int_X (ag + bh) d\mu = a \int_X g d\mu + b \int_X h d\mu$, $\forall a, b \in \mathbb{R}$,*
2. *if $A \in \mathscr{F}$ is a null set, i.e. $\mu(A) = 0$, then $\int_A g d\mu := \int_X g \cdot \mathbb{1}_A d\mu = 0$,*
3. *if $g \geq 0$ and $\int_X g d\mu = 0$ then $\mu(\{x \in X \mid g(x) > 0\}) = 0$,*
4. *if $g \leq h$ then $\int_X g d\mu \leq \int_X h d\mu$,*
5. *$|\int_X g d\mu| \leq \int_X |g| d\mu$.*

The latter property implies that integrability of g can be characterized also as $\int_X |g| d\mu < \infty$.

Example 4.12 Let $g : [0, 1] \to \{0, 1\}$ be defined as

$$g(x) := \begin{cases} 1 \text{ if } x \in \mathbb{R} \setminus \mathbb{Q} \\ 0 \text{ if } x \in \mathbb{Q} \end{cases}$$

It is well-known that g is not Riemann integrable (check this!). We proceed to show that g is Lebesgue-integrable.

It follows from the defining properties of measures that $\lambda(\mathbb{Q}) = 0$, where λ denotes the Lebesgue measure, since $\mathbb{Q}$ is countable and $\lambda(x) = 0$ for all $x \in \mathbb{R}$. Hence

$$\int_{[0,1]} g d\lambda = \lambda([0,1] \setminus \mathbb{Q}) = \lambda([0,1]) - \lambda(\mathbb{Q}) = 1.$$

Example 4.13 (μ-almost sure properties) Let $(X, \mathscr{F}, \mu)$ be a measure space. If a statement or property, depending on a variable $x \in X$, holds under the condition that $x \in A \in \mathscr{F}$ with $\mu(X \setminus A) = 0$, then the statement/property is said to hold *μ-almost surely*, or *for μ-almost all $x \in X$*.

For instance, in Example 4.12,

$$\int_{[0,1]} (g-1) d\lambda = 0 \quad \Leftrightarrow \quad g \text{ is } \lambda\text{-almost surely equal to 1.}$$

Conditional probability

Let $(X, \mathscr{F}, \mu)$ be a probability space and $g : X \to \mathbb{R}$ be a integrable random variable. Then, the *expectation of* g is defined as

$$\mathbb{E}[g] := \int_X g d\mu.$$

Let $\mathscr{G}$ be a σ-subalgebra of $\mathscr{F}$, i.e. $\mathscr{G} \subset \mathscr{F}$ and is a σ-algebra. Then an integrable $\mathscr{G}$-measurable random variable $g' : X \to \mathbb{R}$ is called a *conditional expectation* of g with respect to $\mathscr{G}$, denoted as $g' = \mathbb{E}[g|\mathscr{G}]$, if

$$\int_A g' d\mu = \int_A g d\mu, \ \forall A \in \mathscr{G}.$$

Example 4.14 Let $\mathscr{G}$ be a σ-subalgebra generated by a finite partition of X, such that $X = \cup_{i=1}^n A_i$ and the A_i's are pairwise disjoint. Suppose furthermore that $\mu(A_i) > 0$ for all $i = 1, \dots, n$. Then

$$\mathbb{E}[g|\mathscr{G}] = \sum_{i=1}^n \frac{\mathbb{E}[g \cdot \mathbb{1}_{A_i}]}{\mu(A_i)} \mathbb{1}_{A_i}.$$

In particular, if $\mathscr{G} = \{\emptyset, X, A, A^c\}$ for some $A \in \mathscr{F}$, then

$$\mathbb{E}[g|\mathscr{G}] = \frac{\mathbb{E}[g \cdot \mathbb{1}_A]}{\mu(A)} \mathbb{1}_A + \frac{\mathbb{E}[g \cdot \mathbb{1}_{A^c}]}{\mu(A^c)} \mathbb{1}_{A^c}.$$

The existence of conditional expectation can be viewed as a consequence of the *Radon-Nikodym* Theorem.

Let μ and ν be σ-finite measures on a measurable space $(X, \mathscr{F})$, then ν is *absolutely continuous with respect to* μ, denoted as $\nu << \mu$, if for all $A \in \mathscr{F}$, $\mu(A) = 0$ implies that $\nu(A) = 0$.

Theorem 4.8 (Radon-Nikodym) *Let μ and ν be σ-finite measures on a measurable space $(X, \mathscr{F})$ such that $\nu << \mu$. Then there exists an essentially unique measurable function $g : X \to \mathbb{R}$ such that*

$$\nu(A) = \int_A g d\mu, \ \forall A \in \mathscr{F}.$$

In this theorem, *essential uniqueness* refers to the fact that any two functions g satisfying (4.8) differ at most on a subset $B \subset X$ of μ-measure zero, i.e. $\mu(B) = 0$.

The *density* g in (4.8) is also denoted as $\frac{d\nu}{d\mu}$ and called the *Radon-Nikodym* derivative.

Lemma 4.5 *The conditional expectation $\mathbb{E}[g|\mathscr{G}]$ exists and is essentially unique.*

Proof Let for all $A \in \mathscr{G}$,

$$\nu(A) := \int_A g^+ d\mu,$$

where, as before, $g^+ := \max\{g, 0\}$. Then $\nu << \mu$ as measures on $(X, \mathscr{G})$ and the Radon-Nikodym derivative $\frac{d\nu}{d\mu}$ is precisely the conditional expectation:

$$\int_A g^+ d\mu = \int_A \underbrace{\frac{d\nu}{d\mu}}_{\mathbb{E}[g^+|\mathscr{G}]} d\mu.$$

Essential uniqueness follows from the Radon-Nikodym theorem.

$\mathbb{E}[g^-|\mathscr{G}]$ can be obtained analogously, and

$$\mathbb{E}[g|\mathscr{G}] = \mathbb{E}[g^+|\mathscr{G}] - \mathbb{E}[g^-|\mathscr{G}].$$

To establish essential uniqueness of $\mathbb{E}[g|\mathscr{G}]$, suppose to the contrary that it is not. Then also $\mathbb{E}[g^+|\mathscr{G}] - \mathbb{E}[g|\mathscr{G}]$ is not essentially unique, but this contradicts the fact that $\mathbb{E}[g^+|\mathscr{G}] - \mathbb{E}[g|\mathscr{G}] = \mathbb{E}[g^-|\mathscr{G}]$ is essentially unique. □

Proposition 4.10 *Let $(X, \mathscr{F})$ be a measurable space and $g_i : X \to \mathbb{R}$, $i = 1, 2$, denote random variables. Let $a_i \in \mathbb{R}$, $i = 1, 2$ and $\mathscr{G}$ denote a σ-subalgebra of $\mathscr{F}$. Then*

1. $\mathbb{E}[a_1 g_1 + a_2 g_2|\mathscr{G}] = a_1\mathbb{E}[g_1|\mathscr{G}] + a_2\mathbb{E}[g_2|\mathscr{G}]$,
2. *if* $g_1 \leq g_2$ *then* $\mathbb{E}[g_1|\mathscr{G}] \leq \mathbb{E}[g_2|\mathscr{G}]$,
3. $\mathbb{E}[\mathbb{E}[g|\mathscr{G}]] = \mathbb{E}[g]$.

References

1. R. Adler. Symbolic dynamics and Markov partitions. *Bulletin of the American Mathematical Society (New Series)* **35** (1998), 1–56.
2. J. Banks, J. Brooks, G. Cairns, G. Davis and P. Stacey. On Devaney's definition of chaos. *The American Mathematical Monthly* **99** (1992), 332–334.
3. M. Brin and G. Stuck. *Dynamical systems.* Cambridge University Press (2002).
4. R.L. Devaney, *An introduction to chaotic dynamical systems*. Benjamin Cummings (1986).
5. B. Hasselblatt and A. Katok. *A first course in dynamics.* Cambridge University Press (2012).
6. T.-Y. Li and J.A. Yorke. Period Three Implies Chaos. *The American Mathematical Monthly* **82** (1975), 985–992.
7. C.E. Silva. *Invitation to Ergodic Theory*. The Student Mathematical Library **42**. AMS (2008)
8. A.N. Sharkovskii. Co-existence of cycles of a continuous mapping of the line into itself. *Ukrainian Mathematical Journal* **16** (1964), 61–71.