

2.1

(a) It is necessary that

$$F(x \bmod 1) = F(x) \bmod 1,$$

since otherwise f is not well-defined (in the sense that it would be multiple-valued). The existence of a lift is by construction: let $F(0) = f(0)$ then by varying x continuously from 0 to 1, $F(x)$ is uniquely determined on $[0, 1)$ by $f(x)$, due to the assumption of continuity of f and F .

(b) By (2.7), if F is a lift, then so is $F + n$ for any $n \in \mathbb{Z}$. On the other hand, if F and \tilde{F} are lifts then $F(x) \bmod 1 = f(x) = \tilde{F}(x) \bmod 1$ which implies that $F(x) - \tilde{F}(x) \in \mathbb{Z}$ and constant because of the continuity of F and \tilde{F} . Hence, there is no freedom in choosing the lift, other than the freedom identified above. It also follows that the degree is independent of the chosen lift since if $F(x+1) - F(x) = k$ and $\tilde{F}(x+1) - \tilde{F}(x) = \tilde{k}$ then

$$k - \tilde{k} = F(x+1) - F(x) - (\tilde{F}(x+1) - \tilde{F}(x)) = F(x+1) - \tilde{F}(x+1) - (F(x) - \tilde{F}(x)) = 0.$$

2.2

(a) $k > 0$: The graph of E_k has k intervals $[\frac{n}{k}, \frac{n+1}{k})$ $n = 0, \dots, k-1$ on which it is a straight line from 0 to 1 (in the limit), cf. Fig. 3.1. Each of these lines of this graph intersects the diagonal line (indicating a fixed point), apart from the final line above the domain $[\frac{k-1}{k}, 1)$. So E_k has $k-1$ fixed points. Now $E_k^n = E_{k^n}$ has $k^n - 1$ fixed points.

$k < 0$: The graph of E_k has $E(0) = 0$ and $|k| - 1$ intervals $(\frac{n}{|k|}, \frac{n+1}{|k|}]$ $n = 0, \dots, |k| - 2$ on which it is a straight line from 1 (in the limit) to 0, intersecting the diagonal once, and a final interval $(\frac{|k|-1}{|k|}, 1)$ on which the graph of E_k also intersects the diagonal once. Hence E_k has $|k| + 1 = |k - 1|$ fixed points. Since $E_k^n = E_{k^n}$, E_k^n has $|k^n - 1|$ fixed points, using the result obtained in the first part if n is even and k^n is positive and the results obtained in the second part if n is odd and k^n is negative.

(b) The arguments are essentially identical to the one under (a). Choosing, without loss of generality, coordinates so that $f(0) = 0$ (by labelling one of the fixed points "0"), if $k > 0$ the graph of f has k branches that monotonically increase from 0 to 1, leading to $k - 1$ fixed points. If f has degree k , f^n has degree k^n , and the result follows. One obtains an analogous generalisation from (a) in the case that $k < 0$.

(c) (i) The equation $f_a^n(x) = x$ is polynomial of degree 2^n . This equation thus has at most 2^n roots.

(ii) If $a > 4$, $f_a^{-1}([0, 1])$ consists of two disjoint closed intervals, say I_0 and I_1 , inside $[0, 1]$. Each of these intervals contains a fixed point (as the graph of f_a above these intervals must intersect the diagonal). Iterating this argument, one finds that I_0 and I_1 each also contain two closed intervals that are mapped by f_a^2 to $[0, 1]$ so that f_a^2 has at least 2^2 fixed points. By repetition of this argument one finds that f_a^n has at least 2^n fixed points, so $P_n(f) \geq 2^n$. Then by (i), above, it follows that $P_n(f) = 2^n$.

2.3 If $c > d$, let $r := \sup\{x \in [d, c] \mid f(r) \geq b\}$ and $s := \inf\{x \in [r, c] \mid f(r) \leq a\}$. Then $f([r, s]) = [a, b]$.

2.4 If $f(a) = c$ then $f(c) = b$ and $f(b) = a$. Introducing the new labels $a' = c$, $b' = b$ and $c' = a$, we have If $f(a') = b'$ then $f(b') = c'$ and $f(c') = a'$ with $a' > b' > c'$. We note that the starting geometric assumption is thus essentially identical to the other case, only with the orientation on the line reversed. As the proof does not rely on this orientation, it invariably applies.

2.5 First of all note that we cannot use Theorem 2.1 since the exercise does not concern a continuous map of a closed interval to itself!

We aim to show that there exist n -periodic points for all $n \geq 2$ by showing that $P_n(f) > \sum_{i=1}^{n-1} P_i(f)$.

If the degree of f is $k > 1$ then $P_n(f) = k^n - 1$, see Exercise 2.2(b). We have

$$\sum_{i=1}^{n-1} P_i(f) = \sum_{i=1}^{n-1} (k^i - 1) = 1 - n + \sum_{i=1}^{n-1} k^i = 1 - n + \frac{k - k^n}{1 - k},$$

where we used the geometric series

$$(1 - k) \sum_{i=1}^{n-1} k^i = \sum_{i=1}^{n-1} k^i - \sum_{i=2}^n k^i = k - k^n.$$

So,

$$P_n(f) - \sum_{i=1}^{n-1} P_i(f) = k^n - 1 + (n - 1) - \frac{k - k^n}{1 - k} =: D.$$

It remains to be shown that $D > 0$ for all $n \geq 2$ and $k \geq 2$. Indeed,

$$\underbrace{(1 - k) D}_{<0} = (1 - k)(k^n + n - 2) - k + k^n = \underbrace{k^n(2 - k)}_{\leq 0} + \underbrace{(1 - k)(n - 2)}_{\leq 0} \underbrace{-k}_{<0}.$$

In the case of negative degree ($k < 1$), the results does not hold in general. For instance, if $k = -2$ then $P_1(f) = |k - 1| = 3$ and $P_2(f) = |k^2 - 1| = 3$ too, so all fixed points of f^2 are also fixed points of f and there are no period-2 orbits.

2.6 By Theorem 2.1 f_r has periodic orbits of all periods if and only if it has an orbit of period 3. This turns out to be the case when $r \geq 1 + 2\sqrt{2}$. The proof comes down to checking whether $f_r^3(x) = x$ has a solution that is not a fixed point of f_r . As f_r^3 is a polynomial of degree 8, it may not be immediately obvious how to approach this problem and in the literature there is an interesting collection of short papers on this topic. The latest one contains an elegant proof by C. Zhang, Period three begins, *Mathematics Magazine* **83** (2010), 295-297 (doi:10.4169/002557010X521859). [The proof is actually not entirely complete, although it can be completed with the information provided in the paper.]

2.7 This is a direct consequence of the continuity of f , by which $f(\tilde{x}) = f(\lim_{n \rightarrow \infty} f^n(x)) = \lim_{n \rightarrow \infty} f^{n+1}(x) = \tilde{x}$.

2.8 \subset : Let $y \in \omega(x)$ then $\exists \{n_k\}_{k \in \mathbb{N}_0}$ such that $\lim_{k \rightarrow \infty} f^{n_k}(x) = y$. Hence, $\forall n \in \mathbb{N}_0$, $\exists k_0 \in \mathbb{N}$ such that $\forall k \geq k_0$, $n_k \geq n$. As a result of this, $f^{n_k}(x) \in O_f^+(f^n(x))$ $\forall k \geq k_0$ and hence $y = \lim_{k \rightarrow \infty} f^{n_k}(x) \in \overline{O_f^+(f^n(x))}$.

\supset : Let $y \in \overline{O_f^+(f^n(x))}$ for all $n \in \mathbb{N}_0$. Then $B_{\frac{1}{k}}(y) \cap O_f^+(f^k(x)) \neq \emptyset \forall k \in \mathbb{N}$. Let $\{n_k\}_{k=n}^\infty$ be such that $f^{n_k}(x) \in B_{\frac{1}{k}}(y) \cap O_f^+(f^k(x))$, with $n_k \rightarrow \infty$ as $k \rightarrow \infty$, then $y = \lim_{k \rightarrow \infty} f^{n_k}(x)$ so $y \in \omega(x)$.

2.9 If $a = p/q \in \mathbb{Q}$ (rational), with $p, q \in \mathbb{Z}$ and $\gcd(p, q) = 1$, then all points are period- $|q|$ orbits of R_a . These periodic orbits are then also the ω -limit sets.

If $a \notin \mathbb{Q}$ (irrational), it turns out that the R_a -orbit of every initial point is dense in S^1 . As a result, there is only one ω -limit set (for all initial conditions) and this is the entire circle S^1 . To show that all orbits are dense, we first note that by virtue of the irrationality, periodic orbits cannot exist and thus every orbit consists of an infinite number of points.

To prove density of orbits, it needs to be shown that for every pair of points $x, \tilde{x} \in S^1$ and for every $\varepsilon > 0$, there exists $m \in \mathbb{N}$ such that $d^{S^1}(R_a^n(x), \tilde{x}) < \varepsilon$. By compactness of S^1 , every orbit $\{R_a^n(x)\}_{n \in \mathbb{N}}$ contains a convergent subsequence, so that for every $\varepsilon > 0$, there exist $n, m \in \mathbb{N}$, $n > m$ such that $d^{S^1}(R_a^n(x), R_a^m(x)) < \varepsilon$. As R_a is an isometry, i.e. $d^{S^1}(R_a(x), R_a(\tilde{x})) = d^{S^1}(x, \tilde{x}) \forall x, \tilde{x} \in S^1$, we also have $d^{S^1}(R_a^{n-m}(x), x) < \varepsilon$. This implies that $R_a^{n-m}(x) = R_\delta(x)$ with $|\delta| < \varepsilon$. This inequality holds uniformly for all $x \in S^1$ since $S^1 = \{b \in [0, 1) \mid R_b x\}$ and R_b is an isometry that commutes with R_a , i.e. $R_a R_b = R_b R_a$. Finally, as R_δ is a uniform rotation over the angle δ , for every $\tilde{x} \in S^1$ there exists $p \in \mathbb{N}$ such that $d^{S^1}(R_a^{p(n-m)}(x), \tilde{x}) < \varepsilon$.

2.10 As X is compact, it has a finite cover by open δ -balls for any $\delta > 0$. Let $N(k)$ denote the number of elements of a 2^{-k} -cover $\{U_{k,j}\}_{j \in \{1, \dots, N(k)\}}$. Let $\{U_{k,j}\}_{k \in \mathbb{N}, j \in \{1, \dots, N(k)\}}$ be the corresponding countably infinite collection of open balls which for convenience we relabel as $\{U_i\}_{i \in \mathbb{N}}$. We aim to construct a forward orbit that intersects every U_i . By transitivity, there exist $N_1 \in \mathbb{N}$ such that $f^{N_1}(U_1) \cap U_2 \neq \emptyset$. Let V_1 be an open ball of radius at most $\frac{1}{2}$ such that $V_1 \subset U_1 \cap f^{-N_1}(U_2)$. There exists $N_2 \in \mathbb{N}$ such that $f^{N_2}(V_1) \cap U_3$ is nonempty. Now, take an open ball V_2 of radius at most $\frac{1}{4}$ such that $V_2 \subset V_1 \cap f^{-N_2}(U_3)$. By induction, we thus construct a nested sequence of open balls V_n of radii at most 2^{-n} such that $V_{n+1} \subset V_n \cap f^{-N_{n+1}}(U_{n+2})$. The centres of these balls form a Cauchy sequence whose limit x is the unique point in the intersection $V = \bigcap_{n=1}^\infty V_n$. Then $f^{N_{n+1}}(x) \in U_{n+2}$ for every $n \in \mathbb{N}$.

If there exists $x \in X$ such that $O_f^+(x)$ is dense, then for all $U, V \in X$ open, there exists $n, m \in \mathbb{N}$ such that $f^n(x) \in U$ and $f^m(x) \in V$. If $m \geq n$ then this implies that $f^{m-n}(U) \cap V \neq \emptyset$. We now proceed that such $m \geq n$ exists if X does not contain an isolated point. Namely, suppose for some $m < n$, $f^m(x) \in V$ is not isolated. Then, by density of the forward orbit there exists a sequence of natural numbers $\{n_k\}_{k \in \mathbb{N}}$

with $n_k \rightarrow \infty$ as $k \rightarrow \infty$ such that $\lim_{k \rightarrow \infty} f^{n_k}(x) = f^m(x)$. In particular, there exists $K > 0$ such that for all $k > K$ we have $f^{n_k}(x) \in V$ and $n_k > n$.

To illustrate the necessity of the absence of isolated points, consider $[0, 1] \cup \{-\frac{1}{2}\}$ endowed with a metric induced from the Euclidean metric on \mathbb{R} . This implies, in particular, that $B_r(-\frac{1}{2}) = \{-1/2\}$ for all $r \leq \frac{1}{2}$. Let the map $f : [0, 1] \cup \{-\frac{1}{2}\} \rightarrow [0, 1] \cup \{-\frac{1}{2}\}$ be defined by $f(x) = g(|x|)$ where g is a continuous topologically transitive map of $[0, 1]$ to itself such that $O_g^+(\frac{1}{2})$ is dense in $[0, 1]$. Then f has a dense forward orbit, $O_f^+(\frac{1}{2})$, but is not topologically transitive as for every open $U \subset [0, 1]$ we have $f^n(U) \cap B_r(-\frac{1}{2}) = \emptyset$ for all $n \in \mathbb{N}$ and $r \leq \frac{1}{2}$.

2.11 As $A = \bigcap_{i \in \mathbb{N}_0} f^i(U)$ we have $f(A) = \bigcap_{i \in \mathbb{N}} f^i(U) \subset A$ since $f(\overline{U}) \subset U$ implies that $f^{n+1}(U) \subset f^n(U)$ for all $n \in \mathbb{N}_0$. On the other hand, $A \subset f(A)$ since $A = U \cap f(A)$.

2.12 It is important to note that the spectral condition on L does not imply that L is a contraction, when considering \mathbb{R}^n equipped with the Euclidean metric. However, one can make a coordinate transformation of \mathbb{R}^n so that L becomes a contraction with respect to the Euclidean metric. This is easy to see from the Jordan normal form, choosing all off-diagonal matrix elements sufficiently small. Namely, consider the invariant (generalized) eigenspace E_λ for eigenvalue λ . Then if the associated (complex) Jordan normal of L restricted to E_λ is diagonal, then the linear action of this normal form is obviously a contraction E_λ with the canonical metric if $|\lambda| < 1$. If L restricted to E_λ is not diagonalizable, on generalized eigenspaces it can be put in the following Jordan form:

$$J_{\lambda, \varepsilon} := \begin{pmatrix} \lambda & \varepsilon & & & \\ 0 & \lambda & \varepsilon & & \\ & 0 & 0 & \ddots & \ddots \\ & 0 & 0 & 0 & \lambda & \varepsilon \\ & 0 & 0 & 0 & 0 & \lambda \end{pmatrix},$$

where $\varepsilon > 0$ can be chosen to be arbitrarily small. It is an exercise to show that indeed if $|\lambda| < 1$, $J_{\lambda, \varepsilon}$ is a contraction for sufficiently small ε .

If L is a contraction, then $\lim_{m \rightarrow \infty} L^m(x) = 0$ by the contraction mapping theorem. Note that the coordinate transformation can also be considered a change of the choice of metric.

An alternative argument relies on Gelfand's formula, which states that

$$\lim_{m \rightarrow \infty} \|L^m\|^{1/m} = r(L),$$

where $\|\cdot\|$ is any matrix norm and $r(L)$ denotes the spectral radius (largest norm of any eigenvalue) of L . By Gelfand's formula and the stated spectral condition on L , implying that $0 \leq r(L) < 1$, one finds that

$$\lim_{m \rightarrow \infty} \|L^m(x)\|^{1/m} \leq \lim_{m \rightarrow \infty} \|L^m\|^{1/m} |x|^{1/m} = r(L),$$

where the matrix and vector norms, $\|\cdot\|$ and $|\cdot|$, are chosen compatibly. This implies that $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall m > N$,

$$\begin{aligned} r(L) - \varepsilon \leq \|L^m(x)\|^{1/m} \leq r(L) + \varepsilon &\Leftrightarrow (r(L) - \varepsilon)^m \leq \|L^m(x)\| \leq (r(L) + \varepsilon)^m \\ &\Rightarrow \lim_{m \rightarrow \infty} |L^m(x)| = 0 \Rightarrow \lim_{m \rightarrow \infty} L^m(x) = 0. \end{aligned}$$

If L has an eigenvalue with norm larger than one, all initial conditions x that lie in the corresponding (real) invariant subspace for this eigenvalue satisfy $\lim_{m \rightarrow \infty} |L^m(x)| = \infty$.

2.13 The first argument in the answer to Exercise 2.12 can be extended. Namely, if $Df(\tilde{x})$ is a contraction, then there exists $\varepsilon > 0$ such that f is also a contraction on $B_\varepsilon(\tilde{x})$. This can be proven using estimates involving the Taylor expansion of f at \tilde{x} . Then \tilde{x} is an attractor by virtue of the contraction mapping theorem and its basin of attraction containing $B_\varepsilon(\tilde{x})$.

2.14 Observe that if \tilde{x} has period p under f , then \tilde{x} is a fixed point of f^p . The local attractivity of \tilde{x} under f^p can be guaranteed by the eigenvalue condition of Exercise 2.13 applied to f^p . One may wonder whether this may lead to inconsistencies of conclusions about attractivity along the periodic orbit. We note that $D(f^p)(x) = Df(f^{p-1}(x)) \cdot \dots \cdot Df(x)$ and thus also for any $n \in \{1, \dots, p-1\}$ $D(f^p)(f^n(x)) = Df(f^{n+p-1}(x)) \cdot \dots \cdot Df(f^n(x))$. We note that the matrix products in both expressions are related by a cyclical order preserving permutation of the matrices in the composition. It is a classical result from linear algebra that such a permutation preserves (all trace invariants and hence also) the eigenvalues of the product. Hence, the eigenvalues of $D(f^p)(f^n(x))$ do not depend on the value of n and at each point of the periodic orbit one reaches the same conclusion.

2.15 (a) Due to compactness, for all $x \in X$ the sequence $\{f^n(x)\}_{n \in \mathbb{N}}$ has an accumulation point, so $\omega(x)$ is nonempty. Consider the sequence $\{a_n\}_{n \in \mathbb{N}}$ with

$$a_n := \text{dist}(f^n(x), \omega(x)), \forall n \in \mathbb{N},$$

and assume that this sequence does not converge to zero as $n \rightarrow \infty$. Then there exists an $\varepsilon > 0$ and a subsequence $\{a_{n_k}\}_{k \in \mathbb{N}}$ with $a_{n_k} \geq \varepsilon$ for all $k \in \mathbb{N}$. Since X is compact, we may assume without loss of generality that the sequence $\{f^{n_k}(x)\}_{k \in \mathbb{N}}$ converges with the limit

$$z := \lim_{k \rightarrow \infty} f^{n_k}(x) \in \omega(x).$$

Note that the distance function $\text{dist}(\cdot, A)$ is Lipschitz continuous with Lipschitz constant 1. We apply this fact to obtain

$$\begin{aligned} \lim_{k \rightarrow \infty} a_{n_k} &= \lim_{k \rightarrow \infty} \text{dist}(f^{n_k}(x), \omega(x)) = \text{dist}(\lim_{k \rightarrow \infty} f^{n_k}(x), \omega(x)) \\ &= \text{dist}(z, \omega(x)) = 0, \end{aligned}$$

since $z \in \omega(x)$. This contradicts the fact that $a_{n_k} \geq \varepsilon$ for all $k \in \mathbb{N}$.

It remains to be shown that $\omega(x)$ is the smallest compact set with this property. Suppose there is a compact set $A \subsetneq \omega(x)$ with the property

$$\lim_{n \rightarrow \infty} \text{dist}(f^n(x), A) = 0.$$

There exists a $y \in \omega(x) \setminus A$, and, since A is closed, an $\varepsilon > 0$ with $B_\varepsilon(y) \cap A = \emptyset$. Since $y \in \omega(x)$, we have

$$y = \lim_{k \rightarrow \infty} f^{n_k}(x)$$

for a suitably chosen sequence $\{n_k\}_{k \in \mathbb{N}}$ with $n_k \rightarrow \infty$ as $k \rightarrow \infty$. This yields

$$\varepsilon \leq \text{dist}(y, A) = \text{dist}(\lim_{k \rightarrow \infty} f^{n_k}(x), A) = \lim_{k \rightarrow \infty} \text{dist}(f^{n_k}(x), A) = 0,$$

which is a contradiction that concludes the proof.

(b) First observe that $\overline{f(\omega(x) \setminus S)} = \overline{f(\overline{\omega(x) \setminus S})}$ in the compact metric space X . Now assume for contradiction that there is a nonempty compact $S \subset \omega(x)$ with $S \cap \overline{f(\overline{\omega(x) \setminus S})} = \emptyset$. Then there are open sets G_1, G_2 such that $\overline{\omega(x) \setminus S} \subset G_1$ and $S \subset G_2$ and $\overline{f(G_1)} \cap \overline{G_2} = \emptyset$. As S and $\overline{\omega(x) \setminus S}$ are nonempty compact subsets of $\omega(x)$, we conclude from (a), above, that there is a sequence $\{n_k\}_{k \in \mathbb{N}}$ such that $f^{n_k}(x) \in G_1$ and $f^{n_k+1}(x) \in G_2$ for all $k \in \mathbb{N}$. Without loss of generality, we may assume that $f^{n_k}(x) \rightarrow \xi \in \omega(x) \cap \overline{G_1}$. Using the continuity of f this implies $f(\xi) \in \overline{G_2}$. This is a contradiction to $\overline{f(G_1)} \cap \overline{G_2} = \emptyset$.

(c) Assume that $\omega(x)$ consists of finitely many points and recall from (i) that $f(\omega(x)) = \omega(x)$. This implies the existence of a periodic orbit $C \subset \omega(x)$. Assume now that $S := \omega(x) \setminus C \neq \emptyset$. S is finite and therefore compact. As C is a periodic orbit, we obtain $S \cap \overline{f(C)} = S \cap C = \emptyset$. This contradicts (b) which proves the claim. The other direction is obvious.

- 2.16** (a) $f(x) = x + a \bmod 1$ with $a \in \mathbb{R} \setminus \mathbb{Q}$. All orbits are periodic.
 (b) $f(x) = x + a \bmod 1$ with $a \in \mathbb{Q}$. All orbits are dense.
 (c) $f(x) = T(x + \frac{1}{4}) \bmod 1$, with T defined by [3.6](#). This map has the property that $f([0, \frac{1}{2}]) = [\frac{1}{2}, 1]$ and $f([\frac{1}{2}, 1]) = [0, \frac{1}{2}]$ and hence cannot be topologically mixing. The fact that f is chaotic and in particular has sensitive dependence is best understood through the observation that f is piecewise expanding on a Markov partition, and topologically semi-conjugate to an irreducible topological Markov chain, but this kind of argument will be discussed only in Chapter [3](#), see Exercise [3.9](#).
- 2.17** Let D be the diameter of X . Let $x \in X$. In the absence of isolated points, $\forall \varepsilon > 0$, \exists a pair $y, z \in X$ such that $d^X(y, z) \geq D - \varepsilon$. By the assumption of topological mixing, $\forall \delta > 0$, $\exists N \geq 0$ and points $\tilde{y}, \tilde{z} \in B_\varepsilon(x)$ such that $f^N(\tilde{y}) \in B_\delta(y)$ and $f^N(\tilde{z}) \in B_\delta(z)$. Then

$$\begin{aligned} D - \varepsilon \leq d^X(y, z) &\leq d^X(f^N(\tilde{y}), f^N(\tilde{z})) + 2\delta \\ &\leq d^X(f^N(\tilde{y}), x) + d^X(f^N(\tilde{z}), x) + 2\delta, \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow d^X(f^N(\tilde{y}), x) + d^X(f^N(\tilde{z}), x) \geq D - \varepsilon - 2\delta, \\ &\Rightarrow d^X(f^N(\tilde{y}), x) \geq \frac{D - \varepsilon - 2\delta}{2} \quad \text{or} \quad d^X(f^N(\tilde{z}), x) \geq \frac{D - \varepsilon - 2\delta}{2}. \end{aligned}$$

- 2.18** $d_n^X(x, \tilde{x}) = d_n^X(\tilde{x}, x)$ since $d^X(x, \tilde{x}) = d^X(\tilde{x}, x)$. $d_n^X(x, \tilde{x}) = 0$ if and only if $x = \tilde{x}$ since $d^X(x, \tilde{x}) = 0$ if and only if $x = \tilde{x}$. Finally, using the triangle inequality for d^X we find for any $x, \tilde{x}, \hat{x} \in X$,

$$\begin{aligned} d_n^X(x, \hat{x}) &= \max_{0 \leq k \leq n-1} d^X(f^k(x), f^k(\hat{x})) \\ &\leq \max_{0 \leq k \leq n-1} (d^X(f^k(x), f^k(\tilde{x})) + d^X(f^k(\tilde{x}), f^k(\hat{x}))), \\ &\leq \max_{0 \leq k \leq n-1} d^X(f^k(x), f^k(\tilde{x})) + \max_{0 \leq \tilde{k} \leq n-1} d^X(f^{\tilde{k}}(\tilde{x}), f^{\tilde{k}}(\hat{x})) \\ &= d_n^X(x, \tilde{x}) + d_n^X(\tilde{x}, \hat{x}). \end{aligned}$$

It follows directly from the definitions that $d_1^X = d$ and that $d_{n+1}^X(x, \tilde{x}) \geq d_n^X(x, \tilde{x})$ for all $x, \tilde{x} \in X$.

- 2.19** If X is compact then for all $\varepsilon > 0$ there exists a finite cover of X by ε -balls. If X is compact when endowed with the metric d^X , it is compact with respect to the metric d_n^X for all $n > 1$. Namely, by continuity of f , whenever $\lim_{m \rightarrow \infty} d^X(x_m, y) = 0$ then also $\lim_{m \rightarrow \infty} d_n^X(x_m, y) = 0$. (The metrics d^X and d_n^X are *topologically equivalent*.) The finite cover of X by ε -balls in the d_n^X metric is an (n, ε) -spanning set. As X is covered by a finite number, say $N(\varepsilon)$, ε -balls in the d_n^X metric, then it is impossible for an (n, ε) -separated set to have a cardinality larger than or equal to $N(\varepsilon)$.

- 2.20** If $k > 2$, basically, the same arguments apply with the number 10 replaced by k . To determine the topological entropy of E_{-k} with $k > 2$, it is useful to note

that $E_{-k}^2 = E_{k^2}$. As $k^2 > 2$ it follows that $h_{\text{top}}(E_{-k}^2) = 2 \ln |k|$. From the definition of topological entropy, it follows that $h_{\text{top}}(f^m) = m h_{\text{top}}(f)$ for any $m \in \mathbb{N}$. Hence, $h_{\text{top}}(E_k) = h_{\text{top}}(E_{-k}) = \ln k$ for all $k > 2$.

2.21 If f is an isometry, then $d^X = d_n^X$. This means that the $\text{span}(n, \varepsilon, f)$ does not depend on n . In turn this implies, directly from Definition **2.11**, that $h_{\text{top}}(f) = 0$.

2.22 If there exists a homeomorphism $h : X \rightarrow Y$ such that $f = h^{-1} \circ g \circ h$ then h maps orbits of f on X to orbits of g on Y . This mapping (and its inverse) are one-to-one and continuous. To prove Proposition **2.3**

- For all open $U_1, U_2 \subset X$, there exists $n \in \mathbb{N}_0$ such that $f^n(U_1) \cap U_2 \neq \emptyset$, if and only if for all open $V_1, V_2 \subset Y$ there exists $n \in \mathbb{N}_0$ such that $g^n(V_1) \cap V_2 \neq \emptyset$. This follows from two observations. First, the conjugacy between f and g implies that

$$f^n(U_1) \cap U_2 \neq \emptyset \Leftrightarrow g^n(V_1) \cap V_2 \neq \emptyset$$

with $V_1 = h(U_1)$ and $V_2 = h(U_2)$. Second, h establishes a one-to-one relationship between the set of open subsets of X and the set of open subsets of Y .

- h establishes a one-to-one relationship between periodic orbits of f and g with the same period. Using in addition the same facts highlighted in the previous point, every open subset of X contains a periodic point if and only if every open subset Y contains a periodic point.
- The argument follows in analogy to the answer given concerning the preservation of topological transitivity.
- For $f : X \rightarrow X$ to be chaotic it is necessary and sufficient that f is topologically transitive, has dense periodic orbits and X is not a single periodic orbit of f . Having shown the preservation of the first two properties under topological conjugacy, it only remains to observe that indeed X consists of a single periodic orbit of f if and only if Y consists of a single periodic orbit of g .

The proof of Proposition **2.4** follows directly from the proof of Theorem **2.6** as in the case of a topological conjugacy, f is a factor of g and g is also a factor of f . This implies that at the same time $h_{\text{top}}(f) \leq h_{\text{top}}(g)$ and $h_{\text{top}}(f) \geq h_{\text{top}}(g)$.

2.23 Let $M_1 : \mathbb{R} \rightarrow [0, 1) \simeq S^1$ be defined as $M_1(x) = x \bmod 1$. Then the defining relationship between a circle map f and its lift F , $F(x) \bmod 1 = f(x \bmod 1)$, can be written as $M_1 \circ F = f \circ M_1$, confirming that f is a factor of F .

2.24 We show that $C_{s_0 \dots s_{m-1}} = B_{3^{-m+1}}(s)$ for all $s \in C_{s_0 \dots s_{m-1}}$. Suppose $s, t \in C_{s_0 \dots s_{m-1}}$, then

$$d^{\Sigma^+}(s, t) = \sum_{i \geq 0} \frac{\delta(s_i, t_i)}{3^i} = \sum_{i \geq m} \frac{\delta(s_i, t_i)}{3^i} \leq \sum_{i \geq m} 3^{-i} = 3^{-m} \sum_{i \geq 0} 3^{-i} = \frac{3^{-m}}{1 - 3^{-1}} < 3^{-m+1}.$$

On the other hand, if $s \in C_{s_0 \dots s_{m-1}}$ and $t \notin C_{s_0 \dots s_{m-1}}$, with $j \in \{0, \dots, m-1\}$ least such that $s_j \neq t_j$, then

$$d^{\Sigma^+}(s, t) \geq \frac{\delta(s_j, t_j)}{3^j} = 3^{-j} \geq 3^{-m+1}.$$

Continuity of $h : \Sigma_k^+ \rightarrow S^1$ follows from the observation that

$$|h(C_{s_0 \dots s_{m-1}})| = 10^{-m},$$

and thus that $\forall \varepsilon > 3^{-m} > 0$ and all $s \in \Sigma_k^+$ there exists $\delta = 10^{-m} > 0$ such that $h(B_\delta(s)) \subset B_\varepsilon(h(s))$.

2.25 Let $h : X \rightarrow Y$ denote the continuous map that topologically semi-conjugates $f : X \rightarrow X$ to $g : Y \rightarrow Y$. Then

- h maps periodic orbits of f to periodic orbits of g (but not necessarily with the same period). It also maps surjectively the set of open subsets of X to the set of open subsets of Y . Hence, if every open subset of X contains a periodic point for f , then every open subset of Y contains a periodic point for g .
- If for all open $U_1, U_2 \subset X$, there exists $n \in \mathbb{N}_0$ such that $f^n(U_1) \cap U_2 \neq \emptyset$, then for all open $V_1, V_2 \subset Y$ there exists $m \in \mathbb{N}_0$ such that $g^m(V_1) \cap V_2 \neq \emptyset$. This follows from two observations. First, the semi-conjugacy between f and g implies that

$$f^n(U_1) \cap U_2 \neq \emptyset \Leftrightarrow g^n(V_1) \cap V_2 \neq \emptyset$$

with $V_1 = h(U_1)$ and $V_2 = h(U_2)$. Second, h maps the set of open subsets of X surjectively to the set of open subsets of Y .

- Preservation of topological mixing is proved in analogy to topological transitivity.
- For $f : X \rightarrow X$ to be chaotic it is necessary and sufficient that f is topologically transitive, has dense periodic orbits and X is not a single periodic orbit of f . Having shown the preservation of the first two properties under topological semi-conjugacy, it suffices to observe the fact that indeed if X consists of a single periodic orbit of f then also Y consists of a single periodic orbit of g .

2.26 The analysis of E_{10} extends to E_k for all integer $k > 1$, basically by replacing any mention of the number 10 by k . The corresponding symbolic representation of points in $[0, 1)$ is the k -nary expansion: $x = 0.x_0x_1x_2\dots = \sum_{i=0}^{\infty} x_i k^{-1-i}$.

2.27 The finite set of all fixed points of f^n is f -invariant and hence there exists $\varepsilon(n) > 0$ (namely the minimum distance between any two points in this set) such that this set is (n, ε) -separated for all $\varepsilon \leq \varepsilon(n)$. This means that $\text{sep}(n, \varepsilon, f) \geq P_n(f)$ for all $\varepsilon \leq \varepsilon(n)$. Hence, by application of Theorem **2.5**, it follows that

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

3.1 It suffices to establish the Lipschitz condition:

$$\begin{aligned}
 d_{n,L}^X(f(x), f(\tilde{x})) &= \sum_{m=1}^n L^{-(k-1)/n} d^X(f^m(x), f^m(\tilde{x})) \\
 &= L^{-(n-1)/n} d^X(f^n(x), f^n(\tilde{x})) + L^{1/n} \sum_{m=1}^{n-1} L^{-(k-1)/n} d^X(f^k(x), f^k(\tilde{x})) \\
 &\geq L^{-1+1/n} L d^X(x, \tilde{x}) + L^{1/n} \sum_{m=1}^{n-1} L^{-(k-1)/n} d^X(f^k(x), f^k(\tilde{x})) \\
 &= L^{1/n} d_{n,L}^X(x, \tilde{x}).
 \end{aligned}$$

3.2 If $|f'(x)| > 1$ for all $x \in I$, then due to compactness of I , there exists $C \geq 1$ such that $|f'(x)| > C$ for all $x \in I$. Then by the mean value theorem and continuity of f , there exists $\varepsilon > 0$ such that for all $\tilde{x} \in B_\varepsilon(x)$, $|f(x) - f(\tilde{x})| \geq C|x - \tilde{x}|$, so that f satisfies the condition in Definition 3.1 with $L = C$.

If $|f'(x)| \leq 1$ then $|f(x) - f(x + \delta)| = |f'(x)\delta| + O(\delta^2)$ and there is no $L > 1$ and $\varepsilon > 0$ such that $|f'(x)\delta| + O(\delta^2) \geq L\delta$ for all $|\delta| < |\varepsilon|$ since $(|f'(x)\delta| + O(\delta^2))/\delta \rightarrow |f'(x)| \leq 1 < L$ provides a contradiction.

3.3 T is piecewise expanding on the Markov partition $\mathcal{R} = \{(0, \frac{1}{2}), (\frac{1}{2}, 1)\}$. As $T((0, \frac{1}{2})) = T((\frac{1}{2}, 1)) = [0, 1]$ we find that f is topologically semi-conjugate to the (full) shift on Σ_2^+ . The fact that $\#h^{-1}(x) \leq 2$ follows from the one-dimensional construction, as explained in Example 3.6

3.4 If $k > 1$, $E_k, E_{-k} : S^1 \rightarrow S^1$ are piecewise expanding on the Markov partition $\mathcal{R} = \{(0, \frac{1}{k}), \dots, (\frac{k-1}{k}, 1)\}$ and $E_{\pm k}((\frac{i}{k}, \frac{i+1}{k})) = S^1$ for all $i \in \{0, \dots, k-1\}$, so we find that E_k and E_{-k} are topologically semi-conjugate to the (full) shift on Σ_k^+ . E_k and E_{-k} are not topologically conjugate to each other due to the fact that set of ambiguous coding sequences is not identical. This for instance leads to the fact that for all $k > 2$ and odd values of n , $P_n(E_k) = k^n - 1 \neq P_n(E_{-k}) = k^n + 1$, cf. Example 2.2

3.5 Let f_1 and f_2 be expanding circle maps of degree k , then there exist continuous and surjective $h_1, h_2 : \Sigma_k^+ \rightarrow S^1$ such that $f_i \circ h_i = h_i \circ \sigma$ for $i = 1, 2$. We now observe that while $h_i^{-1}(x)$ is potentially multi-valued, the geometrical construction of the partition and its refinements are such that for each pair of expanding circle maps of the same degree k , the same coding ambiguities arise. Hence, the compositions $h_1 \circ h_2^{-1}, h_2 \circ h_1^{-1} : S^1 \rightarrow S^1$ are well-defined and

$$h \circ f_1 = f_2 \circ h,$$

with $h = h_2 \circ h_1^{-1}$. h is invertible since $h^{-1} = h_1 \circ h_2^{-1}$. h and h^{-1} are continuous by continuity of h_1, h_2 and the continuity of h_1^{-1}, h_2^{-1} on each of their (potentially two) branches. Hence, every continuous expanding circle map of degree k is topologically conjugate to E_k .

In addition, as has been shown in Exercise 3.4, expanding circle maps can only be topologically conjugate if they have equal degree.

3.6 (i) We prove this by induction. The statement is true by the definition of A if $m = 1$. Suppose it is true for m so that $(A^m)_{ij}$ represents the number of distinct paths of length m from vertex i to vertex j . Then

$$(A^{m+1})_{ij} = (A \cdot A^m)_{ij} = \sum_n A_{in}(A^m)_{nj} = \sum_{A_{in} \neq 0} (A^m)_{nj}.$$

The resulting is the sum over n of the number of distinct paths between vertex n to vertex j in m steps, $(A^m)_{nj}$, conditioned on whether there exists a one-step path from vertex i to vertex n . This is precisely the number of distinct paths between vertex i and vertex j .

(ii) The number of distinct paths of length m from a vertex to itself is represented by a diagonal element $(A^m)_{ii}$. The total number of such paths is thus the sum $\sum_i (A^m)_{ii} =: \text{Tr}(A^m)$.

3.7 If a topological Markov chain $\Sigma_{k,A}^+$ is not irreducible then for some $i, j \in \{0, \dots, k-1\}$ we have $(A^m)_{ij} = 0$ for all $m \in \mathbb{N}$. This implies that for the pair of non-empty cylinder sets $U = C_i$ and $V = C_j$, $\sigma^q(U) \cap V = \emptyset$ for all $q \in \mathbb{N}$. Vice-versa, if σ is transitive then for any pair of cylinder sets C_i and C_j then there exists $m \in \mathbb{N}$ such that $\sigma^m(C_i) \cap C_j \neq \emptyset$, implying that $(A^m)_{ij} \neq 0$.

3.8 The connectivity matrix A in (3.14) satisfies

$$A^2 = \begin{pmatrix} 3 & 3 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 7 & 7 & 6 & 6 \\ 2 & 2 & 4 & 4 \\ 5 & 5 & 2 & 2 \\ 5 & 5 & 2 & 2 \end{pmatrix}. \quad (4.2)$$

Hence A is primitive, so that σ_A is chaotic and topologically mixing. $h_{\text{top}}(\sigma_A) = h_{\text{top}}(f) = r(A) = \ln \left(\frac{1+\sqrt{17}}{2} \right)$. (As computed using Wolfram Alfa.)

3.9 The partition $\mathcal{R} = \left\{ \left(0, \frac{1}{4}\right), \left(\frac{1}{4}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{3}{4}\right), \left(\frac{3}{4}, 1\right) \right\}$ is a piecewise expanding Markov partition for S and the corresponding connectivity matrix A of the Markov chain is (3.17) of Example 3.9. As A is irreducible but not primitive, σ_A is chaotic but not topologically mixing and the same holds for S . The topological entropy is given by $h_{\text{top}}(S) = h_{\text{top}}(\sigma_A) = \ln r(A) = \ln 2$.

3.10 It is crucial to note that the partition (3.9) satisfies the condition of a Markov partition, with connectivity matrix (and Markov graph) as in Fig. 3.3. However, f is not piecewise expanding on this partition since it is not expanding on $R_2 = \left(\frac{2}{3}, 1\right)$. However, a closer examination reveals that the lack of expansion in R_2 is overcome by subsequent expansion in R_0 , so that the elements of the refinements of the partition still exponentially shrink. In fact, f^2 is piecewise expanding on the Markov partition refinement \mathcal{R}_2 , see Fig. 4.1.

The exponential shrinking of labelling intervals is crucial for the convergence of labelling intervals to points and the existence of the function h that semi-conjugates

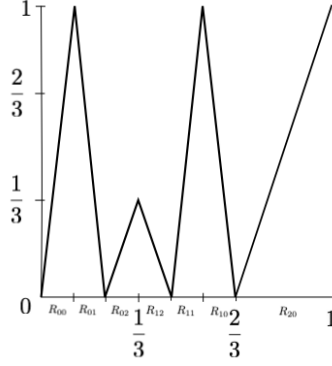


Fig. 4.1: Graph of f^2 (3.21) and refined Markov partition \mathcal{R}_2 , as discussed in the solution to Exercise 3.10.

f to a subshift of finite type on $\Sigma_{3,A}^+$. In particular $|R_{i_0 \dots i_{n-1}}| \leq \sqrt{3}^{-n}$ for any $\Sigma_{3,A}^+$ -admissible sequence $i_0 \dots i_{n-1}$. Hence, f is topologically semi-conjugate to the topological Markov chain $\Sigma_{3,A}^+$. As A is primitive since A^2 has no zero entries, the shift map on the topological Markov chain is chaotic and topologically mixing and so is f . Since $\#(h^{-1}(x)) \leq 2$ for all $x \in [0, 1]$ (as by the arguments provided in Example 3.6), $h_{\text{top}}(f) = h_{\text{top}}(\sigma_A) = \ln r(A) = \ln(1 + \sqrt{2})$.

3.11 The graphs of the maps under consideration are reproduced in Fig. 4.2. Left graph: There exists a piecewise expanding Markov partition $\mathcal{R} = \{R_i\}_{i=0}^4$ as depicted in the graph. The corresponding connectivity matrix A is given by

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}. \quad (4.3)$$

A is not irreducible since for all $m \geq 1$

$$(A^m)_{20} = (A^m)_{21} = (A^m)_{30} = (A^m)_{31} = (A^m)_{40} = (A^m)_{41} = 0,$$

reflecting the fact that points in $(\frac{1}{2}, 1)$ are never mapped into $(0, \frac{1}{2})$. Hence the map is not transitive (and not chaotic or topologically mixing). The topological entropy is positive and given by $\ln r(A) = \ln 2$.

It is possible to be mapped from $[0, \frac{1}{2})$ into $(\frac{1}{2}, 1)$, namely precisely when an orbit lands in $[0, 0.05)$. In order to study this leakage, we can use the partition $\{R_0, R_1\}$ of $U = [0, \frac{1}{2})$ to study the orbits that do not escape. It is readily found that the non-escaping set $N(U)$ (3.15) is a Cantor set on which the dynamics is topologically conjugate to a full shift on two symbols, with connectivity matrix

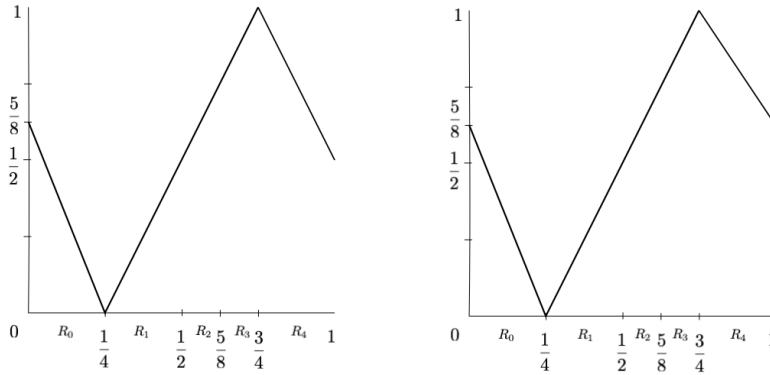


Fig. 4.2: Graphs of maps of $[0, 1]$ with Markov partition, for Exercise 3.11.

$$A_l = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad (4.4)$$

and topological entropy $\ln r(A_l) = \ln 2$.

So almost all orbits that start in $[0, 1]$, namely $[0, 1] \setminus N(U)$, eventually end up in $[\frac{1}{2}, 1]$. $[\frac{1}{2}, 1]$ is an invariant set and the dynamics on this set can be studied using the Markov partition $\{R_2, R_3, R_4\}$, on which the connectivity matrix reads

$$A_r = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \quad (4.5)$$

It turns out that A_r^2 has no zero entries, so A_r is primitive and the dynamics on $[\frac{1}{2}, 1]$ is chaotic and topologically mixing. The topological entropy on $[\frac{1}{2}, 1]$ is given by $\ln r(A_r) = \ln 2$. (This result can also be obtained using the simpler piecewise expanding Markov partition $\{(\frac{1}{2}, \frac{3}{4}), (\frac{3}{4}, 1)\}$ for $[\frac{1}{2}, 1]$, which directly yields a full shift on two symbols.)

We note that the equality $r(A_l) = r(A_r) = r(A)$ is coincidental in this example.

Finally, we note that although $[\frac{1}{2}, 1]$ attracts almost all orbits, it is not an attractor in the sense of Definition 2.4 as $[\frac{1}{2}, 1]$ does not attract an open neighbourhood since $N(U)$ contains points that are arbitrarily close to $\frac{1}{2}$. This explains also the fact that there are also weaker definitions of attractors, cf. footnote 4, that would allow $[\frac{1}{2}, 1]$ to be classified as one.

Right graph: The map in the right graph is very similar to the one on the left and has the same piecewise expanding Markov partition. The only difference is that the last element of the partition is mapped to the final two rather than the final three partition elements. Hence, the corresponding connectivity matrix for the Markov chain is given by

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}. \quad (4.6)$$

The Markov chain is reducible for the same reason as mentioned above and hence the map is not transitive (and not chaotic or topologically mixing). The topological entropy is given by $\ln r(A) = \ln 2$. The non-escaping subset of $[0, \frac{1}{2}]$ is the same as discussed above since the dynamics restricted to $[0, \frac{1}{2}]$ is unaltered.

However, the right map now maps $(\frac{1}{2}, 1]$ into $[\frac{5}{8}, 1]$ which is an invariant set and an attractor in the sense of Definition 2.4, with trapping region $\tilde{U} = (\frac{9}{16}, 1]$ (for instance), the closure of which is directly mapped to $[\frac{5}{8}, 1]$. It follows that the basin of attraction of this attractor $B([\frac{5}{8}, 1]) = [0, 1] \setminus N(U)$.

The attractor $[\frac{5}{8}, 1]$ has a piecewise expanding Markov partition $\{R_3, R_4\}$ for which the connectivity matrix is given by

$$A_r = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}. \quad (4.7)$$

Since A^2 has no zero entries, A is primitive and the dynamics on the attractor is chaotic and topologically mixing. The topological entropy on the attractor is given by $\ln r(A_r) = \ln \left(\frac{1}{2}(1 + \sqrt{5}) \right)$.

We observe that $r(A) = r(A_l) > r(A_r)$. Indeed, the structure of this system implies that $r(A) = \max\{r(A_l), r(A_r)\}$. This illustrates the fact that the topological entropy is not necessarily determined by the attractor $[\frac{5}{8}, 1]$ that attracts almost all orbits and thus also not necessarily a measure of the typical long-term complexity of the dynamics. Dynamically unstable sets equally contribute to the topological entropy.²

² An example may be constructed of a continuous map of the interval with positive topological entropy for which almost every orbit converges to a fixed point.

3.6 (i) We prove this by induction. The statement is true by the definition of A if $m = 1$. Suppose it is true for m so that $(A^m)_{ij}$ represents the number of distinct paths of length m from vertex i to vertex j . Then

$$(A^{m+1})_{ij} = (A \cdot A^m)_{ij} = \sum_n A_{in}(A^m)_{nj} = \sum_{A_{in} \neq 0} (A^m)_{nj}.$$

The resulting is the sum over n of the number of distinct paths between vertex n to vertex j in m steps, $(A^m)_{nj}$, conditioned on whether there exists a one-step path from vertex i to vertex n . This is precisely the number of distinct paths between vertex i and vertex j .

(ii) The number of distinct paths of length m from a vertex to itself is represented by a diagonal element $(A^m)_{ii}$. The total number of such paths is thus the sum $\sum_i (A^m)_{ii} =: \text{Tr}(A^m)$.

3.7 If a topological Markov chain $\Sigma_{k,A}^+$ is not irreducible then for some $i, j \in \{0, \dots, k-1\}$ we have $(A^m)_{ij} = 0$ for all $m \in \mathbb{N}$. This implies that for the pair of non-empty cylinder sets $U = C_i$ and $V = C_j$, $\sigma^q(U) \cap V = \emptyset$ for all $q \in \mathbb{N}$. Vice-versa, if σ is transitive then for any pair of cylinder sets C_i and C_j then there exists $m \in \mathbb{N}$ such that $\sigma^m(C_i) \cap C_j \neq \emptyset$, implying that $(A^m)_{ij} \neq 0$.

3.8 The connectivity matrix A in (3.14) satisfies

$$A^2 = \begin{pmatrix} 3 & 3 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{pmatrix}, \quad A^3 = \begin{pmatrix} 7 & 7 & 6 & 6 \\ 2 & 2 & 4 & 4 \\ 5 & 5 & 2 & 2 \\ 5 & 5 & 2 & 2 \end{pmatrix}. \quad (4.2)$$

Hence A is primitive, so that σ_A is chaotic and topologically mixing. $h_{\text{top}}(\sigma_A) = h_{\text{top}}(f) = r(A) = \ln \left(\frac{1+\sqrt{17}}{2} \right)$. (As computed using Wolfram Alfa.)

3.9 The partition $\mathcal{R} = \left\{ \left(0, \frac{1}{4}\right), \left(\frac{1}{4}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{3}{4}\right), \left(\frac{3}{4}, 1\right) \right\}$ is a piecewise expanding Markov partition for S and the corresponding connectivity matrix A of the Markov chain is (3.17) of Example 3.9. As A is irreducible but not primitive, σ_A is chaotic but not topologically mixing and the same holds for S . The topological entropy is given by $h_{\text{top}}(S) = h_{\text{top}}(\sigma_A) = \ln r(A) = \ln 2$.

3.10 It is crucial to note that the partition (3.9) satisfies the condition of a Markov partition, with connectivity matrix (and Markov graph) as in Fig. 3.3. However, f is not piecewise expanding on this partition since it is not expanding on $R_2 = \left(\frac{2}{3}, 1\right)$. However, a closer examination reveals that the lack of expansion in R_2 is overcome by subsequent expansion in R_0 , so that the elements of the refinements of the partition still exponentially shrink. In fact, f^2 is piecewise expanding on the Markov partition refinement \mathcal{R}_2 , see Fig. 4.1.

The exponential shrinking of labelling intervals is crucial for the convergence of labelling intervals to points and the existence of the function h that semi-conjugates

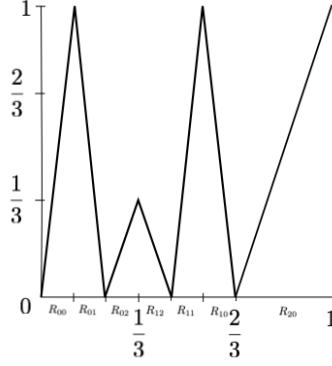


Fig. 4.1: Graph of f^2 (3.21) and refined Markov partition \mathcal{R}_2 , as discussed in the solution to Exercise 3.10.

f to a subshift of finite type on $\Sigma_{3,A}^+$. In particular $|R_{i_0 \dots i_{n-1}}| \leq \sqrt{3}^{-n}$ for any $\Sigma_{3,A}^+$ -admissible sequence $i_0 \dots i_{n-1}$. Hence, f is topologically semi-conjugate to the topological Markov chain $\Sigma_{3,A}^+$. As A is primitive since A^2 has no zero entries, the shift map on the topological Markov chain is chaotic and topologically mixing and so is f . Since $\#(h^{-1}(x)) \leq 2$ for all $x \in [0, 1]$ (as by the arguments provided in Example 3.6), $h_{\text{top}}(f) = h_{\text{top}}(\sigma_A) = \ln r(A) = \ln(1 + \sqrt{2})$.

3.11 The graphs of the maps under consideration are reproduced in Fig. 4.2. Left graph: There exists a piecewise expanding Markov partition $\mathcal{R} = \{R_i\}_{i=0}^4$ as depicted in the graph. The corresponding connectivity matrix A is given by

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}. \quad (4.3)$$

A is not irreducible since for all $m \geq 1$

$$(A^m)_{20} = (A^m)_{21} = (A^m)_{30} = (A^m)_{31} = (A^m)_{40} = (A^m)_{41} = 0,$$

reflecting the fact that points in $(\frac{1}{2}, 1)$ are never mapped into $(0, \frac{1}{2})$. Hence the map is not transitive (and not chaotic or topologically mixing). The topological entropy is positive and given by $\ln r(A) = \ln 2$.

It is possible to be mapped from $[0, \frac{1}{2})$ into $(\frac{1}{2}, 1)$, namely precisely when an orbit lands in $[0, 0.05)$. In order to study this leakage, we can use the partition $\{R_0, R_1\}$ of $U = [0, \frac{1}{2})$ to study the orbits that do not escape. It is readily found that the non-escaping set $N(U)$ (3.15) is a Cantor set on which the dynamics is topologically conjugate to a full shift on two symbols, with connectivity matrix

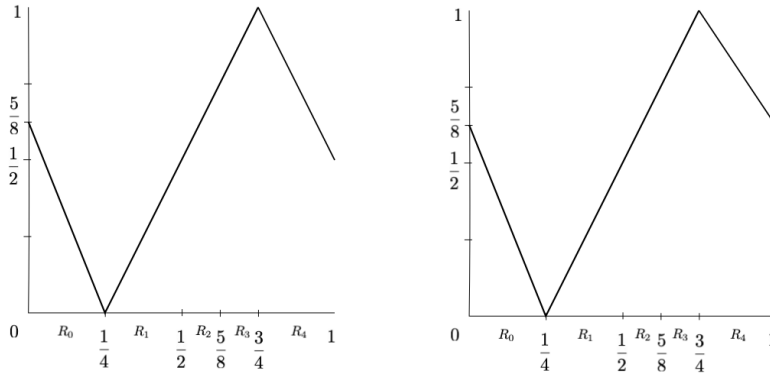


Fig. 4.2: Graphs of maps of $[0, 1]$ with Markov partition, for Exercise 3.11.

$$A_l = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad (4.4)$$

and topological entropy $\ln r(A_l) = \ln 2$.

So almost all orbits that start in $[0, 1]$, namely $[0, 1] \setminus N(U)$, eventually end up in $[\frac{1}{2}, 1]$. $[\frac{1}{2}, 1]$ is an invariant set and the dynamics on this set can be studied using the Markov partition $\{R_2, R_3, R_4\}$, on which the connectivity matrix reads

$$A_r = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \quad (4.5)$$

It turns out that A_r^2 has no zero entries, so A_r is primitive and the dynamics on $[\frac{1}{2}, 1]$ is chaotic and topologically mixing. The topological entropy on $[\frac{1}{2}, 1]$ is given by $\ln r(A_r) = \ln 2$. (This result can also be obtained using the simpler piecewise expanding Markov partition $\{(\frac{1}{2}, \frac{3}{4}), (\frac{3}{4}, 1)\}$ for $[\frac{1}{2}, 1]$, which directly yields a full shift on two symbols.)

We note that the equality $r(A_l) = r(A_r) = r(A)$ is coincidental in this example.

Finally, we note that although $[\frac{1}{2}, 1]$ attracts almost all orbits, it is not an attractor in the sense of Definition 2.4 as $[\frac{1}{2}, 1]$ does not attract an open neighbourhood since $N(U)$ contains points that are arbitrarily close to $\frac{1}{2}$. This explains also the fact that there are also weaker definitions of attractors, cf. footnote 4, that would allow $[\frac{1}{2}, 1]$ to be classified as one.

Right graph: The map in the right graph is very similar to the one on the left and has the same piecewise expanding Markov partition. The only difference is that the last element of the partition is mapped to the final two rather than the final three partition elements. Hence, the corresponding connectivity matrix for the Markov chain is given by

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}. \quad (4.6)$$

The Markov chain is reducible for the same reason as mentioned above and hence the map is not transitive (and not chaotic or topologically mixing). The topological entropy is given by $\ln r(A) = \ln 2$. The non-escaping subset of $[0, \frac{1}{2}]$ is the same as discussed above since the dynamics restricted to $[0, \frac{1}{2}]$ is unaltered.

However, the right map now maps $(\frac{1}{2}, 1]$ into $[\frac{5}{8}, 1]$ which is an invariant set and an attractor in the sense of Definition 2.4, with trapping region $\tilde{U} = (\frac{9}{16}, 1]$ (for instance), the closure of which is directly mapped to $[\frac{5}{8}, 1]$. It follows that the basin of attraction of this attractor $B([\frac{5}{8}, 1]) = [0, 1] \setminus N(U)$.

The attractor $[\frac{5}{8}, 1]$ has a piecewise expanding Markov partition $\{R_3, R_4\}$ for which the connectivity matrix is given by

$$A_r = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}. \quad (4.7)$$

Since A^2 has no zero entries, A is primitive and the dynamics on the attractor is chaotic and topologically mixing. The topological entropy on the attractor is given by $\ln r(A_r) = \ln \left(\frac{1}{2}(1 + \sqrt{5}) \right)$.

We observe that $r(A) = r(A_l) > r(A_r)$. Indeed, the structure of this system implies that $r(A) = \max\{r(A_l), r(A_r)\}$. This illustrates the fact that the topological entropy is not necessarily determined by the attractor $[\frac{5}{8}, 1]$ that attracts almost all orbits and thus also not necessarily a measure of the typical long-term complexity of the dynamics. Dynamically unstable sets equally contribute to the topological entropy.²

² An example may be constructed of a continuous map of the interval with positive topological entropy for which almost every orbit converges to a fixed point.

4.1 We have the setting of continuous maps on a compact metric space S . Such maps are Borel-measurable. We denote the relevant Borel σ -algebra $\mathcal{B}(X)$.

- If $f(x) = x$ then $f^{-1}(x) \ni x$. Hence, if $x \in A \in \mathcal{B}(X)$, then $f_*\delta_x(A) := \delta_x(f^{-1}(A)) = 1 = \delta_x(A)$. Moreover, if $x \notin A \in \mathcal{B}(X)$, then $x \notin f^{-1}(A) \in \mathcal{B}(X)$, in which case $f_*\delta_x(A) := \delta_x(f^{-1}(A)) = 0 = \delta_x(A)$. So, indeed $f_*\delta_x = \delta_x$ if $f(x) = x$.
- Let $P = \{x, f(x), \dots, f^{p-1}(x)\}$ with $f^p(x) = x$. We note that $P \in \mathcal{B}(X)$. With $\mu = \sum_{x \in P} \delta_x$, we observe that $\mu(A) = \#(P \cap A)$ for all $A \in \mathcal{B}(X)$. The key observation is that

$$\#(P \cap A) = \#(P \cap f^{-1}(A)).$$

Hence $f_*\mu(A) := \mu(f^{-1}(A)) = \#(P \cap f^{-1}(A)) = \#(P \cap A) = \mu(A)$.

- It suffices to verify the invariance on the semi-ring of half-open intervals I on the circle. As rigid rotations are isometries, we have $|R_a(I)| = |I|$ for all $a \in [0, 1)$. Since $\lambda(I) = |I|$, it thus follows that $(R_a)_*\lambda(I) = \lambda(R_{-a}(I)) = |R_{-a}I| = |I| = \lambda(I)$. Invariance on the semi-ring implies invariance on $\mathcal{B}(S^1)$.

The demonstration that λ is the unique invariant measure requires some more sophisticated technical results from analysis. Let μ be a (possibly different) invariant probability measure. Then with $e_k(x) := \exp(2\pi i k x)$, $k \in \mathbb{Z}$,

$$\int_{S^1} e_k d\mu = 0.$$

Namely, since μ is R_a -invariant, we have

$$\int_{S^1} e_k d\mu = \int_{S^1} e_k \circ R_a d\mu = \exp(2\pi i k a) \int_{S^1} e_k d\mu,$$

and since $a \in \mathbb{R} \setminus \mathbb{Q}$, $\exp(2\pi i k a) \neq 1$ if $k \neq 0$.

Now let f have Fourier series $\sum_{k \in \mathbb{Z}} a_k e_k$ so that $a_0 = \int_f d\lambda$. Let $f_n := \sum_{|k| \leq n} a_k e_k$ then, by Fejér's theorem¹⁰, $f_n \rightarrow f$ uniformly as $n \rightarrow \infty$. Hence

$$\lim_{n \rightarrow \infty} \int_{S^1} f_n d\mu = \int_{S^1} f d\mu.$$

However, also

$$\int_{S^1} f_n d\mu = a_0 = \int_{S^1} f d\lambda.$$

Hence $\int_{S^1} f d\mu = \int_{S^1} f d\lambda$ for every continuous f . The latter implies via Riesz's representation theorem¹¹ that $\mu = \lambda$.

- It suffices to verify the invariance on the semi-ring of half-open sub-intervals of the closures of the partition elements I_i . Let $A \in I_i$ be such an interval. Then $f^{-1}(A) = \cup_{i=0}^{n-1} f_i^{-1}(A)$ so that

¹⁰ L. Fejér. Untersuchungen über Fouriersche Reihen. *Math. Annalen* **58** (1904), 51–69.

¹¹ F. Riesz. Sur les opérations fonctionnelles linéaires. *C. R. Acad. Sci. Paris*. **149** (1909), 974–977.

$$\lambda(f^{-1}(A)) = \sum_{i=0}^{n-1} \lambda(f_i^{-1}(A)) = \sum_{i=0}^{n-1} |I_i| \lambda(A) = \lambda(A).$$

This result extends to all half-open intervals in the semi-ring, which in turn implies the unique extension to the Borel σ -algebra.

4.2 Let R denote an irrational circle rotation. We assert that λ -almost every point $x \in S^1$ is *recurrent*, i.e. for λ -almost all $x \in S^1$ there exists a strictly increasing sequence of positive integers $\{n_k\}_{k \in \mathbb{N}}$ such that $\lim_{k \rightarrow \infty} d^{S^1}(x, R^{n_k}(x)) = 0$. Let $x \in S^1$ be such that $O_R^+(x)$ accumulates to itself and let R_a denote the rigid rotation (translation) over angle (arclength) $2\pi a$ (a), with $a \in [0, 1)$. Then as $R \circ R_a = R_a \circ R$, it follows that if $O_R^+(x)$ accumulates to x then $O_R^+(R_a x) = R_a(O_R^+(x))$ accumulates to $R_a(x)$. Since $S^1 = \{R_a(x)\}_{a \in [0, 1)}$, every forward orbit accumulates to its initial condition.

It remains to prove the assertion. Let $P_{n,i} := [\frac{i}{n}, \frac{i+1}{n})$, for all $n \in \mathbb{N}$ and $i = 0, \dots, n-1$. Then $X := S^1 = \cup_{i=0}^{n-1} P_{n,i}$. Let

$$X_n := \cup_{i=0}^{n-1} \left(P_{n,i} \setminus \cup_{j=1}^{\infty} R^{-j}(P_{n,i}) \right),$$

denote the set of points $x \in X$ whose forward orbits under R do not return to the set $P_{n,i} \ni x$. Then set of points whose forward orbit returns to every open neighbourhood of its starting point is $X \setminus \cup_{n=1}^{\infty} X_n$. The aim is to thus show that the latter set has full λ -measure zero and we will achieve this by showing that $\lambda(X_n) = 0$ for all $n \in \mathbb{N}$.

By R -invariance of λ and the fact that $\lambda(P_{n,i}) = \frac{1}{n} > 0$, the Poincaré recurrence theorem implies that

$$\lambda \left(P_{n,i} \setminus \bigcup_{j=1}^{\infty} R^{-j}(P_{n,i}) \right) = 0$$

for all $n \in \mathbb{N}$ and $i = 0, \dots, n-1$. From this, it directly follows that $\lambda(X_n) = 0$.

4.3 Let $O_f^+(x)$ be periodic with period p , then $\mu := \frac{1}{p} \sum_{n=1}^p \delta_{f^n(x)}$ is an invariant probability measure for which $\mu(\tilde{x}) \neq 0$ if and only if $\tilde{x} \in O_f^+(x)$, in which case $\mu(\tilde{x}) = \frac{1}{p}$. With $A = \{\tilde{x}\} \subset O_f^+(x)$, we have, with reference to Kac's lemma, $\mu(A^{c*}) = 0$ since the set of points returning to A has full μ -measure. Hence, the expected return time

$$(\mu(A))^{-1} \int_A n_A d\mu = \left(\frac{1}{p}\right)^{-1} = p.$$

This aligns with the (obvious) fact that $n_A = p$ as the return time of any point on a periodic orbit equals its period.

4.4 *Poincaré recurrence*: If $a \in \mathbb{R} \setminus \mathbb{Q}$, it was shown already in Exercise **4.2** that the invariance of Lebesgue measure implies the accumulation of every initial condition by its forward orbit. The same argument also applies if $a \in \mathbb{Q}$ (this fact is also obvious, since in this case every orbit is periodic).

Expected return times: By Kac's lemma, given an invariant measure μ the μ -expected return time to a set A with $\mu(A) > 0$ is equal to $\frac{1-\mu(A^{c*})}{\mu(A)}$. Considering μ to be Lebesgue measure and taking $A = B_\varepsilon(x)$, gives $\mu(A) = 2\varepsilon$.

To obtain $\mu(A^{c*})$, we first note that if a is irrational, all orbits are dense and $A^{c*} = \emptyset$, so $\mu(A^{c*}) = 0$. Consequently, in this case the expected return time to A is $(2\varepsilon)^{-1}$.

In case $a = p/q$ is rational, with $\gcd(p, q) = 1$, A^{c*} (the subset of A^c whose forward orbits never enter A) is equal to q intervals of size 2ε is $2q\varepsilon < 1$ or the empty set if $2q\varepsilon \geq 1$. In the former case, the expectation is $(1 - (1 - 2\varepsilon q))/(2\varepsilon) = q$ (the period of all the orbits), whereas in the latter case the expectation of the return time is $(2\varepsilon)^{-1}$.

The correctness of the latter result can also be verified by direct evaluation of the return times. We have

$$\begin{cases} n_a(x) = 1 & \text{if } 0 < x < 2\varepsilon - \frac{1}{q}, \\ n_a(x) = \lceil q(1-x) \rceil & \text{if } 2\varepsilon - \frac{1}{q} < x < 2\varepsilon. \end{cases}$$

We note that in the latter domain we have (at most) two different first return times (they will be represented by $q + 1 - k$ and $q - k$, below). We change variables. Let $x = 2\varepsilon - \frac{1}{q} + \tilde{x}$ and $k := \lfloor 2\varepsilon q \rfloor$, then $\lceil q(1-x) \rceil = q + 1 - \lfloor q(2\varepsilon + \tilde{x}) \rfloor$. Let $\eta := \frac{k+1}{q} - 2\varepsilon$, then

$$\int_{2\varepsilon - \frac{1}{q}}^{2\varepsilon} \lceil q(1-x) \rceil dx = \int_0^\eta (q+1-k) d\tilde{x} + \int_\eta^{\frac{1}{q}} (q-k) d\tilde{x} = \frac{q-k}{q} + \eta = 1 - 2\varepsilon + \frac{1}{q}.$$

Consequently,

$$\int_A n_A d\lambda = \int_0^{2\varepsilon - \frac{1}{q}} dx + \int_{2\varepsilon - \frac{1}{q}}^{2\varepsilon} \lceil q(1-x) \rceil dx = 2\varepsilon - \frac{1}{q} + 1 - 2\varepsilon + \frac{1}{q} = 1,$$

yielding the expected return time to be indeed $(\mu(A))^{-1} = (2\varepsilon)^{-1}$.

A similar detailed verification of the expected return times on the basis of actual return times could be carried out when a is irrational, but is a little more laborious.

4.5 By Birkhoff's ergodic theorem, if μ is ergodic, for μ -almost all $x \in X$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_A(f^i(x)) = \int_A d\mu = \mu(A) > 0.$$

This means that the set $S \subset X$ of points for which this equality does not hold has zero μ -measure, i.e. $\mu(S) = 0$. But from the definition of A^{c*} we see that $\forall x \in A^{c*}, f^n(x) \notin A \forall n \in \mathbb{N}_0$. Hence, $\forall x \in A^{c*}$ it follows that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_A(f^i(x)) = 0 \neq \mu(A),$$

so that $A^{c*} \subset S$ and $\mu(A^{c*}) = 0$.

4.6 Let (X, \mathcal{F}) a measure space and $f : X \rightarrow X$ measurable.

- (i) Suppose μ_1 and μ_2 are f -invariant ergodic probability measures and $\mu_1 \ll \mu_2$. Consider any bounded integrable function $g : \mathcal{F} \rightarrow \mathbb{R}$. Then ergodicity of μ_2 implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x)) = \int_X g d\mu_2,$$

for all $x \in A \in \mathcal{F}$, where $\mu_2(A) = 1$. If $\mu_1 \ll \mu_2$ and $\mu_2(A) = 1$, then $\mu_2(A^c) = 0 \Rightarrow \mu_1(A^c) = 0 \Rightarrow \mu_1(A) = 1$. Hence,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x)) = \int_X g d\mu_2, \quad \mu_1\text{-almost surely.} \quad (\text{S.7})$$

On the other hand, ergodicity of μ_1 implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g(f^i(x)) = \int_X g d\mu_1, \quad \mu_1\text{-almost surely.} \quad (\text{S.8})$$

Let $g = \mathbb{1}_B$ with $B \in \mathcal{F}$, then (S.7)-(S.8) imply that $\mu_1(B) = \mu_2(B)$ for all $B \in \mathcal{F}$, or - in other words - that $\mu_1 = \mu_2$.

- (ii) Suppose 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 $t \in (0, 1)$. Since for any $A \in \mathcal{F}$, $\mu(A) = 0$ implies that $\mu_1(A) = \mu_2(A) = 0$, it follows that $\mu_1 \ll \mu$ and $\mu_2 \ll \mu$. Also, if $\mu(A) = 1$ then $\mu_1(A) = \mu_2(A) = 1$. Hence, if μ would be ergodic, then so are μ_1 and μ_2 , but by part (i), above, then $\mu = \mu_1 = \mu_2$, contradicting the assumption. In fact, all extremal points of the convex set of f -invariant probability measures are ergodic. Namely, suppose μ is extremal but not ergodic. Then there exists $A \in \mathcal{F}$ with $f^{-1}(A) = A$ and $\mu(A) \in (0, 1)$. Defining the f -invariant probability measures μ_A and μ_{A^c} as

$$\mu_A(B) := \frac{\mu(A \cap B)}{\mu(A)}, \quad \mu_{A^c}(B) := \frac{\mu(A^c \cap B)}{\mu(A^c)}, \quad \forall B \in \mathcal{F},$$

then

$$\mu = \mu(A)\mu_A + \mu(A^c)\mu_{A^c},$$

implies by the previous argument that μ is not an extremal point of the convex set of f -invariant probability measures, contradicting the assumption.

- (iii) Let $\mu_1 \neq \mu_2$ be two ergodic f -invariant probability measures. Let $\mu = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2$. Then by part (ii), above, μ is a non-ergodic invariant probability measure, which implies that there exists $A \in \mathcal{F}$ satisfying $f^{-1}(A) = A$ such that $\mu(A) \notin \{0, 1\}$. As μ_1 and μ_2 are ergodic we have $\mu_1(A) \in \{0, 1\}$ and $\mu_2(A) \in \{0, 1\}$. If $\mu_1(A) = \mu_2(A)$ then $\mu(A) \in \{0, 1\}$, so $\mu_1(A) \neq \mu_2(A)$ and either $\mu_1(A) = 1 - \mu_2(A) = 0$ or $\mu_2(A) = 1 - \mu_1(A) = 0$, i.e. μ_1 and μ_2 are mutually singular.

4.7 λ is the unique invariant Borel probability measure for irrational rigid circle rotations (cf. Example 4.1 and Exercise 4.1). Suppose $A \in \mathcal{B}(S^1)$, $f^{-1}(A) = A$ and $\lambda(A) \notin \{0, 1\}$, then the measure $\mu : \mathcal{B} \rightarrow [0, 1]$, defined for all $B \in \mathcal{B}(S^1)$ as

$$\mu(B) := \frac{\lambda(B \cap A)}{\lambda(A)}$$

is an invariant Borel probability measure, different from λ , contradicting the uniqueness of λ .

4.8 Let $f(x) = E_k(x) = kx \bmod 1$ for some $k \geq 2$.

- (a) If $\lambda(A) > 0$, the Lebesgue density theorem implies that for all $\varepsilon > 0$ and λ -almost all $x \in A$ there exists a $\Delta > 0$ such that

$$\frac{\lambda(A \cap B_\delta(x))}{\lambda(B_\delta(x))} \geq 1 - \varepsilon$$

for all $\delta < \Delta$. This implies that for every x for which this relation holds, there exists an interval $J \in \mathcal{P}_n$ with n large enough such that $J \subset B_\Delta(x)$. Then $\frac{\lambda(A \cap J)}{\lambda(J)} \geq 1 - \varepsilon$ and the subsequent equivalent relation follows directly from using the identity $\lambda(A \cap J) = \lambda(J) - \lambda(J \setminus A)$.

- (b) Note that $f^n : J \rightarrow [0, 1)$ is surjective and invertible. Hence, $f^{-1}(A) = A$ implies that

$$\begin{aligned} f^{-n}(A) \cap J &= A \cap J \Rightarrow \underbrace{f^n(f^{-n}(A) \cap J)}_{=A} = f^n(A \cap J) \\ &\Rightarrow A = f^n(J) \setminus f^n(J \setminus A) = [0, 1) \setminus f^n(J \setminus A) \\ &\Rightarrow f^n(J \setminus A) = [0, 1) \setminus A. \end{aligned}$$

Since f^n is homeogeneously expanding on J , it preserves ratios of Lebesgue measures of subsets. Hence

$$\frac{\lambda([0, 1) \setminus A)}{\lambda([0, 1))} = \frac{\lambda(f^n(J \setminus A))}{\lambda(f^n(J))} = \frac{\lambda(J \setminus A)}{\lambda(J)} \leq \varepsilon,$$

implying that $\frac{\lambda([0, 1) \setminus A)}{\lambda([0, 1))} \leq \varepsilon$. Since $\varepsilon > 0$ was chosen arbitrarily, it follows that $\lambda(A) = 1$.

4.9 Consider the parametrisation $S^1 \simeq [0, 1)$, $f(x) = 10x \bmod 1$. $g = \mathbb{1}_{[0, \frac{1}{2})}$.

(a) For instance $x = 0$ satisfies $f^i(0) = 0$ for all $i \in \mathbb{N}$, so that $g(f^i(0)) = 1$ for all $i \in \mathbb{N}$ and

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

(b) For instance, let $x = 0.07^2 0^6 7^{12} 0^{24} 7^{48} \dots$, then

$$\frac{1}{3 \cdot 2^m} \sum_{i=0}^{3 \cdot 2^m - 1} g(f^i(x)) = \begin{cases} \frac{1}{3} & \text{if } m \text{ is odd,} \\ \frac{2}{3} & \text{if } m \text{ is even.} \end{cases}$$

Hence, the limit does not exist.

(c) In every interval of the form $[\frac{1}{10^k}, \frac{i+1}{10^k})$ with $k \in \mathbb{N}$ and $i \in \{0, \dots, 10^k - 1\}$, there exists a point x whose decimal expansion ends in $\bar{0}$. This set of points is thus dense in S^1 and the corresponding limit of the Birkhoff average is 1, just like in the example under **(a)** above. Similarly, every interval of the form just mentioned also contains a point x whose decimal expansion ends with the sequence $07^2 0^6 7^{12} 0^{24} 7^{48} \dots$. The set of such points is thus again dense in S^1 and - assuming without loss of generality that the described tail commences after an even number of decimals - has the property that

$$\lim_{m \rightarrow \infty} \frac{1}{3 \cdot 2^{2m}} \sum_{i=0}^{3 \cdot 2^{2m} - 1} g(f^i(x)) = \frac{2}{3}, \quad \lim_{m \rightarrow \infty} \frac{1}{3 \cdot 2^{2m+1}} \sum_{i=0}^{3 \cdot 2^{2m+1} - 1} g(f^i(x)) = \frac{1}{3},$$

so that indeed the Birkhoff average does not exist.

4.10 TBA

4.11 TBA

4.12 TBA

4.13 We may represent the state of the system at time j by the number of $x_j \in \{0, \dots, 100\}$ of balls in the first container. The dynamics may be represented as a shift map σ_A for a topological Markov chain $\Sigma_{101,A}^+$ with 101×101 transition matrix A with $A_{ij} = 1$ if $|i - j| = 1$ and $A_{ij} = 0$ otherwise. It is easily seen that A is irreducible (but not primitive). The relevant Markov measure is found from the given transition probabilities in the problem: the probability $P_{i,i+1}$ that there are $i + 1$ balls in the first container after we have i balls in the first container is $\frac{100-i}{100}$. Similarly, $P_{i,i-1} = \frac{i}{100}$. The left probabilistic eigenvector $p = \{p_0, \dots, p_{100}\}$ for eigenvalue 1 of the stochastic matrix P , i.e. $pP = p$, has entries

$$p_i = \frac{1}{2^{100}} \binom{100}{i}.$$

One could have guessed: p_i is the probability to have i balls in the first container if each of the 100 balls is put in either bin with equal probability. Let μ_P denote the corresponding ergodic Markov measure. Then $\mu_P(C_{100}) = p_{100} = 2^{-100}$. Hence, by Kac's lemma, the expected return time to this state is $n_{C_{100}} = 2^{100}$. With the time units being seconds, this amounts to approximately 4.0×10^{22} years. Compare this to the estimated lifetime of the universe of 13.8×10^9 years...

In contrast, the expected return time from the average state 50 to itself is

$$n_{C_{50}} = \frac{1}{\mu_P(C_{50})} = p_{50}^{-1} = 2^{100} \binom{100}{50}^{-1} \approx 12.56.$$

4.14 Invariance: for all $A \in \mathcal{G}$,

$$h_*\mu(g^{-1}(A)) = \mu(h^{-1}(g^{-1}(A))) = \mu((h \circ f)^{-1}(A)) = \mu(f^{-1}(h^{-1}(A))) = \mu(h^{-1}(A)) = h_*\mu(A),$$

where we note that the "-1" superscripts indicate preimages, so that invertibility is not assumed or required. Ergodicity: let $A \in \mathcal{G}$ such that $g^{-1}(A) = A$. Then

$$h^{-1}(g^{-1}(A)) = h^{-1}(A) \Leftrightarrow f^{-1}(h^{-1}(A)) = h^{-1}(A),$$

so $h^{-1}(A) \in \mathcal{F}$ is f -invariant. In turn, this implies by ergodicity of μ for f , that

$$h_*\mu(A) = \mu(h^{-1}(A)) \in \{0, 1\},$$

i.e. that $h_*\mu$ is ergodic for g .

4.15 The Markov partition $\mathcal{R} = \{(0, \frac{1}{4}), (\frac{1}{4}, \frac{1}{2}), (\frac{1}{2}, \frac{3}{4}), (\frac{3}{4}, 1)\}$ is piecewise expanding since $|f'(x)| > 1$ in each of the partition elements. Hence there exists a topological semi-conjugacy between f and $\sigma_A : \Sigma_{4,A} \rightarrow \Sigma_{4,A}$ where the connectivity matrix

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

follows from examining how the images under f of the partition elements overlap those partition elements. Following the prescription in the notes on how to find a Markov measure $\mu_{v,P}$ on $\Sigma_{4,A}$ that projects to an ergodic invariant measure on $[0, 1]$ for f that is absolutely continuous with respect to the Lebesgue measure, we let $P_{ij} = |f'_x|$ with $x \in R_i$ (which is well-defined since f' is constant on each of the partition elements). Thus

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

As P is irreducible it has a unique left probability eigenvector for eigenvalue 1. It is readily verified that this is $v = \frac{1}{8}(1, 1, 3, 3)$.

The corresponding Markov measure is defined as $\mu_{v,P}(C_{i_0 \dots i_{n-1}}) := v_{i_0} p_{i_0, i_1} \dots p_{i_{n-2}, i_{n-1}}$.

Let $h : \Sigma_{4,A}^+ \rightarrow [0, 1]$ denote the map that semi-conjugates σ_A to f . In the lecture notes it has been derived that $h_* \mu_{v,P}(C_{i_0 \dots i_{n-1}}) = \frac{v_{i_0}}{|R_{i_0}|} |h(C_{i_0 \dots i_{n-1}})|$, where v_i denotes the i th component of v . As $|R_0| = |R_1| = |R_2| = |R_3| = \frac{1}{4}$, we have

$$\begin{aligned} h_* \mu_{v,P}(C_{i_0 \dots i_{n-1}}) &= \frac{1}{2} |h(C_{i_0 \dots i_{n-1}})| \text{ if } i_0 \in \{0, 1\}, \\ h_* \mu_{v,P}(C_{i_0 \dots i_{n-1}}) &= \frac{1}{2} |h(C_{i_0 \dots i_{n-1}})| \text{ if } i_0 \in \{2, 3\} \end{aligned}$$

so that $h_* \mu_{v,P} \ll \lambda$ with density

$$\rho(x) = \begin{cases} \frac{1}{2} & \text{if } x \in [0, \frac{1}{2}] \\ \frac{3}{2} & \text{if } x \in (\frac{1}{2}, 1], \end{cases}$$

i.e. $h_* \mu(A) = \int_A \rho(x) dx$ for all $A \in \mathcal{B}([0, 1])$.

4.16

(a) To verify that $f_4 \circ h = h \circ T$ we check:

$$\begin{aligned} f_4 \circ h(x) &= 4 \sin^2\left(\frac{\pi}{2}x\right) \left(1 - \sin^2\left(\frac{\pi}{2}x\right)\right) = 4 \sin^2\left(\frac{\pi}{2}x\right) \cos^2\left(\frac{\pi}{2}x\right) = \sin^2(\pi x), \\ h \circ T(x) &= \begin{cases} \sin^2(\pi x) & \text{if } x \in [0, \frac{1}{2}], \\ \sin^2\left(\frac{\pi}{2}(2-2x)\right) = \sin^2(\pi x) & \text{if } x \in [\frac{1}{2}, 1]. \end{cases} \end{aligned}$$

(b) We have

$$d\mu(x) = d\lambda(h^{-1}(x)) = \frac{d}{dx}(h^{-1}(x))dx = \frac{1}{h'(h^{-1}(x))}dx,$$

where the last equality follows from

$$\frac{d}{dx} (h^{-1}(h(x))) = 1 \Leftrightarrow \left(\frac{d}{dx} h^{-1} \right)(h(x)) \cdot h'(x) = 1 \Rightarrow \frac{d}{dx} (h^{-1})(x) = \frac{1}{h'(h^{-1}(x))}.$$

If $h = \sin^2(\frac{\pi}{2}x)$,

$$\begin{aligned} h'(x) &= \pi \sin(\frac{\pi}{2}x) \cos(\frac{\pi}{2}x) = \pi \frac{\sin^2(\frac{\pi}{2}x)}{\tan(\frac{\pi}{2}x)} \\ &= \pi \sin^2(\frac{\pi}{2}x) \sqrt{\frac{1 - \sin^2(\frac{\pi}{2}x)}{\sin^2(\frac{\pi}{2}x)}} = \pi \sqrt{\sin^2(\frac{\pi}{2}x)(1 - \sin^2(\frac{\pi}{2}x))}, \\ \Rightarrow h'(h^{-1}(x)) &= \pi \sqrt{x(1-x)} \Rightarrow d\mu(x) = \frac{1}{\pi \sqrt{x(1-x)}} dx. \end{aligned}$$

$\mu = h_*\lambda$ is ergodic for f_4 since λ is ergodic for T .

(c) We have

$$g_n(x) := \left| \frac{df_4^n}{dx}(x) \right| = \left| \frac{df_4}{dx}(f_4^{n-1}(x)) \cdot \dots \cdot \frac{df_4}{dx}(x) \right| = \prod_{i=1}^n \left| \frac{df_4}{dx}(f_4^i(x)) \right|,$$

so that with $\tilde{g}(x) := \ln \left| \frac{df_4}{dx}(x) \right|$,

$$\ln g_n(x) = \sum_{i=1}^n \ln \left| \frac{df_4}{dx}(f_4^i(x)) \right| = \sum_{i=1}^n \tilde{g}(f_4^i(x)).$$

The latter is sum the observable \tilde{g} along the first n points on an f_4 -orbit. Accordingly, Birkhoff's ergodic theorem yields that

$$\int_0^1 \tilde{g}(x) d\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \tilde{g}(f_4^i(x)), \text{ for } \mu\text{-almost all } x \in [0, 1].$$

In fact, as $\mu \ll \lambda$, it follows that this equality hold for λ -almost all $x \in [0, 1]$. We evaluate the left-hand side

$$\int_0^1 \tilde{g}(x) d\mu = \int_0^1 \frac{\ln |4 - 8x|}{\pi \sqrt{x(1-x)}} dx = \int_0^{\frac{1}{2}} \ln 2 dx + \int_{\frac{1}{2}}^1 \ln |-2| dx = \ln 2 > 0,$$

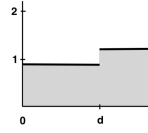
using in the second equality the fact that f_4 is smoothly conjugate to T on $[0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$ and $|T'(x)| = 2$ on this domain. This implies exponential growth of g_n , in the sense that $g_n(x) = C(n)2^n$ for some function C satisfying $\lim_{n \rightarrow \infty} \frac{1}{n} \ln C(n) = 0$. $\ln 2$ is the *Lyapunov exponent* of f_4 associated with the ergodic invariant measure μ .

4.17 The solution follows closely the similar argument for E_{10} in Example 4.4. The tent map T is topologically semi-conjugate to the full shift $\sigma : \Sigma_2^+ \rightarrow \Sigma_2^+$. Consider

the 2×2 stochastic matrix P , with $P_{ij} = \frac{1}{2}$ for all $i, j = 0, \dots, 1$. Then the unique probability left eigenvector for P is $v = \frac{1}{2}(1, 1)$, which gives rise to the Markov (and Bernoulli) measure $\mu_{v,P}$ on $\mathcal{B}(\Sigma_2^+)$, characterised by its definition on cylinder sets: $\mu(C_{i_0 \dots i_{n-1}}) = 2^{-n}$. Let $h : \Sigma_2^+ \rightarrow [0, 1]$ 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_2^+ , from which it follows that $h_*\mu_{v,P} = \lambda$, the Lebesgue measure. Ergodicity of λ follows from ergodicity of $\mu_{v,P}$, by Proposition 4.4

4.18

$$\begin{aligned} (a) \quad \int_0^1 d\mu(x) &= d^{-1} \left(\frac{1}{2} \int_0^d dx + \frac{1}{\sqrt{2}} \int_d^1 dx \right) = d^{-1} \left(\frac{1}{2}d + \frac{1}{\sqrt{2}}(1-d) \right) = \\ &= (2d)^{-1} (d + \sqrt{2}(1-d)) = (2d)^{-1} (2 - \sqrt{2} + \sqrt{2}(\sqrt{2}-1)) = 1. \end{aligned}$$



Sketch of the graph of the density:

- (b) We need to show that $\mu(f^{-1}(A)) = \mu(A)$, $\forall A \in \mathcal{B}([0, 1])$. Note that $\mu(A) = \frac{1}{2}\lambda(A_0) + \frac{1}{\sqrt{2}}\lambda(A_1)$, where λ denotes the Lebesgue measure. Also, $\mu(f^{-1}(A)) = \mu(f^{-1}(A_0)) + \mu(f^{-1}(A_1))$ since $f(\text{Int}(A_0)) = \text{Int}(A_1)$ and $f(\text{Int}(A_1)) = \text{Int}(A_0)$. As f is uniformly expanding, with $|f'(x)| = \sqrt{2}$, we have $\lambda(f^{-1}(A_0)) = \frac{1}{\sqrt{2}}\lambda(A_0)$ and $\lambda(f^{-1}(A_1)) = 2 \cdot \frac{1}{\sqrt{2}}\lambda(A_1)$ (since f^{-1} has two branches in $(d, 1]$.) We have $\mu(f^{-1}(A_0)) = \frac{1}{\sqrt{2}}\lambda(f^{-1}(A_0))$ since $f^{-1}(A_0) \subset [d, 1]$ and $\mu(f^{-1}(A_1)) = \frac{1}{2}\lambda(f^{-1}(A_1))$ since $f^{-1}(A_1) \subset [0, d]$. Hence $\mu(f^{-1}(A)) = \frac{1}{\sqrt{2}}\lambda(f^{-1}(A_0)) + \frac{1}{2}\lambda(f^{-1}(A_1)) = \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\lambda(A_0) + \frac{1}{2} \cdot 2 \cdot \frac{1}{\sqrt{2}}\lambda(A_1) = \mu(A)$.
- (c) The partition $\mathcal{R} = \{(0, c), (c, d), (d, 1)\}$ is a piecewise expanding Markov partition for the map f . The connectivity matrix of the associated topological Markov chain

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

The Markov measure related to μ can be obtained from the connectivity matrix A associated to shift $\sigma_A : \Sigma_{3,A}^+ \rightarrow \Sigma_{3,A}^+$, by transforming (A) into the (irreducible) stochastic matrix

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

with the matrix entries representing the (relative) transition probabilities between the labelling intervals. The left eigenvector π for eigenvalue 1 of P is $v = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$. The corresponding Markov measure is defined as $\mu_{v,P}(C_{i_0 \dots i_{n-1}}) := v_{i_0} p_{i_0, i_1} \dots p_{i_{n-2}, i_{n-1}}$.

Let $h : \Sigma_{3,A}^+ \rightarrow [0, 1]$ denote the map that semi-conjugates σ_A to f . In the notes it has been derived that $h_*\mu_{v,P}(C_{i_0 \dots i_{n-1}}) = \frac{v_{i_0}}{|R_{i_0}|} |h(C_{i_0 \dots i_{n-1}})|$, where v_i denotes

the i th component of v . As $|R_0| = |R_1| = d/2$, $|R_2| = 1 - d = d/\sqrt{2}$, we have

$$\begin{aligned} h_*\mu_{v,P}(C_{i_0\dots i_{n-1}}) &= \frac{v_{i_0}}{|R_{i_0}|} |h(C_{i_0\dots i_{n-1}})| = \frac{1}{4} \frac{2}{d} |h(C_{i_0\dots i_{n-1}})| \\ &= \frac{1}{2} d^{-1} |h(C_{i_0\dots i_{n-1}})|, \text{ if } i_0 \in \{0, 1\}, \\ h_*\mu_{v,P}(C_{i_0\dots i_{n-1}}) &= \frac{v_{i_0}}{|R_{i_0}|} |h(C_{i_0\dots i_{n-1}})| = \frac{1}{2} \frac{\sqrt{2}}{d} |h(C_{i_0\dots i_{n-1}})| \\ &= \frac{1}{\sqrt{2}} d^{-1} |h(C_{i_0\dots i_{n-1}})|, \text{ if } i_0 = 2 \end{aligned}$$

so that indeed $\mu = h_*\mu_{v,P}$.

- (d) By Birkhoff's ergodic theorem, $\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f^i(x)$ is μ -a.s. equal to $\int_{[0,1]} x d\mu(x) = d^{-1} \left(\frac{1}{2} \int_0^d x dx + \frac{1}{\sqrt{2}} \int_d^1 x dx \right) = (4d)^{-1} (d^2 + \sqrt{2}(1 - d^2)) = \frac{2d+1}{4} = \frac{5}{4} - \frac{1}{\sqrt{2}}$ using $d^2 - \sqrt{2}(d^2 - 1) = 2d^2 + d$. As $\mu \ll \lambda$, the Birkhoff sum is also Lebesgue-a.s equal to $\frac{5}{4} - \frac{1}{\sqrt{2}}$. The given integral is the average of the Birkhoff sum with respect to the Lebesgue measure on $[0, 1]$ and thus precisely equal to $\frac{5}{4} - \frac{1}{\sqrt{2}}$.