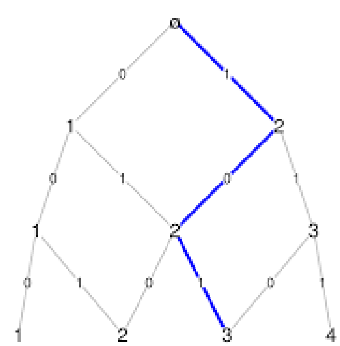
VO Special Topics in Stochastics: Symbolic Dynamic (2019W.25050.1)

Henk Bruin



November 18, 2019

Contents

0.1 Notation and motivation

0.1.1 Symbol sequences

Let $\mathcal{A}$ be a finite or countable **alphabet** of letters. Usually $\mathcal{A} = \{0, \dots, N-1\}$ or $\{0, 1, 2, \dots\}$ but we could use other letters and symbols too. We are interested in the space of infinite or bi-infinite sequences of letters:

$$\Sigma = \mathcal{A}^{\mathbb{N} \text{ or } \mathbb{Z}} = \{x = (x_i)_{i \in \mathbb{N} \text{ or } \mathbb{Z}} : x_i \in \mathcal{A}\}.$$

Such symbol strings find applications in data-transmission and storage, linguistics, theoretical computer science and also dynamical systems (symbolic dynamics). A finite string of letters, say $x_1 \dots x_n in \mathcal{A}^n$ is called a **word** or **block**.

Sets of the form

$$[e_k \dots e_l] = \{x \in \Sigma : x_i = e_i \text{ for } k \leq i \leq l\}$$

are called **cylinder sets**[1] . Intersections of cylinder sets are again cylinder sets. In the **product topology** on Σ, open sets are those sets that can be written as arbitrary unions of cylinder sets, i.e., the cylinder sets are a basis of the topology.

Note that a cylinder set is both open and closed (because it is the complement of the union of complementary cylinders). Sets that are both open and closed are called **clopen**.

Exercise 0.1.1. *Are there open sets in product topology on Σ that are not closed?*

Shift spaces with product topology are metrizable. One of the usual[2] metrics that generates product topology is

$$d(x, y) = 2^{-m} \qquad \text{for } m = \sup\{n \geq 0 : x_i = y_i \text{ for all } |i| < n\},$$

so in particular $d(x, y) = 1$ if $x_0 \neq y_0$, and $\operatorname{diam}(\Sigma) = 1$.

Exercise 0.1.2. *Show that Σ with product topology is compact if and only if $\#\mathcal{A} < \infty$.*

Exercise 0.1.3. *Let x^k be a sequence of sequences. Show that $x^k \to x$ in product topology if and only if x^k stabilizes on every finite window, i.e., for all $m < n$, $x^k_m x^k_{m+1} \dots x^k_n$ is eventually constant.*

---

[1]In greater generality, if X is a topological space and $n \in \mathbb{N} \cup \{\infty\}$, every set of the form $A \times X^{n-k}$ for $A \subset X^k$ is called a cylinder set. If $X = \mathbb{R}$, $n = 3$ and A is a circle in $\mathbb{R}^2$, then $A \times \mathbb{R}$ is indeed a geometrical cylinder, stretching infinitely far in the third direction.

[2]Other metrics are $d'(x, y) = \frac{1}{m}$ or $d'(x, y) = \sum_i |x_i - y_i| 2^{-|i|}$, but they are equivalent to $d(x, y)$, not in the sense that there is some C such that $\frac{1}{C} d(x, y) \leq d'(x, y) \leq C d(x, y)$ for all $x, y \in Sigma$, but in the sense that the embedding $i : (\Sigma, d') \to (\Sigma, d)$ as well as its inverse i^{-1} are uniformly continuous. This means that they generate the same topology.

Lemma 0.1.4. *If $2 \leq \#\mathcal{A} < \infty$, then Σ is a* **Cantor set** *(i.e., compact, without isolated points and its connected components are points).*

Proof. Set $\mathcal{A} = \{0, 1, \dots, N-1\}$ with discrete topology. Clearly $\mathcal{A}$ is compact, because finite. Compactness of Σ then follows from Tychonov's Theorem. No isolated point x is isolated, because, for arbitrary $x \in \Sigma$, the sequence x^n defined as $x^n_i = x_i$ if $i \neq n$ and $x^n_n = N-1-x_n$, converges to x. Finally, if $x \neq y$, say $n = \min\{|i| : x_i \neq y$, then $Z := \{x' \in X : x_i = x'_i$ for all $|i| \leq n$ and $X \setminus Z$ are two clopen disjoint nonempty sets whose union is X. Thus x and y cannot belong to the same connected component. □

0.1.2 Subshifts

The **shift map** or **left-shift** $\sigma : \Sigma \to \Sigma$, defined as

$$\sigma(x)_i = x_{i+1}, \quad i \in \mathbb{N} \text{ or } \mathbb{Z}.$$

is invertible on $\mathcal{A}^{\mathbb{Z}}$ (with inverse $\sigma^{-1}(x)_i = x_{i-1}$) but non-invertible on $\mathcal{A}^{\mathbb{N}}$.

Exercise 0.1.5. *Show that the shift is continuous, and in fact uniformly continuous even if $\#\mathcal{A} = \infty$.*

Exercise 0.1.6. *Let $\mathcal{A} = \{1, 2, \dots, a\}$ for some $a \in \mathbb{N}$. Show that the number of periodic sequences $x \in \mathcal{A}^{\mathbb{Z}}$ of minimal period n equals*

$$Per(n) = \sum_{1 \leq d \leq n, d|n} \mu(\frac{n}{d}) a^d,$$

*where μ denote the Möbius function, see (**??**). In particular, $Per(n) = a^n - a$ if n is a prime.*

Definition 0.1.7. *The* **orbit** *of $x \in X$ is the set*

$$orb(x) = \begin{cases} \{\sigma^n(x) : n \in \mathbb{Z}\} & \textit{if } \sigma \textit{ is invertible;} \\ \{\sigma^n(x) : n \geq 0\} & \textit{if } \sigma \textit{ is non-invertible.} \end{cases}$$

The set $orb^+(x) = \{\sigma^n(x) : n \geq 0\}$ is the **forward orbit** *of x. This is of use if σ is invertible; if σ is non-invertible, then $orb^+(x) = orb(x)$. We call x* **recurrent** *if $x \in \overline{orb^+(\sigma(x))}$. The ω-***limit set** *of x is the set of accumulation points of its forward orbit, or in formula*

$$\omega(x) = \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{m \geq n} \sigma^m(x)} = \{y \in X : \exists\, n_i \to \infty, \lim_{i \to \infty} \sigma^{n_i}(x) = y\}.$$

*Analogously for invertible shifts, the α-***limit set** *of x is the set of accumulation points of its backward orbit of x:*

$$\alpha(x) = \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{m \leq -n} \sigma^m(x)} = \{y \in X : \exists\, n_i \to \infty, \lim_{i \to \infty} \sigma^{-n_i}(x) = y\}.$$

Exercise 0.1.8. *Let $\sigma : \Sigma \to \Sigma$ be invertible. Is there a difference between $x \in \overline{orb(x) \setminus \{x\}}$ and $x \in \overline{orb^+(x) \setminus \{x\}}$?*

Definition 0.1.9. *A subset $X \subset \Sigma$ is a* **subshift** *if it is closed (in product topology) and strongly shift-invariant, i.e., $\sigma(X) = X$. If σ is invertible, then we also stipulate that $\sigma^{-1}(X) = X$.*

In the following examples, we use $\mathcal{A} = \{0, 1\}$ unless stated otherwise.

Example 0.1.10. *The set $X = \{x \in \Sigma : x_i = 1 \Rightarrow x_{i+1} = 0\}$ is called the* **Fibonacci shift**[3]. *It disallows sequences with two consecutive 1s. This Fibonacci shift is an example of a subshift of finite type (SFT) , see Section 2.1.*

Example 0.1.11. *X is a collection of labels of infinite paths through the graph in Figure 1 (left). Labels are given to the vertices of the graph, and no label is repeated.*

Example 0.1.12. *X is a collection of labels of infinite paths through the graph in Figure 1 (right). Labels are given to the arrows of the graph, and labels can be repeated (different arrows with the same label can occur).*

Figure 1: Fibonacci transition graphs: vertex-labeled and edge-labeled.

Example 0.1.13. *$X_{even} \subset \{0, 1\}^{\mathbb{N}}$ is the collection of infinite sequences in which the 1s appear only in blocks of even length, and also $1111 \cdots \in X$. We call $Xeven$ the* **even shift**. *Similarly, the* **odd shift** *X_{odd} is the collection of infinite sequences in which the 0s appear only in blocks of odd length, and also $0000 \cdots \in X$, see Figure 2.*

Example 0.1.14. *Let S be a non-empty subset of $\mathbb{N}$. Let $X \subset \{0, 1\}^{\mathbb{Z}}$ be the collection of sequences in which the appearance of two consecutive 1s occur always s positions apart for some $s \in S$. Hence, sequences in X have the form*

$$x = \dots 10^{s_{-1}-1}10^{s_0-1}10^{s_1-1}10^{s_2-1}1 \dots$$

where $s_i \in S$ for each $i \in \mathbb{Z}$. This space is called the S-gap shift, see Section **??**.

[3]Warning: there is also a Fibonacci substitution shift = Fibonacci Sturmian shift (see Example 3.1.4), which is different from this one

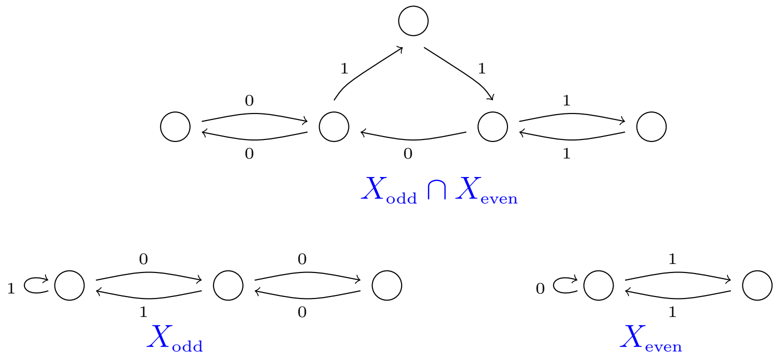


Figure 2: Edge-labeled graphs for X_{odd}, X_{even} and $X_{\text{odd}} \cap X_{even}$.

Example 0.1.15. *X is the closure of the collection of symbolic itineraries of a circle rotation $R : \mathbb{S}^1 \to \mathbb{S}^1$, see Figure 3. That is, if $y \in \mathbb{S}^1$ and $R^n(y) \in [0, \alpha)$ then we write $x_n = 0$. Otherwise $x_n = 1$. The resulting shift is called a* **Sturmian shift**, *see Definition 3.3.14.*

Example 0.1.16. *X is the closure of the collection of symbolic trajectories of β-transformation $T_\beta : [0, 1] \to [0, 1]$, $T_\beta(x) = \beta x \pmod 1$, see Figure 3. These are the* **β-shifts**, *see Section 2.3.*

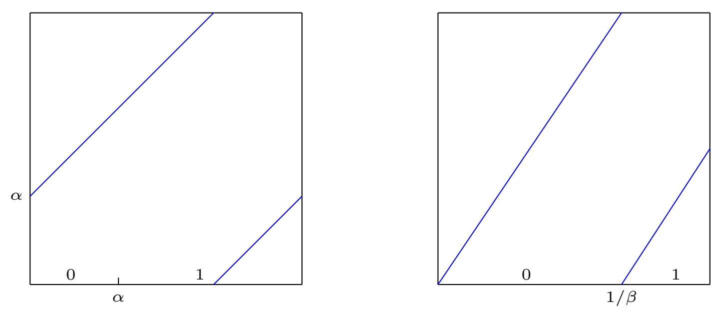


Figure 3: Symbolic dynamics for a circle rotation R_α and a β-transformation T_β

Example 0.1.17. *The alphabet $\mathcal{A}$ consists of brackets $(,), [,]$ and $\mathcal{L}(X)$ (see Definition 1.1.1 below) consists of all words of pairs of brackets that are properly paired and unlinked. So [([ ])] and (( ) [ ]) belong to $\mathcal{L}(X)$, but [( ] and ([ )] do not. This example is called the* **Dyck shift**, *see Section* **??**.

0.1.3 Turing machines

A Turing machine is a formal description of a simple type of computer, named after the British mathematician Alan Turing (1912-1954). He used this in theoretic papers to explore the limits what is computable by computers and what is not. For us, the size of a Turing machine that can recognize words in a language $\mathcal{L}(X)$, or reject words that don't belong to $\mathcal{L}(X)$, is a measure for how complicated a subshift is. In fact, a subshift is called **regularly enumerable** in the Chomsky hierarchy if its language can be recognized by a Turing machine.

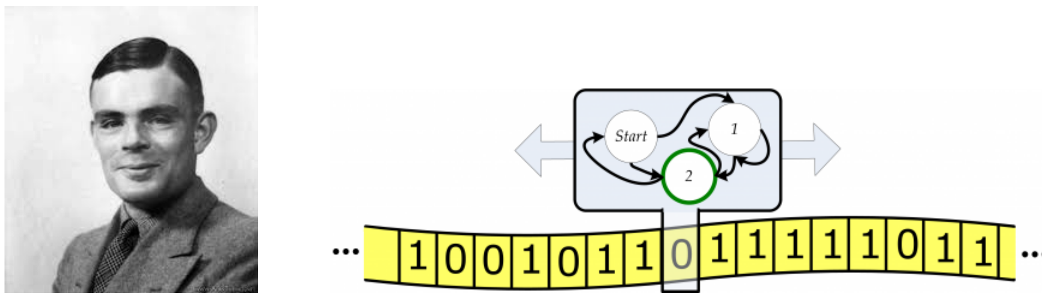


Figure 4: Alan Turing (1912-1954) and his machine.

A Turing machine has the following components:

- A tape on which the input is written as a word in the alphabet $\{0, 1\}$.
- A reading device, that can read a symbol at one position on the tape at the time. It can also erase the symbol and write a new one, and it can move to the next or previous position on the tape.
- A finite collection of states $S_1, \dots, S_N$, so N is the size of the Turing machine. Each state comes with a short list of instructions:
  - read the symbol;
  - replace the symbol or not;
  - move to the left or right position;
  - move to another (or the same) state.

  One state, say S_1, is the **initial state**. One (or several) states are **halting states**. When one of these is reached, the machine stops.

Example 0.1.18. *The following Turing machine rejects tape inputs that do not belong to the language of the Fibonacci shift. Let s be the symbol read at the current position of the tape, starting at the first position. We describe the states:*

S_1: *If $s = 0$, move to the right and go to State S_1. If $s = 1$, move to the right and go to State S_2.*

S_2: *If $s = 0$, move to the right and go to State S_1. If $s = 1$, go to State S_3.*

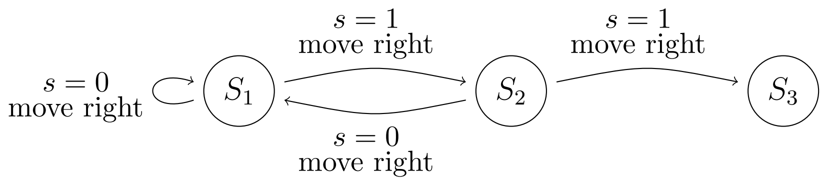


Figure 5: A Turing machine accepting words from the Fibonacci shift.

S_3: Halt. The word is rejected, see Figure 5.

Exercise 0.1.19. *Design a Turing machine that accepts the word in the even shift (Example 0.1.13).*

Exercise 0.1.20. *Suppose two positive integers m and n are coded on a tape by first putting m ones, then a zero, then n ones, and then infinitely many zeros. Design Turing machines that compute $m + n$, $|m - n|$ and mn so that the outcome is a tape with a single block of ones of that particular length, and zeros otherwise.*

Chapter 1

General properties of subshifts

1.1 Word-complexity

Any finite contiguous block of letters is called a word; an n-word is a word of n letters and ϵ is the empty word (of length 0). We use the notation $\mathcal{A}^n = \{n\text{-words in } \Sigma\}$ and

$$\mathcal{A}^* = \{\text{words of any finite length in } \Sigma \text{ including the empty word}\}.$$

Given a subshift X, a finite word u appearing in some $x \in X$ is sometimes called a **factor**[1] of x. If u is concatenated as $u = vw$, then v is a **prefix** and w a **suffix** of u.

Definition 1.1.1. *The collection*

$$\mathcal{L}(X) = \{\textit{words of any finite length in } X\}$$

is called the **language** *of X. We use the notation $\mathcal{L}_n(X)$ for all the length n words in the language.*

Definition 1.1.2. *The function $p : \mathbb{N} \to \mathbb{N}$ defined by*

$$p(n) = \#\{n\textit{-words in } \mathcal{L}(X)\}$$

is called the **word-complexity** *of X.*

Exercise 1.1.3. *Show that for the Fibonacci shift of Example 0.1.10, $p(n) = F_{n+1}$, where $F_1, F_2, F_3, F_4, F_5, \dots = 1, 2, 3, 5, 8, \dots$ are the Fibonacci numbers.*

If $p(n)$ increases exponentially fast, we call the exponential growth rate $h_{top}(\sigma) = \lim_n \frac{1}{n} \log p(n)$ the entropy of the subshift. For subshifts with zero entropy, the **polynomial growth rate** $\lim_{n\to\infty} \frac{\log p(n)}{\log n}$ can take any value in $[1, \infty]$, and also be zero

[1]We will rather not use this word, because of possible confusion with the factor of a subshift (= image under a sliding block code)

(if (X, σ) is periodic). We say that (X, σ) is of **sublinear complexity** if there is a constant C such that $p(n) \leq Cn$, i.e., the polynomial growth rate is 1. Sturmian sequence (see Section 3.3) have $p(n) = n + 1$. For one-dimensional shifts, the minimal complexity of interest is $p(n) = n + 1$, because if $p(n) \leq n$ for some n, then X consists of a single periodic orbit. This was proved by Morse & Hedlund [185], see Proposition 3.3.2.

Example 1.1.4. *The sequences*

$$x = \dots 000.10000 \dots \qquad \textit{and} \qquad y = 00001111.00000 \dots$$

both have $p(n) = n+1$. They are not recurrent, but asymptotically fixed for $n \to \pm\infty$ Ormes & Pavlov [188] showed that for non-recurrent shifts that are not asymptotically periodic, $\liminf_n p(n)/n \geq \frac{3}{2}$, and that this bound is sharp, as is demonstrated by

$$z = 0000.10^{n_0}10^{n_1}10^{n_2}10^{n_3}1 \dots$$

for the increasing sequence of gaps $(n_i)_{i \to \infty}$ is carefully chosen. See [113] for further results along this line.

Definition 1.1.5. *A map $T : [0, 1) \to [0, 1)$ is called an* **interval exchange transformation** *(***IET***) if there is a finite partition into half-open intervals $I_{i=1}^k$ such that $T|_{I_i}$ is a translation, and T is invertible. That is, the collection $\{I_i\}_{i=1}^k$ is put into $[0, 1)$ after a permutation.*

Instead of $[0, 1)$, we can define IETs on the circle $\mathbb{S}^1$. Every IETs on the intervals thus becomes a rotation. Every IET preserves Lebesgue measure.

Symbolic spaces associated with interval exchanges transformations on k intervals have $p(n) = (k-1)n + 1$. The Chacon substitution shift (see Example **??**) has word-complexity $p(n) = 2n + 1$, see [119]. For many subshift, the polynomial complexity can be bounded, but are hard to compute exactly; often $\lim_n p(n)/n$ doesn't exists (and if $\lim_n p(n)/n$, the the limit cannot be strictly between 1 and 2, as was shown in [144]), but $\liminf p(n)/n$ and $\limsup p(n)/n$ can be computed. For instance, the word-complexity of the Thue-Morse shift is

$$p(n) = \begin{cases} 3 \cdot 2^m + 4r & \text{if } 0 \leq r < 2^{m-1}, \\ 4 \cdot 2^m + 2r & \text{if } 2^{m-1} \leq r < 2^m, \end{cases}$$

where $n = 2^m + r + 1$. In [67], the word-complexity of certain (Fibonacci-like) unimodal restrictions to the critical ω-limit set are computed.

All substitution shifts, in fact all linearly recurrent have sublinear complexity, see Theorem 3.1.6.

Subshift with polynomial growth rate $d > 1$ are less studied. They emerge for example as symbolic spaces for polygonal billiards on d-dimensional billiard tables.

There is an analogous conjecture[2] for shifts on $\mathbb{Z}^2$: If $p_x(m,n)$ is the number of different $m \times n$-blocks in the two-dimensional infinite word $X \subset \{0,1\}^{\mathbb{Z}^2}$, and $p(m,n) \leq mn$ for some mn, then x is periodic, i.e., there is an integer vector $\vec{v}$ such that $x_{\vec{\imath}} = x_{\vec{\imath}+\vec{v}}$ for all $\vec{\imath} \in \mathbb{Z}^2$. Epifanio et al. [116] showed

$$p_x(m,n) \leq cmn \Rightarrow x \text{ is periodic} \tag{1.1}$$

for $c = 1/144$. This was later improved to $c = 1/16$ in [200]. Sander & Tijdeman [212] proved (1.1) for $m = 2$. A similar conjecture on $\mathbb{Z}^d$ for $d \geq 3$ is false, as shown in [226] and [210].

1.2 Word-frequencies and shift-invariant measures

In addition to the number of words, we can also study the **frequency** of words w appearing inside infinite sequences:

$$f_w(x) = \lim_{n\to\infty} \frac{1}{n} \#\{0 \leq i < n : x_i \dots x_{i+|w|-1} = w\}. \tag{1.2}$$

The question whether the limit exists and to what extent it depends on x is answer by the Birkhoff's Ergodic Theorem 4.2.3. For this we need a measure μ that assigns a number to every cylinder set, according to the rules

$$\begin{array}{rl} (i) & 0 \leq \mu(Z) \leq 1 \text{ for every cylinder } Z \\ (ii) & \mu(\varnothing) = 0, \ \mu(X) = 1; \\ (iii) & \mu(\bigcup_i Z_i) = \sum_i (Z_i) \quad \text{for all disjoint cylinders } Z_1, Z_2, \dots \end{array}$$

The Kolmogorov Extension Theorem (see Remark 1.2.1) implies that μ can be defined uniquely to every set in the σ-algebra $\mathcal{B}$ generated by the cylinder sets. In particular, $\{x \in X : f_w(x) = \alpha\} \in \mathcal{B}$ for every word w and $\alpha \in [0,1]$.

Remark 1.2.1. *The Kolmogorov Extension Theorem (see e.g. [20, Section 21.10]) is about extending probability measures μ_n on finite Cartesian products X_n to a measure on the infinite product X^∞. That is, if $\mu_{n+1}(A \times X) = \mu_n(A)$ for every $n \in \mathbb{N}$ and μ_n-measurable set $A \subset X^n$, then there is a unique probability measure μ on X^n such that $\mu(A \times X^\infty) = \mu_n(A)$ for every $n \in \mathbb{N}$ and μ_n-measurable set $A \subset X^n$.*

This carries over to indicator sets: Linear combinations of sets 1_A with $A \subset X^n$, $n \in \mathbb{N}$, lie dense in $L^1(\mu)$, i.e., for every $f \in L^1(\mu)$ and $\varepsilon > 0$ there is N and are finitely many sets $A_k \subset X^N$ and $a_k \in \mathbb{R}$ such that $\int_{X^\infty} |f - \sum_k a_k 1_{A_k}| \, d\mu < \varepsilon$.

[2]due to Maurice Nivat, invited talk at the International Colloquium on Automata, Languages, and Programming, Bologna 1987

Definition 1.2.2. *A measure μ on subshift (X, σ) is called* **invariant** *if $\mu(B) = \mu(\sigma^{-1}B)$ for all $B \in \mathcal{B}$.*

A measure is called **ergodic** *if $\sigma^{-1}(A) = A \pmod{\mu}$ for some $A \in \mathcal{B}$ implies that $\mu(A) = 0$ or $\mu(A^c) = 0$. That is, the only shift-invariant sets are nullsets or the whole space up to a nullset.*

Example 1.2.3. *The full shift $\{0,1\}^{\mathbb{N}}$ encodes the outcomes of a sequence of coin-flips, say $x_i = 0$ if the i-th gives a "head", and $x_i = 0$ if the i-th gives a "tail". If the coin has a bias, say "head" come up with probability $p > \frac{1}{2}$ and "tail" with probability $q = 1-p < \frac{1}{2}$, then the probability of a word can be computed by multiplying probabilities, e.g.*

$$\mathcal{P}(x_1x_2x_3x_4 = 0010) = p^3q$$

The corresponding measure μ_p is called the (p,q)-Bernoulli measure.

Definition 1.2.4. *Let $\mathcal{A} = \{1, 2, \dots, d\}$ and $X = \mathcal{A}^{\mathbb{N} \text{ or } \mathbb{Z}}$ be the full shift space. Let $p = (p_1, \dots, p_d)$ be a probability vector, i.e., $p_i \geq 0$ and $p_1 + \dots + p_d = 1$. The* **product measure** *that assigns to every cylinder set*

$$\mu_p([x_k \dots x_l]) = p_{x_k}p_{x_{k+1}} \cdots p_{x_l} \qquad \text{and } p(X) = 1$$

*is called the p-***Bernoulli measure***. The measure can be extended to the Borel σ-algebra by means of the Kolmogorov Extension Theorem. Each p-Bernoulli measure is shift-invariant.*

The Birkhoff Ergodic Theorem (see Theorem 4.2.3) implies that if μ is an ergodic shift-invariant probability measure, then

$$\mu([w]) = f_w(x) \quad \text{ for } \mu\text{-a.e. } x,$$

for every cylinder set $[w] = \{x \in X : x_0 \dots x_{|w|-1} = w\}$.

Definition 1.2.5. *A subshift (X, σ) is* **uniquely ergodic** *if it admits only one invariant probability measure. If (X, σ) is both uniquely ergodic and minimal, it is called* **strictly ergodic**.

This should not be confused with **intrinsically ergodic** *which means that there is a unique measure of maximal entropy, see Definition* **??**

The full shift is obviously not uniquely ergodic; it has for instance a Bernoulli measure for every probability vector p. Neither are SFTs, sofic shifts or β-shifts (which are in fact, intrinsically ergodic). Sturmian shifts on the other hand are uniquely ergodic.

Proposition 1.2.6. *A recurrent subshift (X, σ) is uniquely ergodic if and only if $f_w(x)$ exists and is the same for every $x \in X$. In this case, the convergence in the limit (1.2) is uniform in x.*

Proof. If μ and ν were two different ergodic measures, then there is a word $w \in \mathcal{L}(X)$ such that $\mu([w]) \neq \nu([w])$. Using the Ergodic Theorem for both measures (with their own typical points x and y), we see that

$$f_w(x) = \mu([w]) \neq \nu([w]) = f_w(y),$$

so the limit is not independent of x.

Conversely, we know by the Ergodic Theorem that $f_w(x) = \mu([w])$ is constant μ-a.e. But if the convergence is not uniform, then there are sequences $(x^i), (y^i) \subset X$ and $(m_i), (n_i) \subset \mathbb{N}$, such that

$$\lim_i \frac{1}{m_i} \sum_{k=0}^{m_i-1} \mathbf{1}_w \circ \sigma^k(x^i) \neq \lim_i \frac{1}{n_i} \sum_{k=0}^{n_i-1} \mathbf{1}_w \circ \sigma^k(y^i).$$

Define functionals $\mu_i, \nu_i : C(X) \to \mathbb{R}$ as $\mu_i(g) = \liminf_i \frac{1}{m_i} \sum_{k=0}^{m_i-1} g \circ \sigma^k(x^k)$ and $\nu_i(g) = \liminf_i \frac{1}{n_i} \sum_{k=0}^{n_i-1} g \circ \sigma^k(y^k)$. Both sequences have weak accumulation points μ and ν which are easily shown to be σ-invariant measures, see the proof of Theorem 4.0.2.

More precisely, since $(C(X), \| \ \|_\infty)$ is a separable Banach space, we can find a countable dense subset $(g_j)_{j\in\mathbb{N}}$ and (by a diagonal argument) we can take subsequences of $(m_i)_{i\in\mathbb{N}}$ and $(n_i)_{i\in\mathbb{N}}$ along which $\mu_i(g_j)$ and $\nu_i(g_j)$ converge for all $j \in \mathbb{N}$.

But μ and ν are not the same, because if we take a subsequence $(j_r)_{r\in\mathbb{N}}$ such that $g_{j_r} \to 1_w$, then $\lim_r \mu(g_{j_r}) = \mu(1_w) \neq \nu(1_w) = \lim_r \nu(g_{j_r})$. Hence (X, T) cannot be uniquely ergodic. □

1.3 Basic notions from dynamical systems

A **dynamical system** is a mathematical description of how a physical system evolves in time. It consists of

- a **phase space** X, usually a metric space, or at least topological space, describing the position of the system. For example, $\mathbb{R}^{2n}$ can be used to describe the positions and velocities of n point-particles moving along a line, or $\mathbb{R}^{6n}$ for the positions and velocities of n point-particles moving in $\mathbb{R}^3$.

- a **time space**, which could be $\mathbb{R}$ (for continuous time) or $\mathbb{N}_0$ (or $\mathbb{Z}$) for when the observations are only made a discrete time steps. More complicated (multi-dimensional or group-valued) time can be considered too, but in this text, time is always discrete: $\mathbb{N}_0$ or (when the system is time-invertible) $\mathbb{Z}$.

- an **evolution rule**, which for discrete time take the form of a mapping $T : X \to X$ satisfying

1. $T^0(x) = x$ for all $x \in X$.
2. $T^{m+n}(x) = T^m(T^n(x))$ for all $m, n \in \mathbb{N}_0$ (or $\mathbb{Z}$) and all $x \in X$.

This is realized when we let T^n be the n-fold composition:

$$T^n(x) = \underbrace{T \circ T \circ \cdots \circ T}_{n \text{ times}}$$

and T^{-1} is the inverse transformation when it exists.

Examples on one-dimensional phase space are the **quadratic family**

$$Q_a : [0,1] \to [0,1], \quad x \mapsto ax(1-x), \qquad a \in [0,4],$$

the family of **β-transformations**

$$T_\beta : [0,1] \to [0,1], \quad x \mapsto \beta x \pmod 1, \qquad \beta \in \mathbb{R}$$

and the family of **circle rotations**

$$R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1, \quad x \mapsto x + \alpha \pmod 1 \qquad \alpha \in [0,1).$$

See Figure 1.1

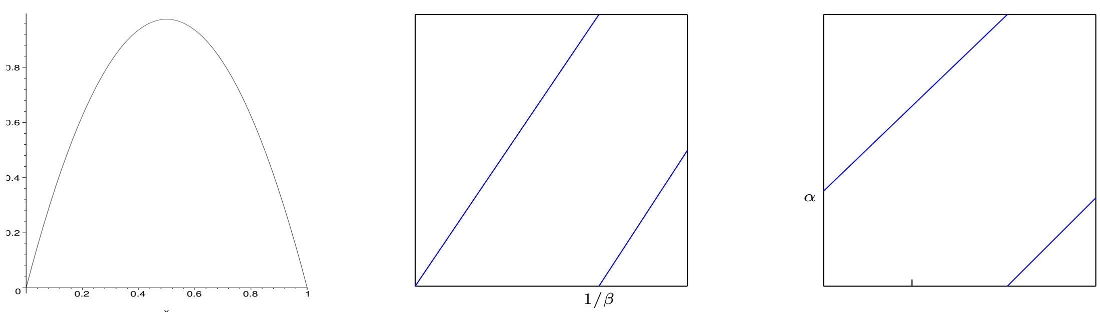


Figure 1.1: A quadratic map, a β-transformation and a circle rotation

The **orbit** of $x \in X$ is the set $\mathrm{orb}(x) = \{T^n(x) : n \in \mathbb{N}_0 \text{ or } \mathbb{Z}\}$. We distinguish several types of orbit. Namely, $\mathrm{orb}(x)$ is

- **periodic** if $T^n(x) = x$ for some $n \in \mathbb{N}$. The smallest such n is called the **period** of x. If the period is 1, then x is called a **fixed point**.

- **preperiodic** if $T^{n+m}(x) = T^m(x)$ for some $m, n \in \mathbb{N}$. The minimal such m, n are called the **preperiod** and **period** of x.

- **asymptotically periodic** if there is a periodic point $y \notin \mathrm{orb}(x)$ such that $d(T^n(x), T^n(y) \to 0$ as $n \to \infty$. In this case, y is an attracting periodic point. More precisely, y is an **(one-sided) attracting periodic point** if it is periodic and has a (one-sided) neighborhood U such that $\lim_{n\to\infty} d(T^n(x), T^n(y) = 0$ for all $x \in U$. If no such neighborhood exists, then y is **repelling**.

For example, for the quadratic family with $a = 3.83187405528332\dots$, the point $x = \frac{1}{2}$ has period 3, and since $Q_a'(\frac{1}{2}) = 0$, it is easy to show that $\frac{1}{2}$ is attracting. The two fixed points are 0 and $1 - \frac{1}{a}$; they are **repelling**. For the circle rotation R_α, every point is periodic if and only if $\alpha \in \mathbb{Q}$: $x = p/q$ in lowest terms, then every point has period q. If $\alpha \notin \mathbb{Q}$, then every orbit is dense in $\mathbb{S}^1$.

Definition 1.3.1. *Two subshifts (X, σ) and (Y, σ) are called* **conjugate** *if there is a homeomorphism $\psi : X \to Y$ such that $\psi \circ \sigma = \sigma \circ \psi$.*

If $\psi : X \to Y$ commutes with σ and is a continuous, onto, but not necessarily one-to-one map, then Y is called a **factor** *of X, and X is called an* **extension***. This extension is* **almost one-to-one** *if there is a dense set of $y \in Y$ such that $\#\psi^{-1}(y) = 1$.*

A conjugacy (or factor map) $\psi : X \to Y$ is called **pointed** *if it sends specified points $x \in X$ and $y \in Y$ to each other.*

Lemma 1.3.2. *Let (X, f) and $Y, g)$ be systems that are conjugate via $g \circ \psi = \psi \circ f$. Then*

1. *If p is a (pre)periodic point for f, then $\psi(p)$ is a (pre)periodic point of g, and the (pre)periods are the same.*
2. *If f, g are continuous, then the conjugacy preserves ω-limit sets: $\psi(\omega(x)) = \omega(\psi(x))$.*
3. *If the periodic point p is attracting/repelling, then $\psi(p)$ is also attracting/repelling.*

Proof. First note that

$$\begin{aligned} \psi \circ f^n &= \psi \circ f \circ \psi^{-1} \circ \psi \circ f \circ \psi^{-1} \circ \psi \circ f \circ \psi^{-1} \circ \cdots \circ f \\ &= g \circ \psi \circ \psi^{-1} \circ g \circ \psi \circ \psi^{-1} \circ g \circ \psi \circ \psi^{-1} \circ \cdots \circ g \circ \psi = g^n \circ \psi. \end{aligned}$$

1. Take p such that $f^n(p) = p$ and $q = \psi(p)$. Then $g^n(q) = g^n \circ \psi(p) = \psi \circ f^n(p) = \psi(p) = q$, so q if n-periodic for g. Next, suppose that $f^{m+n}(p) = f^m(p)$, and set $\psi(p) = q$, $p' = f^m(p)$. Then $g^{m+n}(q) = g^{m+n} \circ \psi(p) = \psi \circ f^{m+n}(p) = \psi \circ f^m(p) = g^m \circ \psi(p) = g^m(q)$.
2. Now assume that $x \in \omega_f(a)$, so there is a sequence $n_k \to \infty$ such that $f^{n_k}(a) \to x$. Set $y = \psi(x)$ and $b = \psi(a)$. Then, by continuity of f, $g^{n_k}(b) = g^{n_k} \circ \psi(a) = \psi \circ f^{n_k}(a) \to \psi(x) = y$, so $y \in \omega_g(b)$.
3. If $p = f(p)$ is asymptotically attracting, then for every a sufficiently close to p, we have $p = \omega_f(a)$. By part 1., $q := \psi(p)$ is a fixed point of g, and by part 2., $q = \omega_g(y)$ for $y = \psi(x)$. □

Exercise 1.3.3. *Is the following true: if X is a factor of Y and Y a factor of X, then X and Y are conjugate?*

Example 1.3.4. *The quadratic map $Q(x) = 2x^2 - 1$ on $[-1, 1]$ (the quadratic* **Chebyshev polynomial** *$Ч_2$) is conjugate to the tent map $T(x) = \min\{2x, 2(\pi - x)\}$ on $[0, \pi]$. Indeed,*

$$Q \circ \psi = \psi \circ T \quad \text{for} \quad \psi(x) = \cos x, \tag{1.3}$$

It is very unusual to find smooth conjugacies between maps, and even here, ψ is not diffeomorphic at the endpoints $0, 1$. But applying (1.3) n times and then differentiating, we find

$$(Q^n)' \circ \psi(x) \cdot \psi'(x) = \psi'(T^n(x)) \cdot (T^n)'(x).$$

If x is an n-periodic point of T, and hence $y = \psi(x)$ and n-periodic point of Q, we see that $|(Q^n)'(y)| = 2^n$. The only periodic point where this fails is $y = 0$, because ψ^{-1} is not differentiable at 0.

Note that the same conjugacy works for the degree n Chebyshev polynomial $Ч_n$ and the slope n tent map with n branches. The characterization of Chebyshev polynomial $Ч_n(x) = \cos(n \arccos x)$ is the cause of this.

Example 1.3.5. *We show that two circle rotations R_α and R_β are not conjugate if $0 \leq \alpha < \beta < 1$. Let $<$ denote the positive orientation on $\mathbb{S}^1$. Choose $n \in \mathbb{N}$ such that $n\alpha \leq k < n\beta$ and $(n-1)\beta \leq k$ for some integer k. Then, setting $y = \psi(0)$,*

$$R_\alpha^n(0) \leq 0 \leq R_\alpha(0) \quad \text{and} \quad y \leq R_\beta^n(y) \leq R_\beta(y). \tag{1.4}$$

The homeomorphism $\psi : \mathbb{S}^1 \to \mathbb{S}^1$ must either preserve or reverse the orientation of the circle, but neither way is comparable with (1.4). Therefore there cannot be any conjugacy.

A more structural way to see this is using lifts and rotation numbers, see Theorem **??**. *Indeed, the rotation number $\rho(f)$ is preserved on conjugacy, and $\rho(R_\alpha) = \alpha \neq \beta = \rho(R_\beta)$.*

1.4 Symbolic itineraries

Symbolic dynamics emerges from a dynamical system (X, T) by coding the T-orbits of the points $x \in X$. To this end, we let $\mathcal{J} = \{J_a\}_{a \in \mathcal{A}}$ (for a finite or countable alphabet $\mathcal{A}$) be a partition of X, and to each $x \in X$ we assign an **itinerary** $\boldsymbol{i}(x) \in \mathcal{A}^{\mathbb{N}_0}$:

$$\boldsymbol{i}_n(x) = a \quad \text{if } T^n(x) \in J_a.$$

If T is invertible, then we can extend sequences to $\mathcal{A}^{\mathbb{Z}}$. It is clear that $\boldsymbol{i} \circ T(x) = \sigma \circ \boldsymbol{i}(x)$. Therefore, if we set $\Sigma = \boldsymbol{i}(X)$, then $\sigma(\Sigma) \subset \Sigma$ and if $T : X \to X$ is surjective, then $\sigma(\Sigma) = \Sigma$. But (Σ, σ) is in general not a subshift, because Σ is not closed.

Example 1.4.1. *Let $X = [0,1]$ and $T(x) = Q_4(x) = 4x(1-x)$. Let $J_0 = [0, \frac{1}{2}]$ and $J_1 = (\frac{1}{2}, 1]$. Then $\boldsymbol{i}(X)$ is not closed, because there is no $x \in [0,1]$ such that $\boldsymbol{i}(x) = 1100000\dots$, while $1100000\dots = \lim_{x \searrow \frac{1}{2}} \boldsymbol{i}(x)$. Naturally, redefining the partition to $J_0 = [0, \frac{1}{2})$ and $J_1 = [\frac{1}{2}, 1]$ doesn't help, because then there is no $x \in [0,1]$ such that $\boldsymbol{i}(x) = 0100000\dots$, while $0100000\dots = \lim_{x \nearrow \frac{1}{2}} \boldsymbol{i}(x)$.*

Other "solutions" that one sees in the literature are:

- *Assigning a different symbol (often $*$ or C) to $\frac{1}{2}$. That is, using the partition $J_0 = [0, \frac{1}{2})$, $J_* = \{\frac{1}{2}\}$ and $J_1 = (\frac{1}{2}, 1]$. This resolves the "ambiguity" about which symbol to give to $\frac{1}{2}$, but it doesn't make the shift space closed.*
- *Assigning the two symbols to $\frac{1}{2}$, so $J_0 = [0, \frac{1}{2}]$ and $J_1 = [\frac{1}{2}, 1]$ are no longer a partition, but have $\frac{1}{2}$ in common. Therefore $\frac{1}{2}$ will have two itineraries, and so will every point in the backward orbit of $\frac{1}{2}$. With all these extra itineraries, $\boldsymbol{i}(X)$ becomes closed. But this doesn't work in all cases, see Exercise 1.4.2.*
- *Taking a quotient space $\boldsymbol{i}(X)/\sim$ where in this case $x \sim y$ if there is $n \in \mathbb{N}_0$ such that*

$$x_0 \dots x_{n-1} = y_0 \dots y_{n-1} \text{ and } \begin{cases} x_n x_{n+1} x_{n+2} x_{n+3} x_{n+4} \dots = 11000\dots, \\ y_n y_{n+1} y_{n+2} y_{n+3} y_{n+4} \dots = 01000\dots \end{cases}$$

  or vice versa. This quotient space adapts the quotient topology (so $\boldsymbol{i}(X)/\sim$ is not a Cantor set anymore), and it turns the coding map $i : [0,1] \to \{0,1\}^{\mathbb{N}_0}/\sim$ into a genuine homeomorphism.

Exercise 1.4.2. *Let $a = 3.83187405528332\dots$ and $T(x) = Q_a(x) = ax(1-x)$. For this parameter, $T^3(\frac{1}{2}) = \frac{1}{2}$. Let $\mathcal{J}' = \{[0, \frac{1}{2}], (\frac{1}{2}, 1]\}$ and $\mathcal{J} = \{[0, \frac{1}{2}], [\frac{1}{2}, 1]\}$, so $\frac{1}{2}$ get two symbols. Let $\Sigma' = \boldsymbol{i}(X)$ w.r.t. $\mathcal{J}'$ and $\Sigma = \boldsymbol{i}(X)$ w.r.t. $\mathcal{J}$. Show that $\overline{\Sigma'} \neq \Sigma$.*

From now on, assume that X is compact metric space without isolated points. We will now discuss the properties of the coding map i itself. First of all, for i to be continuous it is crucial that $T|_{J_a}$ is continuous on each element $J_a \in \mathcal{J}$. But this is not enough: if x is a common boundary of two element of $\mathcal{J}$ then (no matter how you assign the symbol to x in Example 1.4.1), for each neighborhood $U \ni x$, $\operatorname{diam}(\boldsymbol{i}(U)) = 1$, so continuity fails at x. It is only by using quotient spaces of $\boldsymbol{i}(X)$ (so changing the topology of $\boldsymbol{i}(X)$) that can make i continuous. Normally, we choose to live with the discontinuity, because it affects only few points:

Lemma 1.4.3. *Let $\partial\mathcal{J}$ denote the collection of common boundary points of different elements in $\mathcal{J}$. If $orb(x) \cap \partial J = \varnothing$, then the coding map $i : X \to \mathcal{A}^{\mathbb{N}_0}$ or $\mathcal{A}^{\mathbb{Z}}$ is continuous at x.*

Proof. We carry out the proof for invertible maps. Let $\varepsilon > 0$ be arbitrary and fix $N \in \mathbb{N}$ such that $2^{-N} < \varepsilon$. For each $n \in \mathbb{Z}$ with $|n| \leq N$, let $U_n \ni T^n(x)$ be such a small neighborhood that it is contained in a single partition element $J_{\boldsymbol{i}_n(x)}$. Since $\mathrm{orb}(x) \cap \partial J = \varnothing$, this is possible. Then $U := \cap_{|n| \le N} T^n(U_n)$ is an open neighborhood of x and $\boldsymbol{i}_n(y) = \boldsymbol{i}_n(x)$ for all $|n| \le N$ and $y \in U$. Therefore $\mathrm{diam}(\boldsymbol{i}(U)) \le 2^{-N} < \varepsilon$, and continuity at x follows. □

Lemma 1.4.4. *Suppose that T is a continuous expansive dynamical system and injective on each $J_a \in \mathcal{J}$. If the expansivity constant is larger than $\sup_{a \in \mathcal{A}} diam(J_a)$, then the coding map $i : X \to \mathcal{A}^{\mathbb{N}_0}$ or $\mathcal{A}^{\mathbb{Z}}$ is injective.*

Proof. Suppose that there are $x \ne y \in X$ such that $\boldsymbol{i}(x) = \boldsymbol{i}(y)$. Since $T|_{J_a}$ is injective for each $a \in \mathcal{A}$, $T^n(x) \ge T^n(y)$ for all $n \ge 0$. Let $\delta > 0$ be an expansivity constant of T. Thus, there is $n \in \mathbb{Z}$ such that $d(T^n(x), T^n(y)) > \delta$, so, by assumption, they cannot lie in the same element of $\mathcal{J}$. Hence x and y cannot have the same itinerary after all. □

To obtain injectivity of the coding map, it often suffices that T is expanding on each partition element J_a. Expanding (and expansion) should not be confused with expansive (and expansivity) of Definition **??**.

Definition 1.4.5. *Let $T : X \to Y$ be a map between metric spaces. We call T* **expanding** *if there is $\rho > 1$ such that $d_Y(T(x), T(y)) \ge \rho d_X(x, y)$ for all $x, y \in X$. and* **locally expanding** *there are $\varepsilon > 0$ and $\rho > 1$ such that $d(T(x), T(y)) \ge \rho d(x, y)$ for all $x, y \in Y$ with $d(x, y) < \varepsilon$.*

Example 1.4.6. *If $X = Y$ is compact, then it carries no expanding map (non-compact examples exist, e.g. $T : \mathbb{R} \to \mathbb{R}$, $x \mapsto 2x$). Local expandingness is less restrictive:*

Let $T : \mathbb{S}^1 \to \mathbb{S}^1$, $x \mapsto 2x \pmod 1$, be the doubling map, and $J_0 = (\frac{1}{4}, \frac{3}{4})$ and $J_1 = \mathbb{S}^1 \setminus J_0$. Clearly $T'(x) = 2$ for all $x \in \mathbb{S}^1$, but T is not expanding on the whole of $\mathbb{S}^1$, because for instance $d(T(\frac{1}{4}), T(\frac{3}{4})) = 0 < \frac{1}{2} = d(\frac{1}{4}, \frac{3}{4})$. More importantly, T is not expanding on the either J_a; for example $d(T(\frac{1}{4} + \varepsilon), T(\frac{3}{4} - \varepsilon)) = 4\varepsilon < \frac{1}{2} - 2\varepsilon = d(\frac{1}{4} + \varepsilon, \frac{3}{4} - \varepsilon)$ for each $\varepsilon \in (0, \frac{1}{12})$. The corresponding coding map is **not** *injective. The way to see this by noting that the involution $S(x) = 1 - x$ commutes with T and also preserves each J_a. It follows that $\boldsymbol{i}(x) = \boldsymbol{i}(S(x))$ for all $x \in \mathbb{S}^1$, and only $x = 0$ and $x = \frac{1}{2}$ have unique itineraries.*

Proposition 1.4.7 (Gottschalk & Hedlund [134])**.** *Let $T : X \to X$ be a homeomorphism on a compact metric space. If T is locally expanding, then X is finite.*

Proof. Let $\eta > 0$, $\rho > 1$ be as in Definition 1.4.5, and take $\varepsilon \in (0, \eta)$ arbitrary. Since T^{-1} is continuous and X is compact, there is a uniform $\delta > 0$ such that $d(x, y) < \delta$ implies $d(f^{-1}(x), f^{-1}(y)) < \varepsilon$. Let $\{U_i\}_{i=1}^N$ be a finite open cover of X such that $\mathrm{diam}(U_i) < \delta$. Then $\{f^{-1}(U_i)\}_{i=1}^N$ is an open cover of X, and $\mathrm{diam} f^{-1}(U_i) < \varepsilon$, so by local expansion, $\mathrm{diam} f^{-1}(U_i) < \mathrm{diam}(U_i)/\rho \le \delta/\rho$. Repeating this argument, we find that $\{f^{-n}(U_i)\}_{i=1}^N$ is a finite open cover of X with $\mathrm{diam}(f^{-n}(U_i)) < \delta\rho^{-n}$. Since n is arbitrary, $\#X \le N$. □

1.5 Minimal subshifts

The following definition express that all parts of a subshift connect to each other:

Definition 1.5.1. *A subshift X is* **transitive** *or* **irreducible** *if for every $u, w \in \mathcal{L}(X)$, there is $v \in \mathcal{L}(X)$ such that $uvw \in \mathcal{L}(X)$. It is called* **totally transitive** *if σ^n is transitive for each $n \in \mathbb{N}$.*

Clearly totally transitive is stronger than transitive. Conversely, the two-point shift $(\{(10)^\infty, (01)^\infty\}, \sigma)$ is transitive but σ^2 is not.

Proposition 1.5.2. *A subshift is transitive if and only if there exists a dense orbit.*

Remark 1.5.3. *The notion of dense orbit may need further explanation if the subshift is two-sided. Consider the sequence*

$$x = \dots 1010001010000000000101000101.0000000000000000000000000\dots \tag{1.5}$$

which emerges from the Cantor substitution from the seed 1.0*, see Example* **??***. This sequence has a dense backward orbit $orb^-(x)$ within its backward orbit closure $\overline{orb^-(x)}$ as well as a dense forward orbit $orb^+(x)$ within its forward orbit closure $\overline{orb^+(x)}$. However, $orb^+(x)$ is not dense in its two-sided orbit closure.*

Proof. Suppose first that $\mathrm{orb}(x)$ is dense. Then for every $u, w \in \mathcal{L}(X)$ there are $m < m + |u| \le n \in \mathbb{N}$ such that $\sigma^m(x) \in [u]$ and $\sigma^n(x) \in [w]$. (Recall that $[v]$ denotes the cylinder set associated to the word v.) Now let v be the word of length $n - (m + |u|)$ such that $\sigma^{m+|u|}(x) \in [v]$. Then $uvw \in \mathcal{L}(X)$.

Conversely, let $(u^j)_{j\in\mathbb{N}}$ be a denumeration of $\mathcal{L}(X)$. We construct a sequence of words v^j recursively. Assume by induction that $u^1v^1 \dots v^{j-1}u^j \in \mathcal{L}(X)$. By transitivity, we can find v^j such that $u^1v^1 \dots v^{j-1}u^jv^ju^{j+1} \in \mathcal{L}(X)$. Now set $x = u^1v^1u^2v^2 \dots$. Then $\mathrm{orb}(x)$ is dense in X. □

A strong form of transitivity is minimality:

Definition 1.5.4. *A subshift (X, σ) is* **minimal** *if every orbit is dense in X.*

Remark 1.5.5. *It is a straightforward application of Zorn's Lemma that every dynamical systems on a compact space[3] contains at least one minimal subsystem.*

Proposition 1.5.6. *We have the following equivalent characterizations for a subshift (X, σ) to be* **minimal***:*

(i) There is no closed shift-invariant proper subset of X;

[3]Compactness is important, otherwise one could take a single non-recurrent orbit (but without its closure) as the phase-space. An interesting example with only recurrent orbits but no minimal subset is due to Auslander [13, page 27]

(ii) Every orbit is dense in X (see Definition 1.5.4);

(iii) There is one dense orbit and σ is **uniformly recurrent**[4]*, i.e., for every open set $U \subset X$ there is $N \in \mathbb{N}$ such that for every $x \in U$ there is $1 \leq n \leq N$ such that $\sigma^n(x) \in U$.*

Definition 1.5.7. *Uniform recurrence means that the sets of integers n such that $\sigma^n(x) \in U$ has is* **syndetic***, i.e., it has* **bounded gaps** *(from the Greek $\sigma\upsilon\nu\delta\varepsilon\tau\iota\kappa o\varsigma$ = bound together). A set that is not syndetic has a complement that is* **thick***: for every $N \in \mathbb{N}$ contains blocks $\{n, n+1, \dots, n+N\}$.*

Proof. We prove the three implications by the contrapositive.
(i) ⇒ (ii): Suppose that $x \in X$ has an orbit that is not dense. Then $\overline{\text{orb}(x)}$ is a shift-invariant closed proper subset, so (i) fails.

(ii) ⇒ (iii): By (i) every orbit is dense, so there is at least one dense orbit.

Now to prove uniform recurrence, let U be any open set. Due to product topology, U contains a cylinder set U_0; in particular U_0 is clopen. Suppose that for every $N \in \mathbb{N}$ there is $x_N \in U_0$ such that $\sigma^n(x_N) \notin U_0$ for all $1 \leq n \leq N$. Let x be an accumulation point of $(x_N)_{N\in\mathbb{N}}$; since U_0 is closed, $x \in U_0$. Suppose by contradiction that there is $n \geq 1$ such that $\sigma^n(x) \in U_0$. Take an open set $V \ni x$ such that $\sigma^n(V) \subset U_0$. Next take $N \geq n$ so large that $x_N \in V$. But this means that $\sigma^n(x_N) \subset U_0$, which is against the definition of x_N. Hence no such n exists, and therefore $\text{orb}(x)$ is not dense, and (ii) fails.
Now take $y \in U$ arbitrary (so not necessarily in U_0), and $x \in U_0$ with a dense orbit. Find a sequence k_i such that $\sigma^{k_i}(x) \to y$. For each i there is $1 \leq n_i \leq N$ such that $\sigma^{k_i+n_i}(x) \in U_0$. Passing to a subsequence, we may as well assume that $n_i \equiv n$. Then $\sigma^n(y) = \sigma^n(\lim_i \sigma^{k_i}(x)) = \lim_i \sigma^{k_i+n}(x) \in U_0 \subset U$. This proves the uniform recurrence of U.

(iii) ⇒ (i): Let x be a point with a dense orbit. Suppose that Y is a closed shift-invariant proper subset of X and let $U \subset X$ be open such that $\overline{U} \cap Y = \varnothing$. Let $n \geq 0$ be minimal such that $u := \sigma^n(x) \in U$. Let $N = N(U) \geq 1$ be as in the definition of uniform recurrence, and let $y \in Y$ be arbitrary. Since $\text{orb}(y) \subset Y$, there is an open set $V \ni y$ such that $\sigma^i(V) \cap U = \varnothing$ for $0 \leq i \leq N$.

Take $n'' > n$ minimal such that $\sigma^{n''}(u) \in V$, and let $n' < n''$ be maximal such that $\sigma^{n'}(u) =: u' \in U$. Then $\sigma^i(u') \notin U$ for all $1 \leq i \leq n'' - n' + N$. Since N was arbitrary, this contradicts the uniform recurrence and hence such Y cannot exist. □

Definition 1.5.8. *A subshift is called* **periodically recurrent** *if for every open set U, there is N such that $\sigma^{kN}(U) \subset U$ for all $k \in \mathbb{N}$ (or $k \in \mathbb{Z}$ if the shift is invertible).*

Since periodic recurrence is obviously stronger than uniform recurrence, we have the following corollary.

[4] The word "almost periodic" is sometimes used as well, e.g. in [134, 169, 175], but it is not the same with all authors. For instance, in [207] it is used as "periodically recurrent" in Definition 1.5.8.

Corollary 1.5.9. *Every periodically recurrent subshift is minimal.*

Definition 1.5.10. *A dynamical system (X,T) on a metric space (X,d) is* **uniformly rigid** *if for every $\varepsilon > 0$ there is an iterate n such that $d(T^n(x), x) < \varepsilon$,*

Lemma 1.5.11. *A subshift (X,σ) is uniformly rigid if and only if it is periodically recurrent.*

Proof. $\Rightarrow$: Take $\varepsilon > 0$ arbitrary with corresponding iterate n, and let $k \in \mathbb{N}$ such that $2^{-k} < \varepsilon$. Thus the distance between every two distinct k-cylinders Z in X is at least ε. By rigidity $\sigma^n(Z) = Z$, and therefore $\sigma^{in}(Z) = Z$ for all $i \geq 0$, proving periodic recurrence.
$\Leftarrow$: Let $\varepsilon > 0$ be arbitrary. For each $x \in X$, we can find a neighborhood U_x of $\mathrm{diam}(U_x) < \varepsilon$ and iterate n_x such that $\sigma^{n_x}(U) \subset U$. By compactness, there is a finite collection $x_1, \dots, x_N$ such that $X = \bigcup_{i=1}^N U_{x_i}$. Take $n = \mathrm{lcm}\{x_1, \dots x_N\}$. Then $d(\sigma^n(x), x) < \varepsilon$ for each $x \in X$, as required. $\square$

The following weakening of minimality is of importance for e.g. Toeplitz shifts and $\mathcal{B}$-free shifts

Definition 1.5.12. *A subshift (X,σ) is called* **essentially minimal** *if it contains a unique minimal set Y, i.e., a unique non-empty closed shift-invariant set.*

Clearly, essentially minimal shifts can have at most one periodic orbit, but as the subshift $X := \{\sigma_k(\dots 000001000000 \dots)\}_{k\in\mathbb{Z}} \cup \{0^\infty\}$ shows, $X \setminus Y \neq \varnothing$ is possible. However, the two-sided orbit closure of (1.5) does not give an essentially minimal shift.

Proposition 1.5.13. *Given a subshift (X,σ) and a point $y \in X$, the following are equivalent:*

(i) *(X,σ) is essentially minimal and y is contained in its minimal set.*

(ii) *For every $x \in X$, $\omega(x) \ni y$.*

If in addition, σ is invertible, then two further equivalent statements are:

(iii) *For every $x \in X$, $\alpha(x) \ni y$.*

(iv) *For every open set $U \ni y$, $\cup_{n\in\mathbb{Z}} sigma^n(U) = X$.*

Proof. **(i) $\Rightarrow$ (ii)**: $\omega(x)$ is a closed non-empty shift-invariant set, so by Zorn's Lemma, it contains a minimal set. But Y is the unique minimal set, so $y \in \omega(x)$.
(ii) $\Rightarrow$ (i): Assume by contradiction that $y \in Y$ and Y' are minimal sets, and take $x \in Y$, $x' \in Y'$. By assumption $y \in \omega(x) \cap \omega(x')$, so $y \in Y \cap Y'$. Thus $Y \cap Y'$ is a non-empty, closed and shift-invariant subset of both Y and Y'. Since Y and Y' are minimal, $Y = Y \cap Y' = Y'$.

(i) $\Leftrightarrow$ **(iii)**: Use the above with σ^{-1} instead of σ.
(i) $\Rightarrow$ **(iv)**: Let U be an arbitrary neighborhood of y. Since $\cup_{n\in\mathbb{Z}}\sigma^n(U)$ is an open (two-sided!) shift-invariant set, its complement Y' is closed and shift-invariant. If $Y' \neq \varnothing$, then it contains a minimal set that is disjoint from y, contradicting essential minimality. Hence $\cup_{n\in\mathbb{Z}}\sigma^n(U) = X$.
(iv) $\Rightarrow$ **(iii)**: Let $x \in X$ be arbitrary; we can assume without loss of generality that $x \neq \sigma^k(y)$ for some $k \geq 0$, because if y is periodic then $\alpha(x) = \text{orb}(y) \ni y$, and otherwise we replace x by $\sigma^{-(k+1)}(x)$ to get it outside the forward orbit of y. Let $(U_r)_{r\in\mathbb{N}}$ be a nested sequence of neighborhoods of y such that $\cap_r U_r\{y\}$. Since $\cup_{n\in\mathbb{Z}}\sigma^n(U_r) = X$ and X is compact, there is a finite N_r such that $\cup_{n=-N_r}^{N_r}\sigma^n(U) = X$. Applying σ^{N_r} to both sides, we obtain $\cup_{n=0}^{2N_r}\sigma^n(U) = X$. Thus there is $n_r \leq 2N_r$ such that $\sigma^{-n_r}(x) \in U_r$. As we can do this for every r, we have found a sequence (n_r) (and $n_r \to \infty$ because $x \neq \sigma^k(y)$ for any $k \geq 0$) such that $\sigma^{-n_r}(x) \to y$. Thus $y \in \alpha(x)$, as required. □

1.6 Topological entropy

The notion of **topological entropy** by Adler, Konheim & McAndrew [2] dates back to 1969, but nowadays, the definition due to the American mathematician Rufus Bowen [48] and, independently, his Russian colleague Efim Dinaburg [100] is most often used. It is a measure of disorder of the system, and one common definition of chaos is that the topological entropy is positive.

Let T be map of a compact metric space (X, d). If my eyesight is not so good, I cannot distinguish two points $x, y \in X$ if they are at a distance $d(x, y) < \varepsilon$ from one another. I may still be able to distinguish their orbits, if $d(T^k x, T^k y) > \varepsilon$ for some $k \geq 0$. Hence, if I'm willing to wait $n - 1$ iterations, I can distinguish x and y if

$$d_n(x, y) := \max\{d(T^k x, T^k y) : 0 \leq k < n\} > \varepsilon.$$

If this holds, then x and y are said to be (n, ε)**-separated**. Among all the subsets of X of which all points are mutually (n, ε)-separated, choose one, say $E_n(\varepsilon)$, of maximal cardinality. Then $s_{n,\varepsilon} := \#E_n(\varepsilon)$ is the maximal number of n-orbits I can distinguish with ε-poor eyesight.

The **topological entropy** is defined as the limit (as $\varepsilon \to 0$) of the exponential growth-rate of $s_{n,\varepsilon}$:

$$h_{top}(T) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log s_{n,\varepsilon}. \tag{1.6}$$

Note that $s_n(\varepsilon_1) \geq s_n(\varepsilon_2)$ if $\varepsilon_1 \leq \varepsilon_2$, so $\limsup_n \frac{1}{n} \log s_{n,\varepsilon}$ is a decreasing function in ε, and the limit as $\varepsilon \to 0$ indeed exists.

Instead of (n, ε)-separated sets, we can also work with (n, ε)**-spanning** sets, that is, sets that contain, for every $x \in X$, a point y such that $d_n(x, y) \leq \varepsilon$. Note that, due to its maximality, $E_n(\varepsilon)$ is always (n, ε)-spanning, and no proper subset of $E_n(\varepsilon)$ is

(n,ε)-spanning. Each $y \in E_n(\varepsilon)$ must have a point of an $(n,\varepsilon/2)$-spanning set within an $\varepsilon/2$-ball (in d_n-metric) around it, and by the triangle inequality, this $\varepsilon/2$-ball is disjoint from $\varepsilon/2$-ball centered around all other points in $E_n(\varepsilon)$. Therefore, if $r_n(\varepsilon)$ denotes the minimal cardinality among all (n,ε)-spanning sets, then

$$r_n(\varepsilon) \le s_{n,\varepsilon} \le r_n(\varepsilon/2). \tag{1.7}$$

Thus we can equally well define

$$h_{top}(T) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon). \tag{1.8}$$

Example 1.6.1. *Consider the β-transformation $T_\beta : [0,1) \to [0,1)$, $x \mapsto \beta x \pmod 1$ for some $\beta > 1$. Take $\varepsilon < 1/(2\beta^2)$, and $G_n = \{\frac{k}{\beta^n} \ : \ 0 \le k < \beta^n\}$. Then G_n is (n,ε)-separating, so $s_{n,\varepsilon} \ge \beta^n$. On the other hand, $G'_n = \{\frac{2k\varepsilon}{\beta^n} \ : \ 0 \le k < \beta^n/(2\varepsilon)\}$ is (n,ε)-spanning, so $r_n(\varepsilon) \le \beta^n/(2\varepsilon)$. Therefore*

$$\log\beta = \limsup_{n\to\infty} \frac{1}{n}\log\beta^n \le h_{top}(T_\beta) \le \limsup_{n\to\infty}\frac{1}{n}\log(\beta^n/(2\varepsilon)) = \log\beta.$$

Circle rotations, or in general isometries, have zero topological entropy. Indeed, if $E(\varepsilon)$ is an ε-separated set (or ε-spanning set), it will also be (n,ε)-separated (or (n,ε)-spanning) for every $n \ge 1$. Hence $s_{n,\varepsilon}$ and $r_n(\varepsilon)$ are bounded in n, and their exponential growth rates are equal to zero. In more generality:

Proposition 1.6.2. *Every equicontinuous transformation (X,T) has zero entropy.*

Proof. Let $\delta > 0$ be arbitrary and choose $\varepsilon > 0$ as in the definition of equicontinuous. Then $\mathrm{diam}(T^n(B_\varepsilon(x)) \le 2\delta$ for all $x \in X$ and $n \ge 0$ (or $n \in Z$ if T is invertible). Take $M = \lceil \mathrm{diam}(X)/\varepsilon \rceil$. Hence, a single cover of X by M ε-balls constitutes a cover of (δ,n)-balls for all n. Therefore $h_{top}(T) \le \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log M = 0$. □

Finally, let (X,σ) be the full shift on N symbols. Let $\varepsilon > 0$ be arbitrary, and take m such that $2^{-m} < \varepsilon$. If we select a point from each $n+m$-cylinder, this gives an (n,ε)-spanning set, whereas selecting a point from each n-cylinder gives an (n,ε)-separated set. Therefore

$$\begin{aligned} \log N = \limsup_n \frac{1}{n}\log N^n &\le \limsup_n \frac{1}{n}\log s_{n,\varepsilon} \le h_{top}(\sigma) \\ &\le \limsup_{n\to\infty}\frac{1}{n}\log r_n(\varepsilon) \le \limsup_{n\to\infty}\frac{1}{n}\log N^{n+m} = \log N. \end{aligned}$$

Corollary 1.6.3. *Given a continuous map $T : X \to X$, $h_{top}(T^k) = k h_{top}(T)$ for all $k \ge 0$, and if T is invertible, then $h_{top}(T^n k) = |k| h_{top}(T)$ for all $k \in \mathbb{Z}$.*

Proof. For any $k \in \mathbb{N}$, a (kn, ε)-separated set for T is also an (n, ε)-separated set for T^k. Therefore

$$h_{top}(T^k) = \lim_{n\to\infty} \frac{1}{n} \log s_{n,\varepsilon}(T^k) = k \lim_{n\to\infty} \frac{1}{kn} \log s_{n,\varepsilon}(T) = kh_{top}(T).$$

Clearly the identity T^0 has zero entropy. If T is invertible, and $E_n(\varepsilon)$ is an (n, ε)-separated set, then $T^{n-1}(E_n(\varepsilon))$ s an $(n\varepsilon)$-separated set for T^{-1}. Therefore $h_{top}(T^{-1}) = h_{top}(T)$. Combined with the first part, we get $h_{top}(T^k) = |k|h_{top}(T)$ for all $k \in \mathbb{Z}$. □

Corollary 1.6.4. *If (Y, S) is a continuous factor of (X, T) (where X is a compact metric space), then $h_{top}(S) \leq h_{top}(T)$. In particular, conjugate systems on compact metric space have the same topological entropy.*

Proof. Let $\pi : X \to Y$ be a continuous factor map. Since X is compact, π is uniformly continuous, so for $\varepsilon > 0$, we can find $\delta > 0$ such that $d(x, y) < \delta$ implies $d(\pi(x), \pi(y)) < \varepsilon$. Therefore, if $E_n(\delta)$ is an (n, δ)-spanning set for T, then $\pi(E_n(\delta))$ is an (n, ε)-spanning set for S (but possibly not a minimal (n, ε)-spanning set, even if $E_n(\delta)$ is minimal). It follows that $r_n(\delta, T) \geq r_n(\varepsilon, S)$, and hence $h_{top}(T) \geq h_{top}(S)$. □

For subshifts, topological entropy is the exponential growth-rate of the word-complexity, see Proposition 1.6.8 below.

Definition 1.6.5. *Given a dynamical system (X, T), a point $x \in X$ is called* **non-wandering** *if for every neighborhood $U \ni x$ there is n such that $T^n(U) \cap U \neq \varnothing$. The* **nonwandering set***, denoted Ω, is the set of all nonwandering points.*

Note that recurrent points are always nonwandering, but Ω can contain nonrecurrent points. In the one-sided full shift, for instance, $x = 0111111\dots$ is not recurrent but nonwandering. If (X, T) is topologically transitive, then $\Omega = X$.

Proposition 1.6.6. *The entropy of a dynamical system (X, T), $h_{top}(T) = h_{top}(T|_\Omega)$.*

Example 1.6.7. *Let $X \subset \{01\}^{\mathbb{N}}$ be the collection of all sequences x that you can write as $x = 0^{n_1}1^{n_2}0^{n_3}1^{n_4}\dots$, with $0 \leq n_1 \leq \max\{n_1, 1\} \leq n_2 \leq n_3 \leq n_4 \leq \dots$ Then Ω consists of periodic orbits $0^k1^k0^k1^k\dots$ or $1^k0^k1^k0^k\dots$, i.e., with period $2k$. Therefore the number of $2n$-periodic point (not necessarily prime period $2n$) equals twice the number of divisors of n, and hence is $\leq 2n$. In view of Proposition 1.6.6, $h_{top}(\sigma) = 0$.*

1.6.1 Entropy of subshifts

For subshifts, the topological entropy is simply the exponential growth rate of the the word-complexity.

Proposition 1.6.8. *For a subshift (X, σ), the limit $\lim_{n\to\infty} \frac{1}{n} \log p(n)$ exists and is equal to $h_{top}(\sigma)$.*

Proof. First note that $p(m + n) \leq p(m)p(n)$, so $\log p(n)$ is subadditive. Thus by Fekete's LemmaindexLemma!Fekete's, $\lim_n \frac{1}{n} \log p(n) = \inf_n \frac{1}{n} \log p(n)$ exists. Next take $\varepsilon > 0$ arbitrary and $N \in \mathbb{N}$ such that $2^{-N} < \varepsilon$. Then every $n + N$-cylinder is an (n, ε)-ball, and we need exactly $p(n + N)$ of them to cover the space. Therefore, writing $m = n+N$, $h_{top}(\sigma) = \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log p(n+N) = \lim_{m\to\infty} \frac{1}{m} \log p(m)$. □

We devote separate chapters to subshifts of positive and subshifts of zero entropy, because they tend to be quite different in regard to their other topological properties (topological mixing, existence and number of periodic orbits, shadowing and synchronization properties. Proposition 1.6.8 shows that the maximal entropy of a subshift on d letters is $\log d$, and this is achieved by the full shift $(\{1, \dots, d\}^{\mathbb{N}}, \sigma)$. One can ask whether all intermediate values can be achieved as entropy for some subshift. As we shall see later, this is not true for the class of subshift of finite type or the sofic shifts, because the entropy is then equal to the logarithm of the leading eigenvalue of some matrix, so logarithms of algebraic numbers.

Proposition 1.6.9.

On the other hand, the entropy of β-shifts and unimodal subshifts of $\{0, 1\}^{\mathbb{N}}$ can take any value in $(0, \log 2]$, as they equal $\log \eta$ and $\log s$ (for the slope of a tent-map) respectively. Also withing the class of gap shift you can achieve every value of the entropy, as can be derived from Theorem **??**. Some specific constructions of subshifts of a chosen entropy can be found among spacing shifts, see [**?**, 168] and Section **??**.

1.7 Sliding block codes

Definition 1.7.1 (Sliding Block Code). *A map $\pi : \mathcal{A}^{\mathbb{Z}} \to \tilde{\mathcal{A}}^{\mathbb{Z}}$ is called a* **sliding block code** *(also called* **local rule** *of* **window size** $2N + 1$ *if there is a function $f : \mathcal{A}^{2N+1} \to \tilde{\mathcal{A}}$ such that $\pi(x)_i = f(x_{i-N} \dots x_{i+N})$.*

In other words, we have a window[5] of width $2N + 1$ put on the sequence x. If it is centered at position i, then the recoded word $y = \pi(x)$ will have at position i the f-image of what is visible in the window. After that we slide the window to the next position and repeat.

Theorem 1.7.2 (Curtis–Hedlund–Lyndon[6]). *Let X and Y be subshifts over finite alphabets $\mathcal{A}$ and $\mathcal{A}'$ respectively. A continuous map $\pi : X \to Y$ commutes with the shift (i.e., $\sigma \circ \pi = \pi \circ \sigma$) if and only if π is a sliding block code.*

[5]Sometime the window can have memory and anticipation of different lengths, so the window would be $[-m, a]$, but calling their maximum N covers all cases

[6]Curtis and Lyndon were working for the military at the time, so their work was "classified", and the paper was published under Hedlund's name only, [140]

Proof. First assume that π is continuous and commutes with the shift. For each $a \in \mathcal{A}'$, the cylinder $[a] = \{y \in Y : y_0 = a\}$ is clopen, so $V_a := \pi^{-1}([a])$ is clopen too. Since V_a is open, it can be written as the union of cylinders, and since V_a is closed (and hence compact) it can be written as the finite union of cylinders: $V_a = \cup_{i=1}^{r_a} U_{a,i}$. Let N be so large that every $U_{a,i}$ is determined by the symbols $x_{-N} \dots x_N$. This makes $2N + 1$ a sufficient window size and there is a function $f : \mathcal{A}^{2N+1} \to \mathcal{A}'$ such that $\pi(x)_0 = f(x_{-N} \dots x_N)$. By shift-invariance, $\pi(x)_i = f(x_{i-N} \dots x_{i+N})$ for all $i \in \mathbb{Z}$.

Conversely, assume that π is a sliding block code of window size $2N + 1$. Take $\varepsilon = 2^{-M} > 0$ arbitrary, and $\delta = \varepsilon 2^{-N}$. If $d(x, y) < \delta$, then $x_i = y_i$ for $|i| \le M + N$. By the construction of the sliding block code, $\pi(x)_i = \pi(y)_i$ for all $|i| \le M$. Therefore $d(\pi(x), \pi(y)) < \varepsilon$, so π is continuous (in fact uniformly continuous). □

Exercise 1.7.3. *Give the sliding block code between the Fibonacci SFT and the even subshift (see Examples 0.1.10 and 0.1.13.*

Each subshift (X, σ) over an alphabet $\mathcal{A}$ can be described as an n**-block shift**, where the alphabet $\mathcal{A}' \subset \mathcal{A}^n$ are the words in $\mathcal{L}_n(X)$, and $a, b \in \mathcal{A}'$ can only follow each other if the $n-1$-suffix of a coincides with the $n-1$-prefix of b. For instance, if $(X_{\text{even}}, \sigma)$ is the even shift, then $\mathcal{A}' = \{00, 01, 10, 11\}$ and the edge-labeled transition graph is given in Figure 1.2

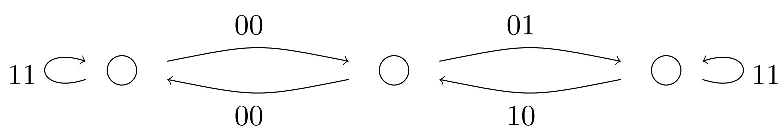


Figure 1.2: The edge-labeled transition graph of the 2-block even shift.

Taking a block shift generally doesn't change the nature of the shift (SFTs remain SFTs, sofic shifts remain sofic, substitution shifts remain substitution shifts, see Section **??**). Block shifts can be used the shrink the window size of sliding block codes, see [175, Proposition 1.5.12].

Proposition 1.7.4. *If π is a sliding block code between X and Y of window size $2N+1$, then there is a sliding block code between the $2N+1$ block shift $\tilde{X}$ of X and Y.*

Proof. We do the proof for invertible shifts; the one-sided shifts works as well, but then we cannot allow a memory in the sliding block code, only anticipation. The letters of the $2N+1$-block shift $\tilde{X}$ correspond exactly with the possible $2N+1$-words on which π is defined. Now define $\tilde{\pi} = \sigma^N \circ \pi$, where the power of the shift is required to move to exactly to the middle of the window. □

Chapter 2

Subshifts of positive entropy

2.1 Subshifts of finite type

2.1.1 Definition of SFTs and transition matrices and graphs

Definition 2.1.1. *A* **subshift of finite type (SFT)** *is a subshift consisting of all string avoiding a* **finite** *list of forbidden words as factors. For example, the Fibonacci shift has* 11 *as forbidden word.*

If $M+1$ is the length of the longest forbidden word, then this SFT is an M**-step SFT**, or an SFT with **memory** M. Indeed, an M-step SFT has the property that if $uv \in \mathcal{L}(X)$ and $vw \in \mathcal{L}(X)$, and $|v| \geq M$, then $uvw \in \mathcal{L}(X)$ as well.

Definition 2.1.2. *A* **synchronizing word** *v of a subshift X is a word such that whenever $uv \in \mathcal{L}(X)$ and $vw \in \mathcal{L}(X)$, then also $uvw \in \mathcal{L}(X)$. A subshift X is* **synchronizing** *if it is transitive, and contains a synchronizing word.*

Lemma 2.1.3. *Every irreducible SFT is synchronizing; in fact, every word of length M (the memory of the SFT) is synchronizing.*

Proof. Let v be any word of length M. If $uv \in \mathcal{L}(X)$ then u has no influence of what happens after v. Hence if $vw \in \mathcal{L}(X)$, then $uvw \in \mathcal{L}(X)$. □

Lemma 2.1.4. *Every SFT X on a finite alphabet can be recoded such that the list of forbidden words consists of* 2*-words only.*

Proof. Assume that the longest forbidden word of X has length $M+1 \geq 2$. Take a new alphabet $\mathcal{B} = \mathcal{A}^M$, say $b_1, \dots, b_n$ are its letters. Now recode every $x \in X$ using a sliding block code π, where for each index i, $\pi(x)_i = b_j$ if b_j is the symbol used for $x_i x_{i+1} \dots x_{i+M-1}$. Effectively, this is replacing X by its M-block code. Then every $M+1$-word is uniquely coded by a 2-word in the new alphabet $\mathcal{B}$, and vice versa, every $b_1 b_2$ such that the $M-1$-suffix of $\pi^{-1}(b_1)$ equals the $M-1$-prefix of $\pi^{-1}(b_2)$ codes a unique M-word in $\mathcal{A}^*$. Now we forbid a 2-word $b_1 b_2$ in $\mathcal{B}^2$ if $\pi^{-1}(b_1 b_2)$ contains a forbidden word of X. Since $\mathcal{A}$ is finite, and therefore $\mathcal{B}$ is finite, this leads to a finite list of forbidden 2-words in the recoded subshift. □

Example 2.1.5. *Let X be the SFT with forbidden words 11 and 101, so $M = 2$. We recode using the alphabet $a = 00$, $b = 01$, $c = 10$ and $d = 11$. Draw the vertex-labeled transition graph, see Figure 2.1; labels at the arrow then just indicate with word in $\{0,1\}^3$ they stand for. For example, the edge $a \to b$ labeled 001 has prefix $a = 00$ and suffix $b = 01$. Each arrow containing a forbidden word is dashed, and then removed in the right panel of Figure 2.1.*

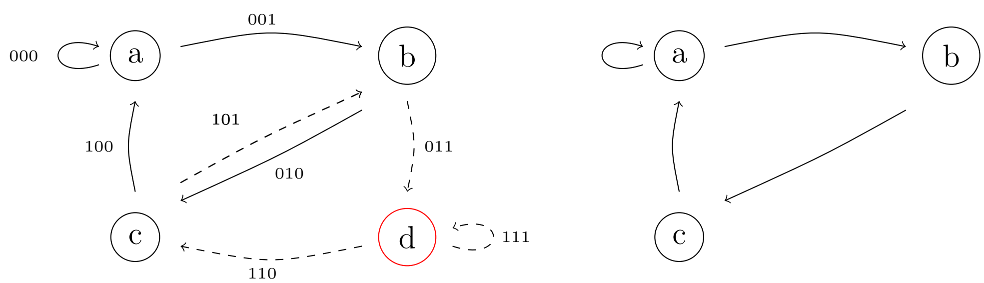


Figure 2.1: Illustrating the recoding of the SFT with forbidden words 11 and 101.

Corollary 2.1.6. *Every SFT X on a finite alphabet can be represented by a finite graph $\mathcal{G}$ with vertices labeled by the letters of $\mathcal{B}$ and arrows $b_1 \to b_2$ only if $\pi^{-1}(b_1 b_2)$ contains no forbidden word of X.*

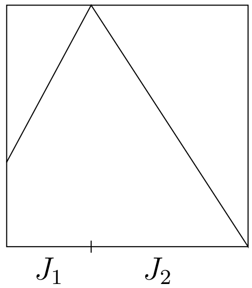
Definition 2.1.7. *The directed graph $\mathcal{G}$ constructed in the previous corollary is called the* **transition graph** *of the SFT. The matrix $A = (a_{ij})_{i,j \in \mathcal{B}}$ is the* **transition matrix** *if $a_{ij} = 1$ if the arrow $i \to j$ exists in $\mathcal{G}$ and $a_{ij} = 0$ otherwise. The graph is vertex-label, which means that each vertex is assigned unique symbol in the alphabet*

Example 2.1.8. *Let $T : [0,1] \to [0,1]$ be the* **piecewise monotone** *map, i.e., there is a finite partition $\{J_k\}_{k=1}^N$ of $[0,1]$ into intervals such that $T|_{J_i}$ is continuous and monotone for each i. Assume also that for each i, $\overline{T(J_i)}$ is the closure of the union of J_ks. In this case we call $\{J_k\}_{k=1}^N$ a* **Markov partition**. *Write*

$$a_{ij} = \begin{cases} 1 & \text{if } \overline{T(J_i)} \supset J_i, \\ 0 & \text{if } T(J_i) \cap J_i^\circ = \varnothing. \end{cases}$$

Then the resulting matrix $A = (a_{i,j})_{i,j=1}^N$ is the transition matrix for the symbolic shift obtained by taking the closure of the collection of itineraries $\{\boldsymbol{i}(x) : x \in [0,1]\}$. This yields a one-sided shift.

The example in Figure 2.2 produces the transition matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, so the corresponding shift is the Fibonacci SFT. It should not come as a surprise that the



$$T(x) = \begin{cases} \gamma(x + \frac{2-\gamma}{\gamma}) & \text{if } x \in J_1 := [0, \frac{\gamma-1}{\gamma}], \\ \gamma(1-x) & \text{if } x \in J_2 := [\frac{\gamma-1}{\gamma}, 1], \end{cases} \qquad \gamma = \frac{\sqrt{5}+1}{2}.$$

Figure 2.2: The tent map with slope equal to the golden ration

leading eigenvalue of A is exactly the slope of T: both equal $e^{h_{top}(T)} = e^{h_{top}(\sigma)}$, see Section 2.1.2.

For the bi-infinite Fibonacci shift, we can look at a toral automorphism.

Definition 2.1.9. *A* **toral automorphism** $T : \mathbb{T}^d \to \mathbb{T}^d$ *is an invertible linear map on the (d-dimensional) torus $\mathbb{T}^d$. Each such T is of the form $T_A(x) = Ax \pmod 1$, where*

- *A is an integer matrix with $\det(A) = \pm 1$;*
- *the eigenvalues of A are not on the unit circle; this property is called* **hyperbolicity***; for toral automorphisms, this is equivalent to $\mathbb{T}^d$ being a hyperbolic set in terms of Definition* **??***.*

The map T_A has a **Markov partition***, that is a partition $\{J_i\}_{i=1}^N$ for sets such that*

1. *The J_i have disjoint interiors and $\cup_i J_i = \mathbb{T}^d$;*
2. *If $T_A(J_i) \cap J_j \neq \varnothing$, then $T_A(J_i)$ stretches across J_j in the unstable direction (i.e., the direction spanned by the unstable eigenspaces of A).*
3. *If $T_A^{-1}(J_i) \cap J_j \neq \varnothing$, then $T_A^{-1}(J_i)$ stretches across J_j in the stable direction (i.e., the direction spanned by the stable eigenspaces of A).*

Every hyperbolic toral automorphism has a Markov partition, see [45] but in general they are fiendishly difficult to find explicitly, especially in dimension ≥ 3 where the boundaries of the J_i might have to be fractal [50]. Therefore we confine ourselves to the simpler case of $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$; it has Markov partition of three rectangles J_i for $i = 1, 2, 3$ can be constructed, see Figure 2.3.

The corresponding transition matrix is

$$B = (b_{i,j}) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \textit{where } b_{ij} = \begin{cases} 1 & \textit{if } T_A(J_i) \cap J_j \neq \varnothing \\ 0 & \textit{if } T_A(J_i) \cap J_j = \varnothing. \end{cases}$$

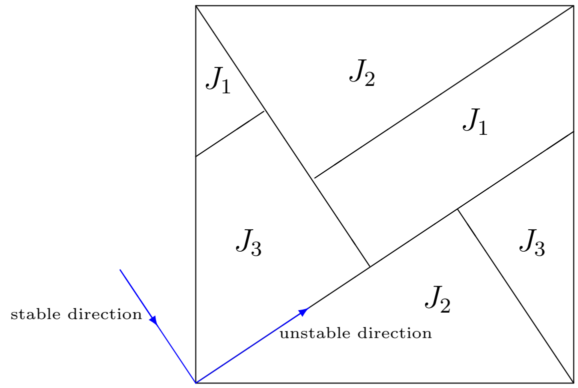


Figure 2.3: The Markov partition for the toral automorphism T_A.

The characteristic polynomial of B is

$$\det(B - \lambda I) = -\lambda^3 + 2\lambda + 1 = -(\lambda + 1)(\lambda^2 - \lambda - 1) = -(\lambda + 1)\det(A - \lambda I).$$

so B has the eigenvalues of A (no coincidence!), together with $\lambda = -1$.

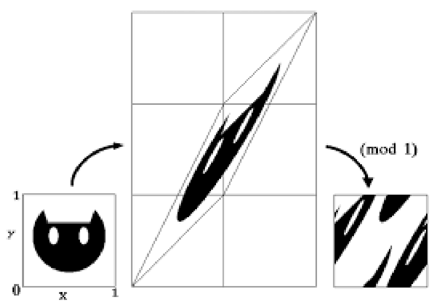


Figure 2.4: Vladimir I. Arnol'd (1937–2010) and his catmap.

Example 2.1.10. *The most "famous" toral automorphism is Arnol'd's catmap, and it has the matrix $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, see Figure 2.4. It is called this way because Arnol'd used this example, including the drawing of a cat's head, in his book(s) to illustrate the nature of hyperbolic maps.*

Exercise 2.1.11. *Show that if $x \in \mathbb{T}^d$ has all coordinates rational, then x is periodic under a toral automorphism. Conclude that, if the pixels in Figure 2.4 have rational coordinates (such as the dyadic coordinates that computers use), then the cat will return intact after a finite number of iterates.*

The following characterization for shadowing subshifts is due to Walters [234] (see also [169, Theorem 3.33]).

Theorem 2.1.12. *A subshift (X, σ) of has the shadowing property if and only if it is a subshift of finite type.*

Proof. We give the proof for $X \subset \mathcal{A}^{\mathbb{N}\cup\{0\}}$ only; the two-sided case follows in a similar way.

The "if"-direction: Let (X, σ) be an SFT of memory M, so M is the length of the longest forbidden word. Let $\varepsilon > 0$ be arbitrary and choose $m \geq M$ so small that $2^{-m} < \varepsilon$. Take $\delta = 2^{-m+2}$. We need to show that every δ-pseudo-orbit $(x^n)_{n\geq 0} \subset X$, that is,

$$\sigma(x^n)_0 \dots \sigma(x^n)_{m-3} = x_1^n \dots x_{m-2}^n = x_0^{n+1} \dots x_{m-2}^{n+1}$$

for every n, there is $y \in X$ that ε-shadows $(x^n)_{n\geq 0}$. To this end, set $y_n = x_0^n$ for each $n \geq 0$. Then for $0 \leq i < m$, we have

$$y_{n+i} = x_0^{n+i} = x_1^{n+i-1} = x_2^{n+i-2} = \dots = x_i^n,$$

so $y_n \dots y_{n+m-1} = x_0^n \dots x_{m-1}^n \in \mathcal{L}(X)$. Since X is an SFT, $y \in X$ and $d(\sigma^n(y), x^n) < \varepsilon$ by construction.

The "only if"-direction: Let (X, σ) be a subshift with the shadowing property, so in articular, for $\varepsilon = 1$, there exists $\delta > 0$ such that every δ-pseudo-orbit in X is ε-shadowed in X. Take $N \in \mathbb{N}$ such that $2^{-N+2} < \delta$, and let $y \in \mathcal{A}^{\mathbb{N}\cup\{0\}}$ be such that $y_n \dots y_{n+N-1} \in \mathcal{L}(X)$ for each n. Then there exists a sequence $(x^n)_{n\geq 0}$ such that $x_0^n \dots x_{N-1}^n = y_n \dots y_{n+N-1}$ for each $n \geq 0$. Therefore

$$\sigma(x^n)_0 \dots \sigma(x^n)_{N-2} = x_1^n \dots x_{N-1}^n = y_{n+1} \dots y_{n+N-1} = x_0^{n+1} \dots x_{N-2}^{n+1}$$

and $d(\sigma(x^n), x^{n+1}) \leq 2^{-N+2} < \delta$. Hence $(x^n)_{n\geq 0}$ is a δ-pseudo-orbit, which can be ε-shadowed by some $z \in X$. But then $z_n = x_0^n = y_n$ for every $n \geq 0$, so $z = y \in X$. Since y was arbitrary, up to the condition that each of its N-blocks belongs to $\mathcal{L}(X)$, it follows that the only restriction of X involve forbidden blocks of length $\leq N$, and therefore X is an SFT. □

2.1.2 Topological entropy for SFTs

Definition 2.1.13. *A non-negative $N \times N$ matrix $A = (a_{ij})_{i,j=1}^N$ is called* **irreducible** *if for every i, j there is k such that A^k has (i,j)-entry $a_{ij}^{(k)} > 0$. For index i, set $per(i) = gcd(k > 1 : a_{ii}^{(k)} > 0)$. If A is irreducible, then $per(i)$ is the same for every i, and we call it the* **period** *of A. We call A* **aperiodic** *if its period is 1. The matrix is called* **primitive** *if there is k such that $a_{ij}^{(k)} > 0$ for all i, j.*

Exercise 2.1.14. *Show that if A is primitive and irreducible, then A is primitive, but irreducibility or aperiodicity alone doesn't imply primitivity. Conversely, if A is primitive, then it is also aperiodic and irreducible.*

Exercise 2.1.15. *If A is irreducible, show that $per(i)$ is indeed independent of i.*

Theorem 2.1.16. *The entropy of an irreducible SFT equals $\log\lambda$ where λ is the leading eigenvalue of the transition matrix.*

Proof. Let $A^n = (a_{ij}^{(n)})_{i,j\in\mathcal{A}}$ be the n-th power of the transition matrix A. Every n-word in $\mathcal{L}(X)$ corresponds to an n-path in the transition graph, and the number of n-paths from i to j is given by $p_{ij}^{(n)}$. From the Perron-Frobenius Theorem **??** we can derive that there is $C > 0$ such that

$$C^{-1}\lambda^n \leq a_{ij}^{(n)} \leq C\lambda^n \quad \text{for all } i,j \in \mathcal{A} \text{ and } n \text{ sufficiently large},$$

provided A is aperiodic. (If A is periodic, then the above estimate holds for every $i \in \mathcal{A}$, n sufficiently large, and some $j \in \mathcal{A}$ depending on i and n. This is enough to complete the argument.) It follows that $C^{-1}\lambda^n \leq p(n) \leq (\#\mathcal{A})^2 C\lambda^n$ and $\lim_n \frac{1}{n}\log p(n) = \log\lambda$. □

Proposition 2.1.17. *If (Y,σ) is a factor of (X,σ), then $h_{top}(Y,\sigma) \leq h_{top}(X,\sigma)$. If (X,σ) and (Y,σ) are conjugate, then $h_{top}(X,\sigma) = h_{top}(Y,\sigma)$.*

The result also holds in general, i.e., not just in the context of subshifts, see Corollary 1.6.4, but using the word-complexity and sliding block codes, the proof is particularly straightforward here.

Proof. Let $\psi : X \to Y$ be the factor map. Since it is continuous, it is a sliding block code by Theorem 1.7.2, say of window length $2N+1$. Therefore the word-complexities relate as $p_Y(n) \leq p_X(n+2N)$, and hence

$$\begin{aligned}
\limsup \frac{1}{n}\log p_Y(n) &\leq \limsup \frac{1}{n}\log p_X(n+2N) \\
&= \limsup \frac{n+2N}{n}\frac{1}{n+2N}\log p_X(n+2N) \\
&= \limsup \frac{1}{n+2N}\log p_X(n+2N).
\end{aligned}$$

This proves the first statement. Using this in both directions, we find $h_{top}(X,\sigma) = h_{top}(Y,\sigma)$. □

As shown by Parry [192], irreducible SFTs are intrinsically ergodic. This follows also from Theorem **??** and Proposition **??**. Weiss [236] showed that factors of irreducible SFTs are intrinsically ergodic as well.

2.1.3 Vertex-splitting and conjugacies between SFTs:

It is natural to ask which SFTs are conjugate to each other. We have seen that having equal topological entropy is a necessary condition for this, but it is not sufficient. The conjugacy problem of SFTs was solved by Bob Williams (1942–) and in this section we discuss the ingredients required for this result. The complete details can be found in [175].

We know that an SFT (X, σ) has a graph representation (as vertex-labeled subshift or edge-labeled subshift, and certainly not unique). The following operation on the graph $\mathcal{G}$, called **vertex splitting**, produces a related subshift.

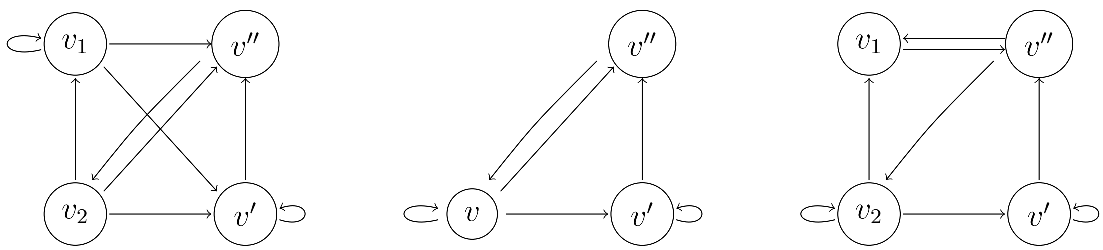


Figure 2.5: Insplit graph Original $\mathcal{G}$ Outsplit graph

Let $\mathcal{G} = (V, E)$ where V is the vertex set and E the edge set. For each $v \in V$, let $E_v \subset E$ be the set of edges starting in v and $E^v \subset E$ be the set of edges terminating in v.

Definition 2.1.18. *Let $\mathcal{G} = (V, E)$, and assume that $\#E^v \geq 2$. An* **elementary insplit graph** $\hat{\mathcal{G}} = (\hat{V}, \hat{E})$ *is obtained by*

- *doubling one vertex $v \in V$ into two vertices $v_1, v_2 \in \hat{V}$;*
- *replacing each $e = (v \to w) \in E_v$ for $w \neq v$ by an edge $\hat{e}_1 = (v_1 \to w)$* **and** *$\hat{e}_2 = (v_2 \to w)$;*
- *replacing each $e = (w \to v) \in E^v$ for $w \neq v$ by a single edge $\hat{e}_1 = (w \to v_1)$* **or** *an edge $\hat{e}_2 = (w \to v_2)$ (but make sure that v_1 and v_2 both have incoming edges);*
- *replacing each loop $(v \to v)$ by two edges $(v_1 \to v_i), (v_2 \to v_i) \in \hat{E}$ (so one of them is a loop) where $i \in \{1, 2\}$.*

An **insplit graph** *is then obtained by successive elementary outsplits.*

(Elementary) outsplit graphs are defined similarly, interchanging the roles of E_v and E^v.

Definition 2.1.19. *Let $\mathcal{G} = (V, E)$, and assume that $\#E_v \geq 2$. An* **elementary outsplit graph** $\hat{\mathcal{G}} = (\hat{V}, \hat{E})$ *is obtained by*

- *doubling one vertex $v \in V$ into two vertices $v_1, v_2 \in \hat{V}$;*
- *replacing each $e = (v \to w) \in E_v$ for $w \neq v$ by a single edge $\hat{e} = (v_1 \to w)$* **or** *$\hat{e} = (v_2 \to w)$ (but make sure that v_1 and v_2 both have outgoing edges);*
- *replacing each $e = (w \to v) \in E^v$ for $w \neq v$ by an edge $\hat{e} = (w \to v_1)$* **and** *an edge $\hat{e} = (w \to v_2)$;*
- *replacing each loop $(v \to v)$ by two edges $(v_i \to v_1), (v_i \to v_2) \in \hat{E}$ (so one of them is a loop) where $i \in \{1, 2\}$.*

An **outsplit graph** *is then obtained by successive elementary outsplits.*

If every $e \in E$ has a unique label, then we will also give each $\hat{e} \in \hat{E}$ a unique label.

Proposition 2.1.20. *Let $\hat{\mathcal{G}}$ be an in- or outsplit graph obtained from $\mathcal{G}$. Then edge-labeled subshift $\hat{X}$ of $\hat{\mathcal{G}}$ and the edge-labeled subshift X of $\mathcal{G}$ are mutually semi-conjugate to each other.*

Proof. We give the proof for an elementary outsplit $\hat{\mathcal{G}}$; the general outsplit and (elementary) insplit graph follow similarly. By Theorem 1.7.2, it suffices to give sliding block code representations for $\pi : \hat{X} \to X$ and $\hat{\pi} : X \to \hat{X}$.

- The factor map $\pi : \hat{X} \to X$ is simple. If $\hat{e} \in \hat{E}$ replaces $e \in E$, then $f(\hat{e}) = e$ and $\pi(x)_i = f(x_i)$.
- Each 2-word $ee' \in \mathcal{L}(X)$ uniquely determines the first edge $\hat{e}$ of the 2-path in $\hat{\mathcal{G}}$ that replaces the 2-path in $\mathcal{G}$ coded by ee'. Set $\hat{f}(e, e') = \hat{e}$ and $\hat{\pi}(x)_i = \hat{f}(x_i, x_{i+1})$.

This concludes the proof. In general, mutual semi-conjugacy is not enough to conclude conjugacy (it is not given that $\hat{\pi} = \pi^{-1}$), but in this situation, conjugacy holds, see Theorem 2.1.25. □

Now let $\hat{\mathcal{G}} = (\hat{V}, \hat{E})$ be an outsplit graph of $\mathcal{G} = (V, E)$ with transition matrices $\hat{A}$ and A respectively. Assume that $N = \#V$ and $\hat{N} = \#\hat{V}$. Then there is $N \times \hat{N}$-matrix $D = (d_{v,\hat{v}})_{v \in V, \hat{v} \in \hat{V}}$ where $d_{v,\hat{v}} = 1$ if $\hat{v}$ replaces v. (Thus D is a sort of rectangular diagonal matrix.)

There also is an $\hat{N} \times N$-matrix $C = (c_{\hat{v},v})_{\hat{v} \in \hat{V}, v \in V}$ where $c_{\hat{v},v}$ is the number of edges $e \in E^v$ that replace an edge $\hat{e} \in \hat{E}_{\hat{v}}$.

Proposition 2.1.21. *With the above notation,*

$$DC = A \quad \text{and} \quad CD = \hat{A}.$$

Sketch of proof. Work it out for an elementary outsplit, and then compose elementary outsplits to a general outsplit. For the first step, we compute the elementary outsplit for the example of Figure 2.5.

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \hat{A} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Also

$$D = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Now do the matrix multiplications to check that $DC = A$ and $CD = \hat{A}$. □

Exercise 2.1.22. *Do the same for the elementary insplit graph in the example of Figure 2.5.*

Definition 2.1.23. *Two matrices A and $\hat{A}$ are* **strongly shift equivalent (of lag ℓ)** *(denoted as $A \approx \hat{A}$) if there are (rectangular) matrices D_i, C_i and A_i, $1 \leq i \leq \ell$ such that*

$$A = A_0, \quad A_{i-1} = D_i C_i, C_i D_i = A_i, i = 1, \dots, \ell, \quad A_\ell = \hat{A}. \tag{2.1}$$

Exercise 2.1.24. *Show that strong shift equivalence $\approx$ is indeed an equivalence relation between nonnegative square matrices. Show that $A \approx \hat{A}$ implies that A and $\hat{A}$ have the same leading eigenvalue $\lambda = \hat{\lambda}$.*

Matrices A and $\hat{A}$ being strongly shift equivalent means, in effect, that their associated graphs $\mathcal{G}$ and $\hat{\mathcal{G}}$ can be transformed into each other by a sequence of elementary vertex-splittings and their inverses (vertex-mergers). This turns out the only mechanism that keeps SFTs conjugate, as shown in Williams' theorem from 1973. The full proof is in [175, Chapter 7].

Theorem 2.1.25 (Williams)**.** *Two SFTs are conjugate if and only if their transition matrices are strongly shift equivalent.*

Strong shift equivalence $A \approx \hat{A}$ may be a complete invariant for the edge-labeled SFTs X_A and $X_{\hat{A}}$ to be conjugate, it is in practice difficult to check if $A \approx \hat{A}$. Even if A and $\hat{A}$ have, say, the same characteristic polynomial, they need not be strongly shift equivalent. An easier to check, but not complete invariant, is a weaker notion called **shift equivalent (with lag ℓ)** (without the "strong") and is denoted as $A \sim \hat{A}$.

This means that the ℓ-th powers of the matrices are strong shift equivalent (with lag 1): there are matrices D, C such that

$$A^\ell = DC, \quad A^\ell = CD \qquad \text{and} \qquad AD = DB \quad C\hat{A} = AC \tag{2.2}$$

If $\ell = 1$, then the first half of (2.2) implies (2.1), whereas the second half is an immediate consequence of the first. Strong shift equivalence implies shift equivalence, see [175, Theorem 7.3.3]. If $A \not\sim \hat{A}$, then X_A and $X_{\hat{A}}$ are not conjugate, but if $A \not{\hat{A}}$, it is still undecided if they are. In turn, $A \sim \hat{A}$ if and only if their dimension groups are isomorphic, see [175, Section 7.4]. We say more about dimension groups in Section **??**.

2.2 Sofic subshifts

Definition 2.2.1. *A subshift (X, σ) is called* **sofic** *if it is the space of paths in an edge-labeled graph. Other than with the vertex-labeling, multiple edge are allowed to be assigned the same symbol in the alphabet.*

Lemma 2.2.2. *Every SFT is sofic.*

Proof. Assume that the SFT has memory M. Let $\mathcal{G}$ be the vertex-labeled M-block transition graph of the SFT i.e., each $a_1 \dots a_M \in \mathcal{L}_M(X)$ is the label of a unique vertex. We have an edge $a_1 \dots a_M \to b_1 \dots b_M$ if and only if $a_1 \dots a_M b_M = a_1 b_1 \dots b_M \in \mathcal{L}_{M+1}(X)$, and then this $M+1$-word is also the label of the edge. Since each infinite vertex-labeled path is in one-to-one correspondence with a an infinite edge-labeled path is in one-to-one correspondence with an infinite word in X, we have represented X as a sofic shift. $\square$

Remark 2.2.3. *Not every sofic shift is an SFT. For example the even shift (Example 0.1.13) has an infinite collection of forbidden words, but it cannot be described by a finite collection of forbidden words. Sofic shifts that are not of finite type are called* **strictly sofic**.

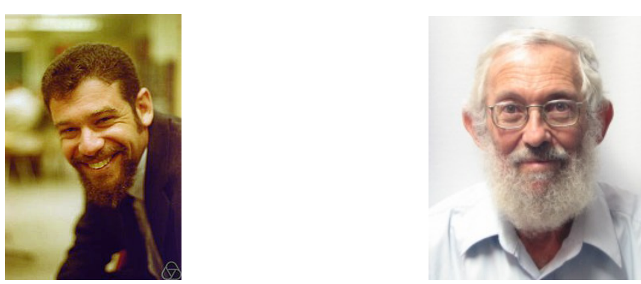


Figure 2.6: Don Ornstein (1934-) and Benji Weiss (1941-)

The word sofic was coined by Benji Weiss; it comes from the Hebrew word for "finite". The following theorem shows that we can equally define the sofic subshifts as those that are a factor of a subshift of finite type.

Theorem 2.2.4. *A subshift X is generated by an edge-labeled graph if and only if it is the factor of an SFT.*

Proof. $\Rightarrow$: Let $\mathcal{G}$ be the edge-labeled graph of X, with edges labeled in alphabet $\mathcal{A}$. Relabel $\mathcal{G}$ in a new alphabet $\mathcal{A}'$ such that every edge has a distinct label. Call the new edge-labeled graph $\mathcal{G}'$. Due to the injective edge-labeling, the edge-subshift X' generated by $\mathcal{G}'$ is isomorphic to an SFT. In fact, we just have to take the dual graph in which the edges of $\mathcal{G}'$ are the vertices, and $a \to b$ if an only if a labels the incoming edge and b the outgoing edge of the **same** vertex in $\mathcal{G}'$. Now $\pi : X' \to X$ is given by $\pi(x)_i = a$ if a is the label in $\mathcal{G}$ of the same edge that is labeled x_i in $\mathcal{G}'$. This π is clearly a sliding block code, so by Theorem 1.7.2, π is continuous and commutes with the shift.

$\Leftarrow$: If X is a factor of an SFT, the factor map is a sliding block code by Theorem 1.7.2, say of window size $2N+1$: $\pi(x)_i = f(x_{i-N}, \dots, x_{i+N})$. Represent the SFT by an edge-labeled graph $\mathcal{G}'$ where the labels are the $2N+1$-words $w \in \mathcal{L}_{2N+1}(X)$. These are all distinct. The factor map turns $\mathcal{G}'$ into an edge-labeled graph $\mathcal{G}$ with labels $f(w)$. Therefore X is sofic. □

Corollary 2.2.5. *Every factor of a sofic shift is again a sofic shift. Every shift conjugate to a sofic shift is again sofic.*

Before we discuss further charaterizations of sofic systems, let us mention that sofic systems with an irreducible transition matrix are always transitive, have a dense set of periodic points, and is mixing if and only if it is totally transitive. , see [16, Theorem 3.3].

Definition 2.2.6. *Given a subshift X and a word $v \in \mathcal{L}(X)$, the* **follower set** *$\mathcal{F}(v)$ is the collection of words $w \in \mathcal{L}(X)$ such that $vw \in \mathcal{L}(X)$.*

Example 2.2.7. *Let X_{even} be the even shift from Example 0.1.13. Then $\mathcal{F}(0) = \mathcal{L}(X_{even})$ because a 0 casts no restrictions on the follower set. Also $\mathcal{F}(011) = \mathcal{L}(X_{even})$, but $\mathcal{F}(01) = 1\mathcal{L}(X) = \{1w : w \in \mathcal{L}(X)\}$. Although each follower set is infinite, there are only these two distinct follower sets. Indeed, $\mathcal{F}(v0) = \mathcal{F}(0)$ for every $v \in \mathcal{L}(X)$, and $\mathcal{F}(0111) = \mathcal{F}(01)$, $\mathcal{F}(01111) = \mathcal{F}(011)$, etc. The follower set $\mathcal{F}(1)$ is not properly defined, but we can ignore this.*

Theorem 2.2.8. *A subshift (X, σ) is sofic if and only if the collection of its follower sets is finite.*

Proof. First assume that the collection $V = \{\mathcal{F}(v) : v \in \mathcal{L}(X)\}$ is finite. We will build an edge-labeled graph representation $\mathcal{G}$ of X as follows:

1. Let V be the vertices of $\mathcal{G}$.

2. If $a \in \mathcal{A}$ and $w \in \mathcal{L}(X)$, then $\mathcal{F}(wa) \in V$; draw an edge $\mathcal{F}(w) \to \mathcal{F}(wa)$, and label t with the symbol a. (Although there are infinitely many $w \in \mathcal{L}(X)$, there are only finitely many follower sets, and we need not repeat arrows between the same vertices with the same label.)

The resulting edge-labeled graph $\mathcal{G}$ represents X.

Conversely, assume that X is sofic, with edge-labeled graph representation $\mathcal{G}$. For every $w \in \mathcal{L}(X)$, consider the collection of paths in $\mathcal{G}$ representing w, and let $T(w)$ be the collection of terminal vertices of these paths. Then $\mathcal{F}(w)$ is the collection of infinite paths starting at a vertex in $T(w)$. Since $\mathcal{G}$ is finite, and there are only finitely many subsets of its vertex set, the collection of follower sets is finite. □

Definition 2.2.9. *An edge-labeled transition graph $\mathcal{G}$ is* **right-resolving** *if for each vertex $v \in \mathcal{G}$, the outgoing arrows all have different labels. It is called* **follower-separated** *if the follower set of each vertex v (i.e., the set of labeled words associated to paths starting in v) is different from the follower set of every other vertex.*

Every sofic shift has a right-resolving follower-separated graph representation and if we minimize the number of vertices in such graph, there is only one such graph with these properties. In fact, the follower set representation $\mathcal{G}$ constructed in the first half of the proof is both right-resolving, follower-separated and minimal. The latter two properties by the choice of V. To see the former, assume that $v \in V$ and $v \to w$ and $v \to w'$ have the same label a. That implies that

$$\mathcal{F}(w) = \{x : ax \in \mathcal{F}(v)\} = \mathcal{F}(w'),$$

so $w = w'$.

Corollary 2.2.10. *Every transitive sofic shift X is synchronizing, and (unless it is a single periodic orbit) has positive entropy. In fact, the entropy $h_{top}(X) = \log \lambda_A$, where λ_A is the leading eigenvalue of the transition graph of the minimal right-resolving representation of X.*

Proof. Let edge-labeled graph $\mathcal{G}$ be the right-resolving follower-separated representation of X. Pick any word $u \in \mathcal{L}(X)$ and let $T(u)$ be the collection of terminal vertices of paths in $\mathcal{G}$ representing u. If $T(u)$ consists of one vertex $v \in V$, then every paths containing u goes through v, and there is a unique follower set $\mathcal{F}(u)$, namely the collection of words representing paths starting in v. In particular, u is a synchronizing word.

If $\#T(u) > 1$, then we show that we can extend u on the right so that it becomes a synchronizing word. Suppose that $v \neq v' \in T(u)$. Since $\mathcal{G}$ is follower-separated, there is $u_1 \in \mathcal{L}(X)$ such that $u_1 \in \mathcal{F}(v)$ but $u_1 \notin \mathcal{F}(v')$ (or vice versa, the argument is the same). Extend u to uu_1. Because $\mathcal{G}$ is right-resolving, u_1 can only represent a single path starting at any single vertex. Therefore $\#T(uu_1) \leq \#T(u)$. But since $u_1 \notin \mathcal{F}(v')$, we have in fact $\#T(uu_1) < \#T(u)$. Continue this way, extending uu_1

until eventually $\#T(uu_1 \dots u_N) = 1$. Then $uu_1 \dots u_N$ is synchronizing. (In fact, what we proved here is that every $u \in \mathcal{L}(X)$ can be extended on the right to a synchronizing word.)

The positive entropy follows from Theorem **??** or Proposition **??**. In fact, since $\mathcal{G}$ is right-resolving, there is an at most $\#V$-to-one correspondence between n-paths starting in $\mathcal{G}$ and words in $\mathcal{L}_n(X)$. Therefore $\#\{n\text{-paths}\} \leq p_X(n) \leq \#V \cdot \#\{n\text{-paths}\}$, and we can use Theorem **??**. □

Remark 2.2.11. *This extends the diagram of Remark* **??** *into:*

$$\text{SFTs} \subset \text{ sofic shifts } \subset \text{ synchronizing subshifts } \subset \text{ coded subshifts } .$$

and irreducible sofic shifts are intrinsically ergodic see [236] and Theorem **??**.

2.3 β-shifts and β-expansions

Throughout this section, we fix $\beta > 1$. A number $x \in [0, 1]$ can be expressed as (infinite) sum of powers of β:

$$x = \sum_{k=1}^{\infty} b_k \beta^{-k} \qquad \text{where} \quad \begin{cases} b_k \in \{0, 1, \dots, \lfloor \beta \rfloor\} & \text{if } \beta \notin \mathbb{N}; \\ b_k \in \{0, 1, \dots, \beta - 1\} & \text{if } \beta \in \{2, 3, 4, \dots\}. \end{cases}$$

For the case $\beta \in \{2, 3, 4, \dots\}$, this is the usual β-ary expansion; it is unique except for the β-adic rationals. For example, if $\beta = 10$, then $0.3 = 0.29999\dots$ If $\beta \notin \mathbb{N}$, then x need not have a unique β-expansion either, but there is a canonical way to do it, called **greedy expansion**:

- Take $b_1 = \lfloor \beta x \rfloor$, that is, we take b_1 as large as we possibly can.
- Let $x_1 = \beta x - b_1$ and $b_2 = \lfloor \beta x_1 \rfloor$, again b_2 is as large as possible.
- Let $x_2 = \beta x_1 - b_2$ and $b_3 = \lfloor \beta x_2 \rfloor$, etc.

In other words, $x_k = T_\beta^k(x)$ for the map $T_\beta : x \mapsto \beta x \pmod 1$, and b_{k+1} is the integer part of βx_k.

Definition 2.3.1. *The closure of the greedy β-expansions of all $x \in [0, 1]$ is a subshift of $\{0, \dots, \lfloor \beta \rfloor\}^{\mathbb{N}}$; it is called the* **β-shift** *and we will denote it as (X_β, σ).*

Note that if $b = (b_k)_{k=1}^{\infty}$ is the β-expansion of some $x \in [0, 1]$, then $\sigma(b)$ is the β-expansion of $T_\beta(x)$. The following lemma ([192]) characterizes the β-shift in terms of $\preceq_{lex}$:

Lemma 2.3.2. *Let $c = c_1 c_2 c_3 \dots$ be the β-expansion of 1. Then $b \in X_\beta$ if and only if*

$$\sigma^n(b) \preceq_{lex} c \text{ for all } n \geq 0,$$

where $\preceq_{lex}$ stands for the lexicographic order.

Example 2.3.3. *Let $\beta = 1.8393\dots$ be the largest root of the equation $\beta^3 = \beta^2+\beta+1$. One can check that $c = 111000000\dots$ Therefore $b \in X_\beta$ if and only if one of*

$$\sigma^n(b) = 0\dots, \qquad \sigma^n(b) = 10\dots, \qquad \sigma^n(b) = 110\dots \qquad \text{or} \qquad \sigma^n(b) = c,$$

holds for every $n \geq 0$. The subshift X_β is itself not of finite type, because there are infinitely many forbidden words 1110^k1, $k \geq 0$, but by some recoding it is easily seen to be conjugate to an SFT (see the middle panel of Figure 2.7), and it has a simple edge-labeled transition graph.

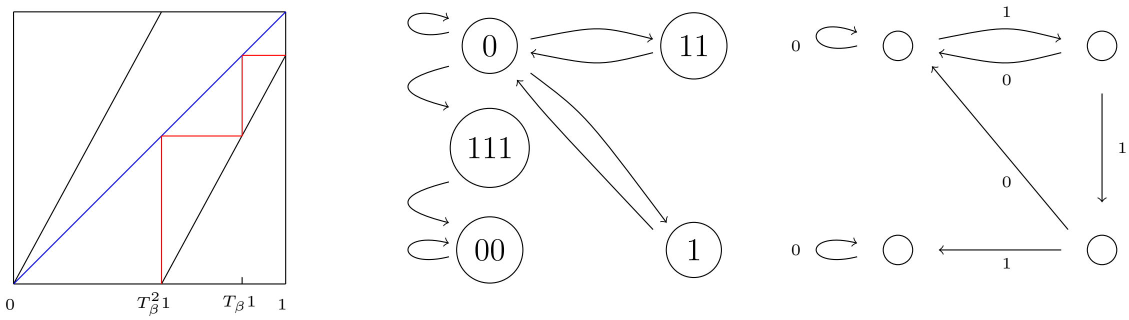


Figure 2.7: Left: The map T_β for $\beta^3 = \beta^2+\beta+1$. Then $T_\beta^3(1) = 0$. Middle: A corresponding vertex-labeled graph. Right: A corresponding edge-labeled graph.

Proof of Lemma 2.3.2. Let $b = (b_k(x))_{k\geq 1}$ be the β-expansion of some $x \in [0,1)$. (If $x = 1$ there is nothing to prove because $b = c$.) Since $x < 1$ we have $b_1 = \lfloor \beta x \rfloor \leq c_1 = \lfloor \beta \cdot 1 \rfloor$. If the inequality is strict, then $b \prec_{lex} c$. Otherwise, $0 \leq x_1 = T_\beta(x) = \beta x - b_1 < \beta \cdot 1 - c_1 = T_\beta(1)$, and we find that $b_2 = \lfloor \beta x_1 \rfloor \leq c_2 = \lfloor \beta T_\beta(1) \rfloor$. Continue by induction.

Conversely, define half-open subintervals of $[0,1]$:

$$\begin{aligned}
A_j &= [\frac{j}{\beta}, \frac{j+1}{\beta}) && 0 \leq j < c_1, \\
A_{c_1 j} &= [\frac{c_1}{\beta} + \frac{j}{\beta^2}, \frac{c_1}{\beta} + \frac{j+1}{\beta^2}) && 0 \leq j < c_1, \qquad (2.3)\\
A_{c_1 c_2 j} &= [\frac{c_1}{\beta} + \frac{c_2}{\beta^2} + \frac{j}{\beta^2}, \frac{c_1}{\beta} + \frac{c_2}{\beta^2} + \frac{j+1}{\beta^2}) && 0 \leq j < c_3, \\
&\vdots \qquad \vdots \qquad \vdots \qquad \vdots && \qquad (2.4)
\end{aligned}$$

They are adjacent and clearly $T_\beta(A_j) = [0,1)$ for $0 \leq j < c_1$. Also $T_\beta(A_{c_1 j}) = [j/\beta, (j+1)/\beta)$ for $0 \leq j < c_2$, and since $\sigma^n((c_k)_{k\geq 1} \preceq_{lex} (c_k)_{k\geq 1}$ by the first part of the proof, we have $c_2 \leq c_1$ and in particular $T_\beta(A_{c_1 j})$ is one of the intervals in the first row of (2.3). Therefore $T_\beta^2(A_{c_1 j}) = [0,1)$. By induction, we obtain

$$T_\beta^{k+1}(A_{c_1 c_2 \dots c_k j}) = [0,1) \qquad \text{for all } k \in \mathbb{N}, 0 \leq k < c_{k+1}. \tag{2.5}$$

In fact, $A_{c_1\dots c_k j} = \{x \in [0,1] : b_n(x) = c_n \text{ for } 1 \le n \le k, b_{k+1}(x) = j\}$.

Now take $(b_k)_{k\ge1} \in \mathcal{A}^{\mathbb{N}}$ such that $(b_k)_{k\ge1} \preceq_{lex} (c_k)_{k\ge1}$, and define $n_0 = 0$ and recursively $n_{r+1} = \min\{k > n_k : b_k \neq c_{k-n_k}\}$. Suppose first that all n_r's are finite. Then $b_{n_r+1} \dots b_{n_{r+1}}$ is the index of one of the intervals in the $n_{r+1} - n_r$'th row of (2.3). The intersection

$$\bigcap_{r\ge0} T_\beta^{-n_r}(A_{b_{n_r+1}\dots b_{n_{r+1}}})$$

(of intervals of length $\le \beta^{-r}$) is a single point x with $(b_k(x))_{k\ge1} = (b_k)_{k\ge1}$. If $n_{s+1} = \infty$ for some $s \ge 0$, and we set $A_{b_{n_s+1}b_{n_s+2}\dots} = \{1\}$, then $\{x\} = \bigcap_{r=0}^s T_\beta^{-n_r}(A_{b_{n_r+1}\dots b_{n_{r+1}}})$ gives again the unique point with $(b_k(x))_{k\ge1} = (b_k)_{k\ge1}$. □

Proposition 2.3.4. *The β-shift is a coded shift.*

Proof. Let $c = c_1c_2c_3\dots$ be the β-expansion of 1. Then we can take as set of code words

$$\begin{aligned} S = \{&\underbrace{0, 1, \dots, (c_1-1)}_{1-\text{words}},\ \underbrace{c_10, c_11, \dots, c_1(c_2-1)}_{2-\text{words}}, \\ &\underbrace{c_1c_20, c_1c_21, \dots, c_1c_2(c_3-1)}_{3-\text{words}}, \quad \dots \\ &\quad\vdots \qquad\qquad \vdots \\ &\underbrace{c_1c_2c_3c_4c_5c_6\dots}_{\text{a single infinite word}}\}. \end{aligned} \tag{2.6}$$

Apart from the single infinite word, these are exactly the indices of the intervals $A_{c_1\dots c_k j}$ in (2.3), and we know from (2.5) that $T_\beta^{k+1}(A_{c_1\dots c_k j}) = [0,1)$, so free concatenations of such code words all represent $(b_k(x))_{k\ge1}$ for some $x \in [0,1]$. Any concatenation in S^* also satisfies Lemma 2.3.2, so that S^* is dense in (and in fact equal to) X_β. □

Corollary 2.3.5. *Every β-transformation is intrinsically ergodic.*

Proof. This was first shown by Hofbauer [146]. Implementing Theorem **??**, we have $\#\{s \in S : |s| = n\} \le \beta$ for each n, so the exponential growth rate of these words is 0. Hence Theorem **??** even implies that every subshift of the β-shift is intrinsically ergodic. □

Example 2.3.6. *In fact, for the β-transformation with slope $\beta > 1$, the measure of maximal entropy is absolutely continuous w.r.t. Lebesgue, and there is an explicit formula for the density:*

$$\frac{d\mu}{dx} = \sum_{n\ge1, T_\beta^n(1)>x} \beta^{-n}$$

see [192].

The following result was probably first stated in [167, Section 6].

Corollary 2.3.7. *For every $\beta \in [1, 2]$, the β-shift (X_β, σ) is hereditary.*

Proof. This follows directly from Lemma 2.3.2 which determines the shape of the code-words in Proposition 2.3.4. Indeed, if $x \in X_\beta$ and $n = \min\{i \geq 1 : x_i \neq c_i\}$. Then $x_n < c_n$ and $x_1 \dots x_n$ is a code word. Now repeat the argument with $\sigma^n(x)$. $\square$

Theorem 2.3.8. *The T_β-orbit of 1*

1. *contains 0 if and only if X_β is conjugate to an SFT;*
2. *preperiodic if and only if X_β is sofic;*
3. *not dense in $[0, 1]$ if and only if X_β is synchronizing.*

Note that, since there are uncountably many choices of $\beta > 1$, all leading to non-conjugate subshifts (see Theorem 2.3.12 below), while there are only countably many sofic shifts, X_β is not sofic for most β.

Proof. First note that if $\beta \in \mathbb{N}$, then X_β is the full shift on β symbols, so clearly an SFT. Assume therefore that β is non-integer.

For statement 1. let $a_j = T_\beta(1)^j$, so $a_0 = 1$ and $a_N = 0$ for some $N \geq 2$. Let $\mathcal{P}$ be the partition given by the branches of T_β^{N-1}. Then $a_j \in \partial\mathcal{P}$ and the image $T_\beta^{N-1}(\partial J) \subset \{a_i\}_{i=0}^N$ for each $J \in \mathcal{P}$. This means that $\mathcal{P}$ is a Markov partition for T_β^{N-1}, and hence (X_β, σ^{N-1}) is a memory $N-1$ SFT over the alphabet $\{0, 1, \dots, \lfloor\beta\rfloor\}$. See Example 2.3.3 for an illustration of this.

For statement 2., and $c = c_1c_2 \dots c_N(c_{N+1} \dots c_{N+p})^\infty$, we claim that X_β only has finitely many different follower sets, see Definition 2.2.6. Let w be a proper prefix of some $s_1s_2s_3 \dots \in S^*$ for the collection of word S from (2.6).. That is, there are $k \geq 1$ and $0 \leq m < |s_k|$ such that $|w| = |s_1 \dots s_{k-1}| + m$. The possible follower sets are

$$\mathcal{F}(w) = \begin{cases} S^* & \text{if } m = 0 \\ \{aS^* : 0 \leq a < c_2\} \cup \{c_2aS^* : 0 \leq a < c_3\} \cup \dots & \text{if } m = 1 \\ \{aS^* : 0 \leq a < c_3\} \cup \{c_3aS^* : 0 \leq a < c_4\} \cup \dots & \text{if } m = 2 \\ \{aS^* : 0 \leq a < c_4\} \cup \{c_4aS^* : 0 \leq a < c_5\} \cup \dots & \text{if } m = 3 \\ \qquad\vdots \qquad\qquad\qquad \vdots & \qquad\vdots \end{cases}$$

Since c is eventually periodic, this list of follower sets becomes periodic as well: for each $i \geq 0$, they are the same for $m = N + i$ and $m = N + p + i$. This proves the claim, so by Theorem 2.2.8, X_β is sofic. If on the other hand, the expansion of 1 is not preperiodic, so the T_β-orbit of 1 is infinite, then there are infinitely many different follower sets by Theorem 2.3.11 below, so X_β cannot be sofic. In fact, it is easy to construct an edge-labeled transtion graph for X_β, see Example 2.3.9.

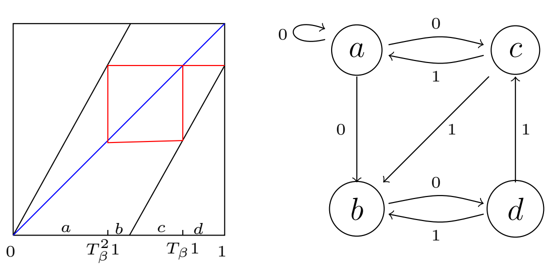
Finally, for statement 3., assume that $\overline{\text{orb}(1)}$ is not dense in $[0,1]$ and let U be an interval that is disjoint from $\overline{\text{orb}(1)}$. Take N so large that the domain Z of an entire branch of T_β^N is contained in U. The set Z is a cylinder set, associated to a unique N-word $v \in \mathcal{L}(X_\beta)$. If $u \in \mathcal{L}(X_\beta)$ is an M-word such that $uv \in \mathcal{L}(X_\beta)$, then the domain Y of the corresponding branch of T_β^M is such that $T_\beta^M(Y) \cap Z \neq \varnothing$. But since $\text{orb}(1) \cap Z = \varnothing$, we have $T_\beta^M(Y) \supset Z$ so that, for every $z \in T_\beta^N(Z)$, there is $y \in Y$ such that $T_\beta^{M+N}(y) = z$. Symbolically, this means that for every word $w \in \mathcal{L}(X)$ such that $vw \in \mathcal{L}(X_\beta)$, also $uvw \in \mathcal{L}(X_\beta)$. In other words, v is synchronizing.

Conversely, suppose that v is some N-word. Then v corresponds to the domain Z of some branch of T_β^N. If orb(1) is dense, then there is $n \in \mathbb{N}$ such that $T_\beta^n(1) \in Z$. Therefore there is a one-sided neighborhood Y of 1 such that $T_\beta^n(Y) = [0, T_\beta^n(1)]$, and there is $x \in Z \setminus T_\beta^n(Y)$. Let w be the itinerary of $T_\beta^N(x)$; since $x \in Y$, $vw \in \mathcal{L}_(X_\beta)$. Similarly, taking $u = c_1 c_2 \dots c_n$, since $T_\beta^n(1) \in Z$, also $uv \in \mathcal{L}(X_\beta)$. However, $uvw \notin \mathcal{L}(X_\beta)$, because there is no $y \in Y$ such that $T_\beta^n(y) = x$. This shows that v is not synchronizing, and since v was arbitrary, X_β is not synchronizing. □

Example 2.3.9. *Let* $\beta = 1.801937735\dots$ *be the largest root of the equation* $\beta^3 = \beta^2 + 2\beta - 1$*. One can check that* $c = 11010101010\dots$ *is preperiodic, and the* T_β*-orbit of* 1 *is* $\{1, \beta - 1, \beta(\beta-1), \beta - 1, \beta(\beta-1), \dots$*. The points* $\{0, \beta(\beta-1), 1/\beta, \beta-1, 1\}$ *define a Markov partition, see Figure 2.8. Therefore the system* $([0,1], T_\beta)$ *can be described as an SFT, but not in the alphabet* $\{0,1\}$*. However, by edge-labeling the transition graph in Figure 2.8, we get* X_β*. Therefore* $b \in X_\beta$ *if and only if one of*

$$\sigma^n(b) = 0\dots, \qquad \sigma^n(b) = 10\dots, \qquad \sigma^n(b) = 110\dots \qquad \text{or} \qquad \sigma^n(b) = c,$$

holds for every $n \geq 0$*. The subshift* X_β *is itself not of finite type, because there are infinitely many forbidden words* 1110^k1*,* $k \geq 0$*, but by some recoding it is easily seen to be conjugate to an SFT (see the middle panel of Figure 2.7), and it has a simple edge-labeled transition graph.*



$$a = [0, \beta(\beta-1)] = [0^\infty, (01)^\infty]$$
$$b = [\beta(\beta-1), 1/\beta] = [0^\infty, (01)^\infty, 01(10)^\infty]$$
$$c = [1/\beta, \beta-1] = [10^\infty, (10)^\infty]$$
$$d = [\beta(\beta-1), 1] = [(10)^\infty, 0(10)^\infty]$$

Figure 2.8: The transition graph for a sofic β-shift

The first two types of β-shifts in Theorem 2.3.8 correspond to certain algebraic properties of β, which we will mention, but not prove. For the definitions of Pisot and Perron number, see Section **??**.

Theorem 2.3.10. *If β is a Pisot number then X_β is sofic. If the subshift X_β is sofic then β is a Perron number.*

See [213] and [30, Chapter 7] for more results in this spirit.

Continuing on the theme of follower sets, let

$$\mathcal{F}(n) := \#\{F : F \text{ is the follower set of some } v \in \mathcal{L}(X_\beta), |v| = n\} \tag{2.7}$$

be the number of distinct follower sets of n-words in $\mathcal{L}(X_\beta)$. Clearly, $\mathcal{F}(n) \leq p(n)$, but in general $\mathcal{F}(n)$ is much smaller. Recall from Theorem 2.2.8 that $\mathcal{F}(n)$ is a bounded sequence if and only if the subshift is sofic. For β-shifts, we see in general linear growth of $\mathcal{F}(n)$.

Theorem 2.3.11. *For every β-shift (X_β, σ) with $\beta > 1$, we have $\mathcal{F}(n) = n + 1$, except when $orb(1)$ is finite; in this case, (X_β, σ) is sofic.*

Proof. This result comes from [188, Theorem 2.25], but we give a different dynamical proof. Set $\beta > 1$, and assume that $c = c_1c_2c_3\dots$ is the β-expansion of 1. Let $\mathcal{D}_0 = [0, 1]$ and in general[1] $\mathcal{D}_n = [0, T_\beta^n(1)]$. First assume that all points $T_\beta^n(1)$ are distinct. The proof will be by induction.

For $n = 0$, there is only one follower set F_0 of the empty word ϵ: $F_0 = \mathcal{L}(X_\beta)$. Therefore $\mathcal{F}(0) = 1$.

For $n = 1$ and $a_1 \neq c_1$, $T_\beta([a_1/\beta, (a_1)/\beta]) = [0, 1] = \mathcal{D}_0$ and the follower set of a_1 is F_0. If $a_1 = c_1$, then $T_\beta([a_1/\beta, 1]) = [0, T_\beta(1)] = \mathcal{D}_1$ and the follower set F_1 of a_1 is equal to the collection of itineraries of points $x \in \mathcal{D}_1$. Therefore $\mathcal{F}(1) = 2$.

For general n, if $v = a_1a_2 \dots a_n$, and k is the smallest index such that $a_{k+1} \dots a_n = c_1 \dots c_{n-k}$, then the corresponding follower set equals F_{n-k}. In particular, if $k = 0$, then the follower set of $a_1 \dots a_n$ is the collection of itineraries of $x \in \mathcal{D}_n$. Hence $\mathcal{F}(n) = n + 1$, proving the statement.

If $\mathcal{D}_n = \mathcal{D}_k$ for some $k < n$ (say n is minimal with this property) then we get no new follower sets anymore, and $\mathcal{F}(m) = n + 1$ for all $m \geq n$. As shown in Theorem 2.2.8, X_β is sofic in this case. □

Theorem 2.3.12. *The β-shift for $\beta > 1$ has topological entropy $\log \beta$.*

Proof. This is a special case of a theorem of interval dynamics saying that every piecewise affine map with slope $\pm\beta$ has entropy $h_{top}(T_\beta) = \log \beta$, but we will give a purely symbolic proof.

[1] This notation is derived the Hofbauer tower construction from Section **??** applied to β-transformations. If the orbit of 1 is infinite, then there are $n + 1$ levels in the tower of height $\leq n$, The image of each n-cylinder under T_β^n is one of these, and therefore $\#\mathcal{F}(n) = n + 1$. The same result holds for unimodal maps, and more general, for interval maps with $d + 1$ branches, $\#\mathcal{F}(n) \leq dn + 1$.

Recall the β-expansion $c = c_1c_2\dots$ of 1 and the set of code words S from (2.6). By Proposition 2.3.4, every word in $\mathcal{L}(X_\beta)$ has the form[2]

$$s_1s_2\dots s_mc_1c_2\dots c_k \quad \text{for some (maximal) } s_1,\dots,s_m \in S,\ k \geq 0. \tag{2.8}$$

Let $p_\beta(n)$ and $p_{S^*}(n)$ be the number of n-words in X_β and S^* respectively. Since every word in S^* is a word in $\mathcal{L}(X_\beta)$, we have $p_{S^*}(n) \leq p_\beta(n)$. Conversely, by (2.8),

$$p_\beta(n) \leq \sum_{0\leq m\leq n} p_{S^*}(m) \leq (n+1) \max_{1\leq m\leq n} p_{S^*}(m).$$

Therefore the exponential growth rates are the same:

$$h_{top}(X_\beta) = \limsup_{n\to\infty} \frac{1}{n}\log p_\beta(n) = \limsup_{n\to\infty} \frac{1}{n}\log p_{S^*}(n).$$

Now to compute the latter, we use generating functions:

$$f_{S^*}(t) = \sum_{n\geq 0} p_{S^*}(n)t^n \quad \text{and} \quad f_S(t) = \sum_{n\geq 1} \#\{s\in S : |s| = n\}t^n.$$

Note that $p_{S^*}(0) = 1$ (the single empty word ϵ) and $\#\{s \in S : |s| = n\} = c_n$. We have $p_{S^*}(n) = \sum_{k=1}^n \#\{s\in S : |s| = k\}p_{S^*}(n-k)$, and this gives for the power series

$$\begin{aligned}
1 + f_{S^*}(t)f_S(t) &= 1 + \sum_{n\geq 0} p_{S^*}(n)t^n \sum_{m\geq 1} \#\{s\in S : |s| = m\}t^m \\
&= 1 + \sum_{N\geq 1}\sum_{k=1}^{N} p_{S^*}(N-k)t^{N-k}\#\{s\in S : |s| = k\}t^k \\
&= 1 + \sum_{N\geq 1} p_{S^*}(N)t^N = f_{S^*}(t).
\end{aligned}$$

Therefore $f_{S^*}(t) = \frac{1}{1-f_S(t)}$. Since $1 = \sum_{n\geq 1} c_n\beta^{-n} = f_S(\beta^{-1})$, β^{-1} is a (simple) pole of f_{S^*} whereas $f_{S^*}(t)$ is well-defined for $|t| < \beta^{-1}$. Hence β^{-1} is the radius of convergence of f_{S^*}, and this means that the coefficients of f_{S^*} satisfy

$$\limsup_{n\to\infty} \frac{1}{n}\log p_{S^*}(n) = \log\beta.$$

This concludes the proof. □

[2] The fact that $\{A_{c_1\dots c_kj} : k\in\mathbb{N}, 0\leq j < c_{k+1}\}$ for a partition of $[0,1)$ show that $(b_k)_{k\geq 1}$ starts with a code word rather than the sufix of a code word for every $x \in [0,1)$.

Chapter 3

Subshifts of zero entropy

3.1 Linear recurrence

Definition 3.1.1. *A subshift (X, σ) is* **linearly recurrent** *if there is $L \in \mathbb{N}$ such that for every k-cylinder Z and every $x \in Z \cap X$, there is $n \leq Lk$ such that $\sigma^n(x) \in Z$.*

This notion is stronger than uniformly recurrent, in that it relates the $N = N(U)$ in the definition of uniform recurrence (in the case that U is a cylinder set) in a "uniform" way to the length of U.

Exercise 3.1.2. *Find minimal subshifts that are not periodically recurrent. Find minimal subshifts that are not linearly recurrent.*

Definition 3.1.3. *Given $u \in \mathcal{L}(X)$, we call w a* **return word** *if*

- *u is a prefix and suffix of wu but u does not occur elsewhere in w;*
- *$wu \in \mathcal{L}(X)$.*

We denote the collection of return words as $\mathcal{R}_u$.

In other words, we can write every $x \in [u]$ as

$$x = w_1 w_2 w_3 w_4 w_5 w_6 \cdots = uw'_1 uw'_2 uw'_3 uw'_4 w'_2 uw'_5 uw'_6 \ldots, \tag{3.1}$$

where $uw'_j = w_j \in \mathcal{R}_u$ for each $j \in \mathbb{N}$., and there no other appearances of u in the rightmost expression. Note that if (X, σ) is minimal (and hence u appears with bounded gaps), then $\mathcal{R}_u$ is finite.

Example 3.1.4. *Construct $\rho \in \{0,1\}^{\mathbb{N}}$ by setting $\rho_1 = 0$, $\rho_2 = 1$ and recursively*

$$\rho_{S_k+1} \ldots \rho_{S_{k+1}} = \rho_1 \ldots \rho_{S_{k-1}}, \qquad k \geq 1$$

for the Fibonacci numbers $S_0, S_1, S_2, S_3, \cdots = 1, 2, 3, 5, \ldots$ This gives

$$\rho = 01\ 0\ 01\ 010\ 01001\ 01001010\ 010010100100\ \ldots$$

(This sequence is in fact the fixed point of the Fibonacci substitution of Example 3.2.2.) If $u = 010010$, then $w = 010 \in \mathcal{R}_u$ because $wu = 010|010010$ starts and ends with u (and these occurrences of u overlap). Note that it is therefore possible that $w \in \mathcal{R}_u$ is shorter than u.

Definition 3.1.5. *A subshift X is called* **square-free** *if $uu \notin \mathcal{L}(X)$ for every $\epsilon \neq u \in \mathcal{L}(X)$. Similarly, X is n-***power free** *if $u^n \notin \mathcal{L}(X)$ for every $\epsilon \neq u \in \mathcal{L}(X)$.*

Theorem 3.1.6 (Durand, Host & Skau [112])**.** *Let (X, σ) is a linearly recurrent subshift with constant L, and which is not periodic under the shift σ. Then*

(i) The word-complexity is sublinear: $p(n) \leq Ln$ for all $n \in \mathbb{N}$.

(ii) X is $L+1$-power free.

(iii) For all $w \in \mathcal{R}_u$, $|u| < L|w|$.

(iv) $\#\mathcal{R}_u \leq L(L+1)^2$.

(v) Every factor (Y, σ) of (X, σ) is linearly recurrent.

Proof. **(i)** Linear recurrence implies that for every $n \in \mathbb{N}$ and every n-word $u \in \mathcal{L}(X)$, the occurrence frequency

$$\liminf_{k\to\infty} \frac{1}{k} \#\{1 \leq i \leq k : x_i \dots x_{i+n-1} = u\} \geq \frac{1}{Ln}$$

for every $x \in X$. Therefore there is no space for more than Ln n-words.
(ii) If an n-word $v \in \mathcal{L}(X)$, then the gap between two occurrences of $v \leq L|v|$, so every word u of length $(L+1)|v| - 1$ contains v at least once. If $v^{L+1} \in \mathcal{L}(X)$, then all n-words are cyclic permutations of v, cf. Proposition 3.3.2. But then $\mathcal{L}(X)$ is shift-periodic.
(iii) Take $u \in \mathcal{L}(X)$ and $w \in \mathcal{R}_u$. If $|u| \geq L|w|$, then the word wu (which starts and ends with u) must in fact have w^{L+1} as prefix. This contradicts (2).
(iv) Take $u \in \mathcal{L}(X)$ and $v \in \mathcal{L}(X)$ of length $(L+1)^2|u|$. By the proof of (2), every word of length $\leq (L+1)|u|$ occurs in v and in particular, every return word $w \in \mathcal{R}_u$ occurs in v. Now return words in v don't overlap (cf. (3.1)), so using the minimal length $|w| \geq |u|/L$ of return words (from item (iii)), we find $\#\mathcal{R}_u \leq |v|/(|u|/L) = L(L+1)^2$.
(v) Finally, suppose that (Y, σ), over alphabet $\mathcal{B}$, is a factor of (X, σ), and $f : \mathcal{A}^{2N+1} \to \mathcal{B}$ is the corresponding sliding block code, so $2N+1$ is its window size. Take $u \in \mathcal{L}(X)$ of length $|u| \geq 2N+1$ and v its image under f. Then $|v| = |u| - 2N$. If $w \in \mathcal{R}_v$, then $|w| \leq \max\{|s| : s \in \mathcal{R}_u\} \leq L|u| \leq L(|v| + 2N) \leq L(2N+1)|v|$. Therefore Y is linearly recurrent with constant $L(2N+1)$. In fact, the proof gives that v will return with gap $\leq L + \varepsilon$ if v is sufficiently long. $\square$

3.2 Substitution shifts

We fix our finite-letter alphabet $\mathcal{A} = \{0, \dots, N-1\}$.

Definition 3.2.1. *A* **substitution** χ *is a map that assigns to every* $a \in \mathcal{A}$ *a single word* $\chi(a) \in \mathcal{A}^*$*:*

$$\chi : \begin{cases} 0 \to \chi(0) \\ 1 \to \chi(1) \\ \quad\vdots \\ N-1 \to \chi(N-1) \end{cases}$$

and extends to $\mathcal{A}^*$ *by concatenation:*

$$\chi(a_1 a_2 \dots a_r) = \chi(a_1)\chi(a_2)\dots\chi(a_r).$$

The substitution is of **constant length** *if* $|\chi(a)|$ *is the same for every* $a \in \mathcal{A}$*.*

Example 3.2.2. *The Fibonacci substitution* χ_{fib} *acts as*

$$0 \to 01 \to 010 \to 01001 \to 01001010 \to 0100101001001 \to \dots$$

The lengths of $\chi^n(0)$ *are exactly the Fibonacci numbers. We will see this word again in Section 3.3 on Sturmian sequences.*

Remark 3.2.3. *As can be seen in Example 3.2.2, if* a *is the first symbol of* $\chi(a)$*, then* $\chi(a)$ *is a prefix of* $\chi^2(a)$*, which is a prefix of* $\chi^3(a)$*, etc. Therefore* $\chi^n(a)$ *tends to a fixed point of* χ *as* $n \to \infty$*.*

Lemma 3.2.4. *For every* $a \in \mathcal{A}$*,* $\chi^n(a)$ *tends to a periodic orbit of* χ *as* $n \to \infty$*.*

Proof. Since $\#\mathcal{A} < \infty$, there must be $p < r \in \mathbb{N} \cup \{0\}$ such that $\chi^p(a)$ and $\chi^r(a)$ start with the same symbol b. Now apply Remark 3.2.3 to χ^{r-p} and b. □

Example 3.2.5. *Take* $\chi(0) = 10$ *and* $\chi(1) = 1$*. Then*

$$0 \to 10 \to 110 \to 1110 \to 11110 \to \dots \to 1^\infty \textit{ fixed by } \chi.$$
$$1 \to 1 \textit{ fixed by } \chi.$$

The second line of this example is profoundly uninteresting, so we will always make the assumption

$$\forall a \in \mathcal{A} \ \lim_{n\to\infty} |\chi^n(a)| = \infty. \tag{3.2}$$

Also we will always take an iterate, and rename symbols, such that

$$\chi(0) \text{ starts with } 0. \tag{3.3}$$

Therefore there is always a fixed point of χ starting with 0.

Example 3.2.6. *The* **Thue-Morse substitution**[1] *is defined by*

$$\chi_{TM}: \begin{cases} 0 \to 01 \\ 1 \to 10 \end{cases}.$$

It has **two** *fixed points*

$$\rho^0 = 01\ 10\ 1001\ 10010110\ 1001011001101001\ \dots$$
$$\rho^1 = 10\ 01\ 0110\ 01101001\ 0110100110010110\ \dots$$

These sequences make its appearance in many settings in combinatorics and elsewhere, cf. [4]. For instance, the n-th entry of rho^0 (where we start counting at $n = 0$) is the parity of the number of 1s in the binary expansion of n. Also, if you have a sequence of objects $(P_k)_{k\geq 1}$ of decreasing quality (e.g. rugby players) which you want to divide over two teams T_0 and T_1, so that the teams are closest in strength as possible, then you assign P_k to team T_i if i is the k-th digit of ρ^0 (or equivalently, of ρ^1). This is the so-called Prouhet-Tarry-Escott problem [41, page 85-96]. The sequences ρ^0 and ρ^1 have also been proved to be binary expansions of transcendental numbers: $\sum_{n\geq 1} \rho^0_n 2^{-n} = 1 - \sum_{n\geq 1} \rho^1_n$ are trancendental, see e.g. [5, Theorem 13.4.2].

Applying the sliding block code $f([01]) = f([10]) = 1$ and $f([00]) = f([11]) = 0$, the images of ρ^0 and ρ^1 are the same:

$$\rho = 10\ 11\ 1010\ 10111011\ 1011101010111010\ \dots \tag{3.4}$$

which is the fixed point of the **period doubling** *or* **Feigenbaum substitution**

$$\chi_{feig}: \begin{cases} 0 \to 11 \\ 1 \to 10 \end{cases}.$$

This sequence appears as the kneading sequence (itinerary of the critical value) of the (infinitely renormalizable) Feigenbaum interval map, see Section **??**. *It is also a Toeplitz sequence, see Example 3.4.3.*

Proposition 3.2.7. *The smallest alphabet size for which square-free subshifts exist is 3. The Thue-Morse sequence is "square+ε-free in the sense that $uuu_1 \notin \mathcal{L}(X)$ for every $u \in \mathcal{L}(X)$ and u_1 is the first letter of u.*

Sketch of Proof. If you try to create a two-letter square-free word you soon get stuck:

$$0 \rightsquigarrow 01 \rightsquigarrow 010 \rightsquigarrow \text{ stuck.}$$

[1]after the Norwegian mathematician Axel Thue (1863-1922) and the American Marston Morse (1892-1977), but the corresponding sequence was used before by the French mathematician Eugène Prouhet (1817-1867), a student of Sturm.

To create a three-letter square-free infinite word, start with a fixed point ρ^0 of the Thue-Morse substitution χ_{TM} and replace the symbol by a 2 if a square threatens to appear:

$$0120\ 1021\ 20210120\ 1021012021201021\ \ldots$$

This turns out to work.

For the Thue-Morse sequence, we work by induction on n in χ^n. At each step, square+εs are avoided, see [5, Theorem 1.6.1] for a complete proof. $\square$

Definition 3.2.8. *A* **substitution subshift** *is any subshift (X, σ) that can be written as $X_\rho = \overline{orb_\sigma(\rho)}$ where ρ is a fixed point (or periodic point) of a substitution satisfying* (3.2)*.*

Lemma 3.2.9. *Each one-sided substitution shift space (X_ρ, σ) allows a two-sided substitution shift extension.*

Proof. First define χ on two-sided sequences as

$$\rho(\ldots x_{-2}x_{-1}x_0.x_1x_2x_3\ldots) = \ldots \rho(x_{-2})\rho(x_{-1})\rho(x_0).\rho(x_1)\rho(x_2)\rho(x_3)\ldots,$$

where the central dot indicates where the zeroth coordinate is.

To create a two-sided substitution shift, take some $i > 1$ such that $\rho_i = 0$, and let $a = \rho_{i-1}$. Similar to the argument of Lemma 3.2.4, there is $b \in \mathcal{A}$ and $p < q \in \mathbb{N}$ such that $\rho^p(a)$ and $\rho^q(a)$ both end in b. Set $N = q - p$, so $\rho^N(b)$ ends with b. Next iterate $\rho^N(b.0)$ repeatedly, so that $\lim_k \rho^{kN}(b.0) =: \hat{\rho}$ is a two-sided fixed point of ρ^N. Finally, set $\hat{X}_\rho = \overline{\{\sigma^n(\hat{\rho}) : n \in \mathbb{Z}\}}$.

Even though $\hat{\rho}$ need not be unique (because the choice of b and N are not unique), due to minimality (see below), the shift space $\hat{X}_\rho$ is unique. $\square$

Definition 3.2.10. *The* **associated matrix** *or* **incidence matrix** *of a substitution χ is the matrix $A = (a_{i,j})_{i,j\in\mathcal{A}}$ such that $a_{i,j}$ is the number of symbols j appearing in $\chi(i)$. We call χ* **aperiodic** *and/or* **irreducible** *if A is* **aperiodic** *and/or* **irreducible***, in the sense of the Perron-Frobenius Theorem, see Definition 2.1.13. The substitution is* **primitive** *if it is both irreducible and aperiodic. Equivalently, χ is irreducible if for every $i, j \in \mathcal{A}$ there exists $n \geq 1$ such that j appears in $\chi^n(i)$.*

Theorem 3.2.11. *Let χ be a substitution satisfying hypotheses* (3.2) *and* (3.3)*. Let ρ be the corresponding fixed point of χ. Then the corresponding substitution subshift (X_ρ, σ) is minimal if and only if for every $a \in \mathcal{A}$ appearing in ρ, there is $k \geq 1$ such that $\chi^k(a)$ contains* 0*.*

Proof. If X_ρ is minimal (i.e., uniformly recurrent according to Proposition 1.5.6), then every word, in particular 0, appears with bounded gaps. Let a be a letter appearing in ρ. Then $\chi^k(a)$ is a word in $\chi^k(\rho) = \rho$, and since $|\chi^k(a)| \to \infty$ by (3.2), $\chi^k(a)$ must contain 0 for k sufficiently large.

Conversely, let $k(a) = \min\{i \geq 1 : \chi^i(a) \text{ contains } 0\}$, and $K = \max\{k(a) : a \text{ appears in } \rho\}$. Set $\Delta_a = \chi^{k(a)}(a)$ and decompose ρ into blocks:

$$\begin{aligned} \rho &= \Delta_{\rho_1}\Delta_{\rho_2}\Delta_{\rho_3}\dots \\ &= \rho_1 \dots \rho_{k(\rho_1)} \quad \rho_{k(\rho_1)+1} \dots \rho_{k(\rho_1)+k(\rho_2)} \quad \rho_{k(\rho_1)+k(\rho_2)+1} \dots \rho_{k(\rho_1)+k(\rho_2)+k(\rho_3)} \quad \dots \end{aligned}$$

By the choice of $k(\rho_j)$, each of these blocks contains a 0, so 0 appears with gap K. Now take $w \in \mathcal{L}(X_\rho)$ arbitrary. There exists $m \in \mathbb{N}$ such that w appears in $\chi^m(0)$. By the above, w appears in each $\chi^m(\Delta_{\rho_j})$ and hence w appears with gap $\max_j |\chi^m(\Delta_{\rho_j})| = \max\{|\chi^{m+k(a)}(a)| : a \text{ appears in } \rho\}$. This proves the uniform recurrence of ρ. □

Theorem 3.2.12 below shows that if χ is primitive, then (X_ρ, σ) is linearly recurrent and hence of linear complexity ($p(n) \leq Ln$) and uniquely ergodic (see Definition 1.2.5). The above theorem doesn't exclude that ρ is periodic. For instance,

$$\chi : \begin{cases} 0 \to 010 \\ 1 \to 101 \end{cases} \tag{3.5}$$

produces two fixed points $\rho^0 = (01)^\infty$ and $\rho^1 = (10)^\infty$. We call a substitution such that its fixed point ρ is not periodic under the shift **aperiodic**. Note that this is different from "the associated matrix of χ is aperiodic", so be aware of this unfortunate confusion of terminology.

A mild assumption dispenses with such periodic examples, and then $p(n) \geq n+1$, see Proposition 3.3.2.

Theorem 3.2.12. *Every primitive substitution shift is linearly recurrent.*

Proof. Let $\chi : \mathcal{A} \to \mathcal{A}^*$ be the substitution with fixed point ρ and (X_ρ, σ) the corresponding shift. Let

$$S_k := \sup\{|\chi^k(a)| : a \in \mathcal{A}\} \quad \text{and} \quad I_k := \inf\{|\chi^k(a)| : a \in \mathcal{A}\}.$$

Note that $I_k \leq S_1 I_{k-1}$ and $I_1 S_{k-1} \leq S_k$ for all $k \in \mathbb{N}$. Since χ is primitive, for every $a, b \in \mathcal{A}$ there exists $N_{a,b}$ such that $\chi^{N_{a,b}}(a)$ contains b. Therefore

$$|\chi^k(b)| \leq |\chi^{k+N_{a,b}}(a)| \leq S_{N_{a,b}} |\chi^k(a)| \quad \text{for all } k \in \mathbb{N}.$$

Hence, taking $N = \sup\{N_{a,b} : a, b \in \mathcal{A}\}$, we find

$$I_k \leq S_k \leq S_N I_k \quad \text{for all } k \in \mathbb{N}.$$

Now let $u \in \mathcal{L}(X_\rho)$ and $v \in \mathcal{R}_u$ be arbitrary. Choose $k \geq 1$ minimal such that $|u| \leq I_k$. Therefore there exists a 2-word $ab \in \mathcal{L}(X_\rho)$ such that u appears in $\chi^k(ab)$. Let R be the largest distance between two occurrences of any 2-word in $\mathcal{L}(X_\rho)$. Then R is finite by minimality of the shift. We have

$$|v| \leq RS_k \leq RS_N I_k \leq RS_N S_1 I_{k-1} \leq RS_N S_1 |u|.$$

This proves linear recurrence with $L = RS_N S_1$. □

Remark 3.2.13. *It turns out (cf. Theorem 3.1.6(v)) that a factor of a substitution subshift is again a substitution subshift. In fact, one of the main results of [112] is that if you keep taking factors of substitution shifts, you will, within a finite number of steps, get a subshift isomorphic to something you saw before.*

A more general result on complexity of substitutions (without the assumption of primitivity) is due to Pansiot [191].

Theorem 3.2.14 (Pansiot)**.** *If $\chi : \mathcal{A} \to A^*$ is a* **non-erasing** *(i.e., $\chi(a) \neq \epsilon$, the empty word, for all $a \in \mathcal{A}$) substitution with $\chi(a) = au$ for some $a \in \mathcal{A}$, $\epsilon \neq u \in \mathcal{A}^*$, then the complexity of $\rho = \lim_{n\to\infty} \chi^n(a)$ is one of the following:*

1. *$p_\rho(n)$ is bounded (when ρ is (pre)periodic);*
2. *$p_\rho(n) \approx n$, including the primitive case;*
3. *$p_\rho(n) \approx n \log\log n$;*
4. *$p_\rho(n) = n \log n$'*
5. *$p_\rho(n) = n^2$.*

Here $p_\rho(n) \approx a(n)$ means that there is $C > 0$ such that $C^{-1}a(n) \leq p_\rho(n) \leq Ca(n)$ for all n sufficiently large.

Example 3.2.15. *If we remove the non-erasing condition in the above theorem, then even more asymptotics for $p(n)$ become possible. Let $\mathcal{A} = \{a, b_0, \dots, b_r\}$ for some $r \neq \mathbb{N}$ and let $\chi : \mathcal{A} \to \mathcal{A}^*$ is given by*

$$\chi : \begin{cases} a \to ab_r, \\ b_k \to b_k b_{k-1}, & \text{for } k = 1, \dots, r, \\ b_0 \to b_0. \end{cases}$$

Then χ has a unique fixed point, which for e.g. $r = 3$ looks like

$$\rho = ab_3.b_3b_2b_2b_1.b_3b_2b_2b_1b_2b_1b_1b_0.b_3b_2b_2b_1b_2b_1b_1b_0b_2b_1b_1b_0b_1b_0b_0\dots$$

Set $v_i = \chi^i(ab_r)$ for $i \geq 0$. The dots separate the blocks w_i, where $w_0 = ab_r$ and w_i is the suffix of v_i of length $|v_i| - |v_{i-1}|$. Then symbol b_k appears exactly $\binom{i}{r-k}$ times in w_i.

Next apply an erasing substitution $\tilde{\chi} : \mathcal{A} \to \{0,1\}^*$ given by

$$\chi : \begin{cases} a \to \epsilon, \\ b_k \to 0, & \text{for } k = 0, \dots, r-1, \\ b_r \to 1. \end{cases}$$

Then

$$\tilde{\rho} := \tilde{\chi}(\rho) = 10^{n_0}1^{n_1}01^{n_2}10^{n_3}1^{n_4}01^{n_5}\dots \qquad \textit{for } n_i = \binom{i}{r} \approx i^r/r!.$$

It can be shown (see [21, Proposition 4.7.2]) that the complexity of $\tilde{\rho}$ is $p_{\tilde{\rho}} \approx n\sqrt[r]{n}$.

For primitive constant length substitutions, there is the following result on their amorphic complexity [124]:

Theorem 3.2.16. *Let $\chi : \{0,1\} \to \{0,1\}^*$ be an aperiodic primitive substitution of constant length ℓ, and let (X,σ) be the associated subshift. Then*

$$ac(\sigma) = \frac{\log \ell}{\log \ell - \log \ell^*} \qquad \textit{for } \ell^* = \#\{1 \le i \le \ell : \chi(0)_i \ne \chi(1)_i\}.$$

In this theorem, $\mathrm{ac}(\sigma) = \infty$ is allowed when the denominator $\log \ell - \log \ell^* = 0$, such as is the case with the Thue-Morse substitution, see Example 3.2.6.

3.2.1 Recognizability

We call a substitution **injective** if $\chi(a) \ne \chi(b)$ for all $a \ne b \in \mathcal{A}$. All the examples above were indeed injective, but in general substitutions are not surjective and hence not invertible, not even as map $\chi : X_\rho \to X_\rho$. But we can still ask:

> Is an injective substitution $\chi : X_\rho \to \chi(X_\rho)$ invertible, and what does the inverse look like?

To illustrate the difficulty here, assume that χ from (3.5) acts on a two-sided shift space. Then what is the inverse of $x = \dots 010101010\dots$. Without putting in the dot to indicate the zeroth position, there are two ways of dividing x into three-blocks,

$$x = \dots|010|101|010|10\dots = \dots 0|101|010|101|0\dots = \dots 01|010|101|010|\dots \tag{3.6}$$

and each with their own inverse. The way to cut x into blocks $\chi(a)$ is called a **1-cutting** of x. The problem is thus: can a sequence $x \in \chi(X_\rho)$ have multiple 1-cuttings if you don't know a priori where the first block starts?

Remark 3.2.17. *We give a brief history of this problem. In 1973, J. C. Martin claimed that any substitution on a two-letter alphabet which is aperiodic is one-sided recognizable (or "rank one determined"). His proof was not convincing. In 1986, Bernard Host proved that a primitive substitution shift X_ρ is one-sided recognizable if and only if $\chi(X_\rho)$ is open in X_ρ. This condition is not so easy to check, though. In 1987, Martine Queffélec announces a short proof of the unilateral recognizability of constant length substitutions due to Gérard Rauzy. Nobody could check this proof. In his 1989 PhD Thesis, M. Mentzen claimed to prove this result, using a paper by T. Kamae of 1972. In 1999, C. Apparicio found a gap in Mentzen proof (Kamae's results* *only works for a particular case of the theorem, namely if the length is a power of a prime number). She solved the problem using a 1978 result by Michel Dekking. In the meantime, in 1992, Brigitte Mossé proved a more general result (also non-constant length), but using a new notion of (two-sided) recognizable substitution. She refined this result in 1996.*

Figure 3.1: Alejandro Maass (1965-) and Brigitte Mossé (1957-).

This problem was tackled by several people (Mentzen, Queffélec [208], Host, Mossé [186, 187]), and it were the results of Mossé that are currently considered as the final answer.

The official terminology is as follows: Fix $x \in X_\rho$ and define the sequences

$$E = \{|\chi(x_1 x_2 \dots x_i)|\}_{i \geq 0}.$$

By convention, the zeroth entry (for $i = 0$) is 0. In short, E_k tells us how to divide x into blocks of length $\chi^k(x_i)$ if we start at 0. Clearly if χ is of constant length M, then $E = \{iM\}_{i \geq 0} \cup \{0\}$.

Definition 3.2.18. *A substitution word $x \in X_\rho$ is*

- **one-sided recognizable** *if there is N such that for every $i, j \in \mathbb{N}$ such that $x_i \dots x_{i+N} = x_j \dots x_{j+N}$ we have $i \in E$ if and only if $j \in E$.*

- **two-sided recognizable** *if there is N such that every $i, j \in \mathbb{N}$ such that $x_{i-N+1} \dots x_{i+N} = x_{j-N+1} \dots x_{j+N}$ we have $i \in E$ if and only if $j \in E$.*

It is **(one- or two-sided) recognizable** *if it is (one- or two-sided) 1-recognizable. We call N the* **recognizability index**.

In this definition, the sequence x from (3.6) is not recognizable, but for example the fixed point of the Fibonacci substitution χ_{fib} is recognizable with recognizability index 2. The Thue-Morse sequence ρ^0 (or ρ^1) is recognizable with recognizability index 4.

Theorem 3.2.19 (Mentzen (1989), Apparicio (1999) [10])**.** *Every primitive injective constant length substitution with aperiodic fixed point is one-sided recognizable.*

For non-constant length substitutions, things are more involved.

Example 3.2.20. *The substitutions*

$$\chi : \begin{cases} 0 \to 0001 \\ 1 \to 01 \end{cases} \qquad \text{and} \qquad \chi_{chac} : \begin{cases} 0 \to 0012 \\ 1 \to 12 \\ 2 \to 012 \end{cases}$$

are not one-sided recognizable. For example, the fixed point of the first one is

$$\rho = 0001\ 0001\ 0001\ \underbrace{01\ 0001}_{u}\ 0001\ 0001\ 01\ 00\underbrace{01\ 0001}_{u}\ 0001\ 01\ 0001\ 01\ \dots$$

and just based on the word $u = u' = 010001$, you cannot say if the cut is directly before its occurrence or not. This problem does not disappear if you take longer words. The latter substitution χ_{chac} is called the **Chacon substitution HB: this needs a reference***.*

Theorem 3.2.21 (Mossé (1992))**.** *Let X_ρ be an aperiodic primitive substitution. If for every $n \in \mathbb{N}$ there exists $v \in \mathcal{L}(X_\rho)$ with $|v| \geq n$ and $a, b \in \mathcal{A}$ such that*

1. *$\chi(a)$ is a proper suffix of $\chi(b)$, and*
2. *$\chi(a)v$ and $\chi(b)v \in \mathcal{L}(X)$ and have the same 1-cutting of v.*

Then χ is **not** *one-sided recognizable.*

Theorem 3.2.22 (Mossé (1992))**.** *Every aperiodic primitive injective substitution is two-sided recognizable.*

Recognizability of aperiodic, but not necessarily primitive, substitution shifts was proved in [24, Theorem 5.17].

3.2.2 S-adic transformations

Mossé [186] proved that for substitution shifts X, $p_X(n+1) - p_X(n)$ is bounded. The same result is true for S-adic shifts [109]; in fact, Bernard Host conjectured that for a subshift $p_X(n+1) - p_X(n)$ is bounded if and only if X is S-adic. For Sturmian shifts, i.e., with $p_X(n+1) - p_X(n) \equiv 1$, see Definition 3.3.14, this is certainly true as explained in Section **??**. More generally, the symbolic itinerary space coming from an N-interval exchange transformation has $p_X(n+1) - p_X(n) \equiv N - 1$.

Definition 3.2.23. *Given a sequence of alphabets $(\mathcal{A}_i)_{i \geq 0}$, symbols $a_i \in \mathcal{A}_i$ and substitutions $\chi_i : \mathcal{A}_i \to \mathcal{A}_{i-1}^*$, assume that*

$$\rho := \lim_{i \to \infty} \chi_1 \circ \chi_2 \circ \cdots \circ \chi_i(a_i)$$

exists. If also the χ_i's are taken from a finite collection $\mathcal{X}$, then we call the subshift $(\overline{orb_\sigma(\rho)}, \sigma)$ an **S-adic shift***.*

3.3 Sturmian subshifts

Sturmian sequences mostly emerge as symbolic dynamics of circle rotations or similar systems. There are however at least three equivalent defining properties, to which we will devote separate sections.

Figure 3.2: Gustav Hedlund (1904-1993) and Marston Morse (1892-1977).

The name **Sturmian** comes from Morse & Hedlund [185], seemingly because they appear in connection with the work of the French mathematician Jacques Sturm (1803-1855), namely in regard to the number of zeroes that $\sin(\alpha x + \beta)\pi$ has in the interval $[n, n+1)$, but the sequences as such were certainly not studied by Sturm. Other ways to obtain Sturmian sequences are manifold. For instance, if you take a piece of paper with a square grid, and draw a line on it with slope α, and write a 0 whenever it crosses a horizontal grid-line and a 1 whenever it crosses a vertical grid-line (see Figure 3.3 left), then you obtain a Sturmian sequence. Or the trajectory of a billiard ball moving frictionless on a rectangular billiard table can be coded symbolically by writing a 0 for each collision with a long edge and a 1 for each collision with a short edge (see Figure 3.3 right). If the motion is not periodic, the resulting sequence is Sturmian.

For simplicity of exposition, we use the property that $p(n) = n+1$ for n as **Sturmian**, see Section 3.3.3. We start with some terminology and a useful proposition.

Definition 3.3.1. *We call an n-word u*

- **left-special** *if both $0u$ and $1u$ belong to $\mathcal{L}(X)$;*
- **right-special** *if both $u0$ and $u1$ belong to $\mathcal{L}(X)$;*
- **bi-special** *if u is both left-special and right-special.*

Note, however, that there are different types of bi-special words u depending on how many of the words $0u0, 0u1, 1u0$ and $1u1$ are allowed.

Proposition 3.3.2. *If the word-complexity of a subshift (X, σ) satisfies $p(n) \le n$ for some n, then (X, σ) is periodic.*

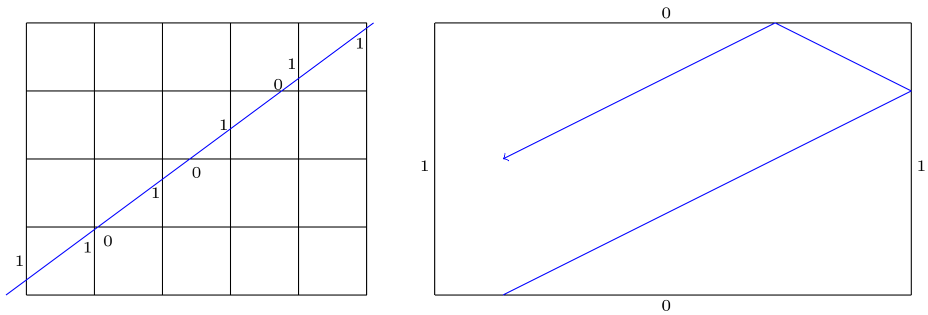


Figure 3.3: Sturmian sequences produced as intersections with horizontal and vertical grid-lines (left) and billiards on a rectangular billiard table (right)

Proof. Let n be maximal such that $p(k) \geq k$ for all $k \leq n$. Then $p(n) = p(n+1) = n$ and there are no right-special words of length n. Start with an n-word u; there is only one way to extend it to the right by one letter, say to ua. Then the n-suffix of ua can also be extended to the right by one letter in only one way. Continue this way, until after at most $p(n) = n$ steps, we end up with suffix u again. Therefore X contains only (shifts of) this word periodically repeated. $\square$

3.3.1 Rotational sequences

Definition 3.3.3. *Let $R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$, $x \mapsto x + \alpha \pmod 1$, be the rotation over an irrational angle α. Let $\beta \in \mathbb{S}^1$ and build the itinerary $\boldsymbol{i}(x) = u = (u_n)_{n \geq 0}$ by*

$$u_n = \begin{cases} 1 & \text{if } R_\alpha^n(x) \in [0, \alpha), \\ 0 & \text{if } R_\alpha^n(x) \notin [0, \alpha). \end{cases} \tag{3.7}$$

Then u is called a **rotational sequence**.

Remark 3.3.4. *The additional sequences obtained by taking the closure can also be obtained by taking the half-open interval the other way around:*

$$u_n = \begin{cases} 1 & \text{if } R_\alpha^n(x) \in (0, \alpha], \\ 0 & \text{if } R_\alpha^n(x) \notin (0, \alpha]. \end{cases}$$

In either way, the resulting two-sided subshift $(X_\alpha = \overline{orb_\sigma(u)}, \sigma)$ is an extension of $(\mathbb{S}^1, R_\alpha)$ where $\boldsymbol{i} : \mathbb{S}^1 \to X_\alpha$ is the inverse factor map $\boldsymbol{i} = \psi^{-1}$. Therefore the $x_n = R_alpha^n(0)$, $n \in \mathbb{Z}$, have fibers $\psi^{-1}(x_n)$ consisting of two point, whereas $\#\psi^{-1}(x) = 1$ for all other x. Thus (X_α, σ) is an almost one-to-one extension of the circle rotation.

Lemma 3.3.5. *Every rotational word u is* **palindromic**: *for every finite subword $w_1 w_2 \dots w_n$ occurring in u, also the reversed word $w_n w_{n-1} \dots w_1$.*

Proof. By symmetry, the two-sided itinerary of $\beta := \alpha/2$ is a palindrome entirely: $u_n = u_{-n}$ for all $n \in \mathbb{Z}$. Since $\{k\alpha + \beta \pmod 1\}_k$ is dense in the circle and uniformly recurrent, every subword $w_1 w_2 \dots w_n$ in every itinerary will have its reversed copy $w_n w_{n-1} \dots w_1$ in the same itinerary. □

Lemma 3.3.6. *If w is a bi-special subword of a rotational sequence, then it coincides with a prefix of $\boldsymbol{i}(2\alpha \pmod 1)$ of length $q_n + aq_{n+1} - 2$ for some $n \in \mathbb{N}$ and $0 \leq a < a_{n+1}$, where p_n/q_n are the convergents of the continued fraction expansion $\alpha = [0; a_1, a_2, a_3, \dots]$ (see Section **??**).*

Proof. Each subword w corresponds to a subinterval J_w of the circle, namely the interval of points x such that $\boldsymbol{i}(x)$ starts with w. If w is left-special, so $0w$ and $1w$ are both allowed, then $R_\alpha^{-1}(J_w)$ contains 0 or α in its interior. In the former case, $\alpha \in J_w^\circ$, so not all $x \in J_w$ have the same first letter in their itinerary. Therefore $\alpha \in R_\alpha^{-1}(J_w^\circ)$ and $R_\alpha^2(0) \in J_w^\circ$.

Let $\hat{J}_w := R_\alpha^{-2}(J_w^\circ) \ni 0$. Now if w is also right-special, then $R_\alpha^{|w|+2}(\hat{J}_w^\circ) = R_\alpha^{|w|}(J_w^\circ) \ni 0$, and therefore $y := R_\alpha^{-(|w|+2)} \in \hat{J}_w^\circ$. This means that y is preimage of 0 such that no preimage of 0 of lower order belongs to $(0, y)$.

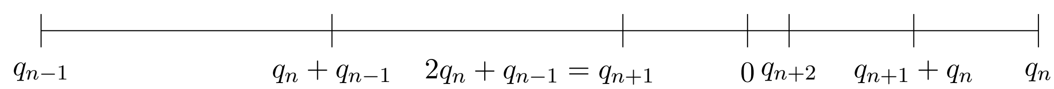


Figure 3.4: Positions of the preimages of 0 under R_α that are closest to 0.

The points y with this property are ordered as in Figure 3.4, where the numbers j refer to the points $R_\alpha^{-j}(0)$. Therefore $|w| + 2 = q_n + aq_{n+1}$ and the lemma follows. □

Exercise 3.3.7. *Show that every bi-special word of a rotational sequence (so Sturmian sequence by Lemma 3.3.17) is a palindrome.*

We give a bit of the wider theory of circle homeomorphism so as also to include Denjoy circle maps which have minimal sets that are exactly conjugate to Sturmian shifts.

Theorem 3.3.8 (Denjoy). *A circle homeomorphism $f : \mathbb{S}^1 \to \mathbb{S}^1$ has a well-defined* **rotation number**

$$\rho(f) = \lim_{n\to\infty} \frac{F^n(x) - x}{n} \pmod 1, \qquad \textit{independent of and uniformly in } x,$$

where $F : \mathbb{R} \to \mathbb{R}$ is a **lift** *of f, i.e., a continuous map of the universal cover $\mathbb{R}$ of $\mathbb{S}^1$ such that $F(x) \pmod 1 = f(\pmod 1)$. Furthermore,*

- $\rho(f) = \frac{p}{q} \in \mathbb{Q}$ *(in lowest terms) if and only if f has a q-periodic point;*

- *if $\rho = \rho(f) \notin \mathbb{Q}$, then f is semi-conjugate to the rotation R_ρ: $h \circ f = R_\rho \circ h$. In fact, h is a conjugacy if and only if f is minimal.*

For the proof we refer to [**?**], but let us give some details on how non-minimal circle homeomorphisms f with irrational rotation numbers can be constructed. Start with the rotation $R_\rho : \mathbb{S}^1 \to \mathbb{S}^1$ and select some $x_1 \in \mathbb{S}^1$ (or in fact any finite or countable set of points $x_j \in \mathbb{S}^1$ such that $|x_j - x_k|/\rho \notin \mathbb{Z}$ for $j \neq k$). For each k and $n \in \mathbb{Z}$, replace $R_\rho^n(x_k)$ by a closed interval $I_{k,n}$ of length $2^{-(k+|n|)}$; this creates a new circle K with circumference $1 + \sum_k \sum_{n\in\mathbb{Z}} 2^{-(k+|n|} = 1 + 3\sum_k 2^{-k}$. Extend $f : I_{k,n} \to I_{k,n+1}$ as an affine (or any orientation preserving) homeomorphism, and for all $x \in \mathbb{S}^1 \setminus \cup_{k,n} R_\rho^n(x_k)$ set $f(x) = R_\rho(x)$. Then $f : K \to K$ is indeed a homeomorphism, and $h : K \to \mathbb{S}^1$,

$$h(x) = \begin{cases} R_\rho^n(x_k) & \text{if } x \in I_{k,n}, \\ x & \text{otherwise,} \end{cases} \tag{3.8}$$

is a semiconjugacy, see Figure 3.5. Such circle homeomorphisms f are called **Denjoy circle maps**. There is some restriction on how smooth such homeomorphisms can be. Denjoy proved that if f is a C^1 diffeomorphism such that $\log f'$ has bounded variation, then f is minimal. On the other hand, for every $\gamma \in [0, \gamma)$, there are $C^{1+\gamma}$ Denjoy circle maps, see [141].

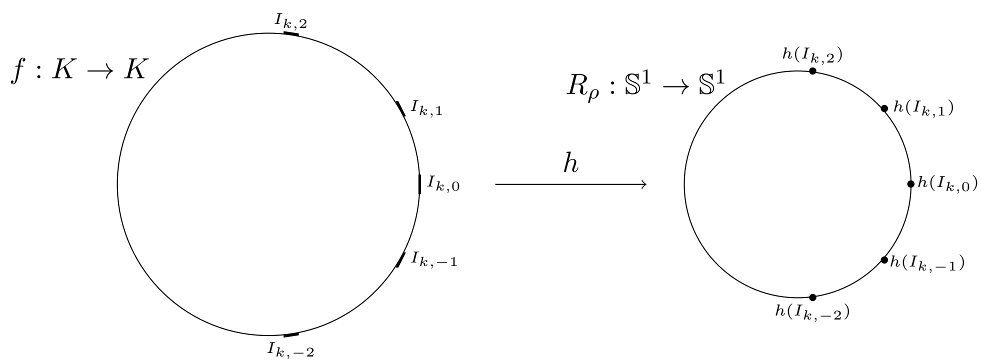


Figure 3.5: The semiconjugacy h from a Denjoy circle map to a rotation

Take R_ρ, split open the orbit of 0, replacing the points $R_\rho^n(0)$ by intervals I_n, and denote the corresponding Denjoy circle map by $f : K \to K$. Then $K \setminus \cup_n I_n^\circ$ is a minimal Cantor set. If we code $[\sup I_0, \inf I_1] \cap X$ by 1 and $[\sup I_1, \inf I_0] \cap X$ by 0, then the coding map $\boldsymbol{i} : X \to \{0,1\}^{\mathbb{Z}}$ is precisely a conjugacy between (X, f) and a two-sided rotational shift X_ρ with frequency $\rho = \rho(f)$.

If we split open $\mathbb{S}^1$ only along the backward orbit of 0, then we obtain a one-sided Sturmian shift.

Theorem 3.3.9. *The amorphic complexity of any non-periodic two-sided rotational subshift (X_ρ, σ) is 1. Equivalently, $ac(f) = 1$ for any Denjoy circle map $f : K \to K$.*

Proof. Since the two-sided shift $\sigma : X_\rho \to X_\rho$ is conjugate to $f : C \to C$ for $C = K \setminus \cup_{k,n} I^\circ_{k,n}$, it suffices to show that $\mathrm{ac}(f_{|C}) = 1$.

Take three points $\xi_1, \xi_2, \xi_3 \in \cup_{k,n} I_{k,n}$ such that $d(h(\xi_i), h(\xi_j)) \geq \frac{1}{4}$ for $i \neq j$. Let $\delta := \min\{|I_{k,n}| : I_{k,n} \ni \xi_j \text{ for some } j\}$ be the minimal length of the intervals corresponding to the ξ_is.

Since $h(\cup_{k,n} I_{k,n})$ is a countable set, we can take $N := \lfloor 1/v$ points in C such that $S := \{h(x_i) : i = 1, \dots, N\}$ is an equidistant lattice in $\mathbb{S}^1$ with minimal mutual distance $1/N$. Set $J = [x_i, x_j]$ for some $i \neq j$, ordered in such a way that $|h(J)| \leq \frac{1}{2}$. Whenever $R^n_\rho(h(J)) \ni \xi_1$, $|f^n(J)| \geq \delta$, but $\mathbb{S}^1 \setminus R^n_\rho(h(J))$ has length $\geq 1/2$, so it must contain ξ_2 and/or ξ_3. Therefore also $|K \setminus f^n(J)| \geq \delta$, and thus $d(f^n(x_i), f^n(z_j) \geq \delta$.

Since

$$\lim_{n\to\infty} \#\{0 \leq k < n : R^k_n(h(J)) \ni \xi_1\} = \mathrm{Leb}(h(J)) \geq \frac{1}{N} \geq v,$$

we obtain $\limsup_{n\to\infty} \#\{0 \leq k < n : d(h^k(x_i), h^k(x_j)) \geq \delta\} \geq v$, so S is (δ, v)-separated. Since $\#S \geq \frac{1}{v} - 1$, it follows that $\underline{ac}(f) \geq 1$.

Now for the other direction, we will use (δ, v)-spanning sets, see Remark **??**. Define a function $\psi_v : \mathbb{S}^1 \to [0, |K|]$ (where $|K|$ is the circumference of K) as

$$\psi_v(x) = \mathrm{Leb}(h^{-1}([x, x+v]).$$

Note that $d(x,y) \leq \mathrm{Leb}(h^{-1}([h(x), h(y)]))$ (because $d(x,y)$ measures the shortest arc between x and y), and $\psi_v(d(h^{-1}([x, x+v]))$ for all v sufficiently small and x outside the countable set $h(\cup_{k,n} I_{k,n})$. Therefore ψ_v is measurable and in fact L^1. The Birkhoff Ergodic Theorem 4.2.3 implies that for Leb-a.e. $y \in \mathbb{S}^1$,

$$\lim_{n\to\infty} \frac{1}{n} \#\{0 \leq k < n : \psi_v(R^k_\rho(y)) \geq \delta |K|\} = \mathrm{Leb}(\{\psi_v \geq \delta|K|\}) =: m_v. \tag{3.9}$$

We claim that $m_v \leq 2v(\lfloor 1/\delta \rfloor + 1)$. Indeed, if $m_v > 2v(\lfloor 1/\delta \rfloor + 1)$, then the set $\{\psi_v \geq \delta|K|\}$ cannot be contained in the union of at most $\lfloor 1/\delta \rfloor + 2$ intervals of length v. Therefore there are $\tilde{N} = \lfloor 1/\delta \rfloor + 2$ points $\xi_i \in \mathbb{S}^1$ such that $\psi_v(\xi_i) \geq \delta|K|$ and of minimal mutual distance $d(\xi_i, \xi_j) \geq v$. Therefore

$$\sum_{i=1}^{\tilde{N}} |h^{-1}([\xi_i, \xi_i + v])| = \sum_{i=1}^{\tilde{N}} \psi_v(\xi_i) \geq \tilde{N}\delta|K| \geq (1+\delta)|K|,$$

contradicting that $h^{-1}([\xi_i, \xi_i + v])$ are $\tilde{N}$ disjoint intervals inside a circle of circumference $|K|$. This proves the claim.

Hence we can find a set $S = \{y_1, y_2, \dots, y_N\}$ for $N = \lfloor 1/v \rfloor$ such that $h(S)$ is an equidistant lattice on $\mathbb{S}^1$ (with minimal mutual distance $1/N$) and (3.9) holds for every $h(y_i)$. Without loss of generality, the y_is can be arranged in circular order on K.

Now take $y \in K$ arbitrary, and i such that $y \in [y_i, y_{i+1 \pmod N})$. Then $h(y) \in [h(y), h(y)+v)$ and $d(f^k(y_i), f^k(y)) \leq \phi_v(R_\rho^k(h(y_i))$. Therefore

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n}\#\{0 \leq k < n : d(f^k(y_i), f^k(y)) \geq \delta\} \\ \leq \limsup_{n\to\infty} \frac{1}{n}\#\{0 \leq k < n : \psi_v(R_\rho^k(h(y_i)) \geq \delta\} = m_v,\end{aligned}$$

which means that S is (δ, m_v)-spanning. Using the spanning set equivalent of (**??**), we obtain

$$\overline{ac}(f) \leq \sup_{\delta|K|>0} \limsup_{v\to 0} \frac{\log 2v(\lfloor 1/\delta \rfloor + 1)}{-\log v} = 1,$$

and the result follows. □

3.3.2 Balanced words

Another characterization of Sturmian words its by it property of being balanced.

Definition 3.3.10. *A subshift X is called* **balanced** *if there exists $N \in \mathbb{N}$ such that for all $a \in \mathcal{A}$ and $n \in \mathbb{N}$, the number of symbols a within any two n-words w, w' in $\mathcal{L}(X)$, differs by at most N. If N is not specified, then* **balanced** *stands for balanced with $N = 1$.*

Definition 3.3.11. *Clearly, a balanced word x contains precisely one of* 00 *and* 11 *as factors, (unless $x = 10101010\dots$ or $x = 01010101\dots$). We say that a balanced word $x \in \{0,1\}^{\mathbb{N} \text{ or } \mathbb{Z}}$ is of* **type** *i is the word ii appears in x.*

Lemma 3.3.12. *Every rotational sequence is balanced.*

Proof. An equivalent way to to define a rotational sequence u is that there is a fixed $\beta \in \mathbb{S}^1$ such that

$$u_n = \lfloor n\alpha + \beta \rfloor - \lfloor (n-1)\alpha + \beta \rfloor \tag{3.10}$$

for all $n \in \mathbb{Z}$. This is easy to check, except that in order to include the sequences mentioned in Remark 3.3.4, we need to add the alternative definition

$$u_n = \lceil n\alpha + \beta \rceil - \lceil (n-1)\alpha + \beta \rceil \tag{3.11}$$

for all $n \in Z$.

Write $|u|_a = \#\{1 \leq i \leq n : u_i = a\}$. By telescoping series,

$$\begin{aligned}|u_{k+1} \dots u_{k+n}|_1 &= \lfloor (k+1)\alpha + \beta \rfloor - \lfloor k\alpha + \beta \rfloor + \\ &\quad \lfloor (k+2)\alpha + \beta \rfloor - \lfloor (k+1)\alpha + \beta \rfloor + \dots \\ &\quad + \lfloor (k+n)\alpha + \beta \rfloor - \lfloor (k+n-1)\alpha + \beta \rfloor \\ &= \lfloor (k+n)\alpha + \beta \rfloor - \lfloor k\alpha + \beta \rfloor = \lfloor n\alpha \rfloor \text{ or } \lfloor n\alpha \rfloor + 1\end{aligned}$$

regardless of what k is. It follows that u is balanced. □

Lemma 3.3.13. *If X is an unbalanced subshift on alphabet $\{0,1\}$, then there is a (possibly empty) word w such that both $0w0,\ 1w1 \in \mathcal{L}(X)$.*

Proof. Given $u = u_1 \dots u_n \in \mathcal{L}(X)$ and $a \in \mathcal{A}$, write $|u|_a = \#\{1 \leq i \leq n : u_i = a\}$. Let K be minimal such there are K-words $a = a_1 \dots a_K$ and $b = b_1 \dots b_K \in \mathcal{L}(X)$ such that $|\ |a|_1 - |b|_1\ | \geq 2$. Since $|a|_1 - |b|_1$ can change by at most 1 if a, b are shortened or expanded by one letter, the minimality of K implies that $a = 0a_2 \dots a_{K-1}0$ and $b = 1b_2 \dots ab_{K-1}1$ (or vice versa) and $|a_2 \dots a_{K-1}|_1 = |b_2 \dots b_{K-1}|_1$. If $a_2 \dots a_{K-1} = b_2 \dots b_{K-1}$, then we have found our word w. Otherwise, take $k = \min\{j > 1 : a_j \neq b_j\}$ and $l = \max\{j < K : a_j \neq b_j\}$. We have four possibilities, all leading to shorter possible words a and b.

$$
\begin{array}{lll}
 & \quad\ \ k \quad\ \ l & \\
a = 0 \dots & 1 \dots 1 & \dots 0 \\
b = 1 \dots & \underbrace{0 \dots 0}_{\text{shorter } a,b} & \dots 1
\end{array}
\qquad
\begin{array}{lll}
 & \ \ k \qquad\ \ l & \\
a = 0 \dots & 1 \dots\ 0 \dots & 0 \\
b = 1 \dots & 0 \dots\ \underbrace{1 \dots 1}_{\text{shorter } a,b} &
\end{array}
$$

$$
\begin{array}{lll}
 & \quad\ \ k \qquad l & \\
a = & 0 \dots 0 & \dots 1 \dots 0 \\
b = & \underbrace{1 \dots 1}_{\text{shorter } a,b} & \dots 0 \dots 1
\end{array}
\qquad
\begin{array}{lll}
 & \ \ k \qquad\ \ l & \\
a = 0 \dots & 0 \dots\ 0 \dots & 0 \\
b = 1 \dots & 1 \dots\ \underbrace{1 \dots 1}_{\text{shorter } a,b} &
\end{array}
$$

This contradicts the minimality of K. The proof is complete, but note that we have proved that $|w| \leq K - 2$ as well. $\square$

3.3.3 Sturmian sequences

Definition 3.3.14. *A sequence $u \in \{0,1\}^{\mathbb{N}}$ or $\{0,1\}^{\mathbb{Z}}$ is called* **Sturmian** *if it is recurrent under the shift σ, and the number of n-words in u equals $p_u(n) = n+1$ for each $n \geq 0$. Take the shift-orbit closure $X = \overline{\mathrm{orb}_\sigma(u)}$. The corresponding subshift (X, σ) for $X = \overline{\mathrm{orb}_\sigma(u)}$ is called a* **Sturmian subshift**.

Remark 3.3.15. *The assumption that u is recurrent is important for the two-sided case. Also $\dots 00000100000 \dots$ has $p(n) = n+1$, but we don't want to consider such asymptotically periodic sequences. In fact, for one-sided infinite words, the recurrence follows from the assumption that $p_u(n) = n+1$.*

Remark 3.3.16. *A Sturmian sequence contains exactly one left-special and one right-special word of length n for each $n \in \mathbb{N}$. If they coincide, then this is a bi-special word, see Lemma 3.3.6.*

Lemma 3.3.17. *Every rotational sequence is Sturmian.*

Proof. Let $u(x)$ denote the itinerary of $x \in \mathbb{S}^1$. If $u_k(x) = u_k(y)$ for $0 \leq k < n$, then $R_\alpha^k(x)$ and $R_\alpha^k(y)$ belong to the same set $[0, \alpha)$ or $[\alpha, 1)$ for each $0 \leq k < n$. In other words, the interval $[x, y)$ contains no point in $Q_n := \{R_\alpha^{-k}(x) : 0 \leq k \leq n\}$. But Q_n consists of exactly $n+1$ points, and it divides the circle in $n+1$ intervals. Each such interval corresponds to a unique n-word in the language, so $p(n) = n + 1$. □

Theorem 3.3.18. *A non-periodic sequence is Sturmian if and only if it is balanced.*

Exercise 3.3.19. *Show that there are balanced periodic sequences. In fact, every finite word in a Sturmian shift, when repeated periodically, is balanced.*

Proof. Let $x \in \mathcal{A}^{\mathbb{N}}$ or $\mathcal{A}^{\mathbb{Z}}$ for $A = \{0, 1\}$.

$\Leftarrow$: We prove by contrapositive, so assume that there is N minimal such that $p(N) \geq N + 2$. (Recall from Proposition 3.3.2 that $p(N) \leq N$ implies that x is periodic.) Since $p(1) = \#\mathcal{A} = 2$ and 00 and 11 cannot both be words of x (otherwise it wouldn't be balanced at word-length 2), $N \geq 3$. In fact, for all $n < N - 1$, there is one right-special word, but there are two distinct right-special words, say u and v, of length $N - 1$. In particular, u and v can only differ at their first symbol, because otherwise there are two distinct right-special words of length $N - 2$. Hence there is w such that $0w = u$ and $1w = v$. But since u and v are right-special, $0w0$ and $1w1$ are both words in x, and x cannot be balanced.

$\Rightarrow$: Again, proof by contrapositive, so assume that $p(n) = n + 1$ for all $n \in \mathbb{N}$, but x is not balanced. Let N be the minimal integer where this unbalance becomes apparent. We have $p(2) = 3$. Since both 01 and 10 occur in x (otherwise it would end in 0^∞ or 1^∞) at least one of 00 and 11 cannot occur in x, and hence $N \geq 3$.

By Lemma 3.3.13, there is a word $w = w_1 \dots w_{N-2}$ such that both $0w0$ and $1w1$ occur in x.

Observe that $w_1 = w_{N-2}$, because otherwise both 00 and 11 occur in x. To be definite, suppose that $w_1 = w_{N-2} = 0$.

If $N = 3$, then $w_1 = w_{N-2}$, so w is a palindrome. If $N \geq 4$, then $w_2 = w_{N-3}$ because otherwise 000 and 101 both occur in x, contradicting that N is the minimal length where the unbalance becomes apparent.

Continuing this way, we conclude that w is a palindrome: $w_k = w_{N-k-1}$ for all $1 \leq k \leq N - 2$.

Since $p(N - 2) = N - 1$ and w is bi-special, exactly one of $0w$ and $1w$ is right-special. Say $0w0$, $0w1$ and $1w1$ occur, but not $1w0$.

Claim: if $1w1$ is a prefix of the $2N - 2$-word $x_{j+1} \dots x_{j+2N-2}$, then $0w$ does not occur in this word.

Suppose otherwise. Since $|1w1| = N$ and $|0w| = N - 1$, the occurrence of $0w$ must overlap with $1w1$, say starting at entry k. Then $w_k \dots w_{N-2}1 = 0w_1 \dots w_{N-k-1}$, so $w_k = 0 \neq 1 = w_{N-k-1}$. This contradicts that w is a palindrome, and proves the claim.

Now $x_{j+1} \dots x_{j+2N-2}$ contains N words of length $N - 1$, but not $0w$, according to the claim. That means that one of the remaining $N - 1$-words must appear twice, and none of these words is right-special. It follows that $x_{j+1} \dots x_{j+2N-2}$ can only be continued to the right periodically, and $p(n) \leq N$ for all n. This contradiction concludes the proof. □

Proposition 3.3.20. *If the infinite sequence u is balanced, then*

$$\alpha := \lim_{n\to\infty} \frac{1}{n}|u_1 \dots u_n|_1$$

exists and is irrational. We call α the **frequency** *of u.*

Proof. Define

$$M_n = \min\{|u_{k+1} \dots u_{k+n}|_1 : k \geq 0\}. \tag{3.12}$$

Since u is balanced, $\max\{|u_{k+1} \dots u_{k+n}|_1 : k \geq 0\} = M_n + 1$, so $|u_{k+1} \dots u_{k+n}|_1 = M_n$ or $M_n + 1$ for every $k \in \mathbb{N}$. For $q, n \in \mathbb{N}$ such that $n > q^2$, we can write $n = kq + r$ for a unique $k \geq q$ and $0 \leq r < q$. We have

$$kM_q \leq M_{kq+r} = M_n \leq k(M_q + 1) + r. \tag{3.13}$$

Dividing by n gives

$$\frac{M_q}{q} - \frac{1}{q}\frac{kM_q}{n} \leq \frac{M_q}{q} + \frac{2}{q}.$$

Since this holds for all $q \leq q^2 < n$, we conclude that $\{\frac{M_n}{n}\}_{n\in\mathbb{N}}$ is a Cauchy sequence, say with limit α.

Now to prove that α is irrational, assume by contradiction that $\alpha = \frac{p}{q}$ and take $k = 2^m$ in (3.13). This gives

$$\frac{M_q}{q} \leq \frac{M_{2^m q}}{2^m q} \leq \frac{M_{2^m q} + 1}{2^m q} \leq \frac{M_q + 1}{q},$$

so $\{\frac{M_{2^m q}}{2^m q}\}_m$ is increasing and $\{\frac{M_{2^m q+1}}{2^m q}\}_m$ is decreasing in m. They converge to $\frac{p}{q}$, so $p = M_q$ or $M_q + 1$. But this can only be if every q-word in u has exactly M_q or exactly $M_q + 1$ ones in it, which is of course not true. This completes the proof. □

Lemma 3.3.21. *If u an u' are balanced words with the same frequency α, then u and u' generate the same language.*

Proof. From the proof of Proposition 3.3.20 we know that $\alpha \in (\frac{M_n}{n}, \frac{M_n+1}{n})$ and $\alpha \in (\frac{M'_n}{n}, \frac{M'_n+1}{n})$ where M_n and M'_n are given by (3.12) for u and u' respectively. This implies that $M_n = M'_n$ for all $n \in \mathbb{N}$. Since for each $n \in \mathbb{N}$, u and u' each have only one right-special n-word, it suffices to prove that these right-special word, say w and w' are the same. Assume by contradiction that there is some minimal n such that $w \neq w'$. Hence there is an $n-1$-word v such that $w = 0v$ and $w' = 1v$ (or vice versa). But v is right-special, so all four of $0v0$, $0v1$, $1v0$ and $1v1$ occur in the combined languages. But then $M_{n+1} = |v|_1 \leq M'_{n+1} - 1$, a contradiction. □

Theorem 3.3.22 (Morse & Hedlund [185]). *Every Sturmian sequence is rotational.*

Proof. Let u be a Sturmian sequence; by Theorem 3.3.18 it is balanced as well. By Proposition 3.3.20, u has an irrational frequency $\alpha = \lim_n \frac{1}{n}|u_1 \dots u_n|_1$, and by Lemma 3.3.21, every Sturmian sequence with frequency α generates the same language as u. It is clear that the rotational sequence $v_n = \lfloor n\alpha \rfloor - \lfloor (n-1)\alpha \rfloor$ has frequency α. Therefore there is a sequence b_j such that $\sigma^{b_j}(v) \to u$. By passing to a subsequence if necessary, we can assume that $\lim_j R_\alpha^{b_j}(0) = \beta$. Then (assuming that $n\alpha + \beta \notin \mathbb{Z}$, so we can use continuity of $x \mapsto \lfloor x \rfloor$ at this point):

$$\begin{aligned} u_n = \lim_j (\sigma^{b_j} v)_n &= \lim_j \lfloor (n+b_j)\alpha \rfloor - \lfloor (n+b_j-1)\alpha \rfloor \\ &= \lfloor n\alpha + \beta \rfloor - \lfloor (n-1)\alpha + \beta \rfloor. \end{aligned}$$

If $n\alpha + \beta \in \mathbb{Z}$, then we need to take the definition (3.11) into account. Note, however, that since $\alpha \notin \mathbb{Q}$, this occurs at most for one value of $n \in \mathbb{Z}$. This proves the theorem. □

3.4 Toeplitz shifts

Definition 3.4.1. *A sequence $x \in \mathcal{A}^{\mathbb{N}}$ (resp. $x \in \mathcal{A}^{\mathbb{Z}}$) is called a* **Toeplitz sequence** *if for every $i \in \mathbb{N}$, there exists $q_i \in N$ such that $x_i = x_{i+kq_i}$ for all $k \in \mathbb{N}$ (resp. $k \in \mathbb{Z}$). The orbit closure $X_{\boldsymbol{q}} = \{\sigma^n(x) : n \geq 0\}$ is called a* **Toeplitz shift**.

The notion as introduced and name by Jacobs & Keane [151]. They took inspiration of construction by Otto Toeplitz (1881 – 1940) [221] to create an almost periodic function on the real line, but otherwise, Toeplitz was not involved.

Figure 3.6: Konrad Jacobs (1928 – 2015) and Mike Keane (1940 –)

Proposition 3.4.2. *If $\chi : \mathcal{A} \to \mathcal{A}^*$ is a constant length substitution such that $\chi(a)$ starts with the same symbol for each $a \in \mathcal{A}$, then the unique fixed point of χ is a Toeplitz sequence.*

Proof. Fix the symbol $a \in \mathcal{A}$ such that $\chi(a)$ starts with a, so $\rho = \rho_1\rho_2\rho_3 \cdots = \lim_n \chi^n(a)$ is the fixed point of χ. Let $N = |\chi(b)|$ for each $b \in \mathcal{A}$. Then clearly $\rho_{1+kN} = a$ for all $k \in \mathbb{N}$, so we can take $q_1 = N$.

It follows that $\chi(\rho_1 \dots \rho_{1+kN})$ (which has length $kN^2 + N$) starts and ends with $\chi(a)$. Therefore $q_i = N^2$ for $i = 2, \dots, N$. Continuing by induction, we find $q_i = N^r$ for $i = N^{r-1} + 1, \dots N^r$. □

Example 3.4.3. *The simplest way to construct a Toeplitz sequence emerges from taking $q_i = 2^i$, the powers of 2, and $x_{q_i/2+kq_i} = \frac{1}{2}(1 - (-1)^i)$ for all $k \geq 0$ and $i = 1, 2, 3, \dots$ The resulting Toeplitz sequence is the Feigenbaum sequence,*

$$\rho_{feig} = 1011101010111011101110101011101010111010101110111011101010111011\dots$$

see Example 3.2.6 for more details on this sequence. Although ρ_{feig} is Toeplitz, not every sequence in $X_{feig} = \overline{orb_\sigma(\rho_{feig})}$ has the Toeplitz property. For example, ρ_{feig} has two preimages in X_{feig}, namely $0\rho_{feig}$ and $1\rho_{feig}$. Of these two, only $0\rho_{feig}$ is a Toeplitz sequence.

*As will be shown in Section **??**, ρ_{feig} is the kneading sequence of an infinitely renormalizable unimodal map. In fact, the kneading sequence of every infinitely renormalizable unimodal map is a Toeplitz sequence. More generally, Alvin [7, 8] classifies all the Toeplitz sequences which appear as a kneading sequence (and for which the unimodal maps act on $\omega(c)$ as (strange) adding machines).*

Proposition 3.4.4. *The Thue-Morse sequence*

$$\rho_{TM} = 1001\,0110\,0110\,1001\,0110\,1001\,1001\,0110$$

is obtained from the Thue-Morse substitution $\chi_{TM} : 0 \mapsto 01, 1 \mapsto 10$. Show that ρ_{TM} not a Toeplitz sequence.

Sketch of Proof. One can show for each $q \in \mathbb{N}$, that ρ_{1+kq} cannot have the same value for all k, i.e., ρ_1 is not the first element of any infinite arithmetic progression. In fact, there are no infinite arithmetic progression $k_j = jq + r$ at all such that ρ_{k_j} is the same for all j. This follows from estimates of the possible lengths of such arithmetic progression in the Thue-Morse sequence by Pashina, see [194]. However, the shift generated by x factorizes to the Feigenbaum substitution shift via the sliding block code $01, 10 \to 1$, $00, 11 \to 1$ (see Example 3.2.6) and the Feigenbaum substitution shift is Toeplitz. □

Lemma 3.4.5. *A Toeplitz shift $(X_{\boldsymbol{q}}, \sigma)$ is uniformly rigid and hence minimal.*

Proof. We give the proof for one-sided Toeplitz sequences; the proof of two-sided sequences goes likewise. Let $[x_1x_2 \dots x_n]$ be any cylinder set. Then every digit x_i reappears with gap q_i. Hence, if $L_n = \text{lcm}(q_1, \dots, q_n)$ is the least common multiple of $q_1, \dots, q_n$, then $\sigma^{kL}([x_1x_2 \dots x_n]) \subset [x_1x_2 \dots x_n]$ for all $k \in \mathbb{N}$. This is uniform rigidity. The minimality of the corresponding subshift follows from Lemma 1.5.11 and Corollary 1.5.9. □

The way to build up a Toeplitz sequence in $\{0,1\}^{\mathbb{N} \text{ or } \mathbb{Z}}$, is to start with $x_1 = 1$, choose q_1 and set $x_{1+k_1n} = 1$ for all $k \in \mathbb{N}$ (or $\mathbb{Z}$ for a two-sided Toeplitz sequence, but we will focus on the one-sided Toeplitz sequences). The rest of the entries get a "temporary $*$" $x_i = *$. Next set $x_2 = 0$, choose q_2 (not coprime with q_1) and set $x_{2+kq_2} = 0$. Continuing this way inductively, let x_i be the first remaining temporary $*$s and choose $q_i - i$ a multiple of the period of the pattern of the remaining $*$s. The periodic sequence $\mathrm{Sk}(q_j) \in \{0,1,*\}^{\mathbb{N}}$ of the j-th line of this construction is called the q_j-**skeleton** of the Toeplitz sequence.

Example 3.4.6. *As an example of building*

$$\begin{array}{rl} q_1 = 3: & 1{*}{*}1{*}{*}1{*}{*}1{*}{*}1{*}{*}1{*}{*}1{*}{*}1{*}{*}\dots \\ q_2 = 6: & 10{*}1{*}{*}10{*}1{*}{*}10{*}1{*}{*}10{*}1{*}{*}10{*}\dots \\ q_3 = 3: & 1011{*}11011{*}11011{*}11011{*}1101\dots \\ q_4 = 12: & 1011011011{*}11011011011{*}1101\dots \\ \vdots & \quad \vdots \qquad \vdots \qquad \vdots \end{array} \tag{3.14}$$

In most cases, q_{j+1} is a multiple of q_j, but (3.14) *shows that this is not necessary. However, if $q = (q_j)_{j\geq 1}$ is such that q_j divides q_{j+1} for all $j \in \mathbb{N}$, then we call q the* **periodic structure** *of the Toeplitz sequence x.*

This construction of skeletons yield an extension of Proposition 3.4.2.

Theorem 3.4.7. *The one-sided sequence $x \in \mathcal{A}^{\mathbb{N}}$ is Toeplitz if and only if there is a sequence of constant length substitutions $\chi_i : \mathcal{A}_i \to \mathcal{A}_{i-1}$ on finite alphabets $\mathcal{A}_i$ with $\mathcal{A} = \mathcal{A}_0$ such that $\chi_i(a)$ starts with the same symbol for each $a \in \mathcal{A}_i$, and $x = \lim_{i\to\infty} \chi_1 \circ \chi_1 \circ \cdots \circ \chi_i(a)$, $a \in \mathcal{A}_i$ arbitrary.*

Proof. Let $N_i = \chi_i(a)$ be the length of the words from the i-th substitution. By the condition that $x_1 = \chi_1(a)$ for all $a \in \mathcal{A}_1$, we find $x_{1+kN_1} = x_1$ for all $k \in \mathbb{N}$. By composing $\chi_1 \circ \chi_2$, we obtain $x_1 \dots x_{N_1} = x_{1+kN_1N_2} \dots x_{N_1+kN_1N_2}$ for all $k \in \mathbb{N}$. In general, the initial block $x_1 \dots x_{N_1N_2\cdots N_r}$ repeats with period $N_1N_2\cdots N_rN_{r+1}$, so x is Toeplitz.

Conversely, if $x = x_1x_2x_3\dots$ is Toeplitz on alphabet $\mathcal{A}_0$, then there is N_1 such that $x_{1+kN_1} = x_1$ for all $k \in \mathbb{N}$, and there is a finite collection of N_1-words b_k, $k = 1,\dots,K_1$, all starting with x_1 such that $x = b_{k_1}b_{k_2}b_{k_3}\dots$ Consider $\{b_k\}_{k=1}^{\mathbb{N}}$ as the letters of alphabet $\mathcal{A}_1$, and define the substitution $\chi_1(b_k(\text{as letter}) = b_k(\text{as } N_1\text{-word})$. Then $x = \chi_1(b_{k_1}b_{k_2}b_{k_3}\dots)$. Since the N_1-words b_{k_i} appear with their own gap, $b_{k_1}b_{k_2}b_{k_3}\dots \in \mathcal{A}_1^{\mathbb{N}}$ is a Toeplitz sequence on its own right, and we can repeat the construction. $\square$

3.4.1 Regular Toeplitz sequences

When constructing a Toeplitz sequence this way, at step n, you have an L_n-periodic sequence, where $L_n = \mathrm{lcm}(q_1,\dots,q_n)$. We call the Toeplitz sequence **regular** if

$\frac{1}{L_n}\#\{0 < i < L_n : x_i = *\} \to 0$ as $n \to \infty$. In fact, the official definition is slightly weaker:

Definition 3.4.8. *A sequence $x \in \mathcal{A}^{\mathbb{N}}$ or $\mathcal{A}^{\mathbb{Z}}$ is a* **regular Toeplitz sequence** *if it is the limit of skeletons $Sk(L_n) \in (\mathcal{A} \cup \{*\})^{\mathbb{N}}$ or $(\mathcal{A} \cup \{*\})^{\mathbb{Z}}$ of period L_n such that*

$$\lim_n \frac{Sk_*(L_n)}{L_n} = 0 \quad \textit{where} \quad Sk_* := \#\{1 \le i \le L_n : Sk(L_n)_i = *\}.$$

Theorem 3.4.9. *A regular Toeplitz shift has zero entropy.*

Proof. We follow [169, Theorem 4.76]. Let $V(i)$ be the L_i-word in $(\mathcal{A}\cup\{*\})^{L_i}$ obtained in the i-th step of the construction of Example 3.4.6, i.e., we have now an L_i-periodic skeleton $\mathrm{Sk}(i) = V(i)^\infty \in (\mathcal{A} \cup \{*\})^{\mathbb{N}}$. Let $r_i = |V(i)|_*$ be the number of $*$s in $V(i)$. Then there are at most $\#\mathcal{A}^{r_i}$ ways to fill in the $*$s later on, and there are at most $\#\mathcal{A}^{r_i}$ L_i-words in the Toeplitz sequence x starting at a position $1 + kL_i$. Therefore $q_x(L_i) \le L_i\#\mathcal{A}^{r_i}$, and

$$\lim_{i\to\infty} \frac{1}{L_i} \log p_x(L_i) \le \lim_{i\to\infty} \frac{\log L_i + r_i \log \#\mathcal{A}}{L_i} \le \log \#\mathcal{A} \lim_{i\to\infty} \frac{r_i}{L_i} = 0.$$

Since $p_x(n)$ is subadditive, Fekete's Lemma **??** implies that $\lim_n \frac{1}{n} \log p_x(n) = 0$. $\square$

The following upper bound for the amorphic complexity of regular Toeplitz sequences was shown in [125].

Theorem 3.4.10. *Let $(X_{\boldsymbol{q}}, \sigma)$ be a Toeplitz sequence with periodic structure $\boldsymbol{q} = (q_j)_{j=1}^\infty$. Then $\overline{ac}(\sigma) \le \limsup_{j\to\infty} \frac{\log q_{j+1}}{-\log Sk_*(q_j)}$. In particular, if $q_{j+1} \le C_1 q_j^t$ and $Sk_*(q_j) \le C_2 q_j^{-u}$, then $\overline{ac}(\sigma) \le \frac{t}{u}$.*

With some more work, and for the two-letter alphabet, one can improve the upper bound to $\overline{ac}(\sigma) \le \limsup_{j\to\infty} \frac{\log q_j}{-\log \mathrm{Sk}_*(q_j)}$. By stipulating further properties on the Toeplitz sequence, one can (see [125, Section 5]) give examples showing that this upper bound is sharp, and also that for a dense set of values $a \in [1, \infty]$ (including $a = 1$ and $a = \infty$), there is a Toeplitz shift with $\mathrm{ac}(\sigma) = a$.

Proof of Theorem 3.4.10. Note that the densities $\mathrm{Sk}_*(q_j)$ are decreasing in j, and by regularity of the Toeplitz shift, $\lim_j \mathrm{Sk}_*(q_j) \to 0$. Choose $\delta > 0$ arbitrary and $m \in \mathbb{N}$ such that $2^{-m} < \delta$. Next choose v arbitrary and j such that $(2m+1)\mathrm{Sk}_*(q_{j+1}) < v \le (2m+1)\mathrm{Sk}_*(q_j)$. Then

$$\mathrm{Sep}(\delta, v) \le \mathrm{Sep}(2^{-m}, (2m+1)\mathrm{Sk}_*(q_j)).$$

We claim that the right hand side is bounded by q_{j+1}. Indeed, assume by contradiction that there is a $(2^{-m}, (2m+1)\mathrm{Sk}_*(q_j))$-separated set S with more than q_{j+1} elements. Then at least two of them, say $x, y \in S$, share the same q_{j+1}-skeleton. This means that x and y differ at most in $q_{j+1}\mathrm{Sk}_*(q_{j+1})$ positions in every q_{j+1}-block. Since $d(\sigma^k(x), \sigma^k(y)) \geq \delta$ only if $x_i \neq y_i$ for some i with $|i-k| \leq m$,

$$\begin{aligned}\#\{0 \leq k < nq_{j+1} &: d(\sigma^k(x), \sigma^k(y)) \geq \delta\} \\ &\leq (2m+1)\#\{0 \leq k < nq_{j+1} : x_k \neq y_k\} \leq (2m+1)\mathrm{Sk}_*(q_{j+1}).\end{aligned}$$

When taking the limit $n \to \infty$, we get a contradiction with the choice of j. This proves the claim.

Therefore $\mathrm{Sep}(\delta, v) \leq p_{j+1}$. Take logarithms and divide left and right hand side by $-\log v \geq -\log(2m+1)\mathrm{Sk}_*(j)$ respectively gives

$$\frac{\log \mathrm{Sep}(\delta, v)}{-\log v} \leq \frac{\log q_{j+1}}{-\log(2m+1) - \log \mathrm{Sk}_*(j)}.$$

Note that m depends only on δ. Thus taking the superior limit $v \to 0$ (and hence $j \to \infty$), we obtain $\overline{ac}(\sigma) \leq \limsup_j \frac{\log q_{j+1}}{-\log \mathrm{Sk}_*(j)}$ as claimed. □

Theorem 3.4.11. *For every real number $K \geq 0$, there is a Toeplitz shift (X, σ) such that $h_{top}(\sigma) = K$. For every real number $K \geq 1$, there is a Toeplitz shift (X, σ) that has polynomial word-complexity with exponent K, i.e., $\lim_{n\to\infty} \frac{\log p(n)}{\log n} = K$.*

Proof. We start with the positive entropy Toeplitz sequence, following [169, Theorem], who in turn follows [237]. Let $\mathcal{A}$ be an alphabet such that $\log \#\mathcal{A} \geq 2K$ and take a sequence $(k_i)_{i\in\mathbb{N}}$ such that $\prod_{i=1}^{\infty}(1 - \frac{1}{k_i}) = \frac{2K}{\log \#\mathcal{A}} \in (0, 1)$. Start with an L_0-word $V(0)$ containing $r_0 = L_0/2$ $*$s. We construct the i-th skeleton $V(i)^\infty$ with $|V(i) = L_i$ recursively. Given $V(i)$, let $W(i)$ be the concatenation of the $(\#\mathcal{A})^{r_i}$ copies of $V(i)$ where the r_i $*$s are replaced by the $(\#\mathcal{A})^{r_i}$ r_i-words n $\mathcal{A}$. Then set

$$V(i+1) := W(i)V(i)^{(\#\mathcal{A})^{r_i}(k_i-1)},$$

so that $|V(i+1)| = k_i(\#\mathcal{A})^{r_i}$, each non-$*$ symbol in $V(i)$ returns with periodic gap $\leq L_i$, and $V(i+1)$ contains $r_{i+1} = r_i\frac{k_i-1}{k_i}$ $*$s.

It follows that $\lim_i \frac{r_i}{L_i} = \frac{r_0}{L_0}\prod_{i=1}^{\infty}(1 - \frac{1}{k_i}) = \frac{r_0}{L_0}\frac{2K}{\log \#\mathcal{A}} > 0$ (so regularity fails). The number $p_x(L_i)$ of L_i in x is bounded below by $(\#\mathcal{A})^{r_i}$ (namely the words that start at a positive $1 + kL_i$) and bounded above by $L_i(\#\mathcal{A})^{r_i}$ (all starting positions). Therefore

$$\frac{r_i \log \#\mathcal{A}}{L_i} \leq \frac{\log p_x(L_i)}{L_i} \leq \frac{\log L_i + r_i \log \#\mathcal{A}}{L_i}$$

whence $\lim_i \frac{\log p(L_i)}{L_i} = \frac{r_0 \log \#\mathcal{A}}{L_0}\prod_{i=1}^{\infty}(1 - \frac{1}{k_i}) = \frac{r_0}{L_0}2K = K$. Fekete's Lemma **??** shows that the entropy $\lim_n \frac{\log p(n)}{n} = K$.

We will not give the examples with logarithmic complexity $\lim_n \frac{\log p_x(n)}{\log n} = K \geq 1$, but the technique is the same. □

3.4.2 Adding machines

Adding machines (as the more general enumeration system in Section **??**) are a class of symbolic systems that are not subshifts. They are also called **odometers**, after the device in a car to measure distance. Such an odometer consists of a number of disks, with the digits $0, \dots, 9$ written on the edge. A single "tick" moves the right-most disk by one unit, and if the 9 is passed (so the disk is back at position 0), it ticks over the second disk by one unit, see Figure 3.7.

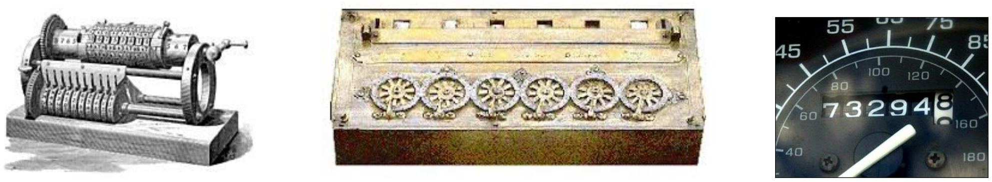


Figure 3.7: Some mechanical adding machines and the decimal odometer of a car.

A mathematical odometer has infinitely many disks, and the number of digits may vary from disk to disk.

The most common one is the **dyadic adding machine** $\boldsymbol{a} : \Sigma \to \Sigma$ for $\Sigma = \{0, 1\}^{\mathbb{N}}$. For $x \in \Sigma$, let $k = \inf\{i : x_i = 0\}$. Then $\boldsymbol{a}$ is defined as

$$\boldsymbol{a}(x)_i = \begin{cases} 0 & i < k, \\ 1 & i = k, \\ x_i & i > k. \end{cases} \tag{3.15}$$

In particular, if $x = 111\dots$, so $k = \infty$, then $\boldsymbol{a}(x) = 000\dots$.

In more generality, we can choose a sequence $\boldsymbol{p} := (p_i)_{i \geq 1}$ of integers $p_i \geq 2$, and define $\boldsymbol{a}$ on $\Sigma_{\boldsymbol{p}} := \{(x_i)_{i \geq 1} : x_i \in \{0, 1, \dots, p_i - 1\}\}$ analogously to (3.15). It is also instructive to view this procedure algorithmically, as "add one and carry".

$$\begin{array}{rl} c := 1 \quad ; & k := 1 \\ \text{Repeat} & \\ & s := x_k + c; \\ & \text{If } s \geq p_k \text{ then } c := 1 \text{ else } c := 0 \\ & x_k := s \pmod{p_k} \ ; \ k := k + 1 \\ \text{Until} & c = 0 \end{array} \tag{3.16}$$

In fact, $\Sigma_{\boldsymbol{p}}$ is a group under the same rule of "add and carry", and $\boldsymbol{a} : \Sigma_{\boldsymbol{p}} \to \Sigma_{\boldsymbol{p}}$ is invertible.

Proposition 3.4.12. *Every odometer $(\Sigma_{\boldsymbol{p}}, \boldsymbol{a})$ is a topological group under addition.*

Proof. The addition $z = x + y$ of two sequences $x, y \in X$ with add-and-carry goes according to the algorithm

$$\begin{array}{rl} c := 0 \quad ; & k := 1 \\ \text{Repeat} & \text{for all } k \in \mathbb{N} \\ & s := x_k + y_k + c; \\ & \text{If } s \geq p_k \text{ then } c := 1 \text{ else } c := 0 \\ & z_k := s \pmod{p_k} \ ; \ k := k+1 \end{array}$$

It is straightforward to check that this is continuous in x and y. □

Exercise 3.4.13. *Show that an odometer $(\Sigma_{\boldsymbol{p}}, \boldsymbol{a})$ is conjugate to its own inverse $(\Sigma_{\boldsymbol{p}}, \boldsymbol{a}^{-1})$.*

Remark 3.4.14. *There is a common alternative way to write adding machines. Given $\boldsymbol{p} = (p_j)_{j \in \mathbb{N}}$, define a sequence $\boldsymbol{q} = (q_j)_{j \in \mathbb{N}_0}$ by $q_0 = 1$ and $q_j = \prod_{k=1}^{j} p_k$. Set*

$$\tilde{\Sigma}_{\boldsymbol{q}} = \{y = (y_j)_{j=1}^{\infty} : y_j \in \{0, \dots, q_j - 1\}, q_{j-1} | (y_j - y_{j-1}) \textit{ for all } j \in \mathbb{N}\},$$

where $y_0 = 0$ by convention. Define $b : \Sigma_{\boldsymbol{p}} \to \tilde{\Sigma}_{\boldsymbol{q}}$ by

$$b(x)_k = \sum_{j=1}^{k} x_j q_{j-1} \quad \textit{with inverse} \quad b^{-1}(y)_k = \frac{y_k - y_{k-1}}{q_{k-1}}. \tag{3.17}$$

Then b is a homeomorphism, and

$$b \circ \boldsymbol{a} = \tilde{\boldsymbol{a}} \circ b \quad \textit{for} \quad \tilde{\boldsymbol{a}}(y)_k = y_k + 1 \pmod{q_k} \textit{ for all } k \in \mathbb{N}.$$

If car odometers were constructed as $\tilde{\Sigma}_{\boldsymbol{q}}$, then $q_j = 10^j$ and the j-th figure on the odometer would be the total number of kilometers driven $\pmod{10^j}$.

Proposition 3.4.15. *Every odometer is uniformly rigid and hence periodically recurrent.*

Proof. Take $\varepsilon > 0$ arbitrary and take k such that $2^{-k} < \varepsilon$. Let $q_k = p_1 p_2 \dots p_k$. Then $\boldsymbol{a}^{q_k}(x)_i = x_i$ for all $i \leq k$, i.e., $d(\boldsymbol{a}^{q_k}(x), x) < \varepsilon$ as required. Periodic recurrence follows by Lemma 1.5.11. □

Proposition 3.4.16. *Every odometer is strictly ergodic, i.e., it is minimal and has a unique invariant probability measure, see Section 4.3.*

Proof. Given any n-cylinder Z, every $x \in \Sigma_{\boldsymbol{p}}$ will visit it exactly once in in every $p_1 p_2 \cdots p_n$ iterates of $\boldsymbol{a}$. Therefore $\mathrm{orb}_{\boldsymbol{a}}(x)$ is dense in $\Sigma_{\boldsymbol{p}}$ and the only $\boldsymbol{a}$-invariant probability measure has $\mu(Z) = (p_1 p_2 \cdots p_n)^{-1}$. □

Proposition 3.4.17. *Every odometer is an isometry, and hence of zero entropy.*

Proof. Let $x, y \in \Sigma_{\boldsymbol{p}}$ and $n = \min\{i \geq 1 : x_i \neq y_i\}$, so $d(x,y) = 2^{-n}$. It is easy to check that $\min_i\{\boldsymbol{a}(x)_i \neq \boldsymbol{a}(y)_i\} = n$ as well. Therefore $\boldsymbol{a}$ is an isometry, and in particular equicontinuous. Proposition 1.6.2 shows that $h_{top}(\boldsymbol{a}) = 0$. $\square$

Proposition 3.4.18. *An odometer has no subshift other than periodic subshifts as continuous factors. However, an odometer can be a factor of a subshift.*

Proof. Clearly the restriction of $\boldsymbol{a}$ to the first n digits gives an $p_1p_2\cdots p_n$-periodic orbit. However, since $\boldsymbol{a}$ is an isometry, it cannot have an expansive continuous factor, and by Proposition **??**, all non-periodic transitive subshifts are expansive.

Conversely, take the Feigenbaum substitution shift (X_{feig}, σ) with $X_{feig} = \overline{\mathrm{orb}_\sigma(\rho)}$ for the fixed point

$$\rho_{\text{feig}} = \rho_0\rho_1\rho_2\cdots = 1011\ 1010\ 10111011\ 1011101010111010\ 1011\ldots$$

The shift is invertible on X_{feig}, except that ρ_{feig} itself has two preimages $0\rho_{\text{feig}}$ and $1\rho_{\text{feig}}$. We define a factor map φ onto the dyadic inverse odometer $(X, \boldsymbol{a}^{-1})$, for $\Sigma = \{0,1\}^{\mathbb{N}}$. Since odometers are conjugate to their own inverses (see Exercise 3.4.13), this gives a factor mp onto $(\Sigma, \boldsymbol{a})$ too.

Follow the following algorithm:

$$\begin{array}{rcl} y_1 & := & \min\{n \geq 1 : x_n = 0\} \pmod 2, \\ y_2 & := & \min\{n \geq 1 : x_{y_1+2n} = 1\} \pmod 2, \\ y_3 & := & \min\{n \geq 1 : x_{y_1+2y_2+4n} = 0\} \pmod 2, \\ y_4 & := & \min\{n \geq 1 : x_{y_1+2y_2+4y_3+8n} = 0\} \pmod 2, \\ \vdots & & \vdots \end{array}$$

and set $\varphi(x) = y$. Note that this is not a sliding block code, since the windows to consider to determine y_i increase with i. However, φ is continuous, and one can check that $\varphi \circ \sigma = \boldsymbol{a}^{-1} \circ \varphi$. Note also that since the above minima are taken over $n \geq 1$, $\varphi(0\rho_{\text{feig}}) = \varphi(1\rho_{\text{feig}})$ and in fact $\varphi(\sigma^{-k}(0\rho_{\text{feig}})) = \varphi(\sigma^{-k}(1\rho_{\text{feig}}))$ for all $k \geq 0$. $\square$

Theorem 3.4.19. *Let $(X_{\boldsymbol{q}}, \sigma)$ be a Toeplitz shift with periodic structure $\boldsymbol{q}$ and assume that $p = (p_i)_{i\geq 1}$ with $p_1 = q_1$, $p_i = q_i/q_{i-1}$ is an integer sequence. Then $(\Sigma_{\boldsymbol{p}}, \boldsymbol{a})$ is the maximal equicontinuous factor of $(X_{\boldsymbol{q}}, \sigma)$, and $(X_{\boldsymbol{q}}, \sigma)$ is a non-trivial almost one-to-one extension of $(\Sigma_{\boldsymbol{p}}, \boldsymbol{a})$.*

Proof. Let $X_{\boldsymbol{q}}$ be the orbit closure of the Toeplitz sequence x with periodic structure $\boldsymbol{q}$. Let $\mathrm{Sk}(j)$ be the j-th skeleton of x, so it is a q_j-periodic sequence in $(\mathcal{A} \cup \{*\})^\infty$. For $y \in X_{\boldsymbol{q}}$, define

$$\pi_j(y) = r \in \{0,\ldots,p_j - 1\} \quad \text{if } y_i = \mathrm{Sk}(j)_{i+r} \text{ whenever } \mathrm{Sk}(j)_{i+r} \neq *.$$

Therefore $\pi_j(\sigma^n y) = \pi_j(y) + n \pmod{q_j}$, so π_j is surjective, and $\pi^{-1}(r), r = 0, \dots, q_j - 1$, are q_j disjoint clopen sets in $X_{\boldsymbol{q}}$. For $y \in X_{\boldsymbol{q}}$, it may not be clear from the first q_j entries what $\pi_j(y)$ is. However, for every j, there is m_j such that the first m_j entries determine the value of $\pi_j(y)$. Therefore π_j is continuous.

Note that $\pi(y)_j - \pi(y)_{j-1}$ is always a multiple of q_{j-1}. Thus we can define $\pi : X_{\boldsymbol{q}} \to \tilde{\Sigma}_{\boldsymbol{q}}$ by

$$\pi(y)_j = \pi_j(y).$$

Then $\pi^{-1}(z) = \cap_j \pi_j^{-j}(z)$, as the intersection of nested non-empty closed sets, is itself non-empty. Thus π is surjective, continuous, and $\pi \circ \sigma = \tilde{\boldsymbol{a}} \circ \pi$, were $\tilde{\boldsymbol{a}}$ is defined in Remark 3.4.14. In fact, via b we can recode $(\tilde{\Sigma}_{\boldsymbol{q}}, \tilde{\boldsymbol{a}})$ to the adding machine $(\Sigma_{\boldsymbol{p}}, \boldsymbol{a})$ as Remark 3.4.14 explains. This adding machine is thus a factor of the Toeplitz shift and, as all adding machines, it is equicontinuous.

If we set $\tilde{\pi} = b \circ \pi$, we see further that $\tilde{\pi}(\sigma^n(x)) = \boldsymbol{a}^n(00000\dots) =: \langle n \rangle$ for each $n \in \mathbb{N}_0$, and that also $\tilde{\pi}^{-1}(\langle n \rangle) = \{\sigma^n(x)\}$. Therefore $(X_{\boldsymbol{p}}, \sigma)$ is an almost one-to-one extension of $(\Sigma_{\boldsymbol{p}}, \boldsymbol{a})$. However, there must be $z \in \Sigma_{\boldsymbol{p}}$ such that $\tilde{\pi}^{-1}(z) \geq 2$, because otherwise $(\Sigma_{\boldsymbol{p}}, \boldsymbol{a})$ would be conjugate to the (expansive) subshift $(X_{\boldsymbol{q}}, \sigma)$, contradicting Proposition 3.4.18. □

Theorem 3.4.20. *Every minimal equicontinuous dynamical system on the Cantor set is conjugate to an adding machine.*

Proof. See [169, Theorem 4.4]. □

Chapter 4

Methods from Ergodic Theory

In ergodic theory, we study dynamical systems $(X, \mathcal{B}, T)$ by means of probability[1] measures $\mu : \mathcal{B} \to [0, 1]$. Here $\mathcal{B}$ is the σ-algebra of measurable sets (usually the Borel algebra generated by the open sets).

Definition 4.0.1. *A measure μ on $(X, \mathcal{B}, T)$ is called* **invariant** *if $\mu(A) = \mu(T^{-1}A)$ for all $A \in \mathcal{B}$, the σ-algebra of measurable sets.*

That there exists invariant measures in the first place is guaranteed by the Krylov-Bogul'jubov Theorem:

Theorem 4.0.2 (Krylov-Bogul'jubov)**.** *If T is a continuous map on a compact space X, then there is at least one T-invariant measure.*

Proof. Let ν be any probability measure and define Césaro means:

$$\nu_n(A) = \frac{1}{n} \sum_{j=0}^{n-1} \nu(T^{-j}A).$$

These are all probability measures. The collection of probability measures on a compact metric space is known to be compact in the weak* topology, i.e., there is limit probability measure μ and a subsequence $(n_i)_{i \in \mathbb{N}}$ such that for every continuous function $\psi : X \to \mathbb{R}$:

$$\int_X \psi \, d\nu_{n_i} \to \int \psi \, d\mu \qquad \text{as } i \to \infty. \tag{4.1}$$

On a metric space, we can, for any $\varepsilon > 0$ and closed set A, find a continuous function $\psi_A : X \to [0, 1]$ such that $\psi_A(x) = 1$ if $x \in A$ and

$$\mu(A) \le \int_X \psi_A d\mu \le \mu(A) + \varepsilon \quad \text{and} \quad \mu(T^{-1}A) \le \int_X \psi_A \circ T \, d\mu \le \mu(T^{-1}A) + \varepsilon.$$

[1]Occasionally also by infinite measures, but not in this text

Here we use outer regularity of the measure μ: $\mu(A) = \inf\{\mu(U) : U \supset A \text{ is open}\}$. We take $U \supset A$ so small that $\mu(U) - \mu(A) < \varepsilon$ and make sure that $\psi_A = 0$ for all $x \notin U$. Note that it is important that A is closed, because if there exists $a \in \partial A \setminus A$, then the above property fails for $\mu = \delta_a$.

Now by the definition of μ

$$\begin{aligned}
|\mu(T^{-1}(A)) - \mu(A)| &\leq \left| \int \psi_A \circ T \, d\mu - \int \psi_A \, d\mu \right| + \varepsilon \\
&= \lim_{i\to\infty} \left| \int \psi_A \circ T \, d\nu_{n_i} - \int \psi_A \, d\nu_{n_i} \right| + \varepsilon \\
&= \lim_{i\to\infty} \frac{1}{n_i} \left| \sum_{j=0}^{n_i-1} \left(\int \psi_A \circ T^{j+1} \, d\nu - \int \psi_A \circ T^j \, d\nu \right) \right| + \varepsilon \\
&\leq \lim_{i\to\infty} \frac{1}{n_i} \left| \int \psi_A \circ T^{n_i} \, d\nu - \int \psi_A \, d\nu \right| + \varepsilon \\
&\leq \lim_{i\to\infty} \frac{2}{n_i} \|\psi_A\|_\infty + \varepsilon = \varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, $\mu(T^{-1}(A)) = \mu(A)$. Because the closed sets generate the σ-algebra of Borel sets, $\mu(T^{-1}(A)) = \mu(A)$ also for arbitrary Borel sets. □

Exercise 4.0.3. *To demonstrate the role of the compactness assumption in Theorem 4.0.2, consider the fixed point ρ of Cantor substitution χ_{Cantor} from Example* **??**, *and let $X = \{\sigma^n(\rho) : n \geq 0\}$ (so no closure taken!). Show that (X, ρ) has no invariant probability measure.*

Invariant measures allow us to study the behavior of **typical** orbits (i.e., all orbit except for a set of μ-measure zero, i.e., up to a **nullset**, and this is denoted as **a.e.** (almost everywhere) or μ-a.e. or $\pmod{\mu}$.

4.1 Ergodicity

The notion of ergodicity says that the space X doesn't fall apart in separate positive measure components.

Definition 4.1.1. *A measure μ is called* **ergodic** *if $A \in \mathcal{B}$ is invariant, i.e., $T^{-1}(A) = A \pmod{\mu}$ then $\mu(A) = 0$ or $\mu(A^c) = 0$. That is, the only T-invariant sets are nullsets or the whole space up to a nullset.*

Corollary 4.1.2. *If (X, T, μ) is ergodic, then the only T-invariant functions (i.e., $v = v \circ T$ μ-a.e.) are constant μ-a.e.*

Proof. If v is a non-constant T-invariant function that is not constant μ-a.e., then there is some $c \in \mathbb{R}$ such that $\mu(\{x \in X : v(x) \leq c\})$ and $\mu(\{x \in X : v(x) > c\})$

both have positive measure. But these sets are T-invariant, proving that μ cannot be ergodic. □

Exercise 4.1.3. *Show that ergodicity of μ is equivalent to: if $\mu = \alpha\mu' + (1-\alpha)\mu''$ for two measures and some $\alpha \in (0,1)$, then $\mu = \mu' = \mu''$. Conclude that if there is only one invariant probability measure, it has to be ergodic.*

The set of all invariant measure $\mathcal{M}(T)$ is a convex set called the **Choquet simplex**. The actual definition of a Choquet simplex is that it is a compact, metrizable, convex set in which every element can be decomposed uniquely as convex combination of extremal points.

The set of probability measures has indeed this property, since, as Exercise 4.1.3 showed, the ergodic measures $\mathcal{M}_{erg}(T)$ are precisely the extremal points of this simplex. Hence for every $\mu \in \mathcal{M}(T)$, there is some probability measure ν on $\mathcal{M}_{erg}(T)$ such that

$$\mu(A) = \int_{\mathcal{M}_{erg}(T)} \mu_{erg}(A) d\nu \qquad \text{for every } A \to B.$$

which is called the **ergodic decomposition** of the measures μ. The Choquet simplex is called **Poulsen simplex if the collection ergodic measures lie dense in the Choquet simplex, see the work of Sigmund [217] and Bowen ????** Downarowicz demonstrated that the family of Toeplitz shifts is so rich that for every simplex Σ, there is a Toeplitz shift whose Choquet simplex equals Σ, see [103]. Later, Cortez & Rivera-Letelier [81] showed that enumeration systems have this same richness.

Definition 4.1.4. *A dynamical system $(X,T)$is called entropy dense if for every invariant measure μ, there is a sequence of ergodic measures μ_n such that $\mu_n \to \mu$ in the weak* topology and the entropies $h(\mu_n) \to h(\mu)$.*

Obviously, uniquely ergodic systems are entropy dense, but there are many more systems which have this property for non-trivial reasons.

Theorem 4.1.5. *A subshift with specification is entropy dense and the collection of equidistributions is dense in the Choquet simplex (which therefore is a Poulsen simplex).*

Due to ???? Kifer & Weiss, see also Pfister & Sullivan ???? $\mathcal{B}$-free shifts are also entropy dense, as are β-shifts [**?**].

4.2 Birkhoff's Ergodic Theorem

A simple consequence of the existence of an invariant probability measure is:

Theorem 4.2.1 (Poincaré Recurrence Theorem)**.** *If $(X, \mathcal{B}, T)$ has an invariant probability measure, then for every set $A \in \mathcal{B}$ and μ-a.e. $x \in A$, there is $n \geq 1$ such that $T^n(x) \in A$. This property of μ is called* **recurrence***, hence the name of the theorem.*

Remark 4.2.2. *Every continuous map on a compact space has an invariant measure, as shown in Theorem 4.0.2. If there is only one invariant measure, it has to be ergodic as well, see Exercise 4.1.3.*

Proof of Theorem 4.2.1. Let A be an arbitrary measurable set of positive measure (if $\mu(A) = 0$, the result is trivially true). As μ is invariant, $\mu(T^{-i}(A)) = \mu(A) > 0$ for all $i \geq 0$. On the other hand, $1 = \mu(X) \geq \mu(\cup_i T^{-i}(A))$, so there must be overlap in the backward iterates of A, i.e., there are $0 \leq i < j$ such that $\mu(T^{-i}(A) \cap T^{-j}(A)) > 0$. Take the j-th iterate and find $\mu(T^{j-i}(A) \cap A) \geq \mu(T^{-i}(A) \cap T^{-j}(A)) > 0$. This means that a positive measure part of the set A returns to itself after $n := j - i$ iterates.

For the part A' of A that didn't return after n steps, assuming A' has positive measure, we repeat the argument. That is, there is n' such that $\mu(T^{n'}(A') \cap A') > 0$ and then also $\mu(T^{n'}(A') \cap A) > 0$.

Repeating this argument, we can exhaust the set A up to a set of measure zero, and this proves the theorem. □

The property demonstrated is this theorem is called **recurrence**. Theorem 4.2.1 is an instance of a very general fact observed in ergodic theory:

Space Average = Time Average (for typical points).

This is expressed in the

Theorem 4.2.3 (Birkhoff Ergodic Theorem)**.** *Let μ be a probability measure and $\psi \in L^1(\mu)$. Then the* **ergodic average**

$$\psi^*(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ T^i(x)$$

exists μ-a.e. (everywhere if ψ is continuous), and ψ^* is T-invariant, i.e., $\psi^* \circ T = \psi^*$ μ-a.e. If in addition μ is ergodic then

$$\psi^* = \int_X \psi \, d\mu \qquad \mu\text{-a.e.} \tag{4.2}$$

Remark 4.2.4. *A point $x \in X$ satisfying* (4.2) *is called* **typical** *for μ. To be precise, the set of μ-typical points also depends on ψ, but for different functions $\psi, \tilde{\psi}$, the (μ, ψ)-typical points and $(\mu, \tilde{\psi})$-typical points differ only on a nullset.*

Exercise 4.2.5. *Let $(X, T, \mathcal{B}, \mu)$ be an ergodic measure preserving system on a compact metric space. Show that T is topologically transitive on supp(mu).*

Definition 4.2.6. *Sometimes μ-typical points are called μ-generic. A point $x \in X$ is called* **quasi-generic** *w.r.t. μ if there are sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ with $b_n - a_n \to \infty$ such that*

$$\lim_{n\to\infty} \frac{1}{b_n - a_n} \sum_{j=a_n}^{b_n-1} \psi \circ T^j(x) = \int_X \psi \; d\mu \qquad \text{for every continuous function } \psi.$$

4.3 Unique ergodicity

Definition 4.3.1. *A transformation (X, T) is* **uniquely ergodic** *if it admits only one invariant probability measure.*

Since X is compact and σ is continuous, there is at least one invariant measure by Theorem 4.0.2. The question we raise in this section is whether there is a unique invariant measure. In this case, the subshift (X, σ) is called **uniquely ergodic**. If (X, σ) is both uniquely ergodic and minimal, we call it **strictly ergodic**.

Lemma 4.3.2. *If (X, σ) is uniquely ergodic, then its measure μ is ergodic.*

Proof. Suppose not, so there is an invariant set $B \in \mathcal{B}$ such that $b := \mu(B) \in (0, 1)$. Clearly also its complement B^c is invariant, and has measure $1 - b$. Construct a measure $\tilde{\mu}$ as $\tilde{\mu}(A) = \frac{1}{2}\mu(A \cap B) + \frac{1-b/2}{1-b}\mu(A \cap B^c)$ for every $A \in B$. Then $\tilde{\mu}$ is invariant as well, contradicting the uniqueness of μ. □

A very useful property of uniquely ergodic systems is that Birkhoff averages conference uniformly, rather than only a.e.

Lemma 4.3.3 (Oxtoby's Theorem)**.** *Let X be a compact space and $T : X \to X$ continuous. A transformation (X, T) is uniquely ergodic if and only if, for every continuous function, the Birkhoff averages $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i(x)$ converge* **uniformly** *to a constant function.*

The main consequence of unique ergodicity is the uniform existence of visit frequencies, i.e., for a uniquely ergodic subshift (X, σ, μ) we have for every word $a_1 \dots a_N$ and **all** $x \in X$

$$\mu([a_1 \dots a_N]) = \lim_{n\to\infty} \frac{1}{n} \#\{0 \le j < n : x_{j+1} \dots x_{j+N} = a_1 \dots a_N\}. \tag{4.3}$$

Proof. If μ and ν were two different ergodic measures, then we can find a continuous function $f : X \to \mathbb{R}$ such that $\int f d\mu \neq \int f d\nu$. Using the Ergodic Theorem for both measures (with their own typical points x and y), we see that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) = \int f d\mu \neq \int f d\nu = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(y),$$

so there is not even convergence to a constant function.

Conversely, we know by the Ergodic Theorem 4.2.3 that $\lim_n \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) = \int f d\mu$ is constant μ-a.e. But if the convergence is not uniform, then there are sequences $(x_i), (y_i) \subset X$ and $(m_i), (n_i) \subset \mathbb{N}$, such that

$$\lim_i \frac{1}{m_i}\sum_{k=0}^{m_i-1} f \circ T^k(x_i) := a \neq b =: \lim_i \frac{1}{n_i}\sum_{k=0}^{n_i-1} f \circ T^k(y_i).$$

Define functionals $\mu_i, \nu_i : C(X) \to \mathbb{R}$ as $\mu_i(g) = \liminf_i \frac{1}{m_i}\sum_{k=0}^{m_i-1} g \circ T^k(x)$ and $\nu_i(g) = \liminf_i \frac{1}{n_i}\sum_{k=0}^{n_i-1} g \circ T^k(x)$. Both sequences have weak accumulation points μ and ν which are easily shown to be T-invariant measures, see the proof of Theorem 4.0.2.

More precisely, since $(C(X), \| \ \|_\infty)$ is a separable Banach space, we can find a countable dense subset $(g_j)_{j\in\mathbb{N}}$ and (by a diagonal argument) we can take subsequences of $(m_i)_{i\in\mathbb{N}}$ and $(n_i)_{i\in\mathbb{N}}$ along which $\mu_i(g_j)$ and $\nu_i(g_j)$ converge for all $j \in \mathbb{N}$.

But μ and ν are not the same. because if we take subsequence a subsequence $(j_r)_{r\in\mathbb{N}}$ such that $g_{j_r} \to f$, then $\lim_r \mu(g_{j_r}) = \mu(f) = a \neq b = \nu(f) = \lim_r \nu(g_{j_r})$. Hence (X, T) cannot be uniquely ergodic. □

Example 4.3.4. *No non-periodic SFT is uniquely ergodic. Indeed, there are infinitely many periodic sequences* $x = (x_1 \dots x_n)^\infty$*, and the equidistribution* $\delta_x(B) = \frac{1}{n}\#(B \cap orb_\sigma(x))$ *is an invariant measure. The same holds for sofic shifts.*

On the other hand, Sturmian shifts (X, σ) *are strictly ergodic, and their unique measure is obtained by lifting Lebesgue measure from the circle, using the itinerary map* $\boldsymbol{i} : \mathbb{S} \to X$*, that is:* $\mu(B) = Leb(\boldsymbol{i}^{-1}(B))$.

Chapter 5

Automata and Coding

Allouche and Shallit: $\lfloor \alpha n + \beta \rfloor$ is an automatic seuqence if and only if $\alpha \in \mathbb{Q}$, and then it is periodic. Definition... [5] ????

5.1 Automata

In this section we discuss some variations on the Turing machine, and ask the question what languages they can recognize or generate. The terminology is not entirely consistent in the literature, so some of the below notions may be called differently depending on which book you read.

5.1.1 Finite automata

A finite automaton (FA) is a simplified type of Turing machine that can only read a tape from left to right, and not write on it. The components are

$$M = \{Q, \mathcal{A}, q_0, F, f\}$$

where

$$\begin{array}{rcl} Q & = & \text{collection of } \textbf{states} \text{ the machine can be in.} \\ \mathcal{A} & = & \text{the alphabet in which the tape is written.} \\ q_0 & = & \text{the initial state in } Q. \\ F & = & \text{collection of final states in } Q;\ \text{the FA halts when it reaches one.} \\ f & = & \text{is the rule how to go from one state to the next when reading} \\ & & \text{a symbol } a \in \mathcal{A} \text{ on the tape. Formally it is a function } Q \times \mathcal{A} \to Q. \end{array}$$

A language is **regular** if it can be recognized by a finite automaton.

Example 5.1.1. *The even shift (Example 0.1.13) is recognized by the following finite automaton with $Q = \{q_0, q_1, q_2, q_3\}$ with initial state q_0 and final states q_2 (rejection)* *and q_3 (acceptance). The tape is written in the alphabet $\mathcal{A} = \{0, 1, b\}$ where b stands for a* blank *at the end of the input word. The arrow $q_i \to q_j$ labeled $a \in \mathcal{A}$ represents $f(q_i, a) = q_j$.*

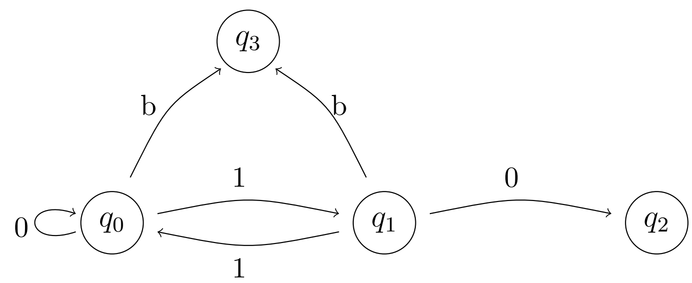


Figure 5.1: Transition graph for a finite automaton recognizing the even shift.

This example demonstrates how to assign a edged-labeled transition graph to a finite automaton, and it is clear from this that the regular languages are precisely the sofic languages.

It is frequently easier, for proofs or constructing compact examples, to allow finite automata with multiple outgoing arrows with the same label. So, if we are in state q, read symbol a on the input tape, and there is more than one outgoing arrow with label a, then we need to make choice. For computers, making choices is somewhat problematic - we don't want to go into the theoretical subtleties of random number generators - but if you take the viewpoint of probability theory, you can simply assign equal probability to every valid choice, and independent of the choices you may have to make elsewhere in the process. The underlying stochastic process is then a discrete Markov process.

Automata of this type are called **non-deterministic finite automata** (NFA), as opposed to **deterministic finite automata** (DFA), where never a choice needs to be made. A word is accepted by an NFA if there is a positive probability that choices are made that parse the word until the end without halting or reaching a rejecting state.

We mention without proof (see [148, page 22] or [5, Chapter 4]):

Theorem 5.1.2. *Let $\mathcal{L}$ be a language that is accepted by a non-deterministic finite automaton. Then there is a deterministic finite automaton that accepts $\mathcal{L}$ as well.*

Corollary 5.1.3. *Let $w^R = w_n \dots w_1$ stand for the* **reverse** *of a word $w = w_1 \dots w_n$. If a language $\mathcal{L}$ is recognized by a finite automaton, then so is its reverse $\mathcal{L}^R = \{w^R : w \in \mathcal{L}\}$.*

Proof. Let $(\mathcal{G}, \mathcal{A})$ the edge-labeled directed graph representing the FA for $\mathcal{L}$. Reverse all the arrows. Clearly the reverse graph $(\mathcal{G}^R, \mathcal{A})$ in which the directions of all arrows are reversed and the final states become initial states and vice versa, recognizes $\mathcal{L}^R$. However, even if in $\mathcal{G}$, every outgoing arrow has a different label (so the FA is deterministic), this is no longer true for $(\mathcal{G}^R, \mathcal{A})$. But by Theorem 5.1.2 there is also an DFA that recognizes $\mathcal{L}^R$. □

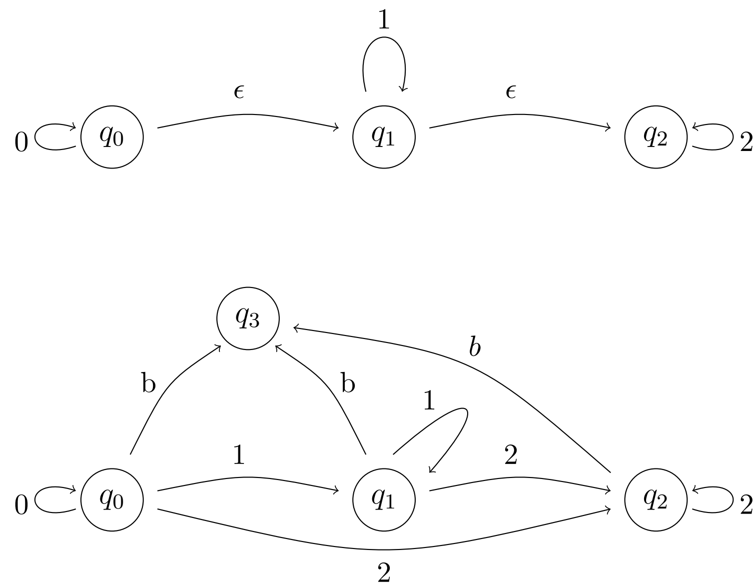


Figure 5.2: Finite automata recognizing $\mathcal{L} = \{0^k 1^l 2^m : k, l, m \geq 0\}$.

Sometimes it is easier, again for proofs or constructing compact examples, to allow finite automata to have transitions in the graph without reading the symbol on the input take (and moving to the next symbol). Such transitions are called **ϵ-moves**. Automata with ϵ-moves are almost always non-deterministic, because if a state q has an outgoing arrow with label a and an outgoing arrow with label ϵ, and the input tape reads a, then still there is the choice to follow that a-arrow or the ϵ-arrow.

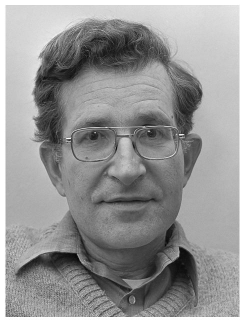
Example 5.1.4. *The follow automata accept the language $\mathcal{L} = \{0^k 1^l 2^m : k, l, m \geq 0\}$, see Figure 5.2. The first is with ϵ-moves, and it stops when the end of the input is reached (regardless which state it is in). That is , if the FA doesn't halt before the end of the word, then the word is accepted. The second is deterministic, but uses a blank b at the end of the input. In either case q_0 is the initial state.*

Again without proof (see [148, page 22]):

Theorem 5.1.5. *Let $\mathcal{L}$ be a language that is accepted by a finite automaton with ϵ-moves. Then there is a non-deterministic finite automaton without ϵ-moves that accepts $\mathcal{L}$ as well.*

5.2 The Chomsky hierarchy

A different approach to complexity of languages is due to Noam Chomsky's (1928–) study to describe grammar of natural languages, based on **production rules**.



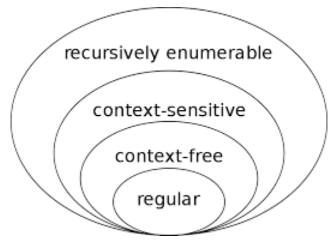


Figure 5.3: Noam Chomsky in 1977 and his hierarchy.

For example, to build sentences in English, you could (repeatedly) use the following rules, until there are no **variables** (i.e., the things within $\langle\ \rangle$) left:

$$
\begin{aligned}
\langle\text{sentence}\rangle &\to \langle\text{articled noun phrase}\rangle\langle\text{transitive verb}\rangle\langle\text{articled noun phrase}\rangle\\
\langle\text{articled noun phrase}\rangle &\to \langle\text{article}\rangle\langle\text{noun phrase}\rangle\\
\langle\text{noun phrase}\rangle &\to \langle\text{adjective}\rangle\langle\text{noun phrase}\rangle\\
\langle\text{noun phrase}\rangle &\to \langle\text{noun}\rangle\\
\langle\text{noun}\rangle &\to \text{mouse, cat, book, decency}\\
\langle\text{article}\rangle &\to \text{the, a}\\
\langle\text{adjective}\rangle &\to \text{big, small, high, low, red, green, orange, yellow}\\
\langle\text{transitive verb}\rangle &\to \text{chases, eats, hits, reads}
\end{aligned}
$$

This produces sentence such as

$$
\begin{array}{l}
\text{a small yellow mouse chases a big green cat}\\
\text{a high low red decency eats a orange book}
\end{array}
\tag{5.1}
$$

Here the first sentence is fine; the second is nonsense. But apart from the fact that "a orange" should be "an orange" it is grammatically correct.

In arithmetic, we can make the following example:

$$
\begin{aligned}
\langle\text{expression}\rangle &\to \langle\text{expression}\rangle * \langle\text{expression}\rangle\\
\langle\text{expression}\rangle &\to \langle\text{expression}\rangle + \langle\text{expression}\rangle\\
\langle\text{expression}\rangle &\to (\langle\text{expression}\rangle)\\
\langle\text{expression}\rangle &\to 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
\end{aligned}
$$

This can generate all kind of arithmetic expressions by repeatedly adding and multiplying the numbers $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$, that a pocket calculator should be able the compute. For instance

$$9 + 5 * 3 + 7, \qquad (9 + 5) * 3 + 7, \qquad 9 + 5 * (3 + 7), \qquad (9 + 5) * (3 + 7),$$

all with different outcomes.

Formally, this **grammar** has the following components

$$G = (V, T, P, S),$$

where

V	$=$	collection of **variables** to which production rules can be applied.
T	$=$	collection of **terminals** which remain unchanged.
P	$=$	collection of **production rules** to replace variable with strings in $V \cup T$.
S	$=$	a special variable, called the **starting symbol**.

The language $\mathcal{L}(G)$ of a grammar G is the collection of all words in T^* that, starting from S, can be generated by repeated application of the production rules until no variables are left.

The **Chomsky hierarchy** is a classification of languages according to how complicated the production rules are. In order of increasing complexity, they are

$$\begin{array}{rcl} \text{regular languages (RL)} & \subset & \text{context-free languages (CFL)} \\ & \subset & \text{context sensitive languages (CSL)} \\ & \subset & \text{recursively enumerable languages (ER)} \end{array}$$

although there are also **unrestricted grammars**, which is a wider class still.

5.2.1 Regular grammars

The **regular grammars** can be brought in a form where the production rules are one of the following types:

left-linear	or	**right-linear**
$A \to Bw$		$A \to wB$
$A \to w$		$A \to w$

where $A, B \in V$ and $w \in T^*$ (possibly w is empty).

Example 5.2.1. *The even shift (Example 0.1.13) is recognized by the following left and right-linear regular grammars on* $T = \{0, 1\}$.

left-linear	**right-linear**
$S \to S0$	$S \to 0S$
$S \to S11$	$S \to 11S$
$S \to \epsilon$	$S \to \epsilon$

Note that the language $\mathcal{L}$ is reversible, i.e., $\mathcal{L}^R = \mathcal{L}$, and this property makes it so simple to convert the left-linear productions into the right-linear productions.

Theorem 5.2.2. *Every regular grammar (left-linear or right-linear) produces a language that can be recognized by a finite automaton and vice versa.*

Hence a regular grammar produced the language of a sofic subshifts.

Proof. First assume that $G = \{V, T, P, S\}$ is a right-regular grammar. Construct a finite automaton with ϵ-moves $\{Q, \mathcal{A}, q_0, F, f\}$ where Q consists of all q such that $q = S$ or q is a (not necessarily proper) suffix of the right hand side of a production rule. Define

$$f(q, a) = \begin{cases} q' & \text{if } q \in V, a = \epsilon, q \to q' \text{ is a production;} \\ q' & \text{if } q = aq' \in T^* \cup T^*V, a \in T, q \to aq' \text{ is a production.} \end{cases}$$

Conversely, if a finite automaton is given by $\{Q, \mathcal{A}, q_0, F, f\}$, then make the right-regular grammar $G = \{V, T, P, qS\}$ where the productions are $p \to aq$ whenever $f(p, a) = q$, and $p \to a$ if $f(p, a) = q$ and q is a final state.

A left-linear grammar is found by first constructing a finite automaton that accepts exactly the reverse $w^R = w_n \dots w_1$ of every $w = w_1 \dots w_n \in \mathcal{L}$ (see Corollary 5.1.3), and then taking the right-linear grammar for this reverse language $\mathcal{L}^R$. Then rewrite every production rule $A \to wB$ to $A \to Bw$ to obtain a left-linear grammar that accepts exactly the original $\mathcal{L}$. □

5.2.2 Context-free grammars

The second sentence in (5.1) makes no sense, because (for example) high does not go together with low, and decencies don't eat. In other words, the grammar rules produces word combinations without looking at the meaning of the particular words, and which words can go together. This is the explanation behind the term **context-free**. Formally, a context-free grammar (V, T, P, S) is one in which the the set P of productions is finite, and each of them has the form $A \to \alpha$, where $\alpha \in (V \cup T)^*$ is a finite string of variables and terminals.

Example 5.2.3. *Consider the language $\mathcal{L} := \{01^n2^n : n \geq 1\}$. That is, every maximal block of* 1*s is succeeded by an equally long word of* 2*s.*

This is a context-free language, generated by the productions

$$\begin{aligned} &S \to 01A2 \\ &A \to 1A2 \\ &A \to \epsilon \qquad \textit{(the empty word)} \end{aligned}$$

Assume by contradiction that $\mathcal{L}$ is sofic. Then there is a finite edge-labeled transition graph $\mathcal{G}$ which generates $\mathcal{L}$. Since the are only finitely many, say r, vertices, *every word 1^n for $n \geq r$ must contain a subword 1^m corresponding to a loop in $\mathcal{G}$. But then we can also take this loop k times. In particular, for each word 01^n2^n, also*

$$01^{n+(k-1)m}2^n = 01^a \underbrace{1^m 1^m 1^m \dots 1^m}_{\text{the } m\text{-loop } k \text{ times}} 1^b 2^n$$

is generated in $\mathcal{G}$. But $01^{n+(k-1)m}2^n \notin \mathcal{L}$, so we have a contradiction.

This example shows that context-free grammars are a strictly wider class than the regular grammars, and it also illustrates the working of a general class of lemmas, called **Pumping Lemmas** that are frequently used in this field as a tool to distinguish grammars. The simplest (which we exploited in Example 5.2.3):

Lemma 5.2.4 (Pumping Lemma for Regular Languages)**.** *Let $\mathcal{L}$ be a regular language. Then there is N such that for every $w \in \mathcal{L}$ of length $|w| \geq N$, we can decompose $w = tuv$ such that $|uv| \leq N$, $v \neq \epsilon$ and $tu^kv \in \mathcal{L}$ for all $k \geq 1$.*

Proof. As in Example 5.2.3. Note that $N \leq \#\{\text{vertices in } \mathcal{G}\}$. □

Corollary 5.2.5. *[[121, Corollary 6.1.11]] The language of a Sturmian sequence x with irrational rotation number is not regular.*

Proof. If the language $\mathcal{L}(x)$ was regular, then by the Pumping Lemma 5.2.4, there are words $tu^kv \in \mathcal{L}(x)$ for some $u \neq \epsilon$ and any $k \geq 1$. But $\lim_{k\to\infty} |tu^kv|_1 \in \mathbb{Q}$ and this contradicts that to rotation number of $x \notin \mathbb{Q}$. □

Exercise 5.2.6. *Let $\mathcal{L} = \{01^{n^2} : n \geq 1\}$. Show that $\mathcal{L}$ is not a regular language. Is it context-free?*

Exercise 5.2.7. *Using the Pumping Lemma 5.2.4 to show that there are β-shifts X_β that are not regular.*

Lemma 5.2.8 (Pumping Lemma for Context-free Languages)**.** *Let $\mathcal{L}$ be a context-free language. Then there is N such that for every $w \in \mathcal{L}$ of length $|w| \geq N$, we can decompose $w = rstuv$ such that $1 \leq |su| \leq |stu| \leq N$, and $rs^ktu^kv \in \mathcal{L}$ for all $k \geq 1$.*

Proof. See [148, Chapter 6]. □

Corollary 5.2.9. *The language $\mathcal{L}(x)$ of a Sturmian word x is not context-free.*

Proof. Take $N \in \mathbb{N}$ and for a given N-word w of x, let $w = rstuv$ be the composition as in Lemma 5.2.8. If $\mathcal{L}(x)$ were context-free, then $rs^ktu^kv \in \mathcal{L}(x)$ as well. But then the limit frequency of 1s is

$$\lim_{k\to\infty} \frac{|rs^ktu^kv|_1}{|rs^ktu^kv|} = \frac{|su|_1}{|su|} \in \mathbb{Q},$$

contradicting that Sturmian sequences have irrational frequencies. □

From the shape of its production rules, it is clear that the language of Example 5.2.3 is context-free. No finite automaton can keep track of the precise number of 1s before starting on the 2s, but there is a simple memory device that can. Imagine that for every 1 you see, you put a card on a stack, until you reach the first 2. At every 2 you read you remove a card again. If at the end of the word no cards are left on the stack, the word is accepted.

This device is simple in construction: you can only add or remove at the top of the stack; what is further down you can not read until you first remove all the cards above it. On the other hand, the stack has no prescribed upper height, so requires infinite memory.

Formally, the **(push-down) stack** has its (finite) stack alphabet $\mathcal{B}$ (think of cards of different color) which is different from $\mathcal{A}$ and a starting stack symbol $Z_0 \in \mathcal{B}$ (the color of the initial card on the stack at the start of the automaton. The moves $f : Q \times \mathcal{A} \times \mathcal{B} \to Q \times \mathcal{B}^*$ now also involve adding cards to the stack (with colors in $\mathcal{B}$) or removing them. The resulting automaton with stack is called a **push-down automaton**.

Theorem 5.2.10. *A language is (not more complicated than) context-free if and only if it is recognized by a push-down automaton.*

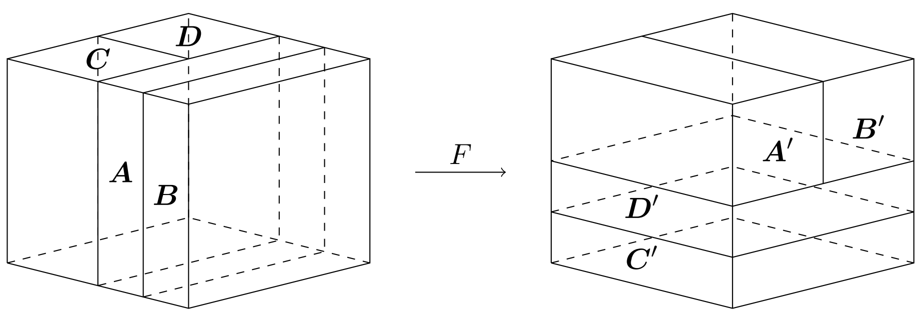


Figure 5.4: A 3-dimensional "heterogeneous" baker transformation.

Exercise 5.2.11. *A piecewise affine and "heterogeneous"*[1] *hyperbolic map* $F : [0,1]^3 \to [0,1]^3$ *(see Figure 5.4) is defined as*

$$F(x,y,z) = \begin{cases} (4x-2, y/2, (1+z)/2) & \text{if } (x,y,z) \in A, \\ (4x-2, (1+y)/2, (1+z)/2) & \text{if } (x,y,z) \in B, \\ (2x, 2y, z/4) & \text{if } (x,y,z) \in C, \\ (2x, 2y, (1+z)/4) & \text{if } (x,y,z) \in D, \end{cases}$$

[1] i.e., stable manifolds don't have the same dimension at every point

and primes indicate the F-images of each of these four boxes, see [3].
(a) Why is the partition into these four boxes not a Markov partition? (b) Why is the symbolic shift (X, σ) associated with this partition (i.e., a subshift of $\{A, B, C, D\}^{\mathbb{Z}}$ not a SFT . **Hint:** *AC can be followed by D but can AAC be followed by D?*
(c) Show instead that (X, σ) is a context-free subshift, as well as synchronizing.

5.2.3 Context-sensitive grammars

A context-sensitive grammar (V, T, P, S) is one in which the set P of productions is finite, and each of them has the form $\alpha \to \beta$, where $\alpha, \beta \in (V \cup T)^*$ and $|\beta| \geq |\alpha|$. The terminals themselves cannot change, but they can swap position with a variable. For example $aA \to Aa$ and $aA \to Ba$ are valid production rules in a context-sensitive grammar.

Remark 5.2.12. *The word* **context-sensitive** *comes from a particular normal form of the productions, in which each of them has the form $\alpha_1 A \alpha_2 \to \alpha_1 B \alpha_2$, where $B \in (V \cup T)^*$ is a* **non-empty** *finite string of variables and terminals, and $\alpha_1, \alpha_2 \in (V \cup T)^*$ are contexts in which the production rule can be applied. Only when A is preceded by α_1 and succeeded by α_2, the production rule can be applied, leaving the* **context** *α_1, α_2 unchanged.*

Example 5.2.13. *Consider the language $\mathcal{L} = \{1^n 2^n 3^n : n \geq 1\}$. Pumping Lemma 5.2.8 can be applied to show that $\mathcal{L}$ is not context-free. However $\mathcal{L}$ is context-sensitive. For example, we can use the productions*

$$\begin{array}{rcl} S & \to & 123 \\ S & \to & 11A23 \\ A2 & \to & 2A \\ 2A3 & \to & 2233 \\ 2A3 & \to & 22B33 \\ 2B & \to & B2 \\ 1B2 & \to & 11A2 \end{array}$$

In practice, A is a marker moving right, doubling 23 when it hits the first 3. The procedure can stop here, or produce marker B that moves to the left, doubling 1 when it hits the first 1.

Example 5.2.14. *The following set of productions produces the language $\mathcal{L} = \{1^{2^n} : n \geq 0\}$, that is: strings of 1s of length equal to a power of 2.*

$$\begin{array}{rclcrcl} S & \to & AC1B & \qquad & 1D & \to & D1 \\ C1 & \to & 11C & & AD & \to & AC \\ CB & \to & DB & & 1E & \to & E1 \\ CB & \to & E & & AE & \to & \epsilon \end{array}$$

Here A and B are begin-marker and end-maker. C is a moving marker, doubling the number of 1*s when it moves to the right. When it reaches the end-marker B, then*

- *it changes to a moving marker D, which just moves to the left until it hits begin-marker A, and changes itself in C again. In this loop, the number of* 1*s is doubled again.*
- *or, it merges with the end-marker B to a new marker E. This marker E moves left until it hits begin-marker A. It then merges with A into the empty word: end of algorithm.*

This language is context-sensitive, although the production rules $CB \to E$ and $AE \to \epsilon$ strictly speaking not of the required form. The trick around it is to glue a terminal 1 *to (pairs of) variables in a clever way, and then call these glued strings the new variables of grammar, see [148, page 224].*

We mentioned Turing machines in the introduction. In effect, a Turing machine is a finite automaton with a memory device in the form of an input tape that can be read, erased and written on, in little steps of one symbol at the time, but otherwise without restrictions on the tape. If the finite automaton part is non-deterministic, then we call it a **non-deterministic Turing machine**. If we put a restriction on the tape that it cannot be used beyond where the initial input is written, then we have a **linearly bounded non-deterministic Turing machine** or **linearly bounded automaton** (LBA). To avoid going beyond the initial input, we assume that the input is preceded by a begin-marker, than cannot be erased, and to the left of which the reading/writing device cannot go. Similarly, the input is succeeded by an end-marker, than cannot be erased, and to the right of which the reading/writing device cannot go.

Theorem 5.2.15. *A language is (not more complicated than) context-sensitive if and only if it is recognized by a linearly bounded non-deterministic Turing machine.*

5.2.4 Recursively enumerable grammars

A grammar is called recursively enumerable if there is no restriction anymore on the type of production rules. For this largest class in the Chomsky hierarchy, there is no restriction on the Turing machine anymore either.

Theorem 5.2.16. *A language is (not more complicated than) recursively enumerable if and only if it is recognized by a Turing machine.*

In summary, we have the table:

Type	Automaton	Productions	Example
regular (sofic shift)	finite automaton	$A \to w, A \to wB$ (right-linear) $A \to w, A \to Bw$ (left-linear)	$\{a^m b^n : m, n \geq 1\}$
context-free	push-down automaton	$A \to \gamma \in (V \cup T)^*$	$\{a^n b^n : n \geq 1\}$
context- sensitive	linearly bounded non-deterministic Turing machine	$\alpha \to \beta,\ \alpha, \beta \in (V \cup T)^*$, $\lvert\beta\rvert \geq \lvert\alpha\rvert$ (ór $\alpha A \beta \to \alpha\gamma\beta$ $\varnothing \neq \gamma \in (V \cup T)^*$)	$\{a^{2^n} : n \geq 0\}$
recursively enumerable	Turing machine	$\alpha \to \beta$ (no restrictions)	

Bibliography

[1] M. Adamska, S. Bezuglyi, O. Karpel, J. Kwiatkowski, *Subdiagrams and invariant measures on Bratteli diagrams,* Ergod. Th. Dynam. Sys. **37** (2017), 2417–2452.

[2] R. L. Adler, A. G. Konheim, M. H. McAndrew, *Topological entropy,* Trans. Amer. Math. Soc. **114** (1965), 309-319.

[3] J. Alexander, J. Yorke, *Fat baker transformations,* Ergod. Th. & Dynam. Sys. **4** (1984), 1–23.

[4] J.-P. Allouche, J. Shallit, *The ubiquitous Prouhet-Thue-Morse sequence,* Sequences and their applications, (Singapore, 1998), 1–16, Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999.

[5] J.-P. Allouche, J. Shallit, *Automatic seuqences: theory, applications, generalizations*, Cambridge Univ. Press, 2nd. edition, Cambridge (2003).

[6] L. Alsedá, J. Llibre, M. Misiurewicz, *Combinatorial dynamics and the entropy in dimension one*, Adv. Ser. in Nonlinear Dynamics **5**, 2nd Edition, World Scientific, Singapore, 2000.

[7] L. Alvin, *The strange star product,* J. Difference Eq. and Appl. **18** (2012), 657–674.

[8] L. Alvin, *Toeplitz kneading sequences and adding machines,* Discrete Contin. Dyn. Syst. 3**3** (2013), 32773287.

[9] T. Apostol, *A proof that Euler missed: evaluating $\zeta(2)$ the easy way,* Math. Intel. **5** (1983), 59–60.

[10] C. Apparicio, *Reconnaissabilité des substitutions de longueur constante*, Stage de Maitrise de l'ENS Lyon, 1999.

[11] P. Arnoux, *????*

[12] P. Arnoux, A. Fischer, *The scenery flow for geometric structures on the torus: the linear setting,* Chinese Ann. Math. **22b** (2000), 427–470.

[13] J. Auslander, *Minimal flows and their extensions,* North Holland Mathematical studies **153** North Holland Press, Amsterdam - New York - Oxford - Tokyo, 1988.

[14] J. Auslander, J. Yorke, *Interval maps, factors of maps, and chaos,* Tôhoku Math. J. **32** (1980) 177–188.

[15] J. Banks, J. Brooks, G. Cairns, G. Davis, P. Stacey, *On Devaney's definition of chaos,* Amer. Math. Monthly **99** (1992) 332–334.

[16] J. Banks, T. Nguyen, P. Oprocha, B. Stanley, B. Trotta, *Dynamics of spacing shifts,* Discrete Contin. Dyn. Syst. **33** (2013), 4207–4232.

[17] G. Barat, T. Downarowicz, A. Iwanik, P. Liardet, *Propriétés topologiques et combinatoires des échelles de numération*, Colloq. Math. **84/85** (2000), 285–306.

[18] G. Barat, T. Downarowicz, P. Liardet, *Dynamiques asociées à une échelle de numération,* Acta Arithmetica **103** (2002), 41–78.

[19] M. Barge, *The Pisot conjecture for β-substitutions,* Ergod. Th. Dynam. Sys. **38** (2018), 444–472.

[20] R. Bass, *Real analysis for graduate students,* CreateSpace Independent Publishing Platform (2016) Version 3.1 online on: http://bass.math.uconn.edu/3rd.pdf

[21] V. Berthé *Combinatorics, automata and number theory,* Ed. V. Berthé, M. Rigo.

[22] J. Berstel, A. de Luca, *Sturmian words, Lyndon words and trees,* Theoretical Computer Science, —bf 178 (1997), 171–203,

[23] A. Besicovitch, *On the density of certain sequences of integers,* Math. Ann. **15** (1935), 336–341.

[24] S. Bezuglyi, J. Kwiatkowski, K. Medynets, *Aperiodic substitution systems and their Bratteli diagrams,* Ergod. Th. Dynam. Sys. **29** (2009), 37–72.

[25] S. Bezuglyi, J. Kwiatkowski, K. Medynets, B. Solomyak, *Invariant measures on stationary Bratteli diagrams,* Ergodic Theory Dynam. Systems **30** (2010), 973–1007.

[26] S. Bezuglyi, J. Kwiatkowski, K. Medynets, B. Solomyak, *Finite rank Bratteli diagrams: structure of invariant measures,* Trans. Amer. Math. Soc. **365** (2013), 2637–2679.

[27] S. Bezuglyi, J. Kwiatkowski, R. Yassawi, *Perfect orderings on finite rank Bratteli diagrams,* Canad. J. Math. **66** (2014), 57–101.

[28] S. Bezuglyi, R. Yassawi, *Orders that yield homeomorphisms on Bratteli diagrams,* Dyn. Syst. **32** (2017), 249–282.

[29] F. Blanchard, E. Glasner, S. Kolyada, A. Maass, *On Li-Yorke pairs,* J. Reine Angew. Math. **547** (2002), 51–68.

[30] *Topics in Symbolic Dynamics and Applications,* London Mathematical Society Lecture Note Series, Editors: F. Blanchard, A. Maass, A. Nogueira, Cambridge Univ. Press 2000, ISBN 9780521796606.

[31] S. Bezuglyi, O. Karpel, J. Kwiatkowski, *Exact number of ergodic measures for Bratteli diagrams,* Preprint 2018, arXiv:1709.00055

[32] S. Bezuglyi, J. Kwiatkowski, K. Medynets, B. Solomyak, *Finite rank Bratteli diagrams: structure of invariant measures,* Trans. Amer. Math. Soc. **365** (2013), no. 5, 2637–2679.

[33] P. Billingsley, *Probability and measure,* John Wiley and Sons, New Delhi (1995).

[34] G. Birkhoff, *Extensions of Jentzsch's Theorem,* Trans. Amer. Math. Soc. (1957), 219–227.

[35] L. Block, J. Guckenheimer, M. Misiurewicz, L.-S. Young, *Periodic points and topological entropy of one-dimensional maps*. In: Z. Nitecki, C. Robinson (eds) *Global Theory of Dynamical Systems* Lecture Notes in Math. **819** Springer Verlag, Berlin (1980) 18–34.

[36] L. Block, J. Keesling, M. Misiurewicz, *Strange adding machines,* Ergod. Th. Dynam. Sys. **26** (2006), 673.

[37] A. Blokh, *Sensitive mappings of an interval,* Uspekhi Mat. Nauk **37** (1982), 189–190.

[38] A. Blokh, *Decomposition of dynamical systems on an interval*, Russ. Math. Surv. **38** (1983), 133–134.

[39] A. Blokh, *The "spectral" decomposition for one-dimensional maps,* Dynamics reported, 1–59, Dynam. Report. Expositions Dynam. Systems (N.S.), **4** Springer, Berlin, 1995.

[40] J. Bobok, H. Bruin, *Constant slope maps and the Vere-Jones classification,* Entropy **18** (2016), 234–250.

[41] P. Borwein, *"The Prouhet-Tarry-Escott problem", Computational Excursions in Analysis and Number Theory*, CMS Books in Mathematics, Springer-Verlag (2002).

[42] R. Bose, D. Ray-Chaudhuri, *On a class of error correcting binary group codes*, Information and Control, **3** (1960), 68–79.

[43] M. Boshernitzan, *A unique ergodicity of minimal symbolic flows with linear block growth,* J. dAnalyse Math. **44** (1985), 77–96.

[44] M. Boshernitzan, *A condition for unique ergodicity of minimal symbolic flow,* Erg. Theory and Dynam. Sys. **12** (1992), 425–428.

[45] R. Bowen, *Markov partitions for Axiom A diffeomorphisms,* Amer. J. of Math. **92**, (1970), 725–747.

[46] R. Bowen, *Markov partitions and minimal sets for Axiom A diffeomorphisms,* Amer. J. Math. 92 (1970), 907–918.

[47] R. Bowen, *Periodic points and invariant measures for Axiom A diffeomorphisms,* Trans. Amer. Mah. Soc. **154** (1971), 377–397.

[48] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms,* Lect. Math. **470** Spring 1974 and second revised edition 2008.

[49] R. Bowen, *Some systems with unique equilibrium states,* Math. Systems Theory, **8** (1974/75), 193–202.

[50] R. Bowen, *Markov partitions are not smooth,* Proc. Amer. Math. Soc. **71** (1978), 130–132.

[51] P. Boyland, A. de Carvalho, T. Hall, *Natural extensions of unimodal maps: virtual sphere homeomorphisms and prime ends of basin boundaries,* Preprint 2017 arXiv:1704.06624

[52] M. Boyle, D. Mike, D. Rudolph, *The automorphism group of a shift of finite type,* Trans. Amer. Math. Soc. **306** (1988), 71–114.

[53] O. Bratteli, *Inductive limits of finite dimensional C^*-algebras,* Trans. Amer. Math. Soc. **171** (1972) 195–234.

[54] X. Bressaud, F. Durand, A. Maass, *Necessary and sufficient conditions to be an eigenvalue for linearly recurrent dynamical Cantor systems*, Journal of the London Mathematical Society **72** (2005), 799–816.

[55] X. Bressaud, F. Durand, A. Maass, *Eigenvalues of finite rank Bratteli-Vershik dynamical systems,* Ergod. Th. & Dynam. Sys. **30** (2010), 639–664.

[56] K. Briggs, *A precise calculation of the Feigenbaum constants,* Math. Comp. **57** (1991) 435–439.

[57] K. Brucks, H. Bruin, *Topics from one-dimensional dynamics,* London Mathematical Society Student Texts **62**, Cambridge University Press, 2004.

[58] H. Bruin, *Invariant measures of interval maps,* PhD-Thesis, University of Delft, 1994.

[59] H. Bruin, *Combinatorics of the kneading map,* Internat. J. Bifur. Chaos Appl. Sci. Engrg. **5** (1995), 1339–1349.

[60] H. Bruin, *Minimal Cantor systems and unimodal maps,* J. Difference Eq. and Appl. **9** (2003) 305–318.

[61] H. Bruin, *Renormalisation in a class of interval translation maps of d branches,* Dynamical Systems, an international journal **22** (2007), 11–24.

[62] H Bruin, J. Hawkins, *Exactness and automorphic factors of unimodal interval maps,* Ergod. Th. & Dyn. Sys. **21** (2001) 1009–1034.

[63] H Bruin, J. Hawkins, *Rigidity of smooth one-sided Bernoulli endomorphisms,* New York Journ. Maths. **15** (2009), 451–483.

[64] H. Bruin, G. Keller, M. St. Pierre, *Adding machines and wild attractors,* Ergod. Th. and Dyn. Sys. **17** (1997) 1267–1287.

[65] H. Bruin, M. Todd, *Transience and thermodynamic formalism for infinitely branched interval maps,* J. London Math. Soc. **86** 2012, 171–194.

[66] H. Bruin, S. Troubetzkoy, *The Gauss map on a class of interval translation mappings,* Isr. J. Math. **137** (2003), 125–148.

[67] H. Bruin, O. Volkova, *The complexity of Fibonacci-like kneading sequences*, Theo. Comp. Science. **377** (2005) 379–389.

[68] J. Buzzi, *Subshifts of quasi-finite type,* Invent. Math. **159** (2005), 369–406.

[69] J. Buzzi, P. Hubert, *Piecewise monotone maps without periodic points: Rigidity, measures and complexity,* Ergod. Th. & Dynam. Sys. **24** (2004), 383–405.

[70] F. Cai, *A characterization of weak-mixing for minimal systems,* Top. Appl. **???** (2019) ???.

[71] M. Campanino, H. Epstein, *On the existence of Feigenbaum's fixed point,* Commun. Math. Phys. **79** (1981), 261–302.

[72] J. Cassaigne, J. Karhumäki, *Toeplitz words, generalized periodicity and periodically iterated morphisms,* European J. Combin. **18** (1997), 497–510.

[73] S. Chowla, *On abundant numbers,* J. Indian Math. Soc. New Ser. **1** (1934), 41–44.

[74] V. Climenhaga, D. Thompson, *Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors,* Israel J. Math,. **192**, (2012), 785–817.

[75] V. Climenhaga, D. Thompson, *Intrinsic ergodicity via obstruction entropies,* Ergod. Th. Dynam. Sys. **34** (2014), 1816–1831.

[76] H. Cohn, *A short proof of the simple continued fraction expansion of* e The American Mathematical Monthly, **113** (2006), 57–62.

[77] E. Coven, I. Kan, J. Yorke, *Pseudo-orbit shadowing in the familiy of tent-maps*, Trans. Amer. Math. Soc. **308** (1988), 217–241.

[78] P. Collet, J.-P. Eckmann, *Iterated maps of the interval as dynamical systems,* Birkhäuser, Boston (1980).

[79] J. Coquet, *Représentation lacunaires des entières naturelles I, II,* Arch. Math. Basel **38** (1982), 184–188 and **41** (1983), 238–242.

[80] M. I. Cortez, F. Durand, B. Host, A. Maass, *Continuous and measurable eigenfunctions of linearly recurrent dynamical Cantor systems,* Journal London Mat. Soc. **67** (2003) 790–804.

[81] M. I. Cortez, J. Rivera-Letelier, *Invariant measures of minimal post-critical sets of logistic maps,* Israel J. Math. **176** (2010), 157–193.

[82] M. I. Cortez, J. Rivera-Letelier, *Topological orbit equivalence classes and numeration scales of logistic maps,* Ergod. Th. Dynam. Sys. **32** (2012), 1501–1526.

[83] E. Coven, I. Kan, J. Yorke. Pseudo-orbit shadowing in the family of tent maps, Trans. Amer. Math. Soc. **308** (1988), 227–241.

[84] V. Cyr, B. Kra, *The automorphism group of a shift of subquadratic growth,* Proc. Amer. Math. Soc. **144** (2016), 613–621.

[85] P. Coullet and C. Tresser, *Itération d'endomorphismes et groupe de renormalization,* C. R. Acad. Sci. Paris Sér. A-B **287** (1978), no. 7, A577–A580.

[86] K. Dajani, C. Kraaikamp, *Ergodic theory of numbers,* Carus Mathematical Monographs **29** Mathematical Association of America, Washington, DC, 2002.

[87] P. Dartnell, F. Durand, A. Maass, *Orbit equivalence and Kakutani equivalence with Sturmian subshifts,* Studia Math. **142** (2000), no. 1, 25–45.

[88] D. Dastjerdi, M. Dabbaghian, *Mixing coded systems,* Georgian Mathematical Journal, **???** (2017), ???

[89] D. Dastjerdi, S. Jangjoo, *Computations on sofic S-gap shifts,* Preprint 2011, arXiv:1108.3414

[90] D. Dastjerdi, S. Jangjoo, *Dynamics and topology of S-gap shifts,* Topology Appl. **159** (2012), 26542661.

[91] D. Dastjerdi, S. Shaldehi, (S, S)*-gap shifts as a generalization of run-length-limited codes,* British J. of Math. & Comp. Science **4** (2014), 2765–2780.

[92] H. Davenport, *Über numeri abundantes,* Sitzungsberichte Preuss. Akad. Wiss. (1933), 830–837.

[93] H. Davenport, P. Erdös, *On sequences of positive integers,* Acta Arithm. **2** (1936), 147–151.

[94] H. Davenport, P. Erdös, *On sequences of positive integers,* J. Indian Math. Soc. New Ser. **15** (1951), 19–24.

[95] M. Dekking, M. Keane, *Mixing properties of substitutions,* Z. Wahrscheinlichkeitstheorie und verw. Gebiete **42** (1978), 23–33.

[96] R. Devaney, *An introduction to chaotic dynamical systems,* Benjamin Cummings 1986.

[97] M. Denker, C. Grillenberger, K. Sigmund, *Ergodic theory on compact spaces,* Lect. Notes Math. **527** (1976) Springer-Verlag Berlin-New York.

[98] B. Derrida, A. Gervois, Y. Pomeau, *Iteration of endomorphisms on the real axis and representation of numbers,* Ann. Inst. H. Poincaré Sect. A (N.S.) **29** (1978) 305–356.

[99] W. Diffie, M. Hellman, —em New directions in cryptography, IEEE Transactions on Information Theory **22** (1976), 644–654.

[100] E. I. Dinaburg, *The relation between topological entropy and metric entropy,* Dokl. Akad. Nauk SSSR., **190** (1970), 19–22, Soviet Math. Dok. **11** (1970), 13-16.

[101] S. Donoso, F. Durand, A. Maass, S. Petite, *On the automorphism group of low complexity subshifts,* Ergodic Theory Dynam. Systems **36** (2016), no. 1, 6495.

[102] A. Douady, *Julia sets and the Mandelbrot set*, in: H.-O. Peitgen, P. Richter: The beauty of fractals, Springer-Verlag, New York (1986), 161–173.

[103] T. Downarowicz, *The Choquet simplex of invariant measures for minimal flows,* Israel J. Math. **74** (1991), 241–256.

[104] T. Downarowicz, *Survey of odometers and Toeplitz flows,* Contemporary Mathematics, Algebraic and Topological Dynamics (Ed. Kolyada, Manin, Wardeds), **385** (2005), 7–38.

[105] T. Downarowicz, *Entropy in Dynamical Systems,* Cambridge University Press (2011).

[106] T. Downarowicz, O. Karpel, *Dynamics in dimension zero: A survey,* Discrete Contin. Dyn. Syst. **38** (2018), 1033–1062.

[107] T. Downarowicz, A. Maass, *Finite-rank Bratteli-Vershik diagrams are expansive,* Ergod. Th. Dynam. Sys. **28** (2008), 739–747.

[108] T. Downarowicz, J. Serafin, *A short proof of the Ornstein theorem,* Ergod. Th. & Dyanm. Sys. **32** Issue 2 (Daniel J. Rudolph in Memoriam) (2012), 587–597.

[109] F. Durand, *A characterization of substitutive sequences using return words,* Discrete Math., **179** (1998), 89–101.

[110] F. Durand, *Linearly recurrent subshifts have a finite number of non-periodic subshift factors,* Ergodic Theory Dynam. Systems **20** (2000), 1061–1078, and corrigendum and addendum, Ergodic Theory Dynam. Systems **23** (2003), 663–669.

[111] F. Durand, A. Frank, A. Maass, *Eigenvalues of minimal Cantor systems,* J. Eur. Math. Soc. **21** (2019), 727–775.

[112] F. Durand, B. Host, C. Skau, *Substitution dynamical systems, Bratteli diagrams and diemnsion groups,* Ergod. Th. Dynam. Sys. **19** (1999), 953–993.

[113] A. Dykstra, N. Ormes, R. Pavlov, *????*

[114] A. Dymek, S. Kasjan, R. Kułaga-Przymus, M. Lemańczyk, *$\mathcal{B}$-free sets and dynamics,* Trans. Amer. Math. Soc. **370** (2018), 5425–5489,

[115] E. Effros, *Dimensions and C-algebras,* CBMS Regional Conference Series in Mathematics, **46** Conference Board of the Mathematical Sciences, Washington, D.C., 1981.

[116] C. Epifanio, M. Koskas, F. Mignosi, *On a conjecture in bi-dimensional words,* Theo. Comput. Sci. **299** (2003), 123–150.

[117] J. Epperlein, D. Kwietniak, P. Oprocha, *Mixing properties in coded systems,* Preprint 2015, arXiv:1503.02838

[118] M. J. Feigenbaum, *Quantitative universality for a class of nonlinear transformations,* J. Stat. Phys. **19** (1978), no. 1, 25–52.

[119] S. Ferenczi, *Les transformations de Chacon: combinatoire, structure géométrique, lien avec les systêmes de complexité* $2n+1$*,* Bull. Soc. Math. France **123** (1995), 271–292.

[120] S. Frick, K. Petersen, S. Shields, *Periodic codings of Bratteli-Vershik systems,* Preprint 2018, arXiv:1812.05680

[121] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics* Lecture Notes in Mathematics, **1794** Springer-Verlag, Berlin, 2002, edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[122] A. Fraenkel, *Systems of enumeration,* Amer. Math. Monthly **92** (1985) 105–114.

[123] G. A. Freiman, *Diofantovy priblizheniya i geometriya chisel (zadacha Markova) [Diophantine approximation and geometry of numbers (the Markov spectrum)]*, Kalininskii Gosudarstvennyi Universitet, Kalinin, 1975.

[124] G. Fuhrmann, M. Gröger, *Constant length substitutions, iterated function systems and amorphic complexity,* Preprint 2018, arXiv:1812.10789

[125] G. Fuhrmann, M. Gröger, T. Oertel-Jäger, *Amorphic complexity,* Nonlinearity **29** (2016), 528–565.

[126] H. Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation,* Math. Systems Theory **1** (1967), 1–49.

[127] H. Furstenberg, B. Weiss, —em Topological dynamics and combinatorial number theory, J. Analyse Math. **34** (1978), 61–85.

[128] T. Giordano, I. Putnam, C. Skau, *Topological orbit equivalence and C-crossed products,* J. Reine Angew. Math. **469** (1995), 51–111.

[129] R. Gjerde, Ø. Johansen, *Bratteli-Vershik models for Cantor minimal systems: applications to Toeplitz flows,* Ergod. Th. Dynam. Sys. **20** (2000), 1687–1710.

[130] E. Glasner, *Structure theory as a tool in topological dynamics*, in London Math. Soc. Lecture Note Ser. **277** Cambridge Univ. Press, (2000), 173–209.

[131] M. Golay, *Notes on Digital Coding,* Proc. IRE. **37** (1947), 657.

[132] B. Gurevič, *Topological entropy for denumerable Markov chains*, Dokl. Akad. Nauk SSSR **10** (1969) 911–915.

[133] D. Gorenstein, N. Zierler, *A class of cyclic linear error-correcting codes in p^m symbols,* Journ. Soc. Ind. App. Math. **9** (1961), 107–214.

[134] W. Gottschalk, G. Hedlund, *Topological dynamics,* Amer. Math. Soc. Colloq. Publ. **36** (1955).

[135] P. Grabner, P. Liardet, R. Tichy, *Odometers and systems of numeration,* Acta Arith. **70** (1995), 103–123.

[136] M. Hall, *On the sum and products of continued fractions,* Annals of Math. **48** (147) 966–993.

[137] G. Hardy, E, Wright, *An introduction to the theory of numbers,* Oxford Univ. Press (1954).

[138] B. Hasselblatt, A. Katok, *Principal structure,* Handbook of dynamical systems, North-Holland, Amsterdam, **Vol. 1A** (2002), 1–203.

[139] N. Haydn, *Multiple measures of maximal entropy and equilibrium states for one-dimensional subshifts,*

[140] G. Hedlund, *Endomorphisms and automorphisms of the shift dynamical system,* Math. Systems Theory **3** (1969), 320–375.

[141] M. Herman, *Sur la conjugaison différentiable des difféomorphismes du cercle,* Publ. IHES. **49** (1979), 5–233.

[142] G. Hedlund, M. Morse, *Symbolic dynamic II: Sturmian trajectories,* Amer. J. Math. **66** 1940 1–42.

[143] R. Herman, Putnam, C. Skau, *Ordered Bratteli diagrams, dimension groups and topological dynamics,* Internat. J. Math. **3** (1992) 827–864.

[144] A. Heynis, *Arithmetics and combinatorics of words of low complexity,* PhD.-thesis Leiden 2001.

[145] Hocquenghem, *Codes correcteurs d'erreurs*, Chiffres, **2** (1959), 147–156.

[146] F. Hofbauer, *On instrinsic ergodicity of piecewise monotone transformations with positive entropy,* Isr. Journ. Math. **34** (1979), 213–237.

[147] F. Hofbauer, *The topological entropy of a transformation* $x \mapsto ax(1-x)$*,* Monatsh. Math. **90** (1980), 117–141.

[148] J. Hopcroft and J. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addision-Wesley Publ. ISBN 0-201-02988-X

[149] B. Host, *Valeurs propres des systèmes dynamiques définis par des substitutions de longueur variable,* Ergodic Theory Dynam. Systems **6** (1986), 529–540.

[150] J. E. Hutchington, *Fractals and self-similarity,* Indiana Univ. Math. Journal **30** (1981), 713–747

[151] K. Jacobs, M. Keane, *0-1-sequences of Toeplitz type,* Z. Wahrscheinlichkeitstheorie verw. Gebiete **13** (1969) 123–131.

[152] U. Jung, *On the existence of open and bi-continuing codes,* Trans. Amer. Math. Soc. **363** (2011), 1399–1417.

[153] M. Keane, ????

[154] M. Keane, *Non-ergodic interval exchange transformations,* Israel J. Math. **26** (1977) 188–196.

[155] M. Keane, M. Smorodinsky, *Bernoulli schemes of the same entropy are finitarily isomorphic,* Ann. of Math. (**109 (1979)**, 397–406.

[156] M. Keane, *A continued fraction titbit,* Symposium in Honor of Benoît Mandelbrot (Curaçao, 1995). Fractals **3** (1995), no. 4, 641–650.

[157] G. Keller, *Generalized heredity in* B*-free systems,* Preprint (2017) arXiv:1704.04079

[158] G. Keller, *Tautness of sets of multiples and applications to* $\mathcal{B}$*-free dynamics,* Studia Math. **247** (2019), 205–216.

[159] J. Kepler, *Harmonices mundi,* Linz 1619.

[160] D. Kerr, H. Li, *Independence in topological and* C^**-dynamics,* Math. Ann. **338** (2007), 869–926.

[161] H. Keynes, J. Robertson, *On ergodicity and mixing in topological transformation groups,* Duke Math. J. **35** (1968), 809–819.

[162] A. A. Ya. Khinchin, *Continued fractions*, Courier Dover Publications (1997).

[163] K. H. Kim, F. Roush, *On automorphism groups of subshifts,* Pure Math. Appl. Ser. B **1** (1990), 203–230.

[164] B. Kitchens, *Symbolic dynamics: one-sided, two-sided and countable state Markov shifts*, Springer Verlag ISBN 3-540 -62738-3

[165] W. Krieger, *On entropy and generators of measure-preserving transformations,* Trans. Amer. Math. Soc. **149** (1970), 453–464 and Erratum **168** (1972), 519.

[166] L. Kuipers, H. Niederreiter, *Uniform distribution of sequences*, Pure & Applied Mathematics, Wiley Interscience series of texts, monographs and tracts, Wiley & Sons, New York, London, Sydney, Toronto, 1974.

[167] D. Kwietniak, *Topological entropy and distributional chaos in hereditary shifts with applications to spacing shifts and beta shifts,* Discrete Contin. Dyn. Syst. **33** (2013), 2451–2467.

[168] M. Kulczycki, D. Kwietniak, J. Li, *Entropy of subordinate shift spaces,* Amer. Math. Monthly, **125** (2018), 141–148.

[169] P. Kůrka, *Topological and Symbolic Dynamics*, Cours Spécialisés **11** Société Mathématique de France, Paris, 2003.

[170] O. Lanford, *A computer-assisted proof of the Feigenbaum conjectures,* Bull. Amer. Math. Soc. **6** (1982), 427–434.

[171] K. Lau, A. Zame, *On weak mixing of cascades,* Math. Systems Theory **6** (1972/73), 307–311.

[172] A. Lasota, J. Yorke, *On the existence of invariant measures for piecewise monotonic transformations,* Trans. Amer. Math. Soc. **186** (1973), 481–488.

[173] F. Ledrappier, *Some properties of absolutely continuous invariant measures on an interval,* Ergodic Theory Dynam. Systems **1** (1981), 77–93.

[174] P. Lévy, *Sur le développement en fraction continue d'un nombre choisi au hasard,* Compositio Math. **3** (1936), 286–303, Reprinted in Œuvres de Paul Lévy, **6** Paris: Gauthier-Villars, (1980), 285–302.

[175] D. Lind, B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge Univ Press, ISBN 0-521-55900-6 Cambridge 1995.

[176] S. Ling, C. Xing, *Coding theorey; a first course,* Cambridge Univ. Press, ISBN 978-0-521=52923-5, Cambridge, 2004.

[177] G. Lochs, *Vergleich der Genauigkeit von Dezimalbruch und Kettenbruch,* Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg **27** (1964), 142–144,

[178] B. Manson, E. Sattler, *S-limited shifts,* Preprint 2017.

[179] K. Medynets, *Cantor aperiodic systems and Bratteli diagrams,* C. R. Math. Acad. Sci. Paris **342** (2006), 43–46.

[180] W. de Melo and S. van Strien, *One-Dimensional Dynamics,* Springer, Berlin Heidelberg New York, (1993).

[181] X. Mena, K. Petersen, *Dynamical properties of the Pascal adic transformation,* Ergod. Th. Dynam. Sys. **25** (2005), 227–256.

[182] M. Metropolis, M.L. Stein, and P.R. Stein, *On finite limit sets for transformations of the interval,* J. Combin. Theory A **15** (1973), 25-44.

[183] J. Milnor and W. Thurston, *On iterated maps of the interval: I, II*, Preprint 1977. Published in Lect. Notes in Math. **1342**, Springer, Berlin New York (1988) 465-563.

[184] T. Moothathu, *Diagonal points having dense orbit,* Colloq. Math. **120** (2010), 127–138.

[185] M. Morse, G. Hedlund, *Symbolic dynamic II: Sturmian trajectories,* Amer. J. Math. **66** 1940 1–42.

[186] B. Mossé, *Puissances de mots et reconnaissabilité des points fixes dune substitution*, Theoretical Computer Science, **99**, (1992), 327-334.

[187] B. Mossé, *Reconnaissabilité des substitutions et complexité des suites automatiques*, Bulletin de la Société Mathématique de France, **124**, (1996), 329-346.

[188] N. Ormes, R. Pavlov, *On the complexity function for sequences which are not uniformly recurrent,* Contemp. Math. to appear.

[189] D. Ornstein, *Ergodic theory, randomness and dynamical systems,* Yale Math. Monographs **5** Yale Univ. New Haven, 1974.

[190] A, Ostrowski, *Bemerkungen zur Theorie der diophantischen Approximationen,* Hamb. Abh. **1** (1921), 77–98.

[191] J. Pansiot, *Complexité des facteurs des mots infinis engendrés par morphismes itérés*, Automata, languages and programming, Springer, Berlin, Lecture Notes in Comput. Sci., **172** (1984), 380–389.

[192] W. Parry, *Intrinsic Markov chains*, Trans. Amer. Math. Soc. **112** (1964), 55–65.

[193] S. Park, Bratteli diagram isomorphic to Chacon homeomorphism, Bull. Korean Math. Soc. **37** (2000), 519–536.

[194] O. Pashina, *On arithmetic progressions in the generalized Thue-Morse word,* In: Manea, F.,Nowotka, D. (eds) WORDS 2015. LNCS, vol. 9304, pp. 191196. Springer, Heidelberg (2015).

[195] R. Pavlov, *On entropy and intrinsic ergodicity of coded subshifts,* Preprint 2018 arxiv:1803.05966v1

[196] K. Petersen, *Ergodic Theory,* Cambridge University Press, 1983 Online ISBN: 9780511608728

[197] C. Preston, *Iterates of maps of an interval,* Lect. Notes. Math. **999** Springer, Berlin New York (1983).

[198] S. Pilyugin, *Shadowing in Dynamical Systems,* Lecture Notes in Mathematics, Springer, **1706** (1999).

[199] W. Pruitt, *Eigenvalues of non-negative matrices,* The Annals of Mathematical Statistics **35** (1964) 1797–1800.

[200] A. Quas, L. Zamboni, *Periodicity and local complexity,* Theo. Comput. Sci. **319** (2004), 129–240.

[201] G. Raney, *On continued fractions and finite automata*, Mathematische Annalen, **206** (1973), 265–283,

[202] I. Reed, G. Solomon, *Polynomial Ccdes over certain finite fields*, Journ. Soc. Ind. Appl. Math., **8** 300–304.

[203] R. Rivest, A. Shamir, L. Adleman, *A method for obtaining digital signatures and public-key cryptosystems,* Communications of the ACM. **21** (1977), 120–126.

[204] H. Davenport, K. Roth, *Rational approximations to algebraic numbers,* Mathematika, (1955), 160–167,

[205] D. Rudolph, —em Fundamentals of measurable dynamics - ergodic theory n Lebesgue space, Oxford Science Publ. Oxford Univ. Press, Oxford, 1990.

[206] S. Ruette, *On the Vere-Jones classification and existence of maximal measures for countable topological Markov chains*, Pacific J. Math. **209** (2003), 366–380.

[207] S. Ruette, *Chaos in the interval,* University Lecture Series **67** Amer. Math. Soc. 2016.

[208] Martine Queffélec, *Substitution dynamical systems and spectral analysis*, Lect. Notes in Math. **1294** Springer-Verlag. ISBN 3-540-18692-1.

[209] V. Salo, *Toeplitz subshift whose automorphism group is not finitely generated*, Colloquium Mathematicum, **146** (2017), 53–76.

[210] J. Sander, R. Tijdeman, *The complexity function on lattices,* Theoret. Comput. Sci. **246** (2000), 195–225.

[211] I. Salama, *Topological entropy and recurrence of countable chains,* Pacific J. Math. **134** (1989) 325–341. Errata **140(2)** (1989).

[212] J. Sander, R. Tijdeman, *The rectangular complexity of functions on two-dimensional lattices,* Theoret. Comput. Sci. **270** (2002), 857–863.

[213] K. Schmidt, *On periodic expansions of Pisot numbers and Salem numbers,* Bull. London Math. Soc. **12** (1980), 269–278.

[214] C. Shannon, *A mathematical theory of communication,* Bell System Technical Journal, (1948), 379–423 & 623–656.

[215] C. Shannon, W. Weaver, *The Mathematical theory of communication,* Urbana, Illinois: University of Illinois Press. (1949).

[216] A. Sharkovskiy, *Coexistence of cycles of a continuous map of the line into itself,* (Russian) Ukrain. Math. Zh. **16** (1964), 61-71.

[217] K. Sigmund, *On dynamical systems with the specification property,* Trans. Amer. Mah. Soc. **170** (1974), 285–299.

[218] S. Silver, *Gosper glider gun,* The Life Lexicon. https://web.archive.org/web/20090418232240/http://www.argentum.freeserve.co.uk/lex_g.htm#gosperglidergun

[219] Y. Sinaĭ, *A weak isomorphism of transformations with invariant measure,* Dokl. Akad. Nauk SSSR **147** (1962), 797–800.

[220] S. Singh, *The code book: the science of secrecy from ancient Egypt to quantum crytography,* Fourth Estate Ltd. London, 1999.

[221] O. Toeplitz, *Beispiele zur Theorie der fastperiodischen Funktionen,* Math. Ann. **98** (1923), 281–295.

[222] J. Smítal, *Chaotic functions with zero topological entropy,* Trans. Amer. Math. Soc. **297** (1986), 269–282.

[223] C. Spandl, *Computing the topological entropy of shifts,* MLQ Math. Log. Q. **53** (2007), 493–510.

[224] H. Thunberg, *A recycled characterization of kneading sequences,* Internat. J. Bifur. Chaos Appl. Sci. Engrg. **9** (1999), 1883–1887.

[225] M. Tsfasman, S. Vlăduţ, Th. Zink, *Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound*, Math. Nachr. **109** (1982), 21–28.

[226] R. Tijdeman, *Periodicity and almost periodicity,* More Sets, Graphs and Numbers, ed. by E. Göri, G. Katona, and L. Lov ász, Springer, Berlin (2006), 381–405.

[227] W. Veech, *Interval exchange transformations,* J. Anal. Math. **33** (1978), 222–272.

[228] A. Vershik, *Uniform algebraic approximation of shifts and multiplication operators,* Soviet Math. Dokl. **24** (1981), 97–100.

[229] A. Vershik, *A theorem on Markov periodic approximation in ergodic theory,* Boundary value problems of mathematical physics and related questions in the theory of functions, 14, Zap. Nauchn. Sem. Leningrad, Otdel. Mat. Inst. Steklov (LOMI) **115** (1982), 72–82.

[230] D. Vere-Jones, *Ergodic properties of non-negative matrices-I*, Pacific J. of Math. **22** (1967), 361–385.

[231] A. Vershik, *????*

[232] P. Walters, *Some results on the classification of non-invertible measure preserving transformations,* Recent advances in topological dynamics (Proc. Conf. Topological Dynamics, Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Hedlund), Lecture Notes in Math., Springer, Berlin, **318** (1973) 266–276.

[233] P. Walters, *An introduction to ergodic theory,* Springer Verlag (1982).

[234] P. Walters, *On the pseudo orbit tracing property and its relationship to stability,* In Lecture Notes in Mathematics book series LNM, **668** (2006), 231–244.

[235] B. Weiss, *Intrinsically ergodic systems,* Bull. Amer. Math. Soc. **76** (1970), 1266–1269.

[236] B. Weiss, *Intrinsically ergodic systems,* Monatsh. für Math. **77** (1973), 462–474.

[237] R. Williams, *Toeplitz minimal flows which are not uniquely ergodic,* Zeits. Wahrs. und verw. Gebiete, **67** (1985), 95–107.

[238] S. Wolfram, *Cellular automata and complexity: collected papers,* Westview Press, Boulder Colorado (1994).

[239] J. Xiong, *A chaotic map with topological entropy zero,* Acta Math. Sci. **6** (1986), 439–443.

[240] R. Yassawi, *Rank one transformations and mixing,* MSc. Thesis McGill University, Montréal, (1993).

Index