NOTES ON SYMBOLIC DYNAMICS.

H. BRUIN

Abstract. These are notes in the making for the course VO 442503-1: Topics in Dynamical Systems: Symbolic Dynamics, Spring Semester 2017, University of Vienna

1. Notation and Introductory Notions

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. After all, the Hebrew, Greek and Roman alphabets and the Chinese characters and many others precede the Arabic numbers by many centuries. 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).

Sets of the form

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

are called **cylinder sets**. In the **product topology** on Σ, open sets are those set that can be written as arbitrary unions of finite intersections of cylinder sets.

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 1. *Are there open sets in product topology that are not closed?*

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

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

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

Exercise 2. *Show that Σ with product topology is compact if and only if $\#\mathcal{A} < \infty$.*

Lemma 3. *If $2 \le \#\mathcal{A} < \infty$, then Σ is a* **Cantor set** *(i.e., compact, without isolated points and its connected components are points). If $\mathcal{A}$ is countable, then Σ is a countable union of Cantor sets (this is called a* **Mycielski set***).*

Date: June 28, 2017.

Subshifts. The **shift** 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-invertibe on $\mathcal{A}^{\mathbb{N}}$.

Exercise 4. *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.*

Exercise 5. *Show that the shift is continuous, and in fact uniformly continuous even if $\#\mathcal{A} = \infty$.*

Definition 6. *The* **orbit** *of $x \in X$ is the set*

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

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 $x \in \overline{orb^+(x) \setminus \{x\}}$. The* **$\omega$-limit set** *of x is the set of accumulation points of its forward orbit, or in formula:*

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

Exercise 7. *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 8. *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 9. *The set $X = \{x \in \Sigma : x_i = 1 \Rightarrow x_{i+1} = 0\}$ is called the* **Fibonacci shift**[1]*. It disallows sequences with two consecutive 1s in it.*

Example 10. *X is a collection of labels of infinite paths through the graph in Figure 11. Labels are given to the vertices of the graph, and no label is repeated.*

Example 11. *X is a collection of labels of infinite paths through the graph in Figure 11. Labels are given to the arrows of the graph, and labels can be repeated (different arrows with the same label can occur).*

Example 12. *X 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 X the* **even shift**.

[1]Warning: there is also a Fibonacci substitution shift = Fibonacci Sturmian shift, which are different from this one

FIGURE 1. Transition graphs: vertex-labeled (left) and edge-labeled (right)

Example 13. *Let S be a non-empty subset of $\mathbb{N}$. Let X 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*

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

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

Example 14. *X is the closure of the collection of symbolic trajectories of a circle rotation $R : \mathbb{S}^1 \to \mathbb{S}^1$, see Figure 2. That is, if $y \in \mathbb{S}^1$ and $R^n(y) \in [0, \alpha)$ then we write $x_n = 0$. Otherwise $x_n = 1$.*

Example 15. *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 2.*

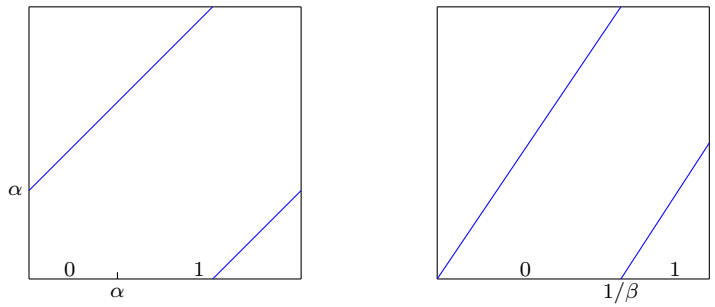

FIGURE 2. Symbolic dynamics for a circle rotation R_α (left) and the β-transformation T_β (right)

Example 16. *The alphabet is $\mathcal{A}$ consists of brackets $(,), [,]$ and $\mathcal{L}(X)$ (see Definition 17 below) consists of all words of pairs of brackets that are properly paired and unlinked. So [([ ])] and (( ) [ ]) belong to $\mathcal{L}(X)$, but [( ] or ([ )] don't.*

Words, prefix, suffix: 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**[2] of x. If u is concatenated as $u = vw$, then v is a **prefix** and w a **suffix** of u.

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

Definition 17. *The collection*

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

is called the **language** *of* X*.*

Definition 18. *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 19. *Show that for the Fibonacci shift of Example 9,* $p(n) = F_{n+1}$*, where* $F_1, F_2, F_3, F_4, F_5, \cdots = 1, 2, 3, 5, 8, \ldots$ *are the Fibonacci numbers.*

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 subshft is called **regularly enumerable** in the Chomsky hierarchy if its language can be recognized by a Turing machine.

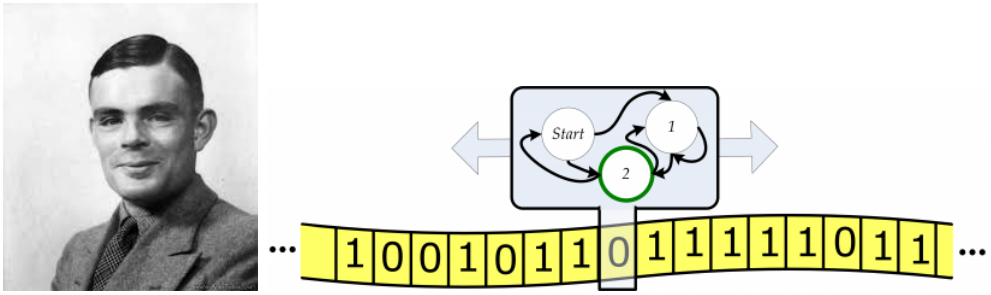

FIGURE 3. Turing and his machine.

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 $S_1, \ldots, S_N$, 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 20. *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 nd 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_2.*

S_3: *Halt. The word is rejected.*

Exercise 21. *Design a Turing machine that accepts the word in the even shift (Example 12).*

Exercise 22. *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.*

2. General properties of subshifts

Definition 23. *A subshift X is* **transitive** *or* **irreducible** *if for every $u, w \in \mathcal{L}(X)$, there is $u \in \mathcal{L}(X)$ such that $uvw \in \mathcal{L}(X)$.*

Proposition 24. *A subshift is transitive if and only if there exists a dense orbit.*

Proof. Suppose first that $\mathrm{orb}(x)$ is dense. Then for every $u, w \in \mathcal{L}(X)$ there are $m < m + |u| \leq n \in \mathbb{N}$ 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 recursevely. 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. □

Definition 25 (Sliding Block Code)**.** *A map $p : \mathcal{A}^{\mathbb{Z}} \to \tilde{A}^{\mathbb{Z}}$ is called a* **sliding block code** *of* **window size** $2N + 1$ *if there is a function $f : \mathcal{A}^{2N+1} \to \tilde{A}$ such that $\pi(x)_i = f(x_{i-N} \dots x_{i+N})$.*

In other words, we have a window 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 26 (Curtis-Hedlund-Lyndon)**.** *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.*

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, $\pi(x)_i = (x_{i-N} \dots x_{i+N})$ 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). □

3. Subshifts of finite type

Definition of SFTs and transition matrices and graphs:

Definition 27. *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 $|v| \ge M$, then $uvw \in \mathcal{L}(X)$ as well.

Definition 28. *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 29. *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 contains no forbidden word, then $uvw \in \mathcal{L}(X)$. □

Lemma 30. *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 $M+1 \ge 2$. Take a new alphabet $\mathcal{B} = \mathcal{A}^M$, say $b_1, \dots, b_n$ are its letters. Now recode all $x \in X$ using a "sliding block code" π, where for each index i, $\pi(x)_i = b_j$ if b_j is the symbol used for $x_i x_{i+1} \dots x_{i+M-1}$. 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 31. *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 4; labels at the arrow then just indicate with word in $\{0,1\}^3$ they stand for. Each arrow containing a forbidden word is colored red, and then removed in the right panel of Figure 4.*

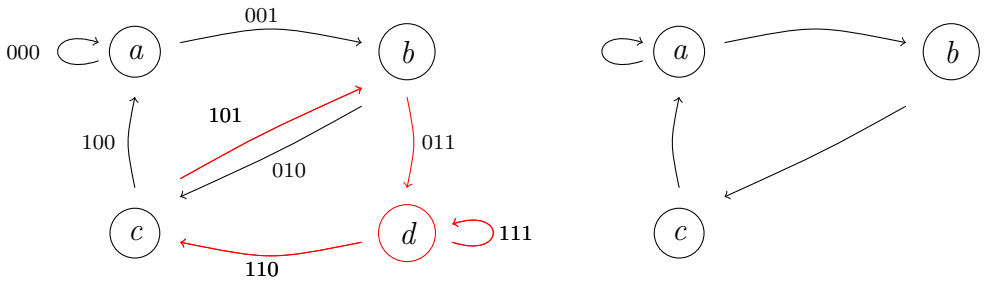

FIGURE 4. Illustrating the recoding of the SFT with forbidden words 11 and 101.

Corollary 32. *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 33. *The 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.*

Definition 34. *A* **coded subshift** *is a subshift X for which there is a countable collection S of words such that the collection of free concatenation of words in S is dense in X.*

Exercise 35. *Is the Fibonacci SFT of Example 9 a coded subshift? Is the even shift of Example 12 a coded shift?*

Exercise 36. *Let X be the SFT with forbidden words* 13, 21 *and* 32. *Is X a coded shift?*

Proposition 37. *Every SFT is a coded shift.*

Proof. Rewrite the SFT to an SFT with memory $M = 1$, i.e., all forbidded words have length ≤ 2. Let $\mathcal{G}$ be the transition graph, and fix vertices a, b such that the arrow $a \to b$ occurs in $\mathcal{G}$. Now let S contain the codes all finite paths $b \to \cdots \to a$; these can be freely concatenated. If $\mathcal{G}$ is not connected, we do the same for every connected component. □

Remark 38. *Naturally, the set S of codes may not be the most economical, but the idea of the proof is quite general. It can also be used to show that synchronized subshift are coded. Therefore we have the inclusion.*

$$\textit{SFTs} \subset \textit{ synchronized subshifts } \subset \textit{ coded subshifts }.$$

Topological entropy:

Definition 39. *A non-negative matrix $A = (a_{ij})$ is called* **irreducible** *if for every i, j there is k such that $a_{ij}^{(k)} > 0$. 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 P. We call A* **aperiodic** *if its period is* 1.

Exercise 40. *If A is irreducible, show that per(i) is indeed independent of i.*

Theorem 41 (Perron-Frobenius)**.** *Let A be an irreducible aperiodic matrix non-negative matrix. Then*

- *There is a real positive eigenvalue λ (called the* **leading** *or* **Perron-Frobenius eigenvalue***), of algebraic multiplicity one, such that $\lambda > |\mu|$ for every other eigenvalue μ of A.*
- *The eigenvector (left and right) associated to λ can be chosen to be strictly positive.*

Definition 42. *The* **topological entropy** *of a subshift is*

$$h_{top}(X,\sigma) = \limsup_{n\to\infty} \frac{1}{n} \log p(n).$$

Theorem 43. *The entropy of an irreducible SFT equals $\log\lambda$ where λ is the leading eigenvalue of the transition matrix.*

Proof. Let $P^n = (p_{ij}^{(n)})_{i,j\in\mathcal{A}}$ be the n-th power of the transition matrix. Every n-word in $\mathcal{L}(X)$ corresponds to an n-path in the transition graph, and the number 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

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

provided P is aperiodic. (If P 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 $C^{-1}\lambda^n \le p(n) \le (\#\mathcal{A})^2 C\lambda^n$ and $\lim_n \frac{1}{n}\log p(n) = \log\lambda$. □

Definition 44. *Two subshifts (X,σ) and (Y,σ) are called* **conjugate** *if there is a homeomorphism $h : X \to Y$ such that $h\circ\sigma = \sigma\circ h$.*

If $h : X \to Y$ commutes with σ and is a continuous, onto, but not necessarily one-to-one map, then Y is called a **factor** *of X.*

Exercise 45. *Show that if X is a factor of Y and Y a factor of X, then X and Y are conjugate.*

Proposition 46. *If (Y,σ) is a factor of (X,σ), then $h_{top}(Y,\sigma) \le h_{top}(X,\sigma)$. If (X,σ) and (Y,σ) are conjugate, then $h_{top}(X,\sigma) = h_{top}(Y,\sigma)$.*

Proof. Let $h : X \to Y$ be the factor map. Since it is continuous, it is a sliding block code by Theorem 26, say of window length $2N+1$. Therefore the word complexities relate as $p_Y(n) \le p_X(n+2N)$, and hence

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

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

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.

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

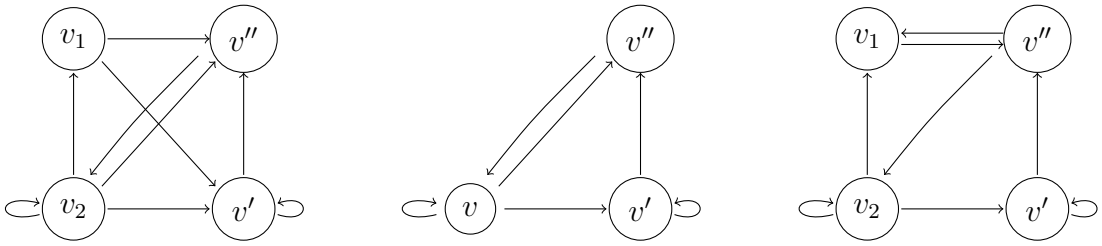

FIGURE 5. Insplit graph Original $\mathcal{G}$ Outsplit graph

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 47. *Let $\mathcal{G} = (V,E)$. An* elementary outsplit graph $\hat{\mathcal{G}} = (\hat{V},\hat{E})$ *is obtained by*

- *doubling one vertex $v \in V$ to 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** *$\hat{e} = (v_2 \to w)$;*
- *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 $(v_i \to v_1), (v_j \to v_2) \in \hat{E}$ where $i,j \in \{1,2\}$.*

An **outsplit graph** *is then obtained by successive elementary outsplits.* (Elementary) insplit graph *are defined similarly, replacing the roles of E_v and E^v.*

If all $e \in E$ had a unique label, we will also give each $\hat{e} \in \hat{E}$ a unique level.

Proposition 48. *Let $\hat{\mathcal{G}}$ be an in- or outsplit graph obtained from $\mathcal{G}$. Then the edge-subshift $\hat{X}$ of $\hat{\mathcal{G}}$ is conjugate to the edge-subshift X of $\mathcal{G}$.*

Proof. We give the proof for an elementary outsplit $\hat{\mathcal{G}}$; the general outsplit and (elementary) insplit graph follow similarly. By Theorem 26, 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 determine 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 $\hat{\pi}(x)_i = \hat{f}(x_i, x_{i+1})$.

□

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 skew-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 49. *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 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 $CD = \hat{A}$. □

Definition 50. *Two matrices A and $\hat{A}$ are* **strongly equivalent** *if there are (rectangular) matrices D and C such that $DC = A$ and $CD = \hat{A}$.*

That is, in effect, that their associated graphs 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.

Theorem 51 (Williams)**.** *Two SFTs are conjugate if and only if their transition matrices are strongly equivalent.*

4. Sofic subshifts

Definition 52. *A subshift (X, σ) is called* **sofic** *if it is the space of paths in an edge-labeled graph.*

Lemma 53. *Every SFT is sofic.*

Proof. Let $\mathcal{G}$ be the vertex-labeled transition graph of the SFT. Create the dual graph by

(1) giving each arrow a distinct label, say in an alphabet $\mathcal{A}$‘;
(2) building a graph $\mathcal{G}'$ with vertices $a \in \mathcal{A}'$;
(3) and arrows $a' \to b'$ in $\mathcal{G}'$ if a' labels the incoming arrow and b' the outgoing arrow of the **same** vertex in $\mathcal{G}$.

The possible paths in $\mathcal{G}$ are in one-to-one correspondence with the paths in $\mathcal{G}'$, so the SFT is sofic. □

Remark 54. *Not every sofic shift is an SFT. For example the even shift (Example 12) 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**.

The word sofic comes from the Hebrew word for finite, and was coined by Benji Weiss. 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 55. *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. 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 26, π is continuous and commutes with the shift.

Conversely, if X is a factor of an SFT, the factor map is a sliding block code by Theorem 26, say of window size $2N+1$: $\pi(x)_i = f(x_{i-N}, \ldots, x_{i+N})$. Represent the SFT by an edge-labeled graph $\mathcal{G}'$ where the labels are the $2N+1$-words $w \in \mathcal{A}^{2N+1} \cap \mathcal{L}(X)$. 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 56. *Every factor of a sofic shift is again a sofic shift. Every shift conjugate to a sofic shift is again sofic.*

Definition 57. *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 58. *Let X be the one-sided* **even shift** *from Example 12. Then $\mathcal{F}(0) = \mathcal{L}(X)$ because a 0 casts no restrictions on the follower set. Also $F(011) = \mathcal{L}(X)$, 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.*

Theorem 59. *A subshift* (X, σ) *is sofic if and only if the collection of its follower sets is finite.*

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

(1) Let C be the vertices of $\mathcal{G}$.
(2) If $a \in \mathcal{A}$ and $v \in \mathcal{L}(X)$, then $F(va) \in C$; draw an edge $F(w) \to F(va)$. (Although there are infinitely many $v \in \mathcal{L}(X)$, there are only finitely many follower sets, and we need not repeat arrows between the same vertices with the same label.)

It is easy to see that 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 be the collection of terminal vertices of these paths. Then $F(w)$ is the collection of infinite paths starting at a vertex in T. Since $\mathcal{G}$ is finite, and there are only finitely many subsets of its vertex set, the collection of follower sets is finite. □

Definition 60. *An edge-labeled transition graph* $\mathcal{G}$ *is* **right-resolving** *if for each vertex* v *of* $\mathcal{G}$*, 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.*

Without proof we mention that every sofic shift has a right-resolving follower-separated graph representation. If we also minimalise the number of vertices in such graph, there is only one such graph with these properties.

Corollary 61. *Every sofic shift* X *is synchronizing.*

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, then every paths containing u goes through v, and there is a unique follower set $F(u)$, 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 $u_1 \in F(v)$ but $u_1 \notin F(v')$ (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 $u_1 \notin F(v')$, 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.) □

Remark 62. *This extends the diagram of Remark 38 into:*

$$\textit{SFTs} \subset \textit{ sofic shifts } \subset \textit{ synchronized subshifts } \subset \textit{ coded subshifts }.$$

5. β-SHIFTS AND β-EXPANSIONS

Throughout this section, we $\beta > 1$. 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 rations. 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 $x_2 = \beta x - b_2$ 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 63. *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 $T_\beta(x)$.

Lemma 64. *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 65. *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 6), and it has a simple edge-labeled transition graph.

Proof of Lemma 64. Let b 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 $b_2 = \lfloor \beta x \rfloor \leq c_2 = \lfloor \beta T_\beta(1) \rfloor$. Continue by induction. □

Proposition 66. *The β-shift is a coded shift.*

Proof. Let $c = c_1 c_2 c_3 \dots$ be the β=expansion of 1. Then we can take as set of codes

$$S = \{\underbrace{0, 1, \dots, (c_1 - 1)}_{1-\text{words}}, \underbrace{c_1 0, c_1 1, \dots, c_1(c_2 - 1)}_{2-\text{words}}, \underbrace{c_1 c_2 0, c_1 c_2 1, \dots, c_1 c_2 (c_3 - 1)}_{3-\text{words}}, \dots\}$$

Any concatenation in S^* then satisfies Lemma 64, so that S^* is dense in X_β. □

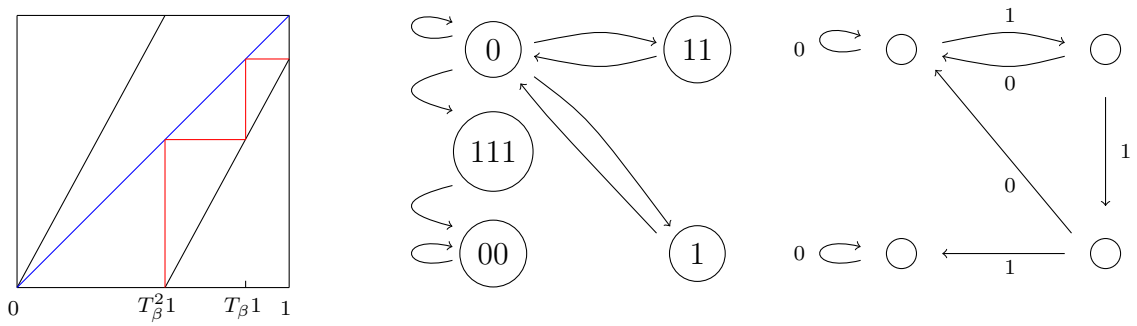

FIGURE 6. Left: The map T_β for $\beta^3 = \beta^2 + \beta + 1$. Then $T_\beta^3(1) = 0$. Middle: A corresponding vertex-labeled graph. Right: A corresponding edge-labeled graph.

Theorem 67. *If* 1 *has a finite expansion, then* X_β *is conjugate to an SFT.*
If 1 *has a preperiodic expansion, then* X_β *is sofic.*
The β*-shift* X_β *is synchronizing if and only if the* T_β*-orbit of* 1 *is not dense in* $[0,1]$.

Note that, since there are uncountably many choices of $\beta > 1$, all leading to non-conjugate subshifts (see Theorem 70 below), while there are only countably many sofic shifts, X_β is not sofic for most β.

Proof. For the first statement, let $a_j = T_\beta(1)$, 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 an SFT over the alphabet $\mathcal{P}$. By enlarging the alphabet appropriately, also (X_β, σ^{N-1}) can be transformed into be an SFT.

If $c = c_1 c_2 \dots c_N (c_{N+1} \dots c_{N+p})^\infty$, we claim that X_β only has finitely many different follower sets, see Definition 57. Let w be a proper suffix of some $s_1 s_2 s_3 \dots \in S^*$. 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

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

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 59, X_β is sofic.

Finally, assume that $\mathrm{orb}(1)$ is not dense in $[0,1]$ and let U be an interval that is disjoint from $\overline{\mathrm{orb}(1)}$. 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 $T_\beta^M(Y) \cap Z \neq \emptyset$. But since $\mathrm{orb}(1) \cap Z = \emptyset$, 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 $\mathrm{orb}(1)$ is dense, then there is $n \in \mathbb{N}$ such that $T_\beta^n(1) \in Z$. Therefore there is a one-sided neighbourhood 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. □

The above types of β-shifts correspond to certain algebraic properties of β, which we will mention, but not prove.

Definition 68. *Let β be an algebraic number and denote its minimal polynomial by f. That is, f has integer coefficients, $f(\beta) = 0$ and the degree of f is minimal w.r.t. the previous properties. The other solutions of $f(x) = 0$ are called the* **algebraic conjugates** *of β. The number $\beta > 1$ is called a* **Pisot number** *if all its algebraic conjugates satisfy $|x| < 1$. It is called a* **Salem number** *if all its algebraic conjugates satisfy $|x| \leq 1$ with equality for at least one algebraic conjugate. Finally, $\beta > 1$ is a* **Perron number** *if all its algebraic conjugates satisfy $|x| < \beta$. (Perron numbers are the leading eigenvalue of some non-negative aperiodic irreducible matrix, see Theorem 41.)*

Theorem 69. *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 [2, Chapter 7] for more results in this spirit.

Theorem 70. *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.

Recall that $c = c_1 c_2 \dots$ denotes the β-expansion of 1. By Proposition 66, every word in $\mathcal{L}(X_\beta)$ has the form

$$s_1 s_2 \dots s_m c_1 c_2 \dots c_k \quad \text{for some (maximal) } s_1, \dots, s_m \in S,\ k \geq 0. \tag{1}$$

Let $p_X(n)$ 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 $p_{S^*}(n) \leq p_X(n)$. Conversely, by (1),

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

Therefore the exponential growth rates are the same.

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

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 $p_{S^*}(n) = \sum_{k=1}^n p_S(k) p_{S^*}(n-k)$, and this gives for the power series

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

Therefore $f_{S^*}(t) = \frac{1}{1-f_S(t)}$. Since $1 = \sum_{n\geq 1} c_n \beta^{-n}$, 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

$$\liminf_{n\to\infty} \frac{1}{n} \log p_S(n) = \log \beta.$$

□

6. Minimal subshifts

Proposition 71. *We have the following equivalent characterizations for a subshift (X, σ) to be* **minimal***:*

1. *There is no closed shift-invariant proper subset of X;*
2. *Every orbit is dense in X;*
3. *There is one dense orbit and σ is* **uniformly recurrent***, 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 72. *Uniform recurrence means that the sets of integers n such that $\sigma^n(x) \in U$ has* **bounded gaps** *or (with a different word) is* **syndetic***.*

Proof. We prove the three implications by the contrapositive.
1. ⇒ 2.: Suppose that $x \in X$ has an orbit that is not dense. Then $\overline{\mathrm{orb}(x)}$ is a shift-invariant closed proper subset, so 1. fails.
2. ⇒ 3.: By 2. 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 $\mathrm{orb}(x)$ is not dense, and 2. fails.
Now take $y \in U$ arbitrary, 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 prioves the uniform recurrence of U. **3.** $\Rightarrow$ **1.:** 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 $\overline{U} \cap Y = \emptyset$. 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 $\mathrm{orb}(y) \subset Y$, there is an open set $V \ni y$ such that $\sigma^i(V) \cap U = \emptyset$ 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 73. *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.

Definition 74. *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

$$x = w_1w_2w_3w_4w_5w_6\cdots = uw'_1uw'_2uw'_3uw'_4w'_2uw'_5uw'_6\ldots, \tag{2}$$

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 75. *If χ_{fib} is the Fibonacci substitution, so*

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

and if $u = 010010$, then $w \in \mathcal{R}_u$ because $wu = 010010010$ 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 76. *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 77 (Duran, Host & Skau [3])**.** *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: $p(n) \le Ln$ 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) *$\#\mathcal{R}_u \le L(L+1)^2$.*
(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

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

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 $v \le L|v|$, 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 98. But then $\mathcal{L}(X)$ is shift-periodic.
(iii) Take $u \in \mathcal{L}(X)$ and $w \in \mathcal{R}_u$. If $|w| < L|u|$, 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 $\le (L+1)|u|$ 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. (2)), so using the minimal length $|w| \ge |u|/L$ of return words (from item (iii)), we find $\#\mathcal{R}_u \le |v|/(|u|/L) = L(L+1)^2$.
(5) 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 $|u| \ge 2N+1$ and v its image under f. Then $|v| = |u| - 2N$. If $w \in \mathcal{R}_v$, then $|w| \le \{|s| : s \in \mathcal{R}_u\} \le L|u| \le L(|v|+2N) \le L(2N+1)|v|$. Therefore Y is linearly recurrent with constant $L(2N+1)$. In fact, the proof gives that v will return with gap $\le L+\varepsilon$ if v is sufficiently long. □

7. Substitution shifts

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

Definition 78. *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:

$$\chi(a_1a_2\dots a_r) = \chi(a_1)\chi(a_2)\dots\chi(a_r).$$

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

Example 79. *The Fibonacci substitution is defined on* $\mathcal{A} = \{0, 1\}$ *by*

$$\chi_{fib} : \begin{cases} 0 \to 01 \\ 1 \to 0 \end{cases}$$

Iterating χ *on symbol* 0 *we find:*

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

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

Remark 80. *As can be seen in Example 79, 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 81. *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 80 to χ^{r-p} and b. □

Example 82. *Take* $\chi(0) = 10$ *and* $\chi(1) = 1$*. Then*

$0 \to 10 \to 110 \to 1110 \to 11110 \to \dots \to 1^\infty$ *fixed by* χ*.*

$1 \to 1$ *fixed by* χ*.*

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

$$\forall a \in \mathcal{A} \quad \lim_{n\to\infty} |\chi^n(a)| = \infty. \tag{3}$$

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

$$\chi(0) \text{ starts with } 0. \tag{4}$$

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

Example 83. *The* **Thue-Morse substitution**[3] *is defined by*

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

It has **two** *fixed points*

$$\rho^0 = 01\ 10\ 1001\ 10010110\ 1001011001101001\ \dots$$

$$\rho^1 = 10\ 01\ 0110\ 01101001\ 0110100110010110\ \dots$$

This sequence makes its appearance in many circumstances in combinatorics and elsewhere. For instance, if you have a sequence of objects $(P_k)_{k\geq 1}$ *(e.g. rugby players) which you want to divide over two teams* T_0 *and* T_1*, so that the teams are closest*

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

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.

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:

$$\rho = 10\ 11\ 1010\ 10111011\ 1011101010111010\ \dots$$

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

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

This sequence appears as the kneading sequence (itinerary of the critical value) of the (infinitely renormalizable) Feigenbaum interval map, see [**?**].

Proposition 84. *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 suqere-free word you get soon stuck:

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

To create a three-letter square-free infinite word, start with ρ^0 and replace the symbol by a 2 if a square threatens to appear:

$$0120\ 1021\ 20210120\ 1021012021201021\ \dots$$

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. □

Definition 85. *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).

Lemma 86. *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

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

where the central dot indicates where the zeroth coordinate is.

Two 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 81, there is $b \in \mathcal{A}$ and $p < q \in \mathbb{N}$ such that $\rho^p(a)$ and $\rho(a)$ both end in b. Set $N = q - p$, so $\rho^N(b)$ ends with b. Next iterate $\rho^N(b.0)$ repeatedly, so that $\lim_k \rho^{kN} =: \hat{\rho}$ 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, due to minimality (see below), the shift space $\hat{X}_\rho$ is unique. □

Definition 87. *The* **associated 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 symbls j appearing in $\chi(i)$. We call chi (***a***)periodic and/or* **irreducible** *if A is (***a***)periodic and/or* **irreducible***, in the sense of the Perron-Frobenius theory, see Definition 39. 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 88. *Let χ be a substitution satisfying hypotheses* (3) *and* (4)*. 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 71), 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), $\chi^k(a)$ 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:

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

By the choice of $k(\rho_j)$, each of these blocks contains a 0, so 0 appears gap K. Now take $w \in \mathcal{L}(X_\rho)$ arbitrary. There exists $m \in \mathbb{N}$ such that w apppears 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 89 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. The above theorem doesn't exclude that ρ is periodic. For instance,

$$\chi : \begin{cases} 0 \to 010 \\ 1 \to 101 \end{cases} \tag{5}$$

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 98.

Theorem 89. *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

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

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

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

Now let $u \in \mathcal{L}(X)$ and $v \in \mathcal{R}_u$ be arbitrary. Choose $k \ge 1$ minimal such that $|u| \le I_k$. 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

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

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

Remark 90. *It turns ut (cf. Theorem 77(v)) that a factor of a substitution subshift is again a substitution subshift. In fact, One of the main results of* [3] *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.*

7.1. **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 in deed injective, but in general they 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 (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,

$$x = \dots|010|101|010|10\cdots = \dots 0|101|010|101|0\cdots = \dots 01|010|101|010|\dots \tag{6}$$

and each with their own inverse. The way to cut x into block $\chi(a)$ 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 91. *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 is 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 unilaterally 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 showed a gap in Mentzen proof (Kamaes 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 nonconstant length), but using a new notion of (two-sided) recognizable substitution. She refined this result in 1996.*

This problem was tackled by several people (Mentzen, Queffélec [13], Host, Mossé [10, 11]), and it the results of Mossé that are predominantly 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 92. *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 $x_i \dots x_{i+N-1} = x_j \dots x_{j+N}$ 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 $x_{i-N+1} \dots x_{i+N-1} = x_{j-N+1} \dots x_{j+N}$ 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 (6) is not recognizable, but for example the fixed point of the Fibonacci substution χ_{fib} is recognizable with recognizability index 2. The Thue-Morse word ρ^0 (or ρ^1) is recognizable with recognizability index 4.

Theorem 93 (Mentzen (1989), Apparicio (1999) [1])**.** *Every primitive injective constant length substitution with aperiodic fixed point is one-sided recognizable.*

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

Example 94. *The substitutions*

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

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 sayingif the is a cut directly before its occurrence or not. This problem does not disappear if you take longer word. The latter substitution is called the **Chacon substitution**.

Theorem 95 (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 $ab \in \mathcal{A}$ 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 96 (Mossé (1992))**.** *Every aperiodic primitive injective substitution is two-sided recognizable.*

8. 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.

The name **Sturmian** comes from Hedlund & Morse [4], 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 7 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 7 right). If the motion is not periodic, the resulting sequence is Sturmian.

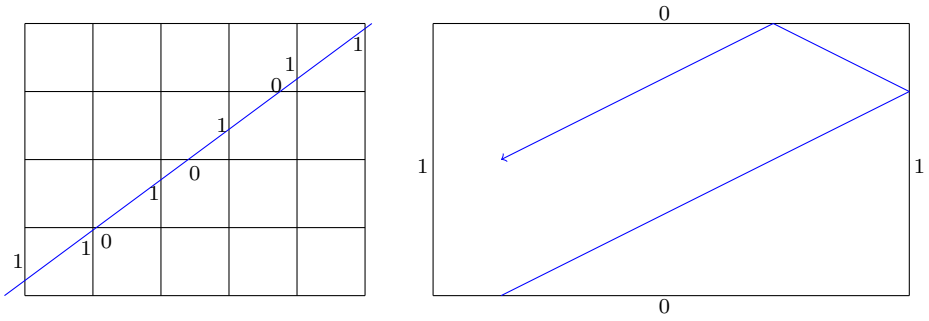

Figure 7. Sturmian sequences produced as intersections with horizontal and vertical grid-lines (left) and billiards on a rectangular billiard table (right)

For simplicity of exposition, we use the property that $p(n)=n+1$ for n as **Sturmian**, see Section 8.3. We start with some terminology and a useful proposition.

Definition 97. *We call an n-word*

- **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 98. *If a recurrent subshift has $p(n)\le n$ for some n, then it is periodic.*

Proof. Let n be maximal such that $p(k)\ge k$ for all $k\le n$. Then $p(n-1)=n$ and there are no right-special words of length $n-1$. Start with an $n-1$-word u; there is

only one way to extend it to the right by one letter, say to ua. Then the $n-1$-suffix of ua can also be extended to the right by one letter in only one way. Continue this way, until after at most $p(n-1) = n$ steps, we end up with suffix u again. Therefore X contain only (shifts of) this word periodically repeated. □

8.1. **Rotational sequences.**

Definition 99. *Let $R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ be the rotation over an irrational angle α. Let $\beta \in \mathbb{S}^1$ and build the itinerary $u = (u_n)_{n\geq 0}$ by*

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

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

Remark 100. *The addition sequences obtain by taken the closure can also be obtained by taking the half-open interval the other way around:*

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

Lemma 101. *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 palindrom 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. □

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

Definition 102. *A subshift X is called* **balanced** *if there exists $N \in \mathbb{N}$ such that for all $a \in \Sigma$ 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 $N = 1$.*

Lemma 103. *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

$$u_n = \lfloor n\alpha + \beta \rfloor - \lfloor (n-1)\alpha + \beta \rfloor \tag{7}$$

for all $n \in \mathbb{Z}$. This is easy to check, except that in order to include the sequences mentioned in Remark 100, we need to add the alternative definition

$$u_n = \lceil n\alpha + \beta \rceil - \lceil (n-1)\alpha + \beta \rceil \tag{8}$$

for all $n \in Z$.

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 104. *If X is an unbalanced subshift on alphabet $\{0,1\}$ with, 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 $|u|_a = \#\{1 \le i \le n : u_i = a\}$. Let N be minimal such there are N-words $a = a_1\dots a_N$ and $b = b_1\dots b_N \in \mathcal{L}(X)$ such that $|\,|a|_1 - |b|_1\,| \ge 2$. Let since $|a|_1 - |b|_1$ can change by at most 1 if a, b are shortened or expanded by one letter, the minimality of N implies that $a = 0a_2\dots a_{N-1}0$ and $b = 1b_2\dots ab_{N-1}1$ (or vice versa) and $|a_2\dots a_{N-1}|_1 = |b_2\dots b_{N-1}|_1$. If $a_2\dots a_{N-1} = b_2\dots b_{N-1}$, then we have found our word w. Otherwise, take $k = \min\{j > 1 : a_j \ne b_j\}$ and $l = \min\{j < N : a_j \ne b_j\}$. We have four possibilities, all leading to shorter possble words a and b.

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

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

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

8.3. **Sturmian sequences.**

Definition 105. *A word $u \in \{0,1\}^N$ 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 $n \ge 0$. Take the shift-orbit closure $X = \overline{orb_\sigma(x)}$. The corresponding subshift (X,σ) for $X = \overline{orb_\sigma(u)}$ is called a* **Sturmian subshift**.

Remark 106. *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 $p_u(n) = n+1$.*

Lemma 107. *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 $Q_n := \{R_\alpha^{-k} : 0 \leq k \leq n\}$. 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 108. *A sequence is Sturmian if and only if it 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 $p(N) \geq N + 2$. 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 104, 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 109. *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

$$M_n = \min\{|u_{k+1}\dots u_{k+n}|_1 : k \geq 0\}. \tag{9}$$

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

$$kM_q \leq M_{kq+r} = M_n \leq k(M_q + 1) + r. \tag{10}$$

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 (10). This gives

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

so $\{\frac{M_{2^m q}}{2^m q}\}_m$ is increasing $\{\frac{M_{2^m q}}{2^m q}\}_m$ 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. $\square$

$\frac{M_q+1}{q}$,

Lemma 110. *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 109 we know that $\alpha \in (\frac{M_n}{n}, \frac{M_n+1}{n})$ and $\alpha \in (\frac{a'_n}{n}, \frac{a'_n+1}{n})$ where M_n and M'_n are given by (9) 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. Asume 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. $\square$

Theorem 111 (Hedlund & Morse)**.** *Every Sturmian sequence is rotational.*

Proof. Let u be a Sturmian seuqence; by Theorem 108 it is balanced as well. By Proposition 109, u has an irrational frequency $\alpha = \lim_n \frac{1}{n}|u_1 \dots u_n|_1$, and by Lemma 110, 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 (8) 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. □

8.4. **Rauzy graphs.** The **Rauzy graph** Γ_n of a Sturmian subshft X is the word-graph in which the vertices are the n-words u in X and there is an arrow $u \to u'$ if $ua = bu'$ for some $a, b \in \{0, 1\}$. Hence Γ_n has $p(n)$ vertices and $p(n+1)$ edges.

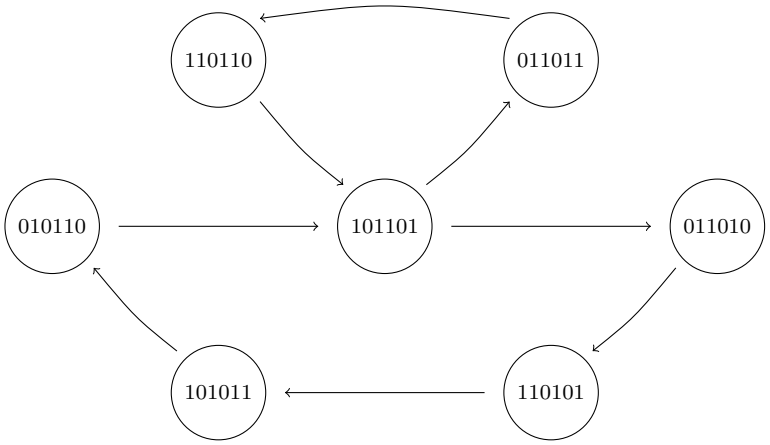

FIGURE 8. The Rauzy graph Γ_6 based on the Fibonacci Sturmian sequence 101 10 101 10110 10110101 1011010110110 ....

Since $p(n) = n + 1$, there is exactly one left-special and one right-special word of length n. They can be merged in a single bi-special word. In the example of Figure 8 below, the word $u = 101101$ is bi-special, but only $0u0, 0u1$ and $1u0 \in \mathcal{L}(X)$.

Since $p(n+1) - p(n) = 1$ for a Sturmian sequence, every Rauzy graph contains exactly one left-special and one right-special word, and they may be merged into a single bi-special word. Hence, there are two types of Rauzy graphs, see Figure 9.

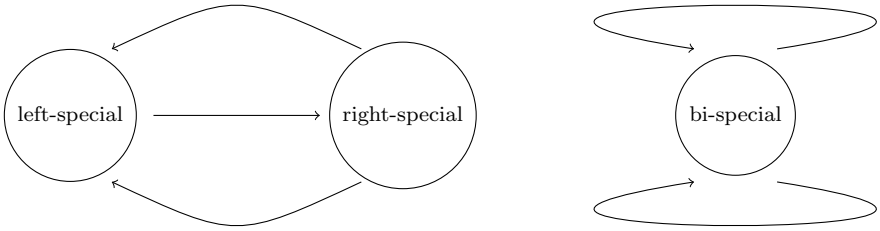

FIGURE 9. The two types of Rauzy graphs for a Sturmian sequence.

The transformation from Γ_n to Γ_{n+1} is as follows:

- If Γ_n is of the first type, then the middle path decreases by one vertex, and the upper and lower path increase by one vertex.
- If Γ_n is of the second type, then one of the two paths becomes the central path in Γ_{n+1}, the other path becomes the upper path of Γ_{n+1}, and there is an extra arrow in Γ_{n+1} from the right-special word to the left-special word.

Theorem 112. *For each $n \in \mathbb{N}$, there are at most three values that the frequency*

$$\lim_{k\to\infty} \frac{1}{k} \#\{1 \le i \le k : x_{i+1} \dots x_{i+n} = w\}$$

can take for an n-word in a Sturmian word x. These three values depend only on n and the rotation angle α.

Remark 113. *This is the symbolic version of what is known as the* **three gap theorem** *which was conjectured by Hugo Steinhaus and eventually proven by Vera Sós:*

> *For every $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and $n \in \mathbb{N}$, the collections $\{j\alpha \pmod 1\}_{j=0}^{n-1}$ divides the circle into intervals of at most three different sizes.*

Indeed, since Lebesgue measure is the only invariant probability measure that is preserved by the rotation $R_\alpha : x \mapsto x + \alpha \pmod 1$, the frequencies in Theorem 112 corresponds to the Lebesgue meaures (i.e., length) of the intervals.

Proof. This is a special case of the more general statement that the frequency can take at most $3(p(n+1) - p(n))$ values, which we will prove here. For Sturmian sequences $3(p(n+1) - p(n)) = 3$.

Let $n \in \mathbb{N}$ be arbitrary and let Γ_n be the word-graph of the language. For every vertex $a \in \Gamma_n$ let a^- and a^+ be the number of incoming and outgoing arrows. That is, a is left-special resp. right-special if $a^- \ge 2$ reps. $a^+ \ge 2$.

Let $V_1 = \#\{a \in \Gamma_n : a^+ \ge 2\}$ be the collection of right-special n-words. Then

$$\#V_1 = \sum_{a^+\ge 2} 1 \le \sum_{a^+\ge 2} a^+ - 1 \le p(n+1) - p(n),$$

Next set $V_2 = \{a \in \Gamma_n : a^+ = 1, a \to b, b^- \ge 2\}$. These are the words $a = a_0 c$ that can be extended to the right in a unique way, say $a_0 c a_{n+1}$, but $b = c a_{n+1}$ is left-special. We have

$$\#V_2 \le \sum_{b^-\ge 2} b^- = \sum_{b^-\ge 2} (b^- - 1) + \sum_{b^-\ge 2} 1 \le 2(p(n+1) - p(n)).$$

Now every $a \in \Gamma_n \setminus V_1 \cup V_2$ is right-special, and if $a \to b$, then b is left-special. That means that a and b appear with the same frequency in infinite words $x \in X$. Every maximal path in $a \in \Gamma_n \setminus V_1 \cup V_2$ is succeeded by a vertex $v \in V_1 \cup V_2$, and no other such maximal path is succeeded by v. Therefore, the number of different frequencies is bounded by $\#(V_1 + V_2) \le 3(p(n+1) - p(n))$ as claimed. □

9. Automata

In this section we discuss some variations on the Turing machine, and ask the question what languages they can recogize 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.

Finite automata. A finite automoton (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 114. *The even shift (Example 12) 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 $f(q_i, a) = q_j$.*

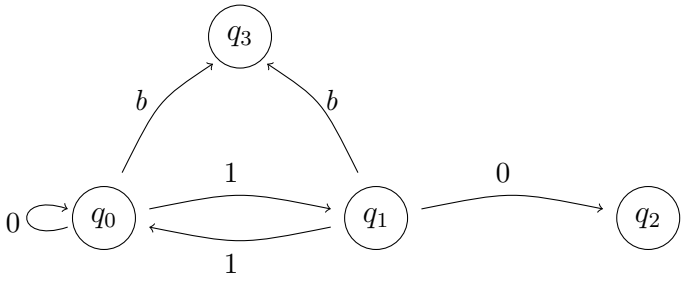

Figure 10. Transition graph for a finite automaton recognizing the even shift.

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 **determinisitc 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 [5, page 22]):

Theorem 115. *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 116. *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-labelled 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 arraws 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 115 there is also an DFA that recognizes $\mathcal{L}^R$. □

Sometimes it is easier, again for proofs or constructing compact examples, to allow finite automata to have transtions 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 117. *The follow automata accept the language $\mathcal{L} = \{0^k 1^l 2^m : k, l, m \geq 0\}$, see Figure 11. 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 [5, page 22]):

Theorem 118. *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.*

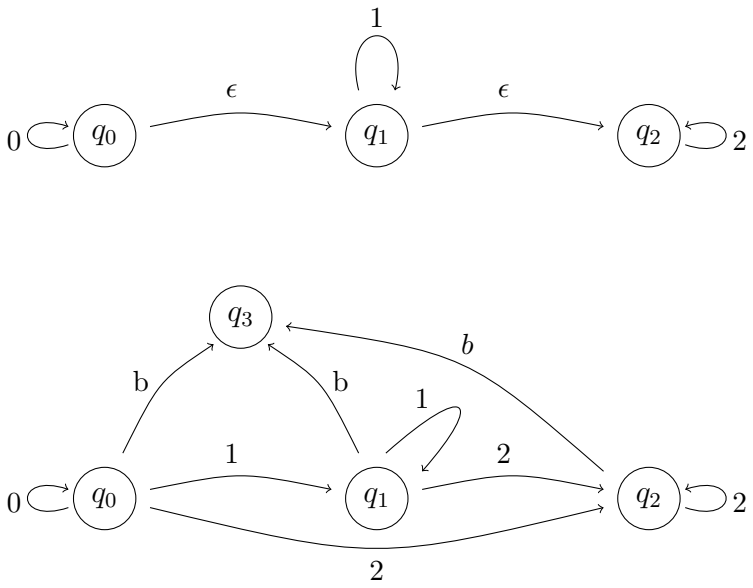

FIGURE 11. Finite automata recognizing $\mathcal{L} = \{0^k 1^l 2^m : k, l, m \geq 0\}$.

10. 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**.

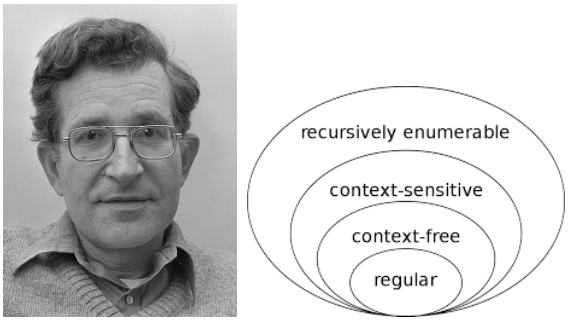

FIGURE 12. Noam Chomsky in 1977 and his hierarchy.

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

(11)
a small yellow mouse chases a big green cat
a high low red decency eats a orange book

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

$$\begin{aligned}
V &= \text{collection of } \textbf{variables} \text{ to which production rules can be applied.}\\
T &= \text{collection of } \textbf{terminals} \text{ which remain unchanged.}\\
P &= \text{collection of } \textbf{production rules} \text{ to replace variable with strings in } V\cup T.\\
S &= \text{a special variable, called the } \textbf{starting symbol}.
\end{aligned}$$

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

$$\begin{aligned} \text{regular languages (RL)} &\subset \text{context-free languages (CFL)} \\ &\subset \text{context sensitive languages (CSL)} \\ &\subset \text{recursively enumerable languages (ER)} \end{aligned}$$

although there are also *unrestricted gammars*, which is a wider class still.

Regular grammars. The **regular grammars** can be brought in a form where the production rules are one of the following types:

left-linear	ór	**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 119. *The even shift (Example 12) 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 120. *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 $w = w_1 \dots w_n \in \mathcal{L}$ (soo Corollary 116), and then taking the right-linear grammar for this reverse language $\mathcal{L}^R$. Then rewite every production rule $A \to wB$ to $A \to Bw$ to obtain a left-linear grammar that accepts exactly the original $\mathcal{L}$. □

Context-free grammars. The second sentence in (11) 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 121. *Consider the language $\mathcal{L} := \{01^n2^n : n \geq 1\}$. That is, every maximal block of 1s is succeeded by an equally long word of 2s.*

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 vertex-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

$$01^{n+(k-1)m}2^n = 01^a \underbrace{1^m1^m1^m\ldots1^m}_{\textit{the } m\textit{-loop } k \textit{ times}} 1^b2^n$$

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 grammers, 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 121):

Lemma 122 (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 121. Note that $N \leq \#\{\text{vertices in } \mathcal{G}\}$. □

Exercise 123. *Let $\mathcal{L} = \{01^{n^2} : n \geq 1\}$. Show that $\mathcal{L}$ is not a regular language. Is it context-free?*

Exercise 124. *Using the Pumping Lemma 122 to show that there are β-shifts X_β that are not regular.*

Lemma 125 (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 [5, Chapter 6]. □

Corollary 126. *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 125. If $\mathcal{L}(x)$ were context-free, then $rs^k tu^k v \in \mathcal{L}(x)$ as well. But then the limit frequency of 1s is

$$\lim_{k\to\infty} \frac{|rs^k tu^k v|_1}{|rs^k tu^k v|} = \frac{|su|_1}{|su|} \in \mathbb{Q},$$

contradicting that Sturmian sequences have irrational frequencies. □

From the shape of its production rules, it is clear that the language of Example 121 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 colour) which is different from $\mathcal{A}$ and a starting stack symbol $Z_0 \in \mathcal{B}$ (the colour 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 colours in $\mathcal{B}$) or removing them. The resultingautomaton with stack is called a **push-down automaton**.

Theorem 127. *A language is (not more complicated than) context-free if and only if it is recognized by a push-down automaton.*

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 128. *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 129. *Consider the language $\mathcal{L} = \{1^n 2^n 3^n : n \geq 1\}$. Pumping Lemma 125 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

$$\begin{aligned} S &\to 012 \\ S &\to 00A12 \\ A1 &\to 1A \\ 1A2 &\to 1122 \\ 1A2 &\to 11B22 \\ 1B &\to B1 \\ 0B1 &\to 00A1 \end{aligned}$$

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 130. *The following set of productions produces the language $\mathcal{L} = \{1^{2^n} : n \geq 0\}$, that is: strings of* 1*s of length equal to a power of* 2*.*

$$\begin{aligned} S &\to AC1B & \qquad 1D &\to D1 \\ C1 &\to 11C & AD &\to AC \\ CB &\to DB & 1E &\to E1 \\ CB &\to E & AE &\to \epsilon \end{aligned}$$

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* [5, 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 131. *A language is (not more complicated than) context-sensitive if and only if it is recognized by a linearly bounded non-deterministic Turing machine.*

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 132. *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 $\alpha A \beta \to \alpha\gamma\beta$ $\emptyset \neq \gamma \in (V \cup T)^*$)	$\{a^{2^n} : n \geq 0\}$
recursively enumerable	Turing machine	$\alpha \to \beta$ (no restrictions)	

References

[1] C. Apparicio, *Reconnaissabilité des substitutions de longueur constante*, Stage de Maitrise de lENS Lyon, 1999.

[2] *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

[3] F. Durand, B. Host, C. Skau, *Substitution dynamical systems, Bratteli diagrams and diemnsion groups,* Ergod. Th. Dynam. Sys. **19** (1999), 953–993.

[4] G. Hedlund, M. Morse, *Symbolic dynamic II: Sturmian trajectories,* Amer. J. Math. **66** 1940 1–42.

[5] John Hopcroft and Jeffrey Ullman, *Introduction to Automata Theory, Languages and Computation*, Addision-Wesley Publ. ISBN 0-201-02988-X

[6] K. Jacobs, M. Keane, *0-1-sequences of Toeplitz type* Z. Wahrscheinlichkeitstheorie verw Gebiete **13** (1969) 123–131. Doi:10.1007/BF00537017

[7] B. Kitchens, *Symbolic dynamics: one-sided, two-sided and countable state Markov shifts*, Springer Verlag. ISBN 3-540 -62738-3

[8] P. Kůrka, *Topological and Symbolic Dynamics*, Cours Spécialisés **11** Société Mathématique de France, Paris, 2003.

[9] D. Lind, B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge Univ Press, ISBN 0-521-55900-6

[10] B. Mossé, *Puissances de mots et reconnaissabilité des points fixes dune substitution*, Theoretical Computer Science, **99**, (1992), 327-334.

[11] B. Mossé, *Reconnaissabilité des substitutions et complexité des suites automatiques*, Bulletin de la Société Mathématique de France, **124**, (1996), 329-346.

[12] 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. Maudit and A. Siegel.

[13] Martine Queffélec, *Substitution dynamical systems and spectral analysis*, Lect. Notes in Math 1294, Springer-Verlag. ISBN 3-540-18692-1.