Chapter 3

Subshifts of Positive Entropy

Most of the subshifts of positive entropy are symbolic version of positive entropy dynamical systems of manifolds, for example dynamical system possessing a Markov partition, β-transformations or unimodal interval maps. Symbolically these correspond to β-shifts, kneading theory and subshifts of finite type (SFT) respectively, and the entropy is given by the exponential growth-rate of periodic points. We discuss also some subshifts that are not in first instance symbolic versions of other dynamical systems, such as density shifts, coded shifts, gap shifts and spacing shifts, and in some cases (such as power-free shift), the entropy is not related to periodic sequences at all.

3.1. Subshifts of Finite Type

Subshifts of finite type are the simplest and most frequently used subshifts in applications. They emerge naturally in hyperbolic dynamical systems such as toral automorphisms, Markov partitions of Anosov diffeomorphisms, Axiom A attractors (including Smale's horseshoe), but also in topological Markov chains.

3.1.1. Definition of SFTs and Transition Matrices and Graphs.

Definition 3.1. A **subshift of finite type (SFT)** is a subshift consisting of all strings avoiding a **finite** list of forbidden words as subwords.

For example, the Fibonacci SFT has 11 as forbidden word. Naturally, then also 110 and 111 are forbidden, but we take only the smallest collection of forbidden words. If $M+1$ is the length of the longest forbidden word, then this SFT is an **M-step SFT**, or an SFT with **memory** M. Indeed, an M-step SFT has the property that if $uv \in \mathcal{L}(X)$ and $vw \in \mathcal{L}(X)$, and $|v| \geq M$, then $uvw \in \mathcal{L}(X)$ as well. The following property is therefore immediate:

Lemma 3.2. *Every irreducible SFT is synchronized; in fact, every word of length M (the memory of the SFT) is synchronizing.*

Lemma 3.3. *Every SFT (X,σ) on a finite alphabet can be recoded such that the list of forbidden words consists of 2-words only.*

Proof. Assume that (X,σ) is a subshift over the alphabet $\mathcal{A}$ and the longest forbidden word has length $M+1 \geq 2$. Take a new alphabet $\tilde{\mathcal{A}} = \mathcal{A}^M$, say $a_1,\dots,a_n$ are its letters. Recode every $x \in X$ using a sliding block code π, where for each index i, $\pi(x)_i = a_j$ if a_j is the symbol used for $x_i x_{i+1} \dots x_{i+M-1}$. Effectively, this is replacing X by its M-block code. Then every $M+1$-word is uniquely coded by a 2-word in the new alphabet $\tilde{\mathcal{A}}$, and vice versa, every a_1a_2 such that the M-suffix of $\pi^{-1}(a_1)$ equals the M-prefix of $\pi^{-1}(a_2)$ encodes a unique $M+1$-word in $\mathcal{A}^*$. Now we forbid a 2-word

a_1a_2 in $\tilde{\mathcal{A}}^2$ if $\pi^{-1}(a_1a_2)$ contains a forbidden word of X. Since $\mathcal{B}$ is finite, and therefore $\mathcal{A}$ is finite, this leads to a finite list of forbidden 2-words in the recoded subshift. □

Example 3.4. 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 3.1; labels at the arrows indicate with word in $\{0,1\}^3$ they stand for. For example, the edge $a \to b$ labeled 001 has prefix $a = 00$ and suffix $b = 01$. Each arrow containing a forbidden word is dashed, and then removed in the right panel of Figure 3.1.

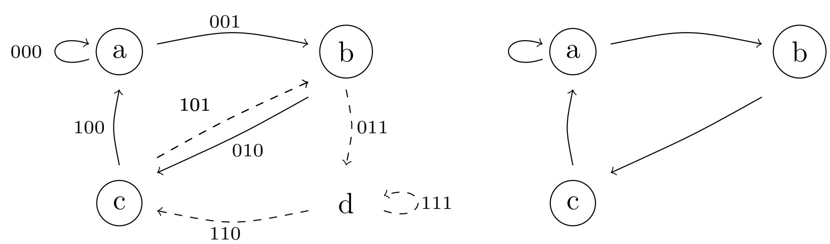


Figure 3.1. The recoding of the SFT with forbidden words 11 and 101.

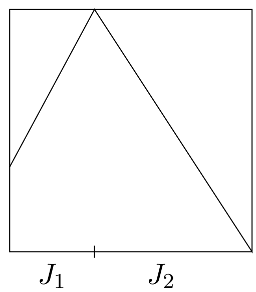
Corollary 3.5. Every SFT (X, σ) on a finite alphabet $\mathcal{A}$ can be represented by a finite graph $\mathcal{G}$ with vertices labeled by the letters in $\mathcal{A}$ and arrows $b_1 \to b_2$ only if $\pi^{-1}(b_1b_2)$ contains no forbidden word of X.

Definition 3.6. The directed graph $\mathcal{G}$ constructed in the previous corollary is called the **transition graph** of the SFT. The matrix $A = (a_{ij})_{i,j\in\mathcal{A}}$ with $a_{i,j} = \#\{\text{arrows } i \to j \text{ in } \mathcal{G}\}$ is its **transition matrix**. The graph is vertex-labeled, which means that each vertex is assigned symbol in the alphabet. We will stipulate throughout this book that the vertex-labels are unique (i.e., no two distinct vertices have the same label), although this assumption is not entirely uniform in the literature.

Example 3.7. Let $T : [0,1] \to [0,1]$ be the **piecewise monotone** map, i.e., there is a finite partition $\{J_i\}_{i\in\mathcal{A}}$ of $[0,1]$ into intervals such that $T|_{J_i}$ is continuous and monotone for each i. Assume also that for each i, $\overline{T(J_i)}$ is the closure of the union of J_k's. In this case we call $\{J_i\}_{i\in\mathcal{A}}$ a **Markov partition**. Write

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

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



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

$$\gamma = \tfrac{\sqrt{5}+1}{2}$$

Figure 3.2. The tent map with slope equal to the golden mean

The example in Figure 3.2 produces the transition matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, so the corresponding subshift is the Fibonacci SFT, see Example 1.3. It should not come as a surprise that the leading eigenvalue of A is exactly the slope of T: both equal to $e^{h_{top}(T)} = e^{h_{top}(\sigma)} = \gamma$, see Section 3.1.2.

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

Definition 3.8. A **toral automorphism** $T : \mathbb{T}^d \to \mathbb{T}^d$ is an invertible linear map on the (d-dimensional) torus $\mathbb{T}^d$. Each such T is of the form $T_M(x) = Mx \bmod 1$, where

- M is an integer matrix with $\det(M) = \pm 1$ (i.e., M is **unimodular**);
- the eigenvalues of M are not on the unit circle; this property is called **hyperbolicity**; for toral automorphisms, this is equivalent to $\mathbb{T}^d$ being a hyperbolic set in terms of Definition 2.79.

The map T_M has a **Markov partition**[1], that is a partition $\{J_i\}_{i \in \mathcal{A}}$ for sets such that

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

Every hyperbolic toral automorphism has a Markov partition, see [**96**], but in general they are fiendishly difficult to find explicitly, especially in dimension ≥ 3 where the boundaries of the J_i might have to be fractal, see [**100**]. Therefore we confine ourselves to the simpler case of $M = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$; it

[1]The construction of Markov partitions for toral automorphisms on $\mathbb{T}^2$ goes back to Berg [**56**] and Adler & Weiss [**10**], extended to more general settings in [**95**, **96**, **255**, **496**] among others.

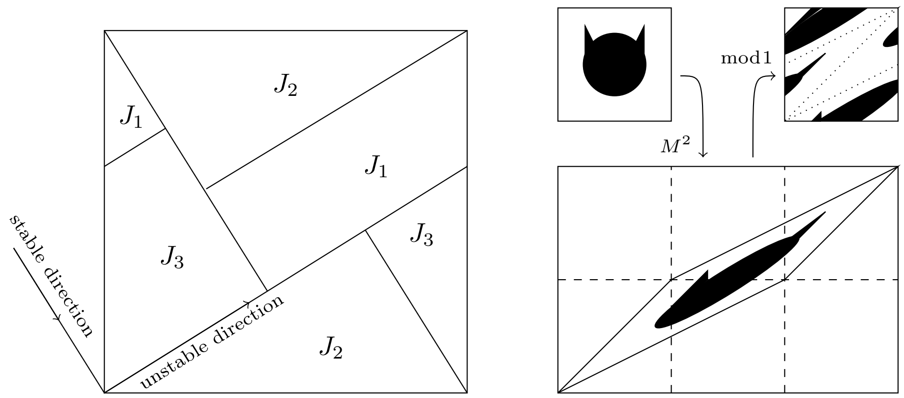


Figure 3.3. The Markov partition for $T_M : \mathbb{T}^2 \to \mathbb{T}^2$; the catmap is T_M^2.

has Markov partition of three rectangles J_i for $i = 1, 2, 3$ can be constructed, see Figure 3.3. The corresponding transition matrix is

$$A = (a_{i,j}) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \text{ where } a_{ij} = \begin{cases} 1 & \text{if } T_M(J_i^\circ) \cap J_j \neq \varnothing, \\ 0 & \text{if } T_M(J_i^\circ) \cap J_j = \varnothing. \end{cases}$$

The characteristic polynomial of A is

$$\begin{aligned} \det(A - \lambda I) &= -\lambda^3 + 2\lambda + 1 \\ &= -(\lambda+1)(\lambda^2 - \lambda - 1) = -(\lambda + 1)\det(M - \lambda I). \end{aligned}$$

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

Example 3.9. The most "famous" toral automorphism is Arnol'd's catmap, and it has the matrix $\binom{2\ 1}{1\ 1} = \binom{1\ 1}{1\ 0}^2$, see Figure 3.3 (right). It is called catmap because Arnol'd used this example, including the drawing of a cat's head, in his book(s) [**29**] to illustrate the nature of hyperbolic maps.

Exercise 3.10. Show that if $x \in \mathbb{T}^d$ has only rational coordinates, then x is periodic under a toral automorphism. Conclude that, if the pixels in Figure 3.3 have rational coordinates (such as the dyadic coordinates that computers use), then the cat will return intact after a finite number of iterates.

The following characterization for shadowing subshifts is due to Walters [**532**] (see also [**369**, Theorem 3.33]).

Theorem 3.11. *A subshift (X, σ) has the shadowing property if and only if it is a subshift of finite type.*

Proof. We give the proof for $X \subset \mathcal{A}^{\mathbb{N}_0}$ only; the two-sided case follows in a similar way.

$\Leftarrow$: Let (X, σ) be an SFT of memory M (see below Definition 3.1) so $M+1$ is the length of the longest forbidden word. Let $\varepsilon > 0$ be arbitrary and choose $m \geq M+1$ so small that $2^{-m} < \varepsilon$. Take $\delta = 2^{2-m}$. We need to show that every δ-pseudo-orbit $(x^n)_{n\geq 0} \subset X$, that is,

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

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

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

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

$\Rightarrow$: Let (X, σ) be a subshift with the shadowing property, so in particular, for $\varepsilon = 1$, there exists $\delta > 0$ such that every δ-pseudo-orbit in X is ε-shadowed in X. Take $N \in \mathbb{N}$ such that $2^{2-N} < \delta$, and let $y \in \mathcal{A}^{\mathbb{N}_0}$ be such that $y_n \dots y_{n+N-1} \in \mathcal{L}(X)$ for each n. Then there exists a sequence $(x^n)_{n\geq 0}$ such that $x_0^n \dots x_{N-1}^n = y_n \dots y_{n+N-1}$ for each $n \geq 0$. Therefore

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

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

3.1.2. Topological Entropy of SFTs.

Definition 3.12. A non-negative $N \times N$ matrix $A = (a_{ij})_{i,j\in\mathcal{A}}$ is called **irreducible** if for every $i, j \in \mathcal{A}$ there is k such that A^k has (i,j)-entry $a_{ij}^{(k)} > 0$. For index i, set $per(i) = \gcd(k > 1 : a_{ii}^{(k)} > 0)$. If A is irreducible, then $per(i)$ is the same for every i, and we call it the **period** of A. We call A **aperiodic** if its period is 1. The matrix is called **primitive** if there is $k \in \mathbb{N}$ such that $a_{ij}^{(k)} > 0$ for all $i, j \in \mathcal{A}$.

Exercise 3.13. Show that if A is aperiodic and irreducible, then A is primitive, but irreducibility or aperiodicity alone doesn't imply primitivity. Conversely, if A is primitive, then it is also aperiodic and irreducible. If A is irreducible, show that $per(i)$ is indeed independent of i.

Theorem 3.14. *The topological entropy of an irreducible SFT equals* $\log \lambda$ *where* λ *is the leading eigenvalue of the transition matrix.*

Proof. Let $A^n = (a_{ij}^{(n)})_{i,j\in\mathcal{A}}$ be the n-th power of the transition matrix A. Every word in $\mathcal{L}_n(X)$ corresponds to an n-path in the transition graph, and the number of n-paths from i to j is given by $p_{ij}^{(n)}$. From the Perron-Frobenius Theorem 8.57 we can derive that there is $C > 0$ such that for all $n \in \mathbb{N}$ there are $i, j \in \mathcal{A}$ such that

$$C^{-1}\lambda^n \leq a_{ij}^{(n)} \leq C\lambda^n. \tag{3.1}$$

It follows that $C^{-1}\lambda^n \leq p(n) \leq (\#\mathcal{A})^2 C\lambda^n$ and $\lim_n \frac{1}{n}\log p(n) = \log\lambda$. If A is periodic (so with the irreducibility assumption already made, A is primitive), then (3.1) holds for every $i, j \in \mathcal{A}$. Also $\lambda > 1$ unless $A = (1)$. $\square$

Proposition 3.15. If (Y, σ) is a factor of (X, σ), then $h_{top}(Y, \sigma) \leq h_{top}(X, \sigma)$. If (X, σ) and (Y, σ) are conjugate, then $h_{top}(X, \sigma) = h_{top}(Y, \sigma)$.

The result also holds in general, i.e., not just in the context of subshifts, see Corollary 2.51, but using the word-complexity and sliding block codes, the proof is particularly straightforward here.

Proof. Let $\psi : X \to Y$ be the factor map. Since it is continuous, it is a sliding block code by Theorem 1.23, say of window length $2N+1$. Therefore the word complexities relate as $p_Y(n) \leq p_X(n+2N)$, and hence

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

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

As shown by Parry [**432**], see Theorem 6.66, irreducible SFTs are intrinsically ergodic. This follows also from Theorem 3.48 and Proposition 3.41. Weiss [**536**] showed that factors of irreducible SFTs are intrinsically ergodic as well.

3.1.3. Vertex-Splitting and Conjugacies between SFTs: It is natural to ask which SFTs are conjugate to each other. We have seen that having equal topological entropy is a necessary condition for this, but it is not sufficient. The conjugacy problem for SFTs was solved by Williams and in this section we discuss the ingredients required for this result. Complete details can be found in [**352**, **386**].

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

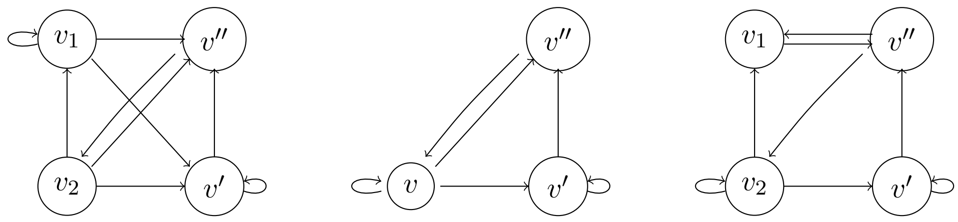


Figure 3.4. Insplit graph Original $\mathcal{G}$ Outsplit graph

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

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

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

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

Definition 3.17. Let $\mathcal{G} = (V, E)$, and assume that $\#E_v \geq 2$. An **elementary outsplit graph** $\hat{\mathcal{G}} = (\hat{V}, \hat{E})$ is obtained by

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

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

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

Proposition 3.18. Let $\hat{\mathcal{G}}$ be an in- or outsplit graph obtained from $\mathcal{G}$. Then the edge-labeled subshift $\hat{X}$ of $\hat{\mathcal{G}}$ and the edge-labeled subshift X of $\mathcal{G}$ are mutually semi-conjugate to each other.

Proof. We give the proof for an elementary outsplit $\hat{\mathcal{G}}$; the general outsplit and (elementary) insplit graph follow similarly. By Theorem 1.23, it suffices to give sliding block code representations for $\pi : \hat{X} \to X$ and $\hat{\pi} : X \to \hat{X}$.

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

This concludes the proof. In general, mutual semi-conjugacy is not enough to conclude conjugacy (it is not given that $\hat{\pi} = \pi^{-1}$), but in this situation, conjugacy holds, see Theorem 3.24. $\square$

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 $\hat{N} = \#\hat{V}$ and $N = \#V$. Then there is an $N \times \hat{N}$-matrix $D = (d_{v,\hat{v}})_{v \in V, \hat{v} \in \hat{V}}$ where $d_{v,\hat{v}} = 1$ if $\hat{v}$ replaces v. (Thus D is a sort of rectangular diagonal matrix.)

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

Proposition 3.19. With the above notation,

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

Sketch of proof. Prove it first 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 3.4.

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

Also

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

Matrix multiplications confirms that $DC = A$ and $CD = \hat{A}$. $\square$

Exercise 3.20. Do the same for the elementary insplit graph in the example of Figure 3.4.

Definition 3.21. Two matrices A and $\hat{A}$ are **strongly shift equivalent (of lag ℓ)** (denoted as $A \approx \hat{A}$) if there are (rectangular) matrices D_i, C_i and A_i, $1 \le i \le \ell$ over $\mathbb{N}_0$ such that

$$A = A_0, \quad A_{i-1} = D_i C_i, \quad C_i D_i = A_i, \quad i = 1, \dots, \ell, \quad A_\ell = \hat{A}. \tag{3.2}$$

Remark 3.22. One important restriction of this definition is that the conjugating matrices must have **non-negative** integer entries. Even if a square matrix has determinant ± 1, its inverse may still have negative integers among its entries. For example

$$A = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \hat{A} = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$$

are similar via $\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right) A = \hat{A} \left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$. From this, we can easily compute that the traces $\operatorname{tr}(A^n) = \operatorname{tr}(\hat{A}^n)$ for all $n \in \mathbb{Z}$, so A and $\hat{A}$ share **$\boldsymbol{\zeta}$-functions** $\zeta_A(t) := \exp(\sum_{n=0}^\infty \operatorname{tr}(A^n))$. However, A and $\hat{A}$ are not (strongly) shift equivalent. This is Williams' [**539**, Example 3] counter-example to Bowen's question whether sharing ζ-functions for SFTs suffices to conclude conjugacy.

Exercise 3.23. Show that strong shift equivalence $\approx$ is indeed an equivalence relation between nonnegative square matrices. Show that $A \approx \hat{A}$ implies that A and $\hat{A}$ have the same leading eigenvalue $\lambda = \hat{\lambda}$.

Strongly shift equivalence between matrices A and $\hat{A}$ means, in effect, that their associated graphs $\mathcal{G}$ and $\hat{\mathcal{G}}$ can be transformed into each other by a sequence of elementary vertex-splittings and their inverses (vertex-mergers). Conjugacy between SFTs can always be reduced to vertex-splittings and vertex-mergers, as shown in Williams' Theorem [**539**] from 1973. The full proof is in [**352**, Chapter 2] and [**386**, Chapter 7, specifically Theorem 7.2.7].

Theorem 3.24. *Two SFTs are conjugate if and only if their transition matrices are strongly shift equivalent.*

Strong shift equivalence $A \approx \hat{A}$ may be a complete invariant for conjugacy between edge-labeled SFTs X_A and $X_{\hat{A}}$. In practice it is difficult to check if $A \approx \hat{A}$. Even if A and $\hat{A}$ have the same characteristic polynomial, they need not be strongly shift equivalent. The following weaker notion may help:

Definition 3.25. Two matrices A and $\hat{A}$ are **shift equivalent (of lag ℓ)** (denoted as $A \sim_\ell \hat{A}$) if there are matrices C, D over $\mathbb{N}_0$ such that

$$A^\ell = CD, \quad \hat{A}^\ell = DC \qquad \text{and} \qquad AC = C\hat{A}, \quad \hat{A}D = DA. \tag{3.3}$$

Said differently, the following diagram commutes:

$$\begin{tikzcd}
\mathbb{Z}^n \arrow[r, "A"] \arrow[d, "D"'] & \mathbb{Z}^n \arrow[r, "A^{\ell-1}"] \arrow[d, "D"'] & \mathbb{Z}^n \arrow[r, "A"] & \mathbb{Z}^n \arrow[d, "D"'] \\
\mathbb{Z}^{\hat n} \arrow[r, "\hat A"] \arrow[urr, "C"] & \mathbb{Z}^{\hat n} \arrow[rr, "\hat A^{\ell}"] \arrow[urr, "C"] & & \mathbb{Z}^{\hat n}
\end{tikzcd}$$

Shift equivalence means that the ℓ-th powers A^ℓ and $\hat A^\ell$ are strong shift equivalent (with lag 1). Shift equivalence is easier to verify than strong shift equivalence, although verification can still be very complicated. But, and this is Williams' Conjecture, it is still not fully[2] known if it is a complete invariant, see [**386**, Section 7.3] and [**102**, Problem 19.1]. If $A \not\sim \hat A$, then X_A and $X_{\hat A}$ cannot be conjugate, but if $A \sim \hat A$, this is insufficient to conclude that (X_A, σ) and $(X_{\hat A}, \sigma)$ are conjugate.

Exercise 3.26. Show that (i) $A \sim_\ell \hat A$ implies $A \sim_k \hat A$ for all $k \geq \ell$, (ii) shift equivalence $\approx$ is an equivalence relation between nonnegative square matrices, and (iii) strong shift equivalence implies shift equivalence, with the same value of ℓ.

Shift equivalence matrices have the same ζ-function, and many other properties coincide too.

Lemma 3.27. *If A and $\hat A$ are shift equivalent (of lag ℓ), then they have the same nonzero eigenvalues (so also $h_{top}(X_A, \sigma) = h_{top}(X_{\hat A}, \sigma)$).*

Proof. We have $A^n C = C \hat A^n$ and $DA^n = \hat A^n D$ for all $n \geq 0$. By linearity, $q(A) \cdot C = C \cdot q(\hat A)$ and $D \cdot q(\hat A) = q(A) D$ for every polynomial. If q is the characteristic polynomial of A (so $q(A) = 0$ by the Cayley-Hamilton Theorem), then $0 = D \cdot q(A) \cdot C = \hat A^\ell \cdot q(\hat A)$. Thus $\hat A$ has no other eigenvalues that those of A, possibly plus 0. On the other hand, if q is the characteristic polynomial of $\hat A$, then $0 = C \cdot q(\hat A) \cdot D = q(A) \cdot A^\ell$, so A has the eigenvalues of $\hat A$, with the possible exception of 0.

Since $h_{top}(X_A, \sigma) = \log \lambda_A$ for the leading eigenvalue λ_A of A, the entropies are the same too. $\square$

In order to say what can be proved with shift equivalence, we define SFTs (X_A, σ) and $(X_{\hat A}, \sigma)$ to be **eventually conjugate** if the n-block shifts are conjugate for all sufficiently large n. Then, see [**386**, Theorem 7.5.15]:

[2]Kim & Roush [**351**] gave a negative answer, but only for reducible matrices.

Theorem 3.28. *Two SFTs (X_A, σ) and $(X_{\hat{A}}, \sigma)$ are eventually conjugate if and only if A and $\hat{A}$ are shift equivalent.*

There remain many open (classification) problems in SFT, as well as in sofic and other subshifts. The survey of Boyle [**102**] contains a long list of open problems, many of which remain open to today.

3.2. Sofic Shifts

Sofic shifts are shifts that can be described by finite of **edge-labeled** (rather than vertex-labeled as needed for SFT) transition graphs. The word sofic was coined by Benji Weiss; it comes from the Hebrew word for "finite". Much of this section can be found in concise form in [**352**, Section 6.1].

Definition 3.29. A subshift (X, σ) is called **sofic** if it is the space of paths in an edge-labeled graph. Other than with the vertex-labeling, in this edge-labeling, more than one edge is allowed to have the same symbol.

Lemma 3.30. *Every SFT is sofic.*

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

Remark 3.31. Not every sofic shift is an SFT. For example the even shift (Example 1.4) 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 following theorem shows that we can equally define the sofic subshifts as those that are a factor of a subshift of finite type.

Theorem 3.32. *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. For this, we can take the dual graph in which the edges of $\mathcal{G}'$ are the vertices, and $a \to b$ if an only if a labels the incoming edge and b the outgoing edge of the **same** vertex in $\mathcal{G}'$. Now $\pi : X' \to X$ is given by $\pi(x)_i = a$ if a is the label in $\mathcal{G}$ of the same edge that is labeled x_i in $\mathcal{G}'$. This π is clearly a sliding block code, so by Theorem 1.23, π is continuous and commutes with the shift.

$\Leftarrow$: If X is a factor of an SFT, then the factor map is a sliding block code by Theorem 1.23, say of window size $2N+1$: $\pi(x)_i = f(x_{i-N}, \dots, x_{i+N})$. Represent the SFT by an edge-labeled graph $\mathcal{G}'$ where the labels are the $2N+1$-words $w \in \mathcal{L}_{2N+1}(X)$. These are all distinct. The factor map turns $\mathcal{G}'$ into an edge-labeled graph $\mathcal{G}$ with labels $f(w)$. Therefore X is sofic. $\square$

Corollary 3.33. Every factor of a sofic shift is again a sofic shift. Every shift conjugate to a sofic shift is again sofic.

In fact, a sofic shift with an irreducible transition matrix are always transitive, has a dense set of periodic points, and is mixing if and only if it is totally transitive, see [**45**, Theorem 3.3].

3.2.1. Follower sets. A further characterizations of sofic shifts relies on the following notion.

Definition 3.34. 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 3.35. Let X_{even} be the even shift from Example 1.4. Then $\mathcal{F}(0) = \mathcal{L}(X_{\text{even}})$ because a 0 casts no restrictions on the follower set. Also $\mathcal{F}(011) = \mathcal{L}(X_{\text{even}})$, but $\mathcal{F}(01) = 1\mathcal{L}(X) = \{1w : w \in \mathcal{L}(X)\}$. Although each follower set is infinite, there are only these two distinct follower sets. Indeed, $\mathcal{F}(v0) = \mathcal{F}(0)$ for every $v \in \mathcal{L}(X)$, and $\mathcal{F}(v0111) = \mathcal{F}(v01)$, $\mathcal{F}(v01111) = \mathcal{F}(v011)$, etc. The follower set $\mathcal{F}(1)$ is not properly defined, but we can ignore this.

The following theorem, appearing in [**536**], is in fact a consequence of the Myhill-Nerode Theorem [**415**, **417**].

Theorem 3.36. *A subshift (X, σ) is sofic if and only if the collection of its follower sets is finite.*

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

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

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

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

Definition 3.37. An edge-labeled transition graph $\mathcal{G}$ is **right-resolving** if for each vertex $v \in \mathcal{G}$, the outgoing arrows all have different labels. It is called **follower-separated** if for each vertex $v \in \mathcal{G}$, the follower set (i.e., the set of labeled words associated to paths starting in v) is different from the follower set of every other vertex.

Every sofic shift has a right-resolving follower-separated graph representation and if we minimize the number of vertices in such graph, there is only one such graph with these properties. In fact, the follower set representation $\mathcal{G}$ constructed in the first half of the proof of Theorem 3.36 is both right-resolving, follower-separated and of smallest size. The latter two properties follow by the choice of V. To see the former, assume that $v \in V$ and $v \to w$ and $v \to w'$ have the same label a. This implies that

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

so $w = w'$.

Corollary 3.38. Every transitive sofic shift X is synchronized, and (unless it is a single periodic orbit) has positive entropy. In fact, $h_{top}(X) = \log \lambda_A$, where λ_A is the leading eigenvalue of the transition graph of the minimal right-resolving representation of X.

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

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

The positive entropy follows from Theorem 1.20 or Corollary 3.47. In fact, since $\mathcal{G}$ is right-resolving, there is an at most $\#V$-to-one correspondence between n-paths starting in $\mathcal{G}$ and words in $\mathcal{L}_n(X)$. Therefore $\#\{n\text{-paths}\} \leq p_X(n) \leq \#V \cdot \#\{n\text{-paths}\}$, and we can use Theorem 3.14. $\square$

Remark 3.39. Irreducible sofic shifts are intrinsically ergodic, see [**536**] and Theorem 3.48.

3.3. Coded Subshifts

Rather than forbid words to appear, as one does in SFTs, we can prescribe which words need to be used, and then these words can be concatenated freely. This type of subshift was first described by Blanchard & Hansel [**80**].

Definition 3.40. A **coded subshift** $(X_{\mathcal{C}}, \sigma)$ is the closure of the collection of free concatenations of a finite or countable collection $\mathcal{C}$.

Of course, this doesn't mean that concatenations of words in $\mathcal{C}$ are the only words in the language $\mathcal{L}(X)$. For example, if $\mathcal{C} = \{10, 01\}$, then $00 \in \mathcal{L}(X) \setminus \mathcal{C}^*$.

Proposition 3.41. Every transitive SFT is a coded shift.

For example, the Fibonacci SFT of Example 1.3 and the even shift of Example 1.4 are both coded subshift, with sets of code words $\mathbb{C} = \{0, 01\}$ and $\mathbb{C} = \{0, 01\}$, respectively. On the other hand, the SFT (X_A, σ) on the alphabet $\{0, 1\}$ with transition matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not transitive, and also not a coded shift, because no code word containing 01 can every be used twice in a concatenation.

Proof. Rewrite the SFT to an SFT with memory $M = 1$, i.e., all forbidden words have length ≤ 2. Let $\mathcal{G}$ be the transition graph; since the SFT is transitive, $\mathcal{G}$ is strongly connected. Fix vertices a, b such that the arrow $a \to b$ occurs in $\mathcal{G}$. Now let S contain the codes of all finite paths $b \to \cdots \to a$; these can be freely concatenated. $\square$

Remark 3.42. Naturally, the set $\mathcal{C}$ of codes may not be the most economical, but the idea of the proof of Proposition 3.41 is quite general. It can also be used to show that sofic and synchronized subshifts are coded. Therefore we have the inclusion.

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

All these inclusions are strict. For instance, Dyck shifts are coded but not synchronized, see Section 3.10. Coded shifts are always transitive, but not always totally transitive; indeed, if the lengths of all code words is a multiple of $N \geq 2$, then σ^N can easily be non-transitive (but not necessarily, see [**177**, Theorem 4.1]). Totally transitive coded subshifts are always weak-mixing (since they have a dense set of periodic orbit, see [**316**, Corollary 3.6]), and also topologically mixing, see [**177**, Theorem 2.2]. Thus for coded systems, these three notions coincide.

It is useful to make some distinction between sequences that are the concatenations of "short" words

$$V_{\mathcal{C}} = \{x \in \mathcal{A}^{\mathbb{Z}} : \exists\, (s_k)_{k\in\mathbb{Z}} \subset \mathbb{Z} \text{ such that } x_{s_k} \dots x_{s_{k+1}-1} \in \mathcal{C}\}, \tag{3.4}$$

and sequences for which every finite subwords appear as subwords of "long" words in $\mathcal{C}$:

$$U_{\mathcal{C}} = \{x \in \mathcal{A}^{\mathbb{Z}} : \forall\, k \in \mathbb{N}, x_{-k} \dots x_k \text{ is a subword of some word in } \mathcal{C}\}. \tag{3.5}$$

We have $X_{\mathcal{C}} = \overline{V_{\mathcal{C}}} \supset U_{\mathcal{C}}$.

Example 3.43. The odd shift X_{odd} (recall from Example 1.4 that in this subshift, blocks of 0s have odd lengths) is a coded shift with collection of code words

$$\mathcal{C} = \{1, 10, 1000, 10000, \dots, 10^{2n-1}, \dots\}.$$

The sequence $\dots 010101010 \dots$ belongs to $V_{\mathcal{C}}$ but not to $U_{\mathcal{C}}$. On the other hand, $\dots 000000 \dots$ belongs to $U_{\mathcal{C}}$ but not to $V_{\mathcal{C}}$. The sequence $\dots 0001000 \dots$ belongs to neither, but lies in the closure $\overline{V_{\mathcal{C}}}$ (but not in $\overline{U_{\mathcal{C}}}$, in fact $\overline{U_{\mathcal{C}}} = U_{\mathcal{C}} = \{0^\infty\}$).

One can view coded shifts by means of (infinite) edge-labeled transition graphs $\mathcal{G}_{\mathcal{C}}$, with a central vertex v_0 from which loops of length ℓ emerge. Here

$$q_\ell = \#\mathcal{C}_\ell \quad \text{ for } \mathcal{C}_\ell := \{C \in \mathcal{C} : |C| = \ell\}.$$

The theory of infinite Markov graphs, as summarized in Section 8.7, should then be applicable. In particular, $\limsup_\ell \frac{1}{\ell} \log q_\ell = h_G(\mathcal{G})$ is the Gurevich entropy, see Definition 8.62. According to Theorem 8.71, unless $\mathcal{G}_{\mathcal{C}}$ is transient, the topological entropy ought to be the leading root of

$$F(h) := \sum_\ell q_\ell e^{-\ell h} = 1. \tag{3.6}$$

This is indeed true in many cases, see e.g. Examples 3.43 (see Exercise 3.114) and Example 3.49 below, but there are two problems.

First, the space of paths on $\mathcal{G}_{\mathcal{C}}$ can multiply code the points in $X_{\mathcal{C}}$, leading to an overestimate of the entropy. We call $x \in X_{\mathcal{C}}$ **recognizable** or **uniquely decipherable** if the sequence (s_k) in (3.4) is unique. The collection of code words $\mathcal{C}$ has the **unique decomposition property** if every finite word $w \in \mathcal{L}(X_{\mathcal{C}})$ can be decomposed in at most one way into words of $\mathcal{C}$.

Example 3.44. Let $\mathcal{C} = \{0, 10, 100\}$, then $X_{\mathcal{C}}$ is the Fibonacci SFT, but clearly the word 100 is superfluous here, since it is the concatenation of the first two. Thus $X_{\mathcal{C}}$ is neither uniquely decipherable nor has the unique decomposition property The entropy is not the logarithm of the silver mean as (3.6) would suggest, but truly the logarithm of the golden mean.

Let $\mathcal{C} = \{1010, 0100\}$. Then $X_{\mathcal{C}}$ doesn't have the unique decomposition property because

$$\underbrace{0100}\underbrace{10\ } = \underbrace{\ 010}\underbrace{010\ }.$$

However, if this word is extended by one symbol (either on the left or on the right), then the decomposition is unique. Therefore $X_{\mathcal{C}}$ is uniquely decipherable.

Let $\mathcal{C} = \{10, 00, 01\}$. In this case, every word containing 11 is uniquely decipherable, and all other words can be deciphered in exactly two ways, e.g.

$$\cdots \underbrace{01}\underbrace{00}\underbrace{00}\underbrace{10}\underbrace{10}\underbrace{01}0 \cdots = \cdots 0\underbrace{10}\underbrace{00}\underbrace{01}\underbrace{01}\underbrace{00}\underbrace{10} \cdots$$

Formula (3.6) suggests that the topological entropy $h_{top}(\sigma) = \frac{1}{2}\log 3$, and this is indeed true.

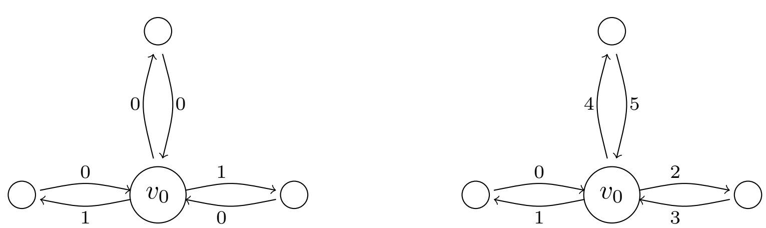


Figure 3.5. The edge-labeled transition graphs of $X_{\mathcal{C}}$ and $X_{\tilde{\mathcal{C}}}$.

We see this by considering $X_{\tilde{\mathcal{C}}}$ for $\tilde{\mathcal{C}} = \{01, 23, 45\}$. These have isomorphic transition graphs (with isomorphic path spaces), see Figure 3.5, but the latter is clearly uniquely decipherable with entropy $\frac{1}{2}\log 3$. Since $(X_{\mathcal{C}}, \sigma)$ is a factor of $(X_{\tilde{\mathcal{C}}}, \sigma)$ via the sliding block-code $\pi : 0 \to 0,\ 1 \to 1,\ 2 \to 1,\ 3 \to 0,\ 4 \to 0,\ 5 \to 0$. Since $\pi : X_{\mathcal{C}} \to X_{\tilde{\mathcal{C}}}$ is at most 2-to-1, π doesn't decrease entropy.

The second problem is that there may not be a good correspondence between the number of loops of length ℓ and the number of subwords of length ℓ. The solution of (3.6) can then underestimate the true value of the entropy, and indeed $h_G(\mathcal{G}) \leq h_{top}(X_\mathcal{C})$. A crude example of this is

$$\mathcal{C} = \{01, 00011011, 000001010011100101110111, \dots\},$$

i.e., the n-th code word is a concatenation of all words in $\{0,1\}^*$ of length n. Then $q_\ell = 1$ if $\ell = n2^n$ and $q_\ell = 0$ otherwise. Since every word appears in $X_\mathcal{C}$, the true entropy is $h_{top}(X_\mathcal{C}) = \log 2$, but (3.6) yields

$$e^{-2h}+e^{-8h}+e^{-24h}+e^{-128h}+\dots = 1, \qquad \text{which gives } h = \log 1.1809\dots < \log 2.$$

Hence, knowing the numbers q_ℓ of length ℓ code words is insufficient to decide on the entropy. Pavlov [**436**] suggests to use the n-subwords W_n inside code words instead. The exponential growth-rate of their number is $\lim_n \frac{1}{n} \log \# W_n = h(U_\mathcal{C})$.

Theorem 3.45. [**436**, Theorems 1.7 and 1.8] *Recall from* (3.6) *that* $F(h) = \sum_\ell q_\ell e^{-\ell h}$.

(i) If $h > h(U_\mathcal{C})$ and $F(h) < 1$, then $h_{top}(X_\mathcal{C}) \leq h$.

(ii) Conversely, if $F(h) > 1$ and $\mathcal{C}$ has the unique decomposition property, then $h_{top}(X_\mathcal{C}) > h$.

Proof. (i) Let Pre_n and Suf_n denote the length n prefixes and suffixes of code words $C \in \mathcal{C}$. Note that $\mathrm{Pre}_n \cup \mathrm{Suf}_n \subset W_n$. Every word in $\mathcal{L}(X_\mathcal{C})$ can be written as concatenation of the one suffix, some code words and one prefix, and therefore

$$\mathcal{L}_n(X_\mathcal{C}) = W_n \cup \bigcup_{k=2}^{n} \bigcup_{\substack{n_1+\dots+n_k=n \\ n_i \geq 1}} \mathrm{Suf}_{n_1} \mathcal{C}_{n_2} \dots \mathcal{C}_{n_{k-1}} \mathrm{Pre}_{n_k},$$

where the inner union really runs over the concatenations of all words in the indicated sets. Note that if the concatenation starts with a full code word, then this counts as a suffix, and similarly if the concatenation ends with a full code. Therefore it is justified to assume that $n_i \geq 1$ for each i.

This gives

$$\#\mathcal{L}_n(X_\mathcal{C}) \leq \# W_n + \sum_{k=2}^{n} \sum_{\substack{n_1+\dots+n_k=n \\ n_i \geq 1}} \#\mathrm{Suf}_{n_1} \cdot q_{n_2} \cdots q_{n_{k-1}} \cdot \#\mathrm{Pre}_{n_k}.$$

Since $\lim_n \frac{1}{n} \log \# W_n = h(U_\mathcal{C})$, our assumption $h > h(U_\mathcal{C})$ implies that there is a constant K such that

$$\max\{\#\mathrm{Pre}_{n_k}, \#\mathrm{Suf}_{n_k}\} \leq \# W_n \leq K e^{nh}.$$

Therefore, setting $m = n_1 + n_k$,

$$\begin{aligned}
\#\mathcal{L}_n(X_\mathcal{C}) &\leq Ke^{nh} + \sum_{k=2}^{n} \sum_{\substack{n_1+\dots+n_k=n \\ n_i\geq 1}} K^2 e^{(n_1+n_k)h} \prod_{j=2}^{k-1} q_{n_j} \\
&= e^{nh}\left(K + K^2 \sum_{m=0}^{n}\sum_{k=2}^{n-m} \sum_{\substack{n_2+\dots+n_{k-1}=n-m \\ n_i\geq 1}} \prod_{j=2}^{k-1} q_{n_j} e^{-n_j h}\right),
\end{aligned}$$

where the empty product counts as 1. All the terms in the last sum are part of the expansion of $F(h))^{k-2} = \left(\sum_{j=1}^{\infty} q_j e^{-jh}\right)^{k-2}$. By the assumption that $F(h) < 1$, we obtain

$$\#\mathcal{L}_n(X_\mathcal{C}) \leq e^{nh}\left(K + (n+1)K^2 \sum_{k=2}^{n} F(h)^{k-2}\right) \leq e^{nh}\left(K + \frac{(n+1)K^2}{1-F(h)}\right).$$

Taking logarithms, dividing by n and taking the limit $n \to \infty$ gives $h_{top}(X_\mathcal{C}) \leq h$.

(ii) If $F(h) > 1$, then for all $t \in \mathbb{N}$ sufficiently large, $S_t := \sum_{j=1}^{t} q_j e^{-jh} > 1$. For $k \in \mathbb{N}$, we have the expansion

$$S_t^k = \left(\sum_{j=1}^{t} q_j e^{-jh}\right)^k = \sum_{n=k}^{tk} e^{-nh} \sum_{\substack{n_2+\dots+n_{k-1}=n \\ n_i\geq 1}} \prod_{j=1}^{k} q_{n_j}.$$

Choose $n = N_k$ such that the second sum is maximized. Obviously $t \leq N_k \leq tk$. Then

$$S_t^k \leq tke^{-N_k h} \sum_{\substack{n_2+\dots+n_{k-1}=N_k \\ n_i\geq 1}} \prod_{j=1}^{k} q_{n_j}.$$

For every choice $n_1, \dots, n_k$ with $\sum_{i=1}^{k} n_i = N_k$, the concatenation of words from $\mathcal{C}_{n_i}$ belongs to $\mathcal{L}(X_\mathcal{C})$. Also, by the unique decomposition property, every different choice of such concatenation gives a different word in $\mathcal{L}(X_\mathcal{C})$. Therefore

$$\#\mathcal{L}_{N_k}(X_\mathcal{C}) \geq \sum_{\substack{n_2+\dots+n_{k-1}=N_k \\ n_i\geq 1}} \prod_{j=1}^{k} q_{n_j} \geq \frac{e^{N_k h} S_t^{N_k/t}}{tk}.$$

Next take logarithms, divide by N_k and let $N_k \to \infty$ to obtain $h_{top}(X_\mathcal{C}) \geq h + \frac{1}{t}\log S_t$. But since $F(h) \geq S_t > 1$ for all sufficiently large t, we get $h_{top}(X_\mathcal{C}) > h$ as required. $\square$

We can now state the consequence for the entropy of coded shifts, paraphrasing results of Pavlov [**436**, Theorems 1.1-1.3].

Corollary 3.46. Let $h(U_{\mathcal{C}}) = \lim_n \frac{1}{n} \log p_{U_{\mathcal{C}}}(n)$ be the exponential growth-rate of words in $U_{\mathcal{C}}$, and recall the function $F(h) = \sum_{\ell \geq 1} q_\ell e^{-h\ell}$ from (3.6).

(a) Assume that $X_{\mathcal{C}}$ has unique decomposition property. If $F(h(U_{\mathcal{C}})) \geq 1$ then $F(h_{top}(X_{\mathcal{C}}, \sigma)) = 1$. In fact, $h = h_{top}(X_{\mathcal{C}})$ is the only solution of $F(h) = 1$.

(b) If $F(h(U_{\mathcal{C}})) < 1$ then $h_{top}(X_{\mathcal{C}}, \sigma) = h(U_{\mathcal{C}})$.

Also $h_{top}(X_{\mathcal{C}}, \sigma) = h(U_{\mathcal{C}})$ if and only if $F(h(U_{\mathcal{C}})) \leq 1$.

Proof. The map $h \mapsto F(h)$ has a critical value h_c such that $F(h) = \infty$ for $h < h_c$ and $F(h) < \infty$ is strictly decreasing for $h > h_c$. At $h = h_c$, $F(h)$ can be finite or infinite.

(a) If $1 < F(h(U_{\mathcal{C}}))$ is finite, then $h_c \leq h(U_{\mathcal{C}})$, as there is a unique $h_1 > h(U_{\mathcal{C}})$ such that $F(h_1) = 1$. Theorem 3.45 gives that $h_{top}(X_{\mathcal{C}}) = h_1$.

(b) If $F(h(U_{\mathcal{C}})) < 1$, then by Theorem 3.45(i) we have $h_{top}(X_{\mathcal{C}}) < h(U_{\mathcal{C}}) + \varepsilon$ for every $\varepsilon > 0$. Since $X_{\mathcal{C}} \supset U_{\mathcal{C}}$, we have $h_{top}(X_{\mathcal{C}}) \geq h(U_{\mathcal{C}})$, so $h_{top}(X_{\mathcal{C}}) = h(U_{\mathcal{C}})$ follows.

Combining (a) and (b) shows that $h_{top}(X_{\mathcal{C}}) = h(U_{\mathcal{C}})$ if and only if $F(h(U_{\mathcal{C}})) \leq 1$. $\square$

Corollary 3.47. Every non-periodic coded shift $(X_{\mathcal{C}}, \sigma)$ has positive entropy.

Proof. If $\mathcal{C}$ is a single word, then $X_{\mathcal{C}}$ is periodic. Let $C, C' \in \mathcal{C}$ be the two shortest words in $\mathcal{C}$. Then by Theorem 8.71, the entropy $h_{top}(X_{\mathcal{C}}, \sigma) \geq \log x$, where x is the largest solution to the equation $x^{-|C|} + x^{-|C'|} = 1$. Clearly $x > 1$. $\square$

The classification also has an analogue for the intrinsic ergodicity of coded shifts. This was studied in several papers by Climenhaga, Thompson and Pavlov, see [**155**, Theorem B] and [**436**]. For countable directed graphs, intrinsic ergodicity is equivalent to positive recurrence, see Theorem 8.66. The results for coded shifts are parallel, except that $h(U_{\mathcal{C}})$ plays the role of $\limsup_\ell \frac{1}{\ell} \log q_\ell$ in the case that the graph $\mathcal{G}$ is formed by a single vertex v_0 from which q_ℓ loops of length ℓ emerge.

That is, if $F(h(U_{\mathcal{C}})) > 1$, then there is a unique measure of maximal entropy μ, and $\mathrm{supp}(\mu) = X_{\mathcal{C}}$. If $F(h(U_{\mathcal{C}})) < 1$, then all invariant measures (if there are any) are of maximal entropy μ, and $\mu(U_{\mathcal{C}}) = 1$. The case $F(h(U_{\mathcal{C}})) = 1$ is a mixture of the two: there may be one or multiple measures of maximal entropy.

Theorem 3.48. *Let (X,σ) be a coded shift with $q_\ell = \#\{c \in \mathcal{C} : |c| = \ell\}$.*

(1) If $\limsup_\ell \frac{1}{\ell}\log q_\ell < h_{top}(X_{\mathcal{C}},\sigma)$, then (X,σ) is intrinsically ergodic.

(2) If $\limsup_\ell \frac{1}{\ell}\log q_\ell = 0$, then every factor of (X,σ) is intrinsically ergodic.

Example 3.49. The next example, taken from [**436**, Example 5.3 and 5.4] shows that in certain cases the theory of countable directed graph does apply to coded shifts. Important for this seems to be that $q_\ell \approx \#\{\text{subwords of the code words from } \mathcal{C}_\ell\}$.

For the alphabet $\mathcal{A} = \{0,1,\dots,d\}$ and some function $\tau : \mathbb{N} \to \mathbb{N}$, take the set of code words

$$\mathcal{C} = \{a_1a_2\dots a_n 0^{\tau(n)} : a_i \in \{1,\dots,d\}, n \geq 2\}.$$

Hence $U_{\mathcal{C}} = \{0,\dots,d\}^{\mathbb{Z}}$, whence $h(U_{\mathcal{C}}) = \log d$, and

$$F(h(U_{\mathcal{C}})) = \sum_{\ell=1}^{\infty} q_\ell d^{-\ell} = \sum_{n=2}^{\infty} d^n \cdot e^{-(n+\tau(n))\log d} = \sum_{n\geq 2} d^{-\tau(n)}.$$

If $d = 2$ and $\tau(n) = n$, then $F(h(U_{\mathcal{C}})) = 1$, so $h_{top}(X_{\mathcal{C}}) = h(U_{\mathcal{C}}) = \log 2$. In fact, this is a situation (see [**436**, Proposition 5.1]) where one can equally well work with the transition graph $\mathcal{G}$, and the Gurevich entropy $h_G(\mathcal{G}) = \log 2$. Since also

$$\frac{d}{dh}F(h)|_{h=h_G(\mathcal{G})} = \sum_{n\geq 2}(n+\tau(n))d^{-\tau(n)-1} = \sum_{n=2}^{\infty} n2^{-n} < \infty,$$

the graph $\mathcal{G}$ is positively recurrent. Thus there is a unique measure of maximal entropy, and it is supported on the whole of $X_{\mathcal{C}}$.

If $d = 4$ and $\tau(n) = \lfloor \log_2 n \rfloor$, then

$$F(h(U_{\mathcal{C}})) = \sum_{n\geq 2} d^{-\tau(n)} = \sum_{n\geq 2} 4^{-\lfloor \log_2 n\rfloor} = \sum_{k=1}^{\infty}\sum_{n=2^k}^{2^{k+1}-1} 4^{-k} = \sum_{k=1}^{\infty} 2^k\, 4^{-k} = 1.$$

Again, $h_{top}(X_{\mathcal{C}}) = h(U_{\mathcal{C}}) = h_G(\mathcal{G})$, and

$$\begin{aligned}\frac{d}{dh}F(h)|_{h=h_G(\mathcal{G})} &= \sum_{n\geq 2}(n+\tau(n))d^{-\tau(n)-1} \geq \frac{1}{4}\sum_{n=2}^{\infty} n4^{-\lfloor \log_2 n\rfloor} \\ &\geq \frac{1}{4}\sum_{k=1}^{\infty}\sum_{n=2^k}^{2^{k+1}-1} 2^k 4^{-k} = \frac{1}{4}\sum_{k=1}^{\infty} 1 = \infty.\end{aligned}$$

Therefore $\mathcal{G}$ is null recurrent, and the measure of maximal entropy is supported on $U_{\mathcal{C}}$. In fact, it is the $(\frac14,\frac14,\frac14,\frac14)$-Bernoulli measure on $\{1,2,3,4\}^{\mathbb{Z}}$, giving no weight to the symbol 0.

3.4. Hereditary and Density Shifts

The natural order on the alphabet $\mathcal{A} = \{0, \dots, N-1\}$ can be used to create shift-invariant rules.

Definition 3.50. A collection $X \subset \mathcal{A}^{\mathbb{N} \text{ or } \mathbb{Z}}$ is **hereditary** if whenever $x \in X$ and $y \leq_{\text{her}} x$ (meaning that $y_n \leq x_n$ for all n), then also $y \in X$.

Hereditary shifts first appeared in [**346**, page 882]. It is clear that this rule is shift-invariant, but it is not necessarily closed under taking limits. For example, the collection

$$X = \{x \in \{0,1\}^{\mathbb{N}} : x_i = 0 \text{ infinitely often}\} \tag{3.7}$$

is hereditary, but contains the sequence 1^∞ in its closure. Therefore, some authors [**370**] make the distinction between hereditary shift and **subordinate shift**, the latter being hereditary and closed. We will write hereditary subshift, meaning it is indeed closed. SFTs are hereditary, if the collection of forbidden words of length M are exactly the largest in the partial order $\leq_{\text{her}}$ on $\mathcal{A}^M$. A similar fact holds for sofic shifts.

Lemma 3.51. *The* **hereditary closure** *of (i.e., smallest hereditary subshift containing) the sofic shift* (X, σ) *is sofic.*

Proof. Extend the edge-labeled transition graph $\mathcal{G}$ of X to $\mathcal{G}'$ so that for each $v \xrightarrow{a} w$, there is also $v \xrightarrow{a'} w$ for each letter $a' < a$. □

We will see later that also β-shifts (Corollary 3.71) and spacing shifts are hereditary. Another way to create hereditary subshifts is by stipulating an upper bound of the frequency of non-zero digits.

Definition 3.52. Let $\mathcal{A} = \{0, 1, \dots, N-1\}$ be the alphabet. The (upper) **density of the subshift** $X \subset \mathcal{A}^{\mathbb{N} \text{ or } \mathbb{Z}}$ is

$$\bar{d}(X) = \sup\{\bar{d}(x) : x \in X\},$$

where $\bar{d}(x)$ is the upper density (see Definition 8.51) of the set of indices j such that $x_j \neq 0$, i.e., $\bar{d}(x) = \limsup_k \frac{1}{k}\{0 \leq j < k : x_j \neq 0\}$. Let $X_\delta := \{x \in \mathcal{A}^{\mathbb{N}} : \bar{d}(x) \leq \delta\}$.

It is clear that $\bar{d}(X_\delta) = \delta$, but the example of (3.7) shows that the property that $\bar{d}(x) \leq \delta$ for every $x \in X_\delta$ is not closed under taking limits.

Remark 3.53. Assume that a collection $X \subset X_\delta$ shift-invariant and *closed*. Then it makes no difference to use Banach density (see Section 8.5) instead of density. Indeed, if there was a sequence $x \in X$ with upper Banach density δ, then there is a sequence n_k such that $\frac{1}{k}\#\{1 \leq j \leq k : x_{n_k+j} \neq 0\} \to \delta$. But then $\frac{1}{k}\#\{1 \leq j \leq k : \sigma^{n_k}(x)_j \neq 0\} \to \delta$, and by compactness, we can find a subsequence of $(n_k)_{k\in\mathbb{N}}$ along which $\sigma^{n_k}(x)$ converges to y. This y has upper density δ. Secondly, if we define a measure μ as accumulation point of $\frac{1}{n_k}\sum_{j=0}^{n_k-1}\delta_{\sigma^j(y)}$ where the sequence $(n_k)_{k\in\mathbb{N}}$ is such that $\lim_k \frac{1}{n_k}\#\{0 \le j < n_k : y_j = 1\} = \delta$, then each μ-typical point satisfies $\lim_n \frac{1}{n}\#\{0 \le j < n : y_j = 1\} = \delta$.

The following entropy estimate is adapted from [**370**].

Theorem 3.54. *A non-periodic hereditary subshift (X, σ) on the alphabet $\mathcal{A} = \{0, 1, \dots, N-1\}$ has positive topological entropy. In fact*[3] $h_{top}(X, \sigma) \ge \bar{d}(X)\log 2$ *and* $h_{top}(X, \sigma) = 0$ *if* $\bar{d}(X) > 0$.

Proof. Let X be a one-sided hereditary shift (the two-sided case goes similarly). Assume that X is not a single periodic orbit, which for hereditary shifts means $X \ne \{0^\infty\}$. If $\bar{d}(X) > 0$, then for every $\varepsilon > 0$ there are $x \in X$ and infinitely many integers n such that $\#\{1 \le i \le n : x_i \ne 0\} \ge (\bar{d}(X) - \varepsilon)n$. Since X is hereditary,

$$\frac{1}{n}\log p(n) \ge \frac{1}{n}\log 2^{(\bar{d}(X)-\varepsilon)n} = (\bar{d}(X) - \varepsilon)\log 2.$$

But $\lim_n \frac{1}{n}\log p(n)$ exists according to Fekete's Lemma 1.15, and $\varepsilon > 0$ is arbitrary, so $h_{top}(\sigma) \ge \bar{d}(X)\log 2$. Note that if X, for every $\varepsilon > 0$, contains sequences x such that $\#\{1 \le i \le n : x_i = N-1\} \ge \bar{d}(X) - \varepsilon$, then we find $h_{top}(\sigma) \ge \bar{d}(X)\log N$.

For the converse, assume that $\bar{d}(X) = 0$, so for every $\varepsilon > 0$ there is n_0 such that for all $n \ge n_0$,

$$p(n) \le \sum_{k=0}^{\lceil n\varepsilon\rceil}\binom{n}{k} \le \lceil n\varepsilon\rceil\binom{n}{\lceil n\varepsilon\rceil}.$$

Using Stirling's formula[4], we obtain

$$\begin{aligned}\frac{1}{n}\log p(n) &\le \frac{1}{n}\log\left(\frac{\varepsilon\sqrt{n}\,n^n e^{-n}}{(n\varepsilon)^{n\varepsilon}e^{-n\varepsilon}\ (n(1-\varepsilon))^{n(1-\varepsilon)}e^{-n(1-\varepsilon)}}\right)\\ &\le \frac{1}{n}\log(\varepsilon\sqrt{n}) - \varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon).\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, it follows that $h_{top}(\sigma) = \lim_n \frac{1}{n}\log p(n) = 0$. □

The drawback of Definition 3.52 above is that the collection

$$X_\delta := \{x \in \mathcal{A}^{\mathbb{N}\text{ or }\mathbb{Z}} : \bar{d}(x) \le \delta\}$$

is not closed. For instance $x^n := 1^n 0^\infty \in X_\delta$ for all $\delta \ge 0$, but $\lim_n x^n = 1^\infty$ belongs only to X_1. To obtain closedness, we need to impose further

[3]But this is not a sharp bound, see Example 3.59.

[4]$n! \sim n^n e^{-n}\sqrt{2\pi n}$

conditions, of the sort that every n-block (for n sufficiently large) contains no more than $\lceil \delta n \rceil$ non-zero symbols. The general approach that we shall present is due to Stanley [**506**].

Definition 3.55. Let $\mathcal{A} = \{0, 1, \dots, N-1\}$ be the alphabet. Given a function $f : \mathbb{N} \to \mathbb{R}$, we define the **density shift** of f as

$$X_f := \{x \in \mathcal{A}^{\mathbb{N} \text{ or } \mathbb{Z}} : \sum_{i=0}^{n-1} x_{k+i} \leq f(n) \text{ for all } k \in \mathbb{N} \text{ or } \mathbb{Z} \text{ and } n \in \mathbb{N}\}.$$

In particular, if $\mathcal{A} = \{0, 1\}$, then

$$X_f := \{x \in \mathcal{A}^{\mathbb{N} \text{ or } \mathbb{Z}} : |x_k \dots x_{k+n-1}|_1 \leq f(n) \text{ for all } k \in \mathbb{N} \text{ or } \mathbb{Z} \text{ and } n \in \mathbb{N}\}.$$

Since the condition in the definition is on finite blocks, X_f is closed, and σ-invariance is clear too. Therefore X_f is a subshift, and it is obviously hereditary. We could define density shifts on the infinite alphabet $\mathcal{A} = \{0, 1, 2, \dots\}$, but as $f(1) < \infty$, we can use only $f(1) + 1$ symbols anyway.

Example 3.56. The odd shift X_{odd} from Example 1.4 is not a density shift, because it is not hereditary. For example, $1011 \in \mathcal{L}(X_{odd})$ but $1001 \notin \mathcal{L}(X_{odd})$.

Definition 3.57. The **canonical function** $\boldsymbol{f}$ of a density shift X is the smallest function such that $X = X_{\boldsymbol{f}}$, in the sense that if $X = X_f$, then $\boldsymbol{f}(n) \leq f(n)$ for all $n \in \mathbb{N}$.

Theorem 3.58. *The canonical function* $\boldsymbol{f}$ *of a density shift satisfies*

(1) $\boldsymbol{f}(\mathbb{N}) \subset \mathbb{N}$;

(2) $\boldsymbol{f}$ *is non-decreasing;*

(3) $\boldsymbol{f}(m+n) \leq \boldsymbol{f}(m) + \boldsymbol{f}(n)$ *(subadditive).*

Conversely, every function $\boldsymbol{f}$ *satisfying (1)-(3) is the canonical function of some density shift.*

Example 3.59. If $f(n) = (n+1)/2$, then the word 11 is forbidden, but no other word is (apart from words that contain 11). Thus X_f is the Fibonacci SFT, and its density $\bar{d}(X) = 1/2$, achieved by $x = 101010\dots$. If we set $\boldsymbol{f}(n) = \lfloor (n+1)/2 \rfloor$, then we get the same shift: $X_f = X_{\boldsymbol{f}}$. In fact, $\boldsymbol{f}$ is the smallest function with this property. This example also shows that the lower bound of the entropy in Theorem 3.54 is not sharp, because $h_{top}(X_f, \sigma) = \log(\frac{1}{2}(1+\sqrt{5}))$ which is larger than the $\frac{1}{2}\log 2$ given by Theorem 3.54.

Proof of Theorem 3.58. For simplicity of exposition, we only consider one-sided shifts. Define the partial order on X as

$$x \preceq_{\text{sum}} y \quad \text{if} \quad \sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} y_i \qquad \text{for all } n \in \mathbb{N}. \tag{3.8}$$

Let $z \in X$ be such that, inductively, for every $n \in \mathbb{N}$, $x_n \in \mathcal{A}$ is the largest symbol such that $x_1 \dots x_n \in \mathcal{L}_n(X)$. We claim that $x \preceq_{\text{sum}} z$ for all $x \in X$.

We prove the claim by induction on the length n. Clearly $x_1 \leq z_1$. Assume by induction that $x_1 \dots x_n \preceq_{\text{sum}} z_1 \dots z_n$ and let $\xi_{n+1} \in \mathcal{A}$ be maximal such that

$$x_1 \dots x_n \xi_{n+1} \in \mathcal{L}(X). \tag{3.9}$$

Set $p = \sum_{i=1}^n z_i - \sum_{i=1}^n x_i \geq 0$. If $\xi_{n+1} \leq p$, then (3.9) clearly holds. If $\xi_{n+1} > p$, then take $a = \xi_{n+1} - p \leq N-1$. For each $1 \leq r \leq n$, we have

$$\begin{aligned}
\sum_{i=r}^n z_i + a &= \sum_{i=1}^n z_i + a - \sum_{i=1}^{r-1} z_i \\
&\leq \sum_{i=1}^n x_i + p + a - \sum_{i=1}^{r-1} x_i && \text{(by the induction hypothesis)} \\
&\leq \sum_{i=r}^n x_i + \xi_{n+1} && \text{(by the choice of } a\text{)},
\end{aligned}$$

which is an allowed sum in X by the choice of ξ_{n+1}. Therefore $z_1 \dots z_n a \in \mathcal{L}(X_{\boldsymbol{f}})$ and because $\sum_{i=1}^n z_i + a = \sum_{i=1}^n x_i + \xi_{n+1}$, we have

$$x_1 \dots x_{n+1} \preceq_{\text{sum}} x_1 \dots x_n \xi_{n+1} \preceq_{\text{sum}} z_1 \dots z_n a \preceq_{\text{sum}} z_1 \dots z_{n+1}.$$

This finishes the induction step. It follows that $\sigma^m(z) \preceq_{\text{sum}} z$ for all $m \geq 0$, i.e., z is shift-maximal with respect to $\preceq_{\text{sum}}$. Define $\boldsymbol{f}(n) = \sum_{i=1}^n z_i$. Then clearly $\boldsymbol{f}$ is integer-valued and non-decreasing. Also $\boldsymbol{f}(m+n) = \sum_{i=1}^n z_i + \sum_{i=1}^n \sigma^m(z)_i \leq \boldsymbol{f}(m) + \boldsymbol{f}(n)$. Hence (1)-(3) hold.

Conversely, suppose that $\boldsymbol{f}$ satisfies (1)-(3) and set $X = X_{\boldsymbol{f}}$. Let z be the maximal sequence with respect to $\preceq_{\text{sum}}$ as before. We will prove by induction that

$$\boldsymbol{f}(r) = \sum_{i=1}^r z_i \quad \text{for all } r \in \mathbb{N}. \tag{3.10}$$

This is clear for $n = 1$, so assume that (3.10) holds for all $1 \leq r \leq n$. Set $a = \boldsymbol{f}(n+1) - \boldsymbol{f}(n)$. We must show that $z_1 \dots z_n a \in \mathcal{L}(X_{\boldsymbol{f}})$ and for this it suffices to show that $\sum_{i=r}^n z_i + a \leq \boldsymbol{f}(n-r+2)$ for each $1 \leq r \leq n$. For $r = 1$ this holds by the choice of a. Otherwise

$$\begin{aligned}
\sum_{i=r}^n z_i + a &= \sum_{i=1}^n z_i - \sum_{i=1}^{r-1} z_i + a \\
&= \boldsymbol{f}(n) - \boldsymbol{f}(f(r-1)) + a && \text{(by the induction hypothesis)} \\
&= \boldsymbol{f}(n+1) - \boldsymbol{f}(r-1) && \text{(by the choice of } a\text{)} \\
&\leq \boldsymbol{f}(n-r+2) && \text{(by property (3))}.
\end{aligned}$$

This concludes the induction step and the entire proof. □

By Fekete's Lemma 1.15, $\lim_n \boldsymbol{f}(n)/n = \inf_n \boldsymbol{f}(n)/n$, so $\inf_n \boldsymbol{f}(n)/n = 0$ if and only if the density shift $(X_{\boldsymbol{f}}, \sigma)$ has zero topological entropy by Theorem 3.54. Without proof we state ([**506**, Theorem 2.10]):

Corollary 3.60. If $\sigma^m(z) \preceq_{\text{sum}} z$ for all $m \geq 0$, then z is the maximal sequence of the density shift $X_{\boldsymbol{f}}$ for $\boldsymbol{f}(n) = \sum_{i=1}^n z_i$.

Theorem 3.61. *Let $X_{\boldsymbol{f}}$ be a non-trivial density shift with canonical function $\boldsymbol{f}$. The following are equivalent:*

(a) *$(X_{\boldsymbol{f}}, \sigma)$ is topologically transitive.*

(b) *$(X_{\boldsymbol{f}}, \sigma)$ is topologically mixing.*

(c) *$\boldsymbol{f}$ is unbounded.*

Proof. $(a) \Rightarrow (b)$: Let $v, w \in \mathcal{L}(X_{\boldsymbol{f}})$ be arbitrary non-empty words. By topological transitivity, there is $u \in \mathcal{L}(X_{\boldsymbol{f}})$ such that $vuw \in \mathcal{L}(X_{\boldsymbol{f}})$ as well. But then $v0^k w \in \mathcal{L}(X_{\boldsymbol{f}})$ for every $k \geq |u|$, proving topological mixing.

$(b) \Rightarrow (a)$: Trivial.

$(a) \Rightarrow (c)$: Since $1 \in \mathcal{L}(X_{\boldsymbol{f}})$, topological transitivity gives a sequence $x \in X_{\boldsymbol{f}}$ containing infinitely many 1s. Thus $\boldsymbol{f}(n) \geq \sum_{i=1}^n x_i \to \infty$ as $n \to \infty$.

$(c) \Rightarrow (a)$: Let $u, v \in \mathcal{L}(X_{\boldsymbol{f}})$ be arbitrary non-empty words. Since $\boldsymbol{f}$ is unbounded, there is $n \in \mathbb{N}$ such that $\boldsymbol{f}(n) \geq \sum_{i=1}^{|u|} u_i + \sum_{i=1}^{|v|} v_i$. Then $u0^n v \in \mathcal{L}(X_{\boldsymbol{f}})$. □

In particular, SFTs (X, σ) that are also density shifts are transitive, because, unless $X = \{0^\infty\}$, there is a non-trivial word v and $x \in X$ that contains v infinitely often as subword. In fact, density SFTs are completely characterized as those for which the canonical function $\boldsymbol{f}$ satisfies $\inf_n \boldsymbol{f}(n)/n = \boldsymbol{f}(p)/p$ for some $p \in \mathbb{N}$, see [**506**, Theorem 4.3]. On the other hand, if $\boldsymbol{f}$ is bounded, then all $x \in X_{\boldsymbol{f}}$ end with 0^∞. They can be represented by a finite edge-labeled transition graph [**506**, Theorem 2.16], and also have a finite collection of follower sets. Hence such density shifts are non-transitive sofic shifts.

Sofic density shifts, in general, are characterized [**506**, Theorem 6.3] as those for which the maximal sequence z is eventually periodic ($z_n = z_{n+p}$ for n sufficiently large), or equivalently $\boldsymbol{f}(n+p) = \boldsymbol{f}(n) + k$ (where $k = \sum_{i=n}^{n+p-1} z_i$ and $k > 0$ if and only if $X_{\boldsymbol{f}}$ is transitive).

Theorem 3.62. *Let $X_{\boldsymbol{f}}$ be a non-trivial density shift with canonical function $\boldsymbol{f}$. The following are equivalent:*

(a) *$X_{\boldsymbol{f}}$ contains a periodic point other than 0^∞.*

(b) *There is $\lambda > 0$ such that $\boldsymbol{f}(n) \geq \lambda n$ for all $n \in \mathbb{N}$.*

(c) *$(X_{\boldsymbol{f}}, \sigma)$ is a coded shift.*

Proof. $(a) \Rightarrow (b)$: If $0^\infty \neq x = \sigma^p(x) \in X_{\boldsymbol{f}}$, then $\sum_{i=1}^n x_i \geq n/p$, so (b) holds for $\lambda = 1/p$.

$(b) \Rightarrow (c)$: Define $s(u) = \lceil \sum_{i=1}^{|u|} u_i \lambda \rceil$, and let

$$\mathcal{C} := \{0^{s(u)}\ u\ 0^{s(u)} : u \in \mathcal{L}(X_{\boldsymbol{f}})\}$$

be the collection of code words. The "padding blocks" $0^{s(u)}$ ensure that the "core words" u are sufficiently apart that the code words can be concatenated freely, see [**506**, Theorem 3.1] for the details. Hence the coded shift $X_{\mathcal{C}} \subset X_{\boldsymbol{f}}$. On the other hand, $\mathcal{L}(X_{\boldsymbol{f}}) \subset \mathcal{L}(X_{\mathcal{C}})$ so the reverse inclusion $X_{\boldsymbol{f}} \subset X_{\mathcal{C}}$ follows.

$(c) \Rightarrow (a)$: If u is a non-trivial code word, then $u^\infty \in X_{\boldsymbol{f}}$. □

Since every infinite subshift is expansive (see below Definition 1.38), Theorems 3.61 and 3.62 allow the following characterizations of chaos for density shifts.

Corollary 3.63. Let $(X_{\boldsymbol{f}}, \sigma)$ be a density shift with canonical function $\boldsymbol{f}$. Then

(1) $(X_{\boldsymbol{f}}, \sigma)$ is Devaney chaotic if and only if $\inf_n \boldsymbol{f}(n)/n > 0$.

(2) $(X_{\boldsymbol{f}}, \sigma)$ is Auslander-Yorke chaotic if and only if $\boldsymbol{f}$ is unbounded.

(3) $(X_{\boldsymbol{f}}, \sigma)$ is Li-Yorke chaotic if and only if $\boldsymbol{f}$ is unbounded.

3.5. β-Shifts and β-Expansions

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

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

For the case $\beta \in \{2, 3, 4, \dots\}$, this is the usual β-ary expansion; it is unique except for the β-adic rationals $\{\frac{m}{\beta^n} : m \in \mathbb{Z}, n \in \mathbb{N}\}$. 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. As summarized in Theorem 3.67, some points have uncountably many different expansions, but there is a canonical way to define an expansion, called the **greedy expansion**:

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

- Let $x_2 = \beta x_1 - b_2$ and $b_3 = \lfloor \beta x_2 \rfloor$, etc.

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

Definition 3.64. 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).

If $b = (b_k)_{k=1}^\infty$ is the β-expansion of some $x \in [0,1]$, then $\sigma(b)$ is the β-expansion of $T_\beta(x)$. The following lemma from **[431]** characterizes the β-shift in terms of the lexicographic order $\preceq_{\text{lex}}$:

Lemma 3.65. *Let $c = c_1 c_2 c_3 \dots$ be the β-expansion of 1, and suppose it is not finite, i.e., $c_i > 0$ infinitely often*[5]*. Then $b \in X_\beta$ if and only if*

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

However, the greedy expansion $(b_i)_{i \geq 1}$ of x is the largest sequence in lexicographical order among all the expansions of x.

Example 3.66. 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

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

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 can be seen to be conjugate to an SFT (see the middle panel of Figure 3.6), and it has a simple edge-labeled transition graph.

Proof of Lemma 3.65. Let $b = (b_k(x))_{k \geq 1}$ be the β-expansion of some $x \in [0,1)$. (If $x = 1$ there is nothing to prove because $b = c$.) Since $x < 1$ we have $b_1 = \lfloor \beta x \rfloor \leq c_1 = \lfloor \beta \cdot 1 \rfloor$. If the inequality is strict, then $b \prec_{\text{lex}} c$. Otherwise, $0 \leq x_1 = T_\beta(x) = \beta x - b_1 < \beta \cdot 1 - c_1 = T_\beta(1)$, and we find that $b_2 = \lfloor \beta x_1 \rfloor \leq c_2 = \lfloor \beta T_\beta(1) \rfloor$. Continue by induction.

[5]This condition is required for the "if" direction. For example, if $c = 1110^\infty$ as in Example 3.66, then $b = (110)^\infty <_{\text{lex}} c$, but there is no point $x \in [0,1]$ with this itinerary. In fact, b is the lazy expansion of the point 1; it is "the other" canonical itinerary that 1 has.

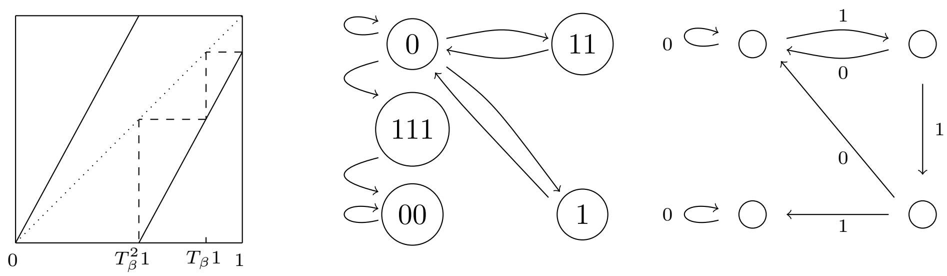


Figure 3.6. Left: The map T_β for $\beta^3 = \beta^2 + \beta + 1$. Then $T_\beta^3(1) = 0$. Middle: A corresponding vertex-labeled graph. Right: A corresponding edge-labeled graph.

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

$$
\begin{aligned}
A_j &= [\frac{j}{\beta}, \frac{j+1}{\beta}) && 0 \le j < c_1, \\
A_{c_1 j} &= [\frac{c_1}{\beta} + \frac{j}{\beta^2}, \frac{c_1}{\beta} + \frac{j+1}{\beta^2}) && 0 \le j < c_2, \qquad (3.11)\\
A_{c_1 c_2 j} &= [\frac{c_1}{\beta} + \frac{c_2}{\beta^2} + \frac{j}{\beta^2}, \frac{c_1}{\beta} + \frac{c_2}{\beta^2} + \frac{j+1}{\beta^2}) && 0 \le j < c_3, \\
&\vdots \qquad \vdots \qquad \vdots && \vdots \qquad (3.12)
\end{aligned}
$$

They are adjacent and clearly $T_\beta(A_j) = [0,1)$ for $0 \le j < c_1$. Also $T_\beta(A_{c_1 j}) = [j/\beta, (j+1)/\beta)$ for $0 \le j < c_2$. Since $\sigma^n((c_k)_{k\ge1}) \preceq_{\text{lex}} (c_k)_{k\ge1}$ by the first part of the proof, we have $c_2 \le c_1$. In particular $T_\beta(A_{c_1 j})$ is one of the intervals in the first row of (3.11). Therefore $T_\beta^2(A_{c_1 j}) = [0,1)$. By induction, we obtain

$$T_\beta^{k+1}(A_{c_1 c_2 \dots c_k j}) = [0,1) \qquad \text{for all } k \in \mathbb{N},\ 0 \le j < c_{k+1}. \tag{3.13}$$

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

Now take $(b_k)_{k\ge1} \in \mathcal{A}^{\mathbb{N}}$ such that $(b_k)_{k\ge1} \preceq_{\text{lex}} (c_k)_{k\ge1}$, and define $n_0 = 0$ and recursively $n_{r+1} = \min\{k > n_r : b_k \ne c_{k-n_r}\}$. Suppose first that all n_r's are finite. Then $b_{n_r+1} \dots b_{n_{r+1}}$ is the index of one of the intervals in the $n_{r+1} - n_r$'th row of (3.11). The intersection

$$\bigcap_{r \ge 0} T_\beta^{-n_r}(A_{b_{n_r+1} \dots b_{n_{r+1}}})$$

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

The greedy expansion above is not the only way of expressing $x = \sum_{k\geq 1} b_k\beta^{-k}$ for b_k in the digit set $\{0,\dots,\lfloor\beta\rfloor\}$. For instance, in the **lazy expansion** we always take the smallest possible[6], digit b_k such that the sum x can still be achieved. For $\beta = 2$, choosing the greedy and lazy expansion make the difference in expressing dyadic rationals in $(0,1)$ as $x = b_1 \dots b_k 1000\dots$ (greedy, with partition $\{[0,\frac{1}{2}), [\frac{1}{2}, 1]\}$) and $x = b_1 \dots b_k 0111\dots$ (lazy, with partition $\{[0,\frac{1}{2}], (\frac{1}{2}, 1]\}$). All other numbers in $[0,1]$ have a unique expansion for $\beta = 2$.

In general, the number of expansions can be much larger, for a larger set of points. This can be shown by counting the number of orbits of the point x under iteration of the **multivalued** map

$$T_\beta : \left[0, \frac{\lfloor\beta\rfloor + 1}{\beta}\right] \to \left[0, \frac{\lfloor\beta\rfloor + 1}{\beta}\right], \qquad x \mapsto \beta x - i \ \text{ if } x \in \Delta_i.$$

Here $\Delta_i = [\frac{i}{\beta}, \frac{\lfloor\beta\rfloor+1+i\beta}{\beta^2}]$ are the domains of the branches of T_β, and the labels i are used as the symbols of the itinerary $\boldsymbol{i}(x) = b_0b_1b_2\dots$ of points, i.e., $b_k = i$ if $T_\beta^k(x) \in \Delta_i$ along some forward orbit is an expansion of x. The intervals where the Δ_i's overlap are called **switch regions**, see Figure 3.7. Points for which infinitely many forward T_β-orbits each visit the switch regions infinitely often have uncountably many expansions.

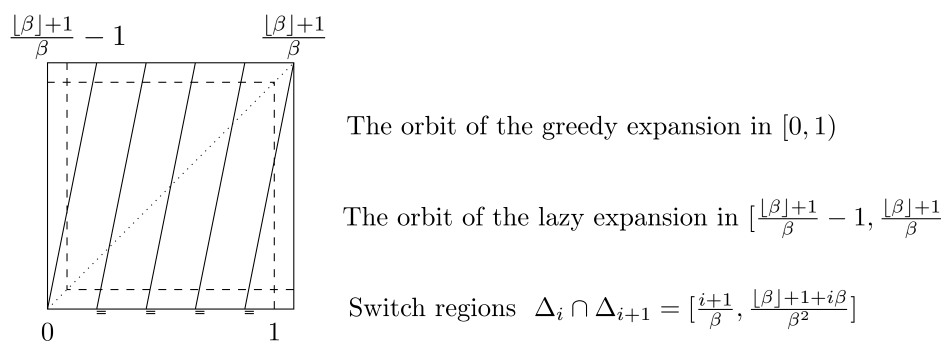


Figure 3.7. The map $T_\beta : [0, \frac{\lfloor\beta\rfloor+1}{\beta}] \to [0, \frac{\lfloor\beta\rfloor+1}{\beta}]$ with switch regions.

On the other hand, points whose forward T_β-orbit avoid switch regions (and then the forward orbit is indeed uniquely defined) have only one expansion. Such points are called **univoque**; we denote the set of univoque points in $(0, \lfloor\beta\rfloor/(\beta-1))$ by U_β. Larger values of β lead to smaller switch

[6]In terms of the algorithm given for the greedy expansion, we need to take $b_k = \lceil \beta x_{k-1} - \lfloor\beta\rfloor/(\beta-1)\rceil$ so that $x_k \leq \sum_{j>k} \lfloor\beta\rfloor\beta^{k-j}$, i.e., x_k (and therefore x) can still be reached choosing the remaining digits b_j maximal.

regions and thus smaller univoque sets; that is, until β becomes integer and the digit set is increased by one. The following theorem is a summary of results from [**273**, **356**], just for the digit set $\{0, 1\}$.

Theorem 3.67. *The set U_β of positive univoque points satisfies:*

(1) $U_\beta = \varnothing$ for $1 < \beta \le \gamma = \frac{1}{2}(1+\sqrt{5}) \approx 1.618\dots$, the golden mean;
(2) $\#U_\beta = 2$ for $\gamma < \beta \le \beta_c \approx 1.755\dots$, the leading root of $x^3 = 2x^2 - x - 1$;
(3) $\#U_\beta = \aleph_0$ for $\beta_c < \beta < \beta_{KL} \approx 1.787\dots$, the so-called Komornik-Loreti constant[7];
(4) $\#U_\beta = 2^{\aleph_0}$ for $\beta_{KL} \le \beta < 2$; it is a Cantor set of positive Hausdorff dimension;
(5) $U_\beta = (0,1) \setminus \{$dyadic rationals$\}$ if $\beta = 2$.

In fact, $Leb(U_\beta) = 0$ for all $\beta \in [1, 2)$.

Further details are given also in [**197**]. Previously, Erdös and coauthors [**235**–**237**], studied the number of β-representations of 1 as function β. For similar results for larger digit sets $\{0, 1, \dots, m\}$, see e.g. [**41**, **196**], among a by now very extensive literature.

Proposition 3.68. The β-shift is a coded shift.

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

$$
\begin{aligned}
S = \{&\underbrace{0, 1, \dots, (c_1 - 1)}_{1-\text{words}},\ \underbrace{c_10, c_11, \dots, c_1(c_2 - 1)}_{2-\text{words}}, \\
&\underbrace{c_1c_20, c_1c_21, \dots, c_1c_2(c_3 - 1)}_{3-\text{words}}, \quad \dots \\
&\quad \vdots \qquad\qquad \vdots \\
&\underbrace{c_1c_2c_3c_4c_5c_6\dots}_{\text{a single infinite word}}\}.
\end{aligned}
\tag{3.14}
$$

Apart from the single infinite word, these are exactly the indices of the intervals $A_{c_1\dots c_kj}$ in (3.11). We know from (3.13) that $T_\beta^{k+1}(A_{c_1\dots c_kj}) = [0, 1)$, so free concatenations of such code words all represent $(b_k(x))_{k\ge1}$ for some $x \in [0, 1]$. Any concatenation in S^* also satisfies Lemma 3.65, so that S^* is dense in (and in fact equal to) X_β. $\square$

[7] This constant is the solution of the equation $\sum_{k\ge0} \rho_{k+1}\beta^{-k} = 1$ for the Thue-Morse sequence $\rho_{\text{TM}} = \rho_0\rho_1\rho_2\dots = 0110\,1001\dots$, see Example 1.6 and [**356**]. The numerical value is $1.7872316501 < \beta_{KL} < 1.7872316505$ and β_{KL} was proven to be transcendental in [**18**].

To illustrate this for the β-shift X_β with $c = c_1c_2c_3\dots$, consider an edge-labeled countable transition graph with vertices $(v_n)_{n\geq 0}$ and arrows[8]

$$\begin{cases} v_{n-1} \xrightarrow{c_n} v_n, & \text{for } n \geq 1, \quad \text{so } c \text{ is written out over the horizontal spine,} \\ v_n \xrightarrow{a} v_0, & \text{for } n \geq 1, \quad 0 \leq a < c_n, \end{cases}$$

see Figure 3.8. The code words are the labels of the simple loops from v_0 to itself, and the infinite paths starting from v_0 are in one-to-one correspondence with the points in X_β.

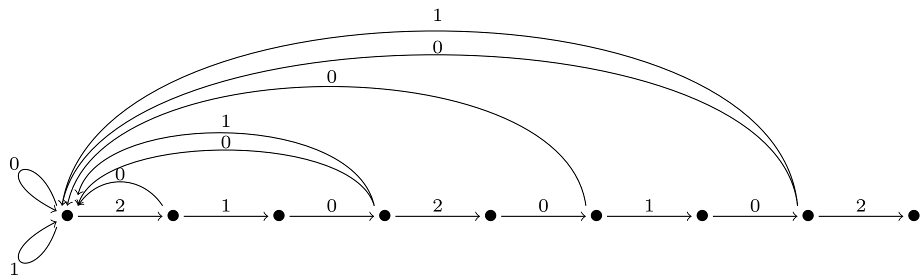


Figure 3.8. The edge-labeled transition graph for a β-shift with $c = 21020102\dots$

Corollary 3.69. Every β-transformation is intrinsically ergodic.

Proof. This was first shown by Hofbauer [**306**], see also [**155**] based on a weakened form of specification[9]. Implementing Theorem 3.48, we have $\#\{s \in S : |s| = n\} \leq \beta$ for each n, so the exponential growth-rate of these words is 0. Hence Theorem 3.48 even implies that every factor of the β-shift is intrinsically ergodic. $\square$

Remark 3.70. For the β-transformation with slope $\beta > 1$, the measure of maximal entropy is absolutely continuous w.r.t. Lebesgue, and there is an explicit formula for the density:

$$\frac{d\mu}{dx} = \frac{1}{\Lambda} \sum_{n\geq 1} \beta^{-n} \mathbf{1}_{[0,T_\beta^n(1)]}$$

for an appropriate normalizing constant Λ, see [**275, 433**].

The following result was probably first stated in [**370**, Section 6].

Corollary 3.71. For every $\beta \in [1,2]$, the β-shift (X_β, σ) is hereditary.

[8] This graph is the edge-labeled version of the Hofbauer tower of the corresponding β-transformation, see Section 3.6.3.

[9] because specification as in Definition 2.81 and hence Lemma 2.85 do not apply

Proof. This follows directly from Lemma 3.65 which determines the shape of the code-words in Proposition 3.68. Indeed, if $x \in X_\beta$ and $n = \min\{i \geq 1 : x_i \neq c_i\}$. Then $x_n < c_n$ and $x_1 \dots x_n$ is a code word. Now repeat the argument with $\sigma^n(x)$. □

Theorem 3.72. *The T_β-orbit of* 1

(1) contains 0 *if and only if X_β is conjugate to an SFT;*

(2) is preperiodic if and only if X_β is sofic[10]*;*

(3) is not dense in $[0,1]$ *if and only if X_β is synchronized;*

(4) is disjoint from $[0,\delta]$ *for some* $\delta > 0$ *if and only if X_β has specification.*

We give a proof below, but refer to [**432**, **473**] for other proofs and related results.

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

For **Statement 1**, let $a_j = T_\beta(1)^j$, so $a_0 = 1$ and $a_N = 0$ for some $N \geq 2$. Let $\mathcal{P}$ be the partition given by the branches of T_β^{N-1}. Then $a_j \in \partial\mathcal{P}$ and the image $T_\beta^{N-1}(\partial J) \subset \{a_i\}_{i=0}^N$ for each $J \in \mathcal{P}$. This means that $\mathcal{P}$ is a Markov partition for T_β^{N-1}, and hence (X_β, σ^{N-1}) is a memory $N-1$ SFT over the alphabet $\{0,1,\dots,\lfloor\beta\rfloor\}$. See Example 3.66 for an illustration of this.

For **Statement 2**, and $c = c_1c_2\dots c_N(c_{N+1}\dots c_{N+p})^\infty$, we claim that X_β only has finitely many different follower sets, see Definition 3.34. Let w be a proper prefix of some $s_1s_2s_3\dots \in S^*$ for the collection of word S from (3.14). That is, there are $k \geq 1$ and $0 \leq m < |s_k|$ such that $|w| = |s_1\dots s_{k-1}| + m$. The possible follower sets are

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

Since c is eventually periodic, this list of follower sets eventually 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 3.36, X_β is sofic. (It is easy to construct an edge-labeled transition graph for X_β, see Example 3.73.)

[10] Since 1 is not in the range of T_β, the orbit of 1 cannot be periodic. If $T^n(1)(j/\beta)$ for some $j \in \mathbb{N}$, then $T^{n+1}(1) = 0$ and case 1. applies, even though $\lim_{y \nearrow j/\beta} T_\beta(y) = 1$.

If on the other hand, the expansion of 1 is not preperiodic, so the T_β-orbit of 1 is infinite, then there are infinitely many different follower sets by Theorem 3.76 below, so X_β cannot be sofic.

For **Statement 3**, assume that $\overline{\mathrm{orb}(1)}$ 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 word $v \in \mathcal{L}_N(X_\beta)$. If $u \in \mathcal{L}_M(X_\beta)$ is such that $uv \in \mathcal{L}(X_\beta)$, then the domain Y of the corresponding branch of T_β^M is such that $T_\beta^M(Y) \cap Z \neq \varnothing$. But since $\mathrm{orb}(1) \cap Z = \varnothing$, we have $T_\beta^M(Y) \supset Z$. Therefore, 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 $v \in \mathcal{L}(X)$ is some 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 neighborhood Y of 1 such that $T_\beta^n(Y) = [0, T_\beta^n(1)]$, and there is $x \in Z \setminus T_\beta^n(Y)$. Let w be the itinerary of $T_\beta^N(x)$; since $x \in Y$, $vw \in \mathcal{L}(X_\beta)$. Similarly, taking $u = c_1 c_2 \dots c_n$, since $T_\beta^n(1) \in Z$, also $uv \in \mathcal{L}(X_\beta)$. However, $uvw \notin \mathcal{L}(X_\beta)$, because there is no $y \in Y$ such that $T_\beta^n(y) = x$. This shows that v is not synchronizing, and since v was arbitrary, X_β is not synchronized.

Finally, for **Statement 4**, take N such that the cylinder set $[0^N]$ corresponds to a subinterval Z_N contained in $[0,\delta]$. Then $T_\beta^N(Z_N) = [0,1]$. Also, for any k-cylinder $[x]$ corresponding to an interval $Z_x \subset [0,1]$, we have $T_\beta^k(Z_x) \supset [0,\delta] \supset Z_N$. Specification follows from this.

On the other hand, if 0 is an accumulation point of $\mathrm{orb}(1)$, then for any $M, N \in \mathbb{N}$, there is some word $x \in \mathcal{L}_M(X_\beta)$ corresponding to a interval Z_x such that $T_\beta^M(Z_x) \subset [0, \beta^{-N+1}]$. Then there is no word $y \in \mathcal{L}_N(X_\beta)$ such that $xy1 \in \mathcal{L}(X_\beta)$, and thus specification fails. $\square$

Example 3.73. Let $\beta = 1.801937735\dots$ be the largest root of the equation $\beta^3 = \beta^2 + 2\beta - 1$. One can check that $c = 11010101010\dots$ is preperiodic, and the T_β-orbit of 1 is $\{1, \beta-1, \beta(\beta-1), \beta-1, \beta(\beta-1), \dots$. The points $\{0, \beta(\beta-1), 1/\beta, \beta-1, 1\}$ define a Markov partition, see Figure 3.9. Therefore the dynamical system $([0,1], T_\beta)$ can be described as an SFT, but not in the alphabet $\{0,1\}$. However, by edge-labeling the transition graph in Figure 3.9, we get X_β. Therefore $x \in X_\beta$ if and only if for every $n \geq 0$ one of

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

holds. 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 can be seen to be conjugate to an SFT (see the middle panel of Figure 3.6), and it has a simple edge-labeled transition graph. Also, X_β is the image of the length one sliding block code $\pi(a) = \pi(b) = 0$, $\pi(c) = \pi(d) = 1$, because $a, b \subset [0, 1/\beta]$ and $c, d \subset [1/\beta, 1]$.

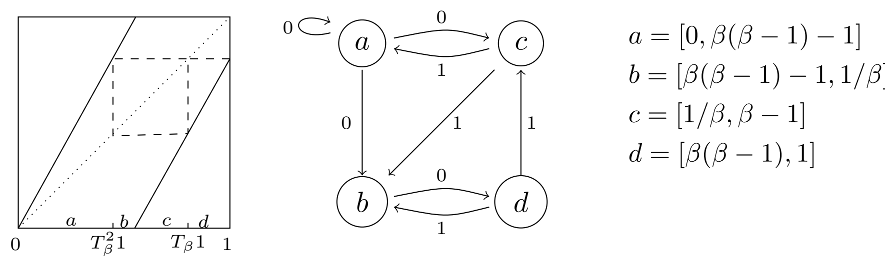


Figure 3.9. The transition graph for a sofic β-shift for $\beta = 1.801937735\dots$

The first two types of β-shifts in Theorem 3.72 correspond to certain algebraic properties of β, which we will mention, but not prove. For the definitions of Pisot and Perron numbers, see Section 8.1.

Theorem 3.74. *If β is a Pisot number then X_β is sofic. If the subshift X_β is sofic then β is a Perron number.*

Remark 3.75. We refer to [**475**] and [**81**, Chapter 7] for more results in this spirit. If X_β is sofic, then the T_β-orbit of 1 is a finite set, say $0 = x_0 < x_1 < x_2 < \cdots < x_d = 1$, where $x_0 = 0$ is added for convenience, also if it is not part of $\mathrm{orb}_{T_\beta}(1)$. The intervals $\tau_i = [x_{i-1}, x_i]$ form a Markov partition with associated matrix $M = (m_{ij})_{i=1}^d$ where $m_{ij} = 1$ if $T_\beta(\tau_i) \supset \tau_j$ and $m_{ij} = 0$ otherwise. This also defines a substitution $\chi_\beta(a) = a_1 \dots a_t$ (with the letters a_i in increasing order) if $T_\beta(\tau_a) = \tau_{a_1} \cup \cdots \cup \tau_{a_t}$ with fixed point $\rho = \lim_n \chi_\beta^n(a_1)$, and substitution shift (X_ρ, σ) for $X_\rho = \overline{\mathrm{orb}_\sigma(\rho)}$. See [**15, 16**] for studies of these kind of substitution systems. The Pisot substitution conjecture states that this subshift has a purely point spectrum (see Section 6.8.3) if and only if β is a Pisot number. This special version of the Pisot substitution conjecture was proved by Barge [**48**].

Continuing on the theme of follower sets, let

$$\mathcal{F}(n) := \#\{F : F \text{ is the follower set of some } v \in \mathcal{L}_n(X_\beta)\} \tag{3.15}$$

be the number of distinct follower sets of words in $\mathcal{L}_n(X_\beta)$. Clearly, $\mathcal{F}(n) \le p(n)$, but in general $\mathcal{F}(n)$ is much smaller. Recall from Theorem 3.36 that $\mathcal{F}(n)$ is a bounded sequence if and only if the subshift is sofic. For β-shifts, we see in general linear growth of $\mathcal{F}(n)$.

Theorem 3.76. *For every β-shift (X_β, σ) with $\beta > 1$, we have $\mathcal{F}(n) = n+1$, except when* orb(1) *is finite; in this case, (X_β, σ) is sofic.*

Proof. This result comes from [**422**, Theorem 2.25], but we give a different dynamical proof. Set $\beta > 1$, and assume that $c = c_1c_2c_3\dots$ is the β-expansion of 1. Let $\mathcal{D}_0 = [0,1]$ and in general[11] $\mathcal{D}_n = [0, T_\beta^n(1)]$. First assume that all points $T_\beta^n(1)$ are distinct. The proof will be by induction.

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

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

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

If $\mathcal{D}_n = \mathcal{D}_k$ for some $k < n$ (say n is minimal with this property) then we get no new follower sets anymore, and $\mathcal{F}(m) = n+1$ for all $m \geq n$. As shown in Theorem 3.36, X_β is sofic in this case. □

Theorem 3.77. *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 topological entropy $h_{top}(T_\beta) = \max\{\log\beta, 0\}$, but we will give a purely symbolic proof.

Recall the β-expansion $c = c_1c_2\dots$ of 1 and the set of code words S from (3.14). By Proposition 3.68, every word in $\mathcal{L}(X_\beta)$ has the form[12]

$$s_1s_2\dots s_mc_1c_2\dots c_k \quad \text{for some (maximal) } s_1,\dots,s_m \in S,\ k \geq 0. \tag{3.16}$$

Let $p_\beta(n)$ and $p_{S^*}(n)$ be the number of words in $\mathcal{L}_n(X_\beta)$ and $\mathcal{L}_n(S^*)$ respectively. Since every word in S^* is a word in $\mathcal{L}(X_\beta)$, we have $p_{S^*}(n) \leq p_\beta(n)$. Conversely, by (3.16),

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

Therefore the exponential growth-rates are the same:

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

[11]This notation is derived the Hofbauer tower construction from Section 3.6.3 applied to β-transformations. If the orbit of 1 is infinite, then there are $n+1$ levels in the tower of height $\leq n$. The image of each n-cylinder under T_β^n is one of these, and therefore $\#\mathcal{F}(n) = n+1$. The same result holds for unimodal maps. More generally, for interval maps with $d+1$ branches, we have $\#\mathcal{F}(n) \leq dn+1$.

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

Now to compute the latter, we use generating functions:

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

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

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

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

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

This concludes the proof. □

One can ask whether β-shifts are density shifts and vice versa. After all, the one-sided β-shift (X_β, σ) is characterized as $\{x \in \mathcal{A}^{\mathbb{N}} : x \preceq_{\text{lex}} c\}$ for the $\preceq_{\text{lex}}$-maximal sequences c of Lemma 3.65 and the one-sided density shift $(X_{\boldsymbol{f}}, \sigma)$ is characterized as $\{x \in \mathcal{A}^{\mathbb{N}} : x \preceq_{\text{sum}} z\}$ for the $\prec_{sumord}$-maximal sequence z of (3.8). If $\sigma^n(y) \preceq_{\text{sum}} x$ for all $n \geq 0$ then $\sigma^n(y) \preceq_{\text{lex}} x$ for all $n \geq 0$, see [**506**, Lemma 8.1]. Therefore the shift-maximal sequence of a density shift is also shift-maximal for a β-shift, and every one-sided density shift is also a β-shift. The converse, however, is false. For example, $c = 302^\infty$ is shift-maximal w.r.t. $\preceq_{\text{lex}}$ but not w.r.t. $\preceq_{\text{sum}}$ because $\sigma^2(c) \not\preceq_{\text{sum}} c$ (in fact, these two sequences are not comparable). A way of finding (non-sofic) density β-shifts with $\beta \in [0,1]$[13] is as follows: Given a β-transformation $T_\beta : [01,1] \to [0,1]$, define $\bar{T}_\beta : [1-\frac{1}{\beta}, 1] \to [1-\frac{1}{\beta}, 1]$ by

$$\bar{T}_\beta(x) = \begin{cases} T_\beta(x) = \beta x & \text{if } 1-\frac{1}{\beta} \leq x \leq \frac{1}{\beta}, \\ 1-\frac{1}{\beta} & \text{if } \frac{1}{\beta} < x < \frac{2\beta-1}{\beta^2}, \\ T_\beta(x) = \beta x - 1 & \text{if } \frac{2\beta-1}{\beta^2} \leq x \leq 1, \end{cases} \tag{3.17}$$

see Figure 3.10.

[13] This is for simplicity of exposition; similar construction for $\beta > 2$ are of course possible.

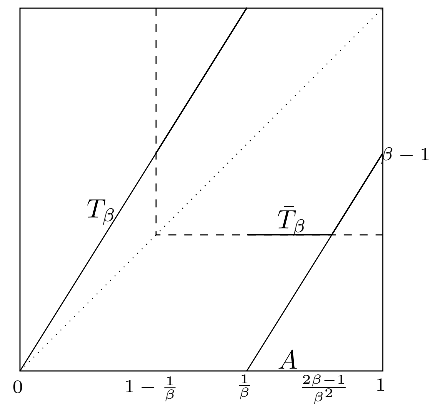


Figure 3.10. Turning a β-transformation into a Sturmian shift.

Since $\bar{T}_\beta(1 - \frac{1}{\beta}) = \bar{T}(1) = \beta - 1$, this map can be considered as a non-decreasing circle endomorphism on $[1 - \frac{1}{\beta}, 1]/_{1-\frac{1}{\beta}\sim 1}$, with plateau $A = [\frac{1}{\beta}, \frac{2\beta-1}{\beta^2}]$. If $\bar{T}^n_\beta(1) \notin A$ for all $n \geq 1$, then the rotation number $\alpha := \rho(\bar{T}_\beta) \notin \mathbb{Q}$, and $c = \boldsymbol{i}(1)$ is shift-maximal both w.r.t. $\preceq_{\text{lex}}$ and $\preceq_{\text{sum}}$, and it is in particular a sequence with maximum frequency of 1s. It is also a Sturmian sequence; more specifically, the itinerary of α for the circle rotation $R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ w.r.t. the partition $(0, \alpha]$ (with symbol 1) and $(\alpha, 1]$ (with symbol 0), cf. Definition 4.48. The canonical function of the density shift equals the Beatty sequence $\boldsymbol{f}(n) = \lceil n\alpha \rceil$.

3.6. Unimodal Subshifts

A **unimodal map** is a continuous map $f : [0,1] \to [0,1]$ with a single point $c \in (0,1)$, called **critical** or **turning point** such that $f|_{[0,c]}$ is increasing and $f|_{[c,1]}$ is decreasing[14]. This makes the **critical value** $f(c)$ the largest value that f assumes. Examples are the **quadratic family** $f_a(x) = ax(1-x)$, $a \in (0,4]$ and the family **tent maps** $T_s(x) = \min\{sx, s(1-x)\}$, $s \in (0,2]$, see Figure 3.11. It is customary to scale unimodal maps so that $f(0) = f(1) = 0$, but the interesting dynamics takes place on the **core** $[f^2(c), f(c)]$, provided $f^2(c) < c < f(c)$.

Unimodal maps are simple to define, but difficult to analyze. Before starting on the symbolic description, i.e., kneading theory, we give some background on the topological properties of unimodal maps[15].

[14] in our definition; in the frequently used family $\mathfrak{f}_c(z) = z^2 + c$, $c \in [-2, \frac{1}{2}]$, the roles of increasing and decreasing are reversed.

[15] Also for multimodal maps (i.e., continuous intervals with multiple critical points), symbolic dynamics have been studied. Much of the structure presented here has a direct analogue, albeit