Chapter 1

# First Examples and General Properties of Subshifts

Symbolic dynamics is concerned with spaces of (infinite) sequences of symbols. Such sequences can come from the symbolic description of a dynamical system, but they also have intrinsic interest. Symbol sequences are used to code messages, digitally process sound and images, and as the objects that computers process. The "dynamics", usually, but not exclusively, refers to a transformation $\sigma$ of such sequences in the form of a shift by one unit to the left. For example,

$$\begin{cases} \sigma(10011\dots) = 0011\dots & \text{for one-sided sequences,} \\ \sigma(011.10011\dots) = 0111.0011\dots & \text{for two-sided sequences.} \end{cases}$$

That is, for a right-infinite sequence, the first symbol disappears, and all other symbols move a place to the left. For a bi-infinite sequence, the dot that indicates position zero moves one place to the right. A closed $\sigma$-invariant subset of sequences over some fixed set of symbols (the alphabet), combined with this left-shift operation $\sigma$, is called a subshift. In this chapter, we give the basic notions and examples of subshifts and discuss the number and frequency of their subwords.

## 1.1. Symbol Sequences and Subshifts

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

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

Such symbol strings find applications in data-transmission and storage, linguistics, theoretic computer science, and also dynamical systems (symbolic dynamics). A finite string of letters, say $x_1 \cdots x_n \in \mathcal{A}^n$, is called a **word** or **block**. A $k$-word is a word of $k$ letters and $\epsilon$ is the empty word (of length 0). We use the notation $\mathcal{A}^k = \{k\text{-words in } \Sigma\}$ and

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

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

A **cylinder set**[2] is any set of the form

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

Intersections of cylinder sets are again cylinder sets. The cylinder sets form a basis of the **product topology** on $\Sigma$; i.e. a set is open in the product topology precisely if it can be written as a union of cylinder sets.

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

**Lemma 1.1.** *If* $2 \le \#\mathcal{A} < \infty$, *then* $\Sigma = \mathcal{A}^{\mathbb{N} \text{ or } \mathbb{Z}}$ *is a* **Cantor set** *(that is,* $\Sigma$ *is* (i) *compact,* (ii) *has no isolated points, and* (iii) *its connected components are points). If* $\#\mathcal{A} = \infty$, *then* $\Sigma$ *is not compact, but* (ii) *and* (iii) *still hold.*

**Proof.** (i) Set $\mathcal{A} = \{0, 1, \dots, N-1\}$ with discrete topology. Clearly $\mathcal{A}$ is compact, because it is finite. Compactness of $\Sigma$ then follows from Tychonov's Theorem.

(ii) No point is isolated, because, for arbitrary $x \in \Sigma$, the sequence $x^n$ defined as $x^n_i = x_i$ if $i \ne n$ and $x^n_n = x_n + 1 \pmod 1$ converges to $x$.

(iii) If $x \ne y$, set $n = \min\{|i| : x_i \ne y\}$; then $Z := \{x' \in X : x_i = x'_i \text{ for all } |i| \le n\}$ and $X \setminus Z$ are two clopen disjoint non-empty sets whose union is $X$. Thus $x$ and $y$ cannot belong to the same connected component.

If $\#\mathcal{A} = \infty$, then the collection $\{[a]\}_{a \in \mathcal{A}}$ is an open cover without finite subcover, so $\Sigma$ is not compact. $\square$

[1] We will rather not use this word, because of possible confusion with the factor of a subshift (= image under a sliding block code; see Section 1.4).

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

Shift spaces with product topology are metrizable. One of the usual[3] metrics that generates the product topology is

$$(1.1) \qquad d(x,y) = \begin{cases} 0 & \text{if } x = y \text{ or} \\ 2^{-m} & \text{for } m = \sup\{n \geq 0 : x_i = y_i \text{ for all } |i| < n\}; \end{cases}$$

so in particular $d(x,y) = 1$ if $x_0 \neq y_0$, and $\operatorname{diam}(\Sigma) = 1$. If $(x^k)_{k\in\mathbb{N}}$ is a sequence of sequences such that $x^k \to x$, then there is $k_0 \in \mathbb{N}$ such that $d(x^k, x) < 2^{-m}$ for every $k \geq k_0$. The definition of the metric $d$ implies that $x_i^k = x_i$ for all $|i| \leq m$. In other words, $x^k \to x$ means that $x^k_{[a,b]}$ is eventually equal to $x_{[a,b]}$ on every finite window $[a,b]$.

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

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

is invertible on $\mathcal{A}^{\mathbb{Z}}$ (with inverse $\sigma^{-1}(x)_i = x_{i-1}$) but non-invertible on $\mathcal{A}^{\mathbb{N}}$. We can use the $\varepsilon$-$\delta$ definition of continuity for $\delta = \varepsilon/2$ to show that $\sigma$ is uniformly continuous. This is even true if $\#\mathcal{A} = \infty$.

**Definition 1.2.** A pair $(X, \sigma)$ with $X \subset \Sigma$ and $\sigma$ the left-shift is a **subshift** (often called simply **shift**) if $X$ is closed (in product topology) and strongly shift-invariant; i.e. $\sigma(X) = X$. If $\sigma$ is invertible, then we also stipulate that $\sigma^{-1}(X) = X$. For example, if $\Sigma = \{0,1\}^{\mathbb{Z}}$ and

$$x = \ldots 000.111111\ldots,$$

then $X = \overline{\{\sigma^n(x) : n \geq 0\}}$ is not a subshift, because $x \in X$ but $\sigma^{-1}(x) \notin X$.

In Examples 1.3–1.6, we use $\mathcal{A} = \{0,1\}$.

**Example 1.3.** The set $X = \{x \in \Sigma : x_i = 1 \Rightarrow x_{i+1} = 0\}$ is called the **Fibonacci shift**[4]. It disallows sequences with two consecutive 1's. This Fibonacci shift is an example of a subshift of finite type (SFT); see Section 3.1. The collection $X$ can be represented by a graph in multiple ways:

- $X$ is the collection of labels of infinite paths through the **vertex-labeled** graph in Figure 1.1 (left). Labels are given to the vertices of the graph, and no label is repeated.
- $X$ is the collection of labels of infinite paths through the **edge-labeled** graph in Figure 1.1 (right). Labels are given to the arrows of the graph, and labels can be repeated (different arrows with the same label can occur).

---

[3]Other metrics are $d'(x,y) = \sum_i |x_i - y_i| 2^{-|i|}$ or $d'(x,y) = \frac{1}{m}$ with $m$ as in (1.1). They are equivalent to $d(x,y)$: the former in the sense that there is some $C$ such that $\frac{1}{C} d(x,y) \leq d'(x,y) \leq C d(x,y)$ for all $x, y \in \Sigma$, the latter in the weaker sense that the embedding $i : (\Sigma, d') \to (\Sigma, d)$ as well as its inverse $i^{-1}$ are uniformly continuous. In either case, they generate the same topology.

[4]Warning: There is also a Fibonacci substitution shift = Fibonacci Sturmian shift (see Example 4.3), which is different from this one.


**Figure 1.1.** Transition graphs: vertex-labeled and edge-labeled.

**Example 1.4.** $X_{\text{even}} \subset \{0,1\}^{\mathbb{N}}$ is the collection of infinite sequences in which the 1's appear only in blocks of even length and also $1111 \cdots \in X$. We call $X_{\text{even}}$ the **even shift**. Similarly, the **odd shift** $X_{\text{odd}}$ is the collection of infinite sequences in which the 0's appear only in blocks of odd length and also $0000 \cdots \in X$; see Figure 1.2.

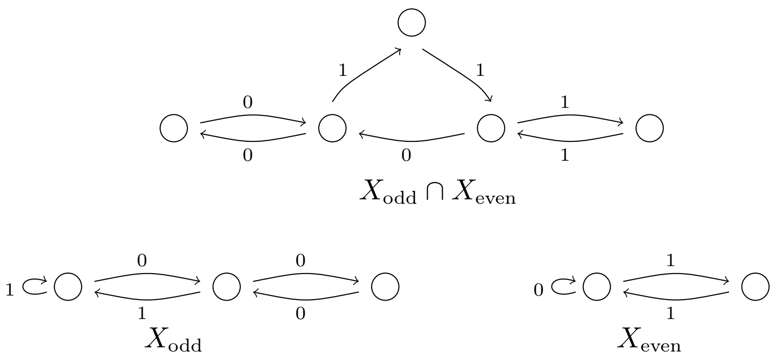


**Figure 1.2.** Edge-labeled graphs for $X_{\text{odd}}$, $X_{\text{even}}$, and $X_{\text{odd}} \cap X_{\text{even}}$.

**Example 1.5.** Let $S$ be a non-empty subset of $\mathbb{N}$. Let $X \subset \{0,1\}^{\mathbb{Z}}$ be the collection of sequences in which the appearance of two consecutive 1's occur always $s$ positions apart for some $s \in S$. Hence, sequences in $X$ have the form

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

where $s_i \in S$ for each $i \in \mathbb{Z}$. If $\#S = \infty$, then allowed sequence can also end and/or start with $0^\infty$. This space is called the $S$-gap shift; see Section 3.7. For $S = \{2,3,4,\dots\}$ we get the Fibonacci SFT, and for $S = \{1,2\dots\}$ we get the Fibonacci SFT with symbols 0 and 1 interchanged.

**Example 1.6.** The **Thue-Morse substitution**[5] $\chi_{\mathrm{TM}} : \{0,1\} \to \{0,1\}^*$ is a special substituion, see Section 4.2, defined by

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

and extended on longer words by concatenation. It has **two** fixed points:

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

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

These sequences make their appearance in many settings in combinatorics and elsewhere; cf. [**19**, **561**]. For instance, the $n$-th entry of $\rho^0$ (where we start counting at $n = 0$) is the parity of the number of 1's in the binary expansion of $n$. The Thue-Morse sequence $\rho^i$ can be defined by the relation $\rho^i_0 = i$, $\rho^i_{2n} = \rho^i_n$, and $\rho^i_{2n+1} = 1 - \rho^i_n$. Also, if we have a sequence $(P_k)_{k\geq 0}$ of decreasing quality (e.g. rugby players) which we want to divide over two teams $T_0$ and $T_1$, so that the teams are as close in strength as possible, then we assign $P_k$ to team $T_i$ if $i$ is the $k$-th digit of $\rho^0$ (or equivalently, of $\rho^1$). Also the series $\sum_{n\geq 1} \rho^0_n 2^{-n} = 1 - \sum_{n\geq 1} \rho^1_n$ has been proved to sum to a transcendental number; see e.g. [**20**, Theorem 13.4.2].

**Example 1.7.** The alphabet $\mathcal{A}$ consists of brackets $(, ), [, ]$ and the allowed words are those (that can be extended to words) consisting of brackets that are properly paired and unlinked. So [ ( [ ] ) ] and ( ( ) [ ] ) are allowed, but [ ( ] and ( [ ) ] are not. The subshift $(X, \sigma)$ of which these are the allowed subwords is called the **Dyck shift**; see Section 3.10.

## 1.2. Word-Complexity

**Definition 1.8.** Given a subshift $X$, the collection

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

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

**Definition 1.9.** The function $p := p_X : \mathbb{N} \to \mathbb{N}$ defined by $p(n) = \#\mathcal{L}_n(X)$ is called the **word-complexity** of $X$.

**Example 1.10.** For the Fibonacci SFT of Example 1.3, let

$$F_n = \#\{w \in \mathcal{L}_n(X) : w_n = 0\}.$$

Then $F_1 = 1$, $F_2 = 2$, and $F_{n+1} = F_n + F_{n-1}$ for $n \geq 3$ because $F_n$ is the cardinality of the set of $n+1$-words ending in 00 and $F_{n-1}$ is the

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

cardinality of the set of $n+1$-words ending in 010. Therefore the $F_n$'s are the Fibonacci numbers. The same argument gives $p(1) = 2 = F_2$ and $p(n) = F_n + F_{n-1} = F_{n+1}$ for $n \geq 2$.

**1.2.1. Sublinear and Polynomial Complexity.** We start with some terminology and a useful proposition.

**Definition 1.11.** We call a word $u \in \mathcal{L}_n(X)$ over the alphabet $\mathcal{A} = \{0,1\}$

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

Note, however, that there are different types of bi-special words $u$ depending on how many of the words $0u0, 0u1, 1u0$, and $1u1$ are allowed. If only one choice of $0u$ or $1u$ is right-special and only one choice of $u0$ and $u1$ is left-special, then $u$ is a **regular bi-special word**. For larger alphabets, we can formulate similar definitions and, naturally, there are more types of left/right/bi-special words.

Clearly

$$\begin{aligned} p(n+1) - p(n) &= \#\{\text{left-special words of length } n\} \\ &= \#\{\text{right-special words of length } n\}. \end{aligned}$$

The following result goes back to Morse & Hedlund [**425**].

**Proposition 1.12.** If the word-complexity of a subshift $(X, \sigma)$ satisfies $p(n) \leq n$ for some $n$, then $(X, \sigma)$ consists of finitely many periodic sequences.

**Proof.** If $p(1) = 1$, then $X = \{a^\infty\}$ is obviously periodic. So assume $p(1) \geq 2$. Since $p$ is non-decreasing, the assumption of this proposition implies that there is a minimal $n$ such that $p(n-1) = p(n) = n$. Hence there are no right-special words of length $n-1$. Start with a word $u \in \mathcal{L}_{n-1}(X)$; there is only one way to extend it to the right by one letter, say to $ua$. Then the $n-1$-suffix of $ua$ can also be extended to the right by one letter in only one way. Continue this way, until, after at most $p(n) = n$ steps, we obtain an $n$-suffix that we have seen before. All strings need to be extendible to the left in some allowed way (otherwise $\sigma(X) \neq X$). When extending $u$ to the left symbol by symbol, we need to arrive at the same periodic pattern, because otherwise $p(n+1) > n+1$. Therefore $X$ consists of (at most $n$) periodic sequences. □

This proposition shows that the minimal complexity of interest is $p(n) = n+1$, because if $p(n) \leq n$ for some $n$, then $X$ consists of at most $n$ periodic sequences. We say that $(X, \sigma)$ is of **sublinear complexity** if there is a constant $C$ such that $p(n) \leq Cn$. Sturmian sequences (see Section 4.3) have $p(n) = n+1$; in fact all recurrent words with this word-complexity are Sturmian. There are further possibilities for non-recurrent subshifts. The sequences

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

both have $p(n) = n+1$. They are not uniformly recurrent, but asymptotically fixed for $n \to \pm\infty$. Ormes & Pavlov [**435**, Theorems 1.2 & 1.3] showed that for non-recurrent shifts $(X, \sigma)$ that are not asymptotically periodic in both directions, $\liminf_n p(n)/n \geq \frac{3}{2}$ and that this bound is sharp, as is demonstrated by

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

for a carefully chosen increasing sequence of gaps $(n_i)_{i\geq 1}$. In fact, given any non-decreasing function $g : \mathbb{N} \to \mathbb{N}$ that tends to infinity, there is $x \in X$ such that $p_{\{x\}}(n) := \#\{w \text{ is subword of } x : |w|\} = n < \frac{3}{2}n + g(n)$. In further detail, if a transitive[6] shift $(X, \sigma)$ with a recurrent point contains $m$ minimal subsystems, of which $m_\infty$ are infinite, then

$$\limsup_{n\to\infty} p_X(n) - (m + m_\infty + 1)n = \infty, \qquad \liminf_{n\to\infty} p_X(n) - (m + m_\infty)n = \infty,$$

and these bounds are sharp. The second estimate holds also without the existence of a recurrent point. See [**230**], specifically Theorems 1.2 and 1.3.

Symbolic spaces associated with interval exchanges transformations on $k$ intervals have $p(n) = (d-1)n + 1$; see Proposition 4.80. The Chacon substitution shift and primitive Chacon substitution shift (see Example 1.27) have word-complexity $p(n) = 2n-1$ (for $n \geq 2$) and $p(n) = 2n+1$; see [**243**]. For many subshifts, $p_X(n)/n$ is bounded in $n$ but hard to compute exactly; often $\lim_n p(n)/n$ doesn't exist. For instance, the word-complexity of the Thue-Morse shift (i.e. the closure $\overline{\{\sigma^n(\rho_{\mathrm{TM}}) : n \in \mathbb{N}_0\}}$ of Example 1.6) is

$$p(n) = \begin{cases} 3 \cdot 2^m + 4r & \text{if } 0 \leq r < 2^{m-1}, \\ 4 \cdot 2^m + 2r & \text{if } 2^{m-1} \leq r < 2^m, \end{cases} \tag{1.2}$$

where $n = 2^m + r + 1$; see [**115, 406**]. In [**129**], the word-complexity of certain (Fibonacci-like) unimodal restrictions to the critical $\omega$-limit set are computed.

The following curious result is due to Heinis; see [**150, 311**].

**Proposition 1.13.** If $\lim_n p_X(n)/n$ exists and is finite, then it has to be an integer.

All substitution shifts, in fact all linearly recurrent shifts, have sublinear complexity; see Theorem 4.4.

[6] See Definition 1.18 below.

The **polynomial growth rate** is defined as $r = \lim_n \frac{\log p(n)}{\log n}$. Naturally, linear complexity implies $r = 1$, but every $r \in \{0\} \cup [1, \infty]$ is possible. Subshifts with polynomial growth rate $r > 1$ are less studied, but for example symbolic spaces for polygonal billiards on $d$-dimensional billiard tables can have polynomial growth rate $r = d$.

**1.2.2. Exponential Complexity.** Anticipating the definition for general dynamical systems in Section 2.4, for subshifts, the **topological entropy** is the **exponential growth rate** of the word-complexity:

$$h_{\text{top}}(\sigma) = \lim_{n \to \infty} \frac{1}{n} \log p_X(n). \tag{1.3}$$

To show that the limit in (1.3) exists, we need one more notion and one well-known lemma.

**Definition 1.14.** We call a real-valued sequence $(a_n)_{n \geq 1}$ **subadditive** if

$$a_{m+n} \leq a_m + a_n \qquad \text{for all } m, n \geq 1.$$

Analogously, $(a_n)_{n \geq 1}$ is **superadditive** if $a_{m+n} \geq a_m + a_n$ for all $m, n \in \mathbb{N}$.

**Lemma 1.15** (Fekete's Subadditive Lemma)**.** *If $(a_n)_{n \geq 1}$ is subadditive, then $\lim_n \frac{a_n}{n} = \inf_{r \geq 1} \frac{a_r}{r}$ (possily $-\infty$). Analogously, if $(a_n)_{n \geq 1}$ is superadditive, then $\lim_n \frac{a_n}{n} = \sup_{r \geq 1} \frac{a_r}{r}$ (possily $-\infty$).*

**Proof.** Every integer $n$ can be written as $n = i \cdot r + j$ for $0 \leq j < r$. Therefore

$$\limsup_{n \to \infty} \frac{a_n}{n} = \limsup_{i \to \infty} \frac{a_{i \cdot r + j}}{i \cdot r + j} \leq \limsup_{i \to \infty} \frac{i a_r + a_j}{i \cdot r + j} = \frac{a_r}{r}.$$

This holds for all $r \in \mathbb{N}$, so we obtain

$$\inf_{r \in \mathbb{N}} \frac{a_r}{r} \leq \liminf_{n \to \infty} \frac{a_n}{n} \leq \limsup_{n \to \infty} \frac{a_n}{n} \leq \inf_{r \in \mathbb{N}} \frac{a_r}{r},$$

as required. The proof for superadditive sequences goes likewise. □

**Remark 1.16.** A positive sequence $(a_n)_{n \in \mathbb{N}}$ is **submultiplicative** if $a_{m+n} \leq a_m a_n$ (and **supermultiplicative** if $a_{m+n} \geq a_m a_n$). By taking logarithms, we can turn a sub/supermultiplicative sequence into a sub/superadditive one, and this suffices for our purposes.

We devote separate chapters to subshifts of positive and subshifts of zero entropy, because most[7] tend to have different topological properties such as topological mixing, existence and number of periodic orbits, shadowing; see

[7] At least most shifts we encounter in this book, but it is not a strict rule. For example, Petersen's shifts mentioned below Theorem 2.77 has zero entropy and is topologically mixing, while Grillenberger's **[287]** construction gives minimal shifts of positive entropy (and therefore lacking periodic orbits), with further examples among Toeplitz shifts (see Theorem 4.94) and $\mathcal{B}$-free shifts (Section 4.6).

Chapter 2. The maximal entropy of a subshift on $N$ letters is $\log N$, and this is achieved by the full shift $(\{0,\dots,N-1\}^{\mathbb{N}},\sigma)$. One can ask whether all intermediate values between 0 and $\log N$ can be achieved as topological entropy for some subshift. As we shall see later, this is not true for the class of subshift of finite type or the sofic shifts, because the entropy is then equal to the logarithm of the leading eigenvalue of some integer matrix, so logarithms of algebraic numbers and, in fact Perron numbers; see [**397**] and (the text below) Definition 8.4.

On the other hand, the topological entropy of $\beta$-shifts $(X_\beta,\sigma)$ can take any non-negative value $\geq 0$, because $h_{\text{top}}(X_\beta)=\log\beta$. Also within the class of gap shift you can achieve every value of the entropy, as can be derived from Theorem 3.114. Some specific constructions of subshifts of a chosen entropy can be found among spacing shifts; see [**380**] and Section 3.8.

**Remark 1.17.** For many subshifts in $\mathcal{A}^{\mathbb{N}\text{ or }\mathbb{Z}}$, the topological entropy can be computed exactly, but not so for subshift in $\mathcal{A}^{\mathbb{Z}^d}$, i.e. cellular automata. Even for the simplest direct generalization of the Fibonacci SFT, namely 0-1-patterns on $\mathbb{Z}^2$ where no two 1's occur directly next to each other (horizontally or vertically), the entropy $\lim_{m,n\to\infty}\frac{1}{mn}\log p_x(m,n)$ is unknown. There are however numerical approximations (e.g. for this example, the entropy equals $0.5878116\dots$ which these digits certainly correct; see [**251**]) and characterizations of which values can occur; see e.g. [**259**, **260**, **289**, **313**, **314**, **399**].

## 1.3. Transitive and Synchronized Subshifts

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

**Definition 1.18.** A subshift $X$ is **transitive** or **irreducible** if for every $u,w\in\mathcal{L}(X)$, there is $v\in\mathcal{L}(X)$ such that $uvw\in\mathcal{L}(X)$.

This definition does not automatically produce periodic sequences, because also if $u=w$, so $uvu\in\mathcal{L}(X)$, then it doesn't follow that $uvuvu\in\mathcal{L}(X)$.

**Definition 1.19.** A subshift $(X,\sigma)$ is called **synchronized** if it is transitive and there is a word $v\in\mathcal{L}(X)$ (called **(intrinsically) synchronizing word**[8]) such that $uv,vw\in\mathcal{L}(X)$ implies $uvw\in\mathcal{L}(X)$. In other words, the appearance of $v$ cancels the influence of the past.

**Theorem 1.20.** *A synchronized shift $(X,\sigma)$ has a dense set of periodic points. If $X$ is not periodic itself, then the entropy $h_{\text{top}}(X,\sigma)>0$.*

[8] Kitchens in [**364**] calls it a **magic Markov word**.

**Proof.** Let $v$ be a synchronizing word and let $x \in \mathcal{L}(X)$ be arbitrary. Since a synchronized $X$ is, by definition, transitive, there are words $u, w \in \mathcal{L}(X)$ such that $xuv \in \mathcal{L}(X)$ and $vwx \in \mathcal{L}(X)$. Now the infinite periodic word $(xuvw)^\infty$ belongs to $X$. Since $x \in \mathcal{L}(X)$ was arbitrary, denseness of periodic words follows.

Next use transitivity again to find distinct words $u, u', v \in \mathcal{L}(X)$ such that $vuv, vu'v \in \mathcal{L}(X)$. Let $X'$ be the subshift constructed by free concatenations of $vu$ and $vu'$; clearly $X'$ is a subshift of $X$, and for $N = \max\{|v| + |u|, |v| + |u'|\}$ we find $p_{X'}(nN) \geq 2^n$. Hence, $h_{\text{top}}(X', \sigma) > \frac{1}{N} \log 2$. □

**Example 1.21.** The Fibonacci SFT (see Example 1.3) has synchronizing word 0. In this case, every 1 is preceded and succeeded by a 0. If we swap 0's and 1's, then we obtain the $S$-gap shift with gap sizes 1 and 2. Hence $h_{\text{top}}(X, \sigma) = \log \lambda$ where $\lambda^{-1} + \lambda^{-2} = 1$, so $\lambda = \frac{1+\sqrt{5}}{2}$. This is in agreement with Example 1.10.

## 1.4. Sliding Block Codes

**Definition 1.22** (Sliding Block Code)**.** Let $\mathcal{A}$ and $\tilde{\mathcal{A}}$ be alphabets. A map $\pi : \mathcal{A}^{\mathbb{Z}} \to \tilde{\mathcal{A}}^{\mathbb{Z}}$ is called a **sliding block code** or **local rule** of **window size** $2N + 1$ if there is a function $f : \mathcal{A}^{2N+1} \to \tilde{\mathcal{A}}$ such that $\pi(x)_i = f(x_{i-N} \cdots x_{i+N})$.

In other words, we have a window[9] of width $2N + 1$ put on the sequence $x$. If it is centered at position $i$, then the recoded word $y = \pi(x)$ will have at position $i$ the $f$-image of what is visible in the window. After that we slide the window to the next position and repeat.

**Theorem 1.23** (Curtis–Hedlund–Lyndon[10])**.** *Let $X$ and $Y$ be subshifts over finite alphabets $\mathcal{A}$ and $\tilde{\mathcal{A}}$, respectively. A continuous map $\pi : X \to Y$ commutes with the shift (i.e. $\sigma \circ \pi = \pi \circ \sigma$) if and only if $\pi$ is a sliding block code.*

If $\pi : X \to Y$ is a homeomorphism, then we call $(X, \sigma)$ and $(Y, \sigma)$ **conjugate**.

**Proof.** First assume that $\pi$ is continuous and commutes with the shift. For each $a \in \tilde{\mathcal{A}}$, the cylinder $[a] = \{y \in Y : y_0 = a\}$ is clopen, so $V_a := \pi^{-1}([a])$ is clopen too. Since $V_a$ is open, it can be written as the union of cylinders, and since $V_a$ is closed (and hence compact) it can be written as the finite union of cylinders: $V_a = \bigcup_{i=1}^{r_a} U_{a,i}$. Let $N$ be so large that every $U_{a,i}$ is

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

[10] Curtis and Lyndon were working for the military at the time, so their work was "classified", and the paper was published under Hedlund's name only, **[308]**.

the union of $2N+1$-cylinders $U_{a,i,j}$; i.e. each $U_{a,i,j}$ is determined by a word $x_{-N} \cdots x_N$. This makes $2N+1$ a sufficient window size and there is a function $f : \mathcal{A}^{2N+1} \to \tilde{\mathcal{A}}$ such that $\pi(x)_0 = f(x_{-N} \cdots x_N)$. By shift-invariance, $\pi(x)_i = f(x_{i-N} \cdots x_{i+N})$ for all $i \in \mathbb{Z}$.

Conversely, assume that $\pi$ is a sliding block code of window size[11] $2N+1$. Take $\varepsilon = 2^{-M} > 0$ arbitrary and $\delta = \varepsilon 2^{-N}$. If $d(x,y) < \delta$, then $x_i = y_i$ for $|i| \le M+N$. By the construction of the sliding block code, $\pi(x)_i = \pi(y)_i$ for all $|i| \le M$. Therefore $d(\pi(x), \pi(y)) < \varepsilon$, so $\pi$ is continuous (in fact uniformly continuous). □

**Exercise 1.24.** Give a surjective sliding block code between the Fibonacci SFT and the even subshift (see Examples 1.3 and 1.4).

**Corollary 1.25.** If $(X,\sigma)$ and $(Y,\sigma)$ are conjugate shifts, then there is $N$ such that $p_X(k-N) \le p_Y(k) \le p_X(k+N)$ for $k > N$.

**Proof.** Let $2N+1$ be the maximal window size among the sliding block codes from $X$ to $Y$ and from $Y$ to $X$. Then every $k$-word in $Y$ is obtained from an $N+k$-word in $X$, so $p_Y(k) \le p_X(N+k)$. Replacing the role of $X$ and $Y$ gives the other inequality. □

**Exercise 1.26.** If $\psi : X \to Y$ is an onto sliding block code which is $k$-to-one for some fixed $k$, show that $h_{\mathrm{top}}(X,\sigma) = h_{\mathrm{top}}(Y,\sigma)$.

**Example 1.27.** The following substitutions (see Section 4.2) are called the **Chacon substitution** and **primitive Chacon substitution**:

$$\chi_{\mathrm{chac}} : \begin{cases} 0 \to 0010, \\ 1 \to 1 \end{cases} \qquad \text{and} \qquad \chi_{\mathrm{Chac}} : \begin{cases} 0 \to 0021, \\ 1 \to 021, \\ 2 \to 21, \end{cases} \tag{1.4}$$

with fixed points

$$\begin{aligned} \rho_{\mathrm{chac}} &= 0010\ 0010\ 1\ 0010\ 0010001010\ 1\ 0010\ldots, \\ \rho_{\mathrm{Chac}} &= 0021\ 0021\ 21\ 021\ 0021002121021\ldots. \end{aligned}$$

They can be transformed into each other using the sliding block code

$$f : \begin{cases} 0\underline{0}a \to 0, \\ 1\underline{0}a \to 1, \\ \ \underline{1}\ \to 2, \end{cases} \quad a \in \{0,1\} \qquad \text{and} \qquad f^{-1} : \begin{cases} \underline{0} \to 0, \\ \underline{1} \to 0, \\ \underline{2} \to 1, \end{cases}$$

and this extends to the shift orbit closures

$$X_{\mathrm{chac}} = \overline{\{\sigma^n(\rho_{\mathrm{chac}}) : n \ge 0\}} \qquad \text{and} \qquad X_{\mathrm{Chac}} = \overline{\{\sigma^n(\rho_{\mathrm{Chac}}) : n \ge 0\}}.$$

[11] If $(X,\sigma)$ is a one-sided subshift, with window size $[0,N]$ (so no memory, only anticipation), then this part of the proof still works. The first part of the proof fails: one must first extend $(X,\sigma)$ to a two-sided shift before the Curtis–Hedlund–Lyndon Theorem can be applied in full.

Therefore, these substitution shifts are topologically conjugate, although the word-complexities are different: $p_{X_{\text{chac}}}(1) = 2$, $p_{X_{\text{chac}}}(n) = 2n - 1$ for $n \geq 2$ and $p_{X_{\text{Chac}}}(n) = 2n + 1$ for $n \geq 1$; see [**243**].

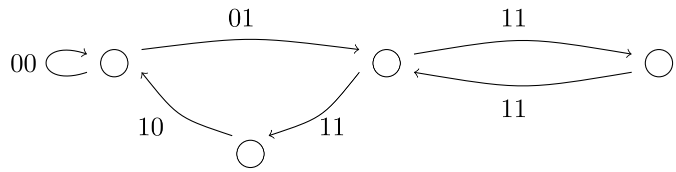


**Figure 1.3.** The edge-labeled transition graph of the 2-block even shift.

Each subshift $(X, \sigma)$ over an alphabet $\mathcal{A}$ is conjugate to an **$\ell$-block shift**, where the alphabet $\tilde{\mathcal{A}} \subset \mathcal{A}^\ell$ consists of the words in $\mathcal{L}_\ell(X)$ and $a, b \in \tilde{\mathcal{A}}$ can only follow each other if the $\ell - 1$-suffix of $a$ coincides with the $\ell - 1$-prefix of $b$. For instance, if $(X_{\text{even}}, \sigma)$ is the even shift, then $\tilde{\mathcal{A}} = \{00, 01, 10, 11\}$ and the edge-labeled transition graph is given in Figure 1.3. Note that to recover the coding of paths in the original shift, we use only the first letters of the codes at the edges.

Taking a block shift generally doesn't change the nature of the shift (SFTs remain SFTs, substitution shifts remain substitution shifts, see Section 6.3.2, etc.). Block shifts can be used to shrink the window size of sliding block codes; see [**398**, Proposition 1.5.12].

**Proposition 1.28.** If $\pi$ is a sliding block code between $X$ and $Y$ of window size $2N + 1$, then there is a sliding block code $\tilde{\pi}$ (of window size 1) between the $2N + 1$ block shift $\tilde{X}$ of $X$ and $Y$.

**Proof.** We do the proof for invertible shifts; the one-sided shifts work as well, but then we cannot allow a memory in the sliding block code, only anticipation. Let $\phi : X \to \tilde{X}$ be the sliding block code that recodes the $2N + 1$-blocks in $X$ into the letters of $\tilde{\mathcal{A}}$; i.e. $\phi(x)_i = f(x_{i-N} \cdots x_{i+N})$. Then $\tilde{\pi} = \pi \circ \phi^{-1}$ is the required sliding block code. □

## 1.5. Word-Frequencies and Shift-Invariant Measures

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

$$f_w(x) = \lim_{n\to\infty} \frac{1}{n} \#\{0 \leq i < n : x_i \cdots x_{i+|w|-1} = w\}. \tag{1.5}$$

The question of whether the limit exists and to what extent it depends on $x$ is answered by Birkhoff's Ergodic Theorem 6.13. For this we need a measure $\mu$

that assigns a number to every cylinder set, according to the following rules:

(i) $0 \le \mu([w]) \le 1$ for every cylinder $[w]$;

(ii) $\mu(\varnothing) = 0, \; \mu(X) = 1$;

(iii) $\mu(\bigcup_i [w_i]) = \sum_i \mu([w_i])$ for all disjoint cylinders $[w_1], [w_2], \dots$.

The Kolmogorov Extension Theorem (see e.g. [**56**, Section 21.10]) implies that $\mu$ can be extended uniquely for every set in the $\sigma$-algebra $\mathcal{B}$ generated by the cylinder sets. Thus, if $x \in X$ is such that $f_w(x)$ exists for every $w \in \mathcal{L}(X)$, then there is a shift-invariant probability measure $\mu$ such that $\mu([w]) = f_w(x)$ for all $w \in \mathcal{L}(X)$.

**Remark 1.29.** The Kolmogorov Extension Theorem is about extending probability measures $\mu_n$ on finite Cartesian products $X^n$ (equipped with an $n$-fold-product $\sigma$-algebra) to a measure on the infinite product $X^\infty$ (equipped with an infinite-product $\sigma$-algebra). That is, if $\mu_{n+1}(A \times X) = \mu_n(A)$ for every $n \in \mathbb{N}$ and measurable set $A \subset X^n$, then there is a unique probability measure $\mu$ on $X^\infty$ such that $\mu(A \times X^\infty) = \mu_n(A)$ for every $n \in \mathbb{N}$ and measurable set $A \subset X^n$.

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

**Definition 1.30.** A measure $\mu$ on a subshift $(X, \sigma)$ is called **invariant** or **shift-invariant** if $\mu(B) = \mu(\sigma^{-1}B)$ for all $B \in \mathcal{B}$.

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

Birkhoff's Ergodic Theorem 6.13 implies that if $\mu$ is an ergodic shift-invariant probability measure on $(X, \sigma)$, then for $\mu$-a.e. $x \in X$, $f_w(x) = \mu([w])$ for all $w \in \mathcal{L}(X)$. However, if $f_w(x)$ exists for every $w \in \mathcal{L}(X)$, the associated measure need not be ergodic. For example, the sequence

$$x = 1001110000111110000001111111 \cdots 0^n 1^{n+1} \cdots$$

is associated to a combination of Dirac measures $\frac{1}{2}(\delta_{0^\infty} + \delta_{1^\infty})$, and this measure is clearly not ergodic.

**Remark 1.31.** Regardless of whether $\mu$ is ergodic or not, we call it a **generic measure** if there is a point $x \in X$ such that the frequency $f_w(x) = \mu([w])$ for all $w \in \mathcal{L}(X)$.

**Definition 1.32.** Let $\mathcal{A} = \{1, 2, \dots, d\}$ and $X = \mathcal{A}^{\mathbb{N} \text{ or } \mathbb{Z}}$ be the full shift space. Let $\boldsymbol{p} = (p_1, \dots, p_d)$ be a probability vector; i.e. $p_i \geq 0$ and $p_1 + \cdots + p_d = 1$. The **product measure** that assigns to every cylinder set

$$\mu_p([x_k \cdots x_l]) = p_{x_k} p_{x_{k+1}} \cdots p_{x_l}$$

is called the **$\boldsymbol{p}$-Bernoulli measure**. The measure can be extended to the Borel $\sigma$-algebra by means of the Kolmogorov Extension Theorem. Each $\boldsymbol{p}$-Bernoulli measure is shift-invariant.

Bernoulli measures[12] are a basic tool in probability theory. For example, encode a sequence of coin-flips by, say, $x_i = 0$ if the $i$-th flip gives a "head", and $x_i = 1$ if the $i$-th flip gives a "tail". This gives a sequence $x \in \{0, 1\}^{\mathbb{N}}$. If the coin has a bias, say "head" come up with probability $p > \frac{1}{2}$ and "tail" with probability $q = 1 - p < \frac{1}{2}$, then the probability of a word can be computed by multiplying probabilities; e.g. the probability $\mathbb{P}(x_1 x_2 x_3 x_4 = 0010) = p^3 q$.

**Definition 1.33.** A subshift $(X, \sigma)$ is **uniquely ergodic** if it admits only one invariant probability measure. If $(X, \sigma)$ is both uniquely ergodic and minimal, it is called **strictly ergodic**. (This should not be confused with **intrinsically ergodic** which means that there is a unique measure of maximal entropy; see Definition 6.70.)

The full shift is obviously not uniquely ergodic; it has for instance a Bernoulli measure for every probability vector $\boldsymbol{p}$ and neither are SFTs, sofic shifts, or $\beta$-shifts (which are, in fact, intrinsically ergodic). The Thue-Morse shift on the other hand is uniquely ergodic. Clearly, unique ergodicity implies intrinsic ergodicity, but not the other way around. It follows from Oxtoby's Theorem 6.20 that a recurrent subshift $(X, \sigma)$ is uniquely ergodic if and only if $f_w(x)$ exists and is the same for every $x \in X$. In this case, the convergence in the limit (1.5) is uniform in $x$.

## 1.6. Symbolic Itineraries

An important application of symbol sequences is to use them to represent trajectories of dynamical systems (see Section 2.1 for an introduction to dynamical systems). It was probably Hadamard who first used this idea in his studies of geodesic flows [**296**]. Over 40 years later, Morse & Hedlund's [**425**] wrote the first monograph on symbolic dynamics. If $T : X \to X$ is some map on a topological space, denote the $n$-fold compositions by $T^n = T \circ \cdots \circ T$ (and $T^{-n}$ is the $n$-fold composition of $T^{inv}$ if $T$ is invertible). Symbolic

[12]Named after Jacob Bernoulli, one of the mathematicians' family originating from Basel who wrote the book "Ars conjectandi", one of the first books on probability theory.

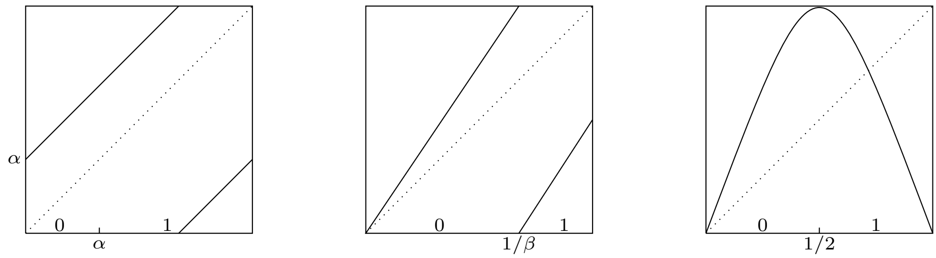


**Figure 1.4.** A circle rotation $R_\alpha(x) = x + \alpha \pmod 1$, a $\beta$-transformation $T_\beta(x) = \beta x \pmod 1$, and the quadratic map $f_4(x) = 4x(1-x)$.

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

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

If $T$ is invertible, then we can extend itineraries to sequences in $\mathcal{A}^{\mathbb{Z}}$. It is clear that $\boldsymbol{i} \circ T(x) = \sigma \circ \boldsymbol{i}(x)$. Therefore, $\boldsymbol{i}(X)$ is $\sigma$-invariant and if $T : X \to X$ is onto, then $\sigma(\boldsymbol{i}(X)) = \boldsymbol{i}(X)$. In general, however, $\boldsymbol{i}(X)$ is not closed, so we need to take the closure before it can be called a subshift. Using this subshift, we can often show the abundance of different trajectories (periodic or with other properties) of the original system $(X, T)$.

**Example 1.34.** Let $X$ be the closure of the collection of symbolic itineraries of a circle rotation $R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ over angle $\alpha \in [0,1] \setminus \mathbb{Q}$; see Figure 1.4 (left). We use the partition $\mathcal{J} = \{J_0, J_1\}$ with $J_0 = [0, \alpha)$ and $J_1 = [\alpha, 1)$. Hence, if $y \in \mathbb{S}^1$ and $n \in \mathbb{Z}$, then

$$\boldsymbol{i}(y)_n = \begin{cases} 0 & \text{if } R^n(y) \in [0, \alpha), \\ 1 & \text{if } R^n(y) \in [\alpha, 1). \end{cases}$$

Slightly different coding comes from the partition $\{(0, \alpha], (\alpha, 1]\}$, but the closure of $\boldsymbol{i}(\mathbb{S}^1)$ is the same for both partitions. The resulting shift is called a **Sturmian shift**; see Definition 4.60.

**Example 1.35.** Consider the $\beta$-transformation $T_\beta : [0,1] \to [0,1]$, $T_\beta(x) = \beta x \bmod 1$ (see Figure 1.4 (middle)), and $\boldsymbol{i}(x)_n = a$ if $T_\beta^n(x) \in J_a := [\frac{a}{n}, \frac{a+1}{\beta})$. The closure of $\boldsymbol{i}([0,1])$ is called a **$\beta$-shift**; see Section 3.5.

**Example 1.36.** Let $X = [0,1]$ and $T(x) = f_4(x) = 4x(1-x)$; see Figure 1.4 (right). Let $J_0 = [0, \frac{1}{2}]$ and $J_1 = (\frac{1}{2}, 1]$. Then $\boldsymbol{i}(X)$ is not closed, because there is no $x \in [0,1]$ such that $\boldsymbol{i}(x) = 1100000\dots$, while $1100000\cdots = \lim_{x \searrow \frac{1}{2}} \boldsymbol{i}(x)$. Naturally, redefining the partition to $J_0 = [0, \frac{1}{2})$ and $J_1 = [\frac{1}{2}, 1]$ doesn't help, because then there is no $x \in [0,1]$ such that $\boldsymbol{i}(x) = 0100000\dots$, while $0100000\cdots = \lim_{x \nearrow \frac{1}{2}} \boldsymbol{i}(x)$. This shows that we have to take the closure to obtain a subshift.

There are other ways of coding in the literature to obtain a subshift:

- Assign a different symbol (often $*$ or $C$) to $\frac{1}{2}$. That is, using the partition $J_0 = [0, \frac{1}{2})$, $J_* = \{\frac{1}{2}\}$ and $J_1 = (\frac{1}{2}, 1]$. This resolves the "ambiguity" about which symbol to give to $\frac{1}{2}$, but it doesn't make the shift space closed.
- Assign the two symbols to $\frac{1}{2}$, so $J_0 = [0, \frac{1}{2}]$ and $J_1 = [\frac{1}{2}, 1]$ are no longer a partition but have $\frac{1}{2}$ in common. Therefore $\frac{1}{2}$ will have two itineraries and so will every point in the backward orbit of $\frac{1}{2}$. With all these extra itineraries, $\boldsymbol{i}(X)$ becomes closed. But this doesn't work in all cases; see Exercise 1.37.
- Take a quotient space $\boldsymbol{i}(X)/\sim$ where in this case $x \sim y$ if there is $n \in \mathbb{N}_0$ such that

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

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

**Exercise 1.37.** Let $a$=3.83187405528332... and $T(x)=f_a(x)=ax(1-x)$. For this parameter, $T^3(\frac{1}{2})=\frac{1}{2}$. Let $\mathcal{J}'=\{[0,\frac{1}{2}],(\frac{1}{2},1]\}$ and $\mathcal{J}=\{[0,\frac{1}{2}],[\frac{1}{2},1]\}$, so $\frac{1}{2}$ gets two symbols. Let $\Sigma' = \boldsymbol{i}(X)$ w.r.t. $\mathcal{J}'$ and $\Sigma = \boldsymbol{i}(X)$ w.r.t. $\mathcal{J}$. Show that $\overline{\Sigma'} \neq \Sigma$.

From now on, assume that $X$ is a compact metric space without isolated points. We will now discuss the properties of the coding map $\boldsymbol{i}$ itself. First of all, for $\boldsymbol{i}$ to be continuous, it is crucial that $T|_{J_a}$ is continuous on each element $J_a \in \mathcal{J}$. But this is not enough: if $x$ is a common boundary of two elements of $\mathcal{J}$, then (no matter how you assign the symbol to $x$ in Example 1.36), for each neighborhood $U \ni x$, $\text{diam}(\boldsymbol{i}(U)) = 1$, so continuity fails at $x$. It is only by using quotient spaces of $\boldsymbol{i}(X)$ (so changing the topology of $\boldsymbol{i}(X)$) that we can make $\boldsymbol{i}$ continuous. Normally, we choose to live with the discontinuity, because it affects only a few points:

**Lemma 1.38.** *Let $\partial\mathcal{J}$ denote the collection of common boundary points of different elements in a partition $\mathcal{J}$. If* $\text{orb}(x) \cap \partial J = \varnothing$ *for all $J \in \mathcal{J}$, then the coding map $\boldsymbol{i} : X \to \mathcal{A}^{\mathbb{N}_0}$ or $\mathcal{A}^{\mathbb{Z}}$ is continuous at $x$.*

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

**Definition 1.39.** A transformation $T : X \to X$ of a metric space $(X, d)$ is called **expansive** if there exists $\delta > 0$ such that for all distinct $x, y \in X$, there is $n \geq 0$ (or $n \in \mathbb{Z}$ if $T$ is invertible) such that $d(T^n(x), T^n(y)) > \delta$. We call $\delta$ the **expansivity constant**.

Every subshift $(X, \sigma)$ is expansive. Indeed, if $x \neq y$, then there is $n \in \mathbb{N}$ (or $n \in \mathbb{Z}$ if $(X, \sigma)$ is a two-sided shift) such that $x_n \neq y_n$, so $d(\sigma^n(x), \sigma^n(y)) = 1$. This makes every $\delta \in (0, 1)$ an expansivity constant.

**Lemma 1.40.** *Suppose that $T$ is a continuous expansive map and injective on each $J_a$ of some partition $\mathcal{J}$. If the expansivity constant $\delta >$ $\sup_{a \in \mathcal{A}} \operatorname{diam}(J_a)$, then the coding map $\boldsymbol{i} : X \to \mathcal{A}^{\mathbb{N}_0 \text{ or } \mathbb{Z}}$ is injective.*

**Proof.** Suppose that $\boldsymbol{i}$ is not injective, so there are $x \neq y \in X$ such that $\boldsymbol{i}(x) = \boldsymbol{i}(y)$. Since $T|_{J_a}$ is injective for each $a \in \mathcal{A}$, $T^n(x) \neq T^n(y)$ for all $n \in \mathbb{Z}$. By expansiveness, there is $n \in \mathbb{Z}$ such that $d(T^n(x), T^n(y)) > \delta$. By assumption, they cannot lie in the same element of $\mathcal{J}$. Hence $x$ and $y$ cannot have the same itinerary after all. □

To obtain injectivity of the coding map, it often suffices (but not always; see Example 1.43 below) that $T$ is expanding on each partition element $J_a$. Expanding (and expansion) should not be confused with expansive (and expansivity) of Definition 1.39.

**Definition 1.41.** Let $T : X \to Y$ be a map between metric spaces. We call $T$ **expanding** if there is $\rho > 1$ such that $d_Y(T(x), T(y)) \geq \rho d_X(x, y)$ for all $x, y \in X$ and **locally expanding** if there are $\varepsilon > 0$ and $\rho > 1$ such that $d(T(x), T(y)) \geq \rho d(x, y)$ for all $x, y \in Y$ with $d(x, y) < \varepsilon$.

**Proposition 1.42** (Gottschalk & Hedlund [**284**]). Let $T : X \to X$ be a homeomorphism on a compact metric space $(X, d)$. If $T$ is locally expanding, then $X$ is finite.

Compact is important. For example $T : \mathbb{R} \to \mathbb{R}$, $x \mapsto 2x$, would be a counterexample without the compactness assumption.

**Proof.** Let $\varepsilon > 0$ and $\rho > 1$ be as in Definition 1.41. Since $T^{-1}$ is continuous and $X$ is compact, there is a $\delta > 0$ such that $d(x, y) < \delta$ implies $d(T^{-1}(x), T^{-1}(y)) < \varepsilon$. Let $\{U_i\}_{i=1}^N$ be a finite open cover of $X$

such that $\operatorname{diam}(U_i) < \delta$. Then $\{T^{-1}(U_i)\}_{i=1}^N$ is an open cover of $X$, and $\operatorname{diam} T^{-1}(U_i) < \varepsilon$, so by local expansion, $\operatorname{diam} T^{-1}(U_i) < \operatorname{diam}(U_i)/\rho \leq \delta/\rho$. Repeating this argument, we find that $\{T^{-n}(U_i)\}_{i=1}^N$ is a finite open cover of $X$ with $\operatorname{diam}(T^{-n}(U_i)) < \delta\rho^{-n}$. Since $n$ is arbitrary, $\#X \leq N$. □

**Example 1.43.** In this example, we show that despite $T$ being expanding on partition elements $J_a$, $a \in \mathcal{A}$, this may still not result in an injective coding map $\boldsymbol{i} : X \to \mathcal{A}^{\mathbb{N}_0}$ if the diameter of some of the $J_i$'s is too big.

Let $T : \mathbb{S}^1 \to \mathbb{S}^1$, $x \mapsto 2x \bmod 1$, be the doubling map, and let $J_0 = (\frac{1}{4}, \frac{3}{4})$ and $J_1 = \mathbb{S}^1 \setminus J_0$. Clearly $T'(x) = 2$ for all $x \in \mathbb{S}^1$, but $T$ is not expanding on the whole of $\mathbb{S}^1$, because for instance $d(T(\frac{1}{4}), T(\frac{3}{4})) = 0 < \frac{1}{2} = d(\frac{1}{4}, \frac{3}{4})$. More importantly, $T$ is not expanding on $J_0$ or $J_1$ either; for example $d(T(\frac{1}{4} + \varepsilon), T(\frac{3}{4} - \varepsilon)) = 4\varepsilon < \frac{1}{2} - 2\varepsilon = d(\frac{1}{4} + \varepsilon, \frac{3}{4} - \varepsilon)$ for each $\varepsilon \in (0, \frac{1}{12})$. The corresponding coding map is **not** injective. The way to see this is by noting that the involution $S(x) = 1 - x$ commutes with $T$ and also preserves each $J_a$. It follows that $\boldsymbol{i}(x) = \boldsymbol{i}(S(x))$ for all $x \in \mathbb{S}^1$, and only $x = 0$ and $x = \frac{1}{2}$ have unique itineraries. For the more general partition $J_0^b = (b, b + \frac{1}{2})$ and $J_1^b = \mathbb{S}^1 \setminus J_0^b$ for $b \in [0, \frac{1}{2})$, see Remark 3.102.