CHAPTER 1

SHIFT SPACES

Shift spaces are to symbolic dynamics what shapes like polygons and curves are to geometry. We begin by introducing these spaces, and describing a variety of examples to guide the reader's intuition. Later chapters will concentrate on special classes of shift spaces, much as geometry concentrates on triangles and circles. As the name might suggest, on each shift space there is a shift map from the space to itself. Together these form a "shift dynamical system." Our main focus will be on such dynamical systems, their interactions, and their applications.

In addition to discussing shift spaces, this chapter also connects them with formal languages, gives several methods to construct new shift spaces from old, and introduces a type of mapping from one shift space to another called a sliding block code. In the last section, we introduce a special class of shift spaces and sliding block codes which are of interest in coding theory.

§1.1. Full Shifts

Information is often represented as a sequence of discrete symbols drawn from a fixed finite set. This book, for example, is really a very long sequence of letters, punctuation, and other symbols from the typographer's usual stock. A real number is described by the infinite sequence of symbols in its decimal expansion. Computers store data as sequences of 0's and 1's. Compact audio disks use blocks of 0's and 1's, representing signal samples, to digitally record Beethoven symphonies.

In each of these examples, there is a finite set $\mathcal{A}$ of *symbols* which we will call the *alphabet*. Elements of $\mathcal{A}$ are also called *letters*, and they will typically be denoted by a, b, c, ..., or sometimes by digits like 0, 1, 2, ..., when this is more meaningful. Decimal expansions, for example, use the alphabet $\mathcal{A} = \{0, 1, \ldots, 9\}$.

Although in real life sequences of symbols are finite, it is often extremely useful to treat long sequences as infinite in both directions (or *bi-infinite*).

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

This is analogous to using real numbers, continuity, and other ideas from analysis to describe physical quantities which, in reality, can be measured only with finite accuracy.

Our principal objects of study will therefore be collections of bi-infinite sequences of symbols from a finite alphabet $\mathcal{A}$. Such a sequence is denoted by $x = (x_i)_{i\in\mathbb{Z}}$, or by

$$x = \ldots x_{-2}x_{-1}x_0x_1x_2\ldots,$$

where each $x_i \in \mathcal{A}$. The symbol x_i is the ith *coordinate* of x, and x can be thought of as being given by its coordinates, or as a sort of infinite "vector." When writing a specific sequence, you need to specify which is the 0th coordinate. This is conveniently done with a "decimal point" to separate the x_i with $i \geqslant 0$ from those with $i < 0$. For example,

$$x = \ldots 010.1101\ldots$$

means that $x_{-3} = 0$, $x_{-2} = 1$, $x_{-1} = 0$, $x_0 = 1$, $x_1 = 1$, $x_2 = 0$, $x_3 = 1$, and so on.

Definition 1.1.1. If $\mathcal{A}$ is a finite alphabet, then the *full $\mathcal{A}$-shift* is the collection of all bi-infinite sequences of symbols from $\mathcal{A}$. The *full r-shift* (or simply *r-shift*) is the full shift over the alphabet $\{0, 1, \ldots, r-1\}$.

The full $\mathcal{A}$-shift is denoted by

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

Here $\mathcal{A}^{\mathbb{Z}}$ is the standard mathematical notation for the set of all functions from $\mathbb{Z}$ to $\mathcal{A}$, and such functions are just the bi-infinite sequences of elements from $\mathcal{A}$. Each sequence $x \in \mathcal{A}^{\mathbb{Z}}$ is called a *point* of the full shift. Points from the full 2-shift are also called *binary sequences*. If $\mathcal{A}$ has size $|\mathcal{A}| = r$, then there is a natural correspondence between the full $\mathcal{A}$-shift and the full r-shift, and sometimes the distinction between them is blurred. For example, it can be convenient to refer to the full shift on $\{+1, -1\}$ as the full 2-shift.

Blocks of consecutive symbols will play a central role. A *block* (or *word*) over $\mathcal{A}$ is a finite sequence of symbols from $\mathcal{A}$. We will write blocks without separating their symbols by commas or other punctuation, so that a typical block over $\mathcal{A} = \{a, b\}$ looks like $aababbabbb$. It is convenient to include the sequence of *no* symbols, called the *empty block* (or *empty word*) and denoted by ε. The *length* of a block u is the number of symbols it contains, and is denoted by $|u|$. Thus if $u = a_1a_2\ldots a_k$ is a nonempty block, then $|u| = k$, while $|\varepsilon| = 0$. A *k-block* is simply a block of length k. The set of all k-blocks over $\mathcal{A}$ is denoted $\mathcal{A}^k$. A *subblock* or *subword* of $u = a_1a_2\ldots a_k$ is a block

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

of the form $a_i a_{i+1} \dots a_j$, where $1 \leqslant i \leqslant j \leqslant k$. By convention, the empty block ε is a subblock of every block.

If x is a point in $\mathcal{A}^{\mathbb{Z}}$ and $i \leqslant j$, then we will denote the block of coordinates in x from position i to position j by

$$x_{[i,j]} = x_i x_{i+1} \dots x_j \, .$$

If $i > j$, define $x_{[i,j]}$ to be ε. It is also convenient to define

$$x_{[i,j)} = x_i x_{i+1} \dots x_{j-1} \, .$$

By extension, we will use the notation $x_{[i,\infty)}$ for the *right-infinite sequence* $x_i x_{i+1} x_{i+2} \dots$, although this is not really a block since it has infinite length. Similarly, $x_{(-\infty,i]} = \dots x_{i-2} x_{i-1} x_i$. The *central* $(2k+1)$*-block of* x is $x_{[-k,k]} = x_{-k} x_{-k+1} \dots x_k$. We sometimes will write $x_{[i]}$ for x_i, especially when we want to emphasize the index i.

Two blocks u and v can be put together, or *concatenated*, by writing u first and then v, forming a new block uv having length $|uv| = |u| + |v|$. Note that uv is in general not the same as vu, although they have the same length. By convention, $\varepsilon u = u\varepsilon = u$ for all blocks u. If $n \geqslant 1$, then u^n denotes the concatenation of n copies of u, and we put $u^0 = \varepsilon$. The law of exponents $u^m u^n = u^{m+n}$ then holds for all integers $m, n \geqslant 0$. The point $\dots uuu.uuu \dots$ is denoted by u^∞.

The index i in a point $x = (x_i)_{i \in \mathbb{Z}}$ can be thought of as indicating time, so that, for example, the time-0 coordinate of x is x_0. The passage of time corresponds to shifting the sequence one place to the left, and this gives a map or transformation from a full shift to itself.

Definition 1.1.2. The *shift map* σ on the full shift $\mathcal{A}^{\mathbb{Z}}$ maps a point x to the point $y = \sigma(x)$ whose ith coordinate is $y_i = x_{i+1}$.

The operation σ, pictured below, maps the full shift $\mathcal{A}^{\mathbb{Z}}$ onto itself. There

$$\begin{array}{lcccccccccccc} x & = & \cdots & x_{-3} & x_{-2} & x_{-1} & \cdot & x_0 & x_1 & x_2 & x_3 & \cdots \\ \downarrow \sigma & & & \searrow & \searrow & \searrow & & \searrow & \searrow & \searrow & & \\ y = \sigma(x) & = & \cdots & x_{-2} & x_{-1} & x_0 & & \cdot x_1 & x_2 & x_3 & x_4 & \cdots \end{array}$$

is also the inverse operation σ^{-1} of shifting one place to the right, so that σ is both one-to-one and onto. The composition of σ with itself $k > 0$ times $\sigma^k = \sigma \circ \dots \circ \sigma$ shifts sequences k places to the left, while $\sigma^{-k} = (\sigma^{-1})^k$ shifts the same amount to the right. This shifting operation is the reason $\mathcal{A}^{\mathbb{Z}}$ is called a full shift ("full" since all sequences of symbols are allowed).

The shift map is useful for expressing many of the concepts in symbolic dynamics. For example, one basic idea is that of codes, or rules, which

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

transform one sequence into another. For us, the most important codes are those that do not change with time. Consider the map $\phi\colon \{0,1\}^{\mathbb{Z}} \to \{0,1\}^{\mathbb{Z}}$ defined by the rule $\phi(x) = y$, where $y_i = x_i + x_{i+1} \pmod 2$. Then ϕ is a coding rule that replaces the symbol at index i with the sum modulo 2 of itself and its right neighbor. The coding operation ϕ acts the same at each coordinate, or is *stationary*, i.e., independent of time.

Another way to say this is that applying the rule ϕ and then shifting gives exactly the same result as shifting and then applying ϕ. Going through the following diagram to the right and then down gives the same result as going down and then to the right.

$$\begin{array}{ccc} x & \xrightarrow{\quad\sigma\quad} & \sigma(x) \\ \Big\downarrow{\scriptstyle \phi} & & \Big\downarrow{\scriptstyle \phi} \\ \phi(x) & \xrightarrow{\quad\sigma\quad} & \sigma(\phi(x)) = \phi(\sigma(x)) \end{array}$$

We can express this as $\sigma \circ \phi = \phi \circ \sigma$, or in terms of the coordinates by $\sigma(\phi(x))_{[i]} = \phi(\sigma(x))_{[i]}$, since both equal $x_{i+1} + x_{i+2} \pmod 2$. Recall that when two mappings f and g satisfy $f \circ g = g \circ f$, they are said to *commute*. Not all pairs of mappings commute (try: $f =$ "put on socks" and $g =$ "put on shoes"). Using this terminology, a code ϕ on the full 2-shift is stationary if it commutes with the shift map σ, which we can also express by saying that the following diagram commutes.

$$\begin{array}{ccc} \{0,1\}^{\mathbb{Z}} & \xrightarrow{\quad\sigma\quad} & \{0,1\}^{\mathbb{Z}} \\ \Big\downarrow{\scriptstyle \phi} & & \Big\downarrow{\scriptstyle \phi} \\ \{0,1\}^{\mathbb{Z}} & \xrightarrow{\quad\sigma\quad} & \{0,1\}^{\mathbb{Z}} \end{array}$$

We will discuss codes in more detail in §1.5.

Points in a full shift which return to themselves after a finite number of shifts are particularly simple to describe.

Definition 1.1.3. A point x is *periodic for* σ if $\sigma^n(x) = x$ for some $n \geqslant 1$, and we say that x has *period* n under σ. If x is periodic, the smallest positive integer n for which $\sigma^n(x) = x$ is the *least period* of x. If $\sigma(x) = x$, then x is called a *fixed point* for σ.

If x has least period k, then it has period $2k, 3k, \ldots$, and every period of x is a multiple of k (see Exercise 1.1.5). A fixed point for σ must have the form a^∞ for some symbol a, and a point of period n has the form u^∞ for some n-block u.

Iteration of the shift map provides the "dynamics" in symbolic dynamics (see Chapter 6). Naturally, the "symbolic" part refers to the symbols used to form sequences in the spaces we will study.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

EXERCISES

1.1.1. How many points $x \in \mathcal{A}^{\mathbb{Z}}$ are fixed points? How many have period n? How many have least period 12?

1.1.2. For the full $\{+1, -1\}$-shift and $k \geqslant 1$, determine the number of k-blocks having the property that the sum of the symbols is 0.

1.1.3. Let ϕ be the coding rule from this section.

(a) Prove that ϕ maps the full 2-shift *onto* itself, i.e., that given a point y in the 2-shift, there is an x with $\phi(x) = y$.

(b) Find the number of points x in the full 2-shift with $\phi^n(x) = 0^\infty$ for $n = 1, 2$, or 3. Can you find this number for every n?

*(c) Find the number of points x with $\phi^n(x) = x$ for $n = 1, 2$, or 3. Can you find this number for every n?

1.1.4. For each k with $1 \leqslant k \leqslant 6$ find the number of k-blocks over $\mathcal{A} = \{0, 1\}$ having no two consecutive 1's appearing. Based on your result, can you guess, and then prove, what this number is for every k?

1.1.5. Determine the least period of u^∞ in terms of properties of the block u. Use your solution to show that if x has period n, then the least period of x divides n.

1.1.6. (a) Describe those pairs of blocks u and v over an alphabet $\mathcal{A}$ such that $uv = vu$.

*(b) Describe those sequences $u_1, u_2, \ldots, u_n$ of n blocks for which all n concatenations $u_1u_2 \ldots u_n,\ u_2 \ldots u_nu_1,\ \ldots,\ u_nu_1u_2 \ldots u_{n-1}$ of the cyclic permutations are equal.

§1.2. Shift Spaces

The symbol sequences we will be studying are often subject to constraints. For example, Morse code uses the symbols "dot," "dash," and "pause." The ordinary alphabet is transmitted using blocks of dots and dashes with length at most six separated by a pause, so that any block of length at least seven which contains no pause is forbidden to occur (the only exception is the SOS signal). In the programming language Pascal, a program line such as `sin(x)***2 := y` is not allowed, nor are lines with unbalanced parentheses, since they violate Pascal's syntax rules. The remarkable error correction in compact audio disks results from the use of special kinds of binary sequences specified by a finite number of conditions. In this section we introduce the fundamental notion of shift space, which will be the subset of points in a full shift satisfying a fixed set of constraints.

If $x \in \mathcal{A}^{\mathbb{Z}}$ and w is a block over $\mathcal{A}$, we will say that w *occurs in* x if there are indices i and j so that $w = x_{[i,j]}$. Note that the empty block ε occurs in every x, since $\varepsilon = x_{[1,0]}$. Let $\mathcal{F}$ be a collection of blocks over $\mathcal{A}$, which we will think of as being the *forbidden blocks*. For any such $\mathcal{F}$, define $X_{\mathcal{F}}$ to be the subset of sequences in $\mathcal{A}^{\mathbb{Z}}$ which do *not* contain any block in $\mathcal{F}$.

Definition 1.2.1. A *shift space* (or simply *shift*) is a subset X of a full shift $\mathcal{A}^{\mathbb{Z}}$ such that $X = X_{\mathcal{F}}$ for some collection $\mathcal{F}$ of forbidden blocks over $\mathcal{A}$.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

The collection $\mathcal{F}$ may be finite or infinite. In any case it is at most countable since its elements can be arranged in a list (just write down its blocks of length 1 first, then those of length 2, and so on). For a given shift space there may be many collections $\mathcal{F}$ describing it (see Exercise 1.2.4). Note that the empty set $\varnothing$ is a shift space, since putting $\mathcal{F} = \mathcal{A}$ rules out every point. When a shift space X is contained in a shift space Y, we say that X is a *subshift* of Y.

In the equation $X = \mathsf{X}_{\mathcal{F}}$, the notation X refers to the operation of forming a shift space, while X denotes the resulting set. We will sometimes use similar typographical distinctions between an operation and its result, for example in §2.2 when forming an adjacency matrix from a graph. By use of such distinctions, we hope to avoid the type of nonsensical equations such as "$y = y(x)$" you may have seen in calculus classes.

Example 1.2.2. X is $\mathcal{A}^{\mathbb{Z}}$, where we can take $\mathcal{F} = \varnothing$, reflecting the fact that there are no constraints. □

Example 1.2.3. X is the set of all binary sequences with no two 1's next to each other. Here $X = \mathsf{X}_{\mathcal{F}}$, where $\mathcal{F} = \{11\}$. This shift is called the *golden mean shift* for reasons which will surface in Chapter 4. □

Example 1.2.4. X is the set of all binary sequences so that between any two 1's there are an even number of 0's. We can take for $\mathcal{F}$ the collection

$$\{10^{2n+1}1 : n \geqslant 0\}.$$

This example is naturally called the *even shift*. □

In the following examples, the reader will find it instructive to list an appropriate collection $\mathcal{F}$ of forbidden blocks for which $X = \mathsf{X}_{\mathcal{F}}$.

Example 1.2.5. X is the set of all binary sequences for which 1's occur infinitely often in each direction, and such that the number of 0's between successive occurrences of a 1 is either 1, 2, or 3. This shift is used in a common data storage method for hard disk drives (see §2.5). For each pair (d, k) of nonnegative integers with $d \leqslant k$, there is an analogous (d, k) *run-length limited shift*, denoted by $\mathsf{X}(d, k)$, and defined by the constraints that 1's occur infinitely often in each direction, and there are at least d 0's, but no more than k 0's, between successive 1's. Using this notation, our example is $\mathsf{X}(1, 3)$. □

Example 1.2.6. To generalize the previous examples, fix a nonempty subset S of $\{0, 1, 2, \dots\}$. If S is finite, define $X = \mathsf{X}(S)$ to be the set of all binary sequences for which 1's occur infinitely often in each direction, and such that the number of 0's between successive occurrences of a 1 is an integer in S. Thus a typical point in $\mathsf{X}(S)$ has the form

$$x = \dots 1\, 0^{n_{-1}}\, 1\, 0^{n_0}\, 1\, 0^{n_1}\, 1 \dots,$$

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

where each $n_j \in S$. For example, the (d, k) run-length limited shift corresponds to $S = \{d, d+1, \dots, k\}$.

When S is infinite, it turns out that to obtain a shift space we need to allow points that begin or end with an infinite string of 0's (see Exercise 1.2.8). In this case, we define $X(S)$ the same way as when S is finite, except that we do *not* require that 1's occur infinitely often in each direction. In either case, we refer to $X(S)$ as the *S-gap shift*.

Observe that the full 2-shift is the S-gap shift with $S = \{0, 1, 2, \dots\}$, the golden mean shift corresponds to $S = \{1, 2, 3, \dots\}$, and the even shift to $S = \{0, 2, 4, \dots\}$. As another example, for $S = \{2, 3, 5, 7, 11, \dots\}$ the set of primes, we call $X(S)$ the *prime gap shift*. □

Example 1.2.7. For each positive integer c, the *charge constrained shift*, is defined as the set of all points in $\{+1, -1\}^{\mathbb{Z}}$ so that for every block occurring in the point, the algebraic sum s of the $+1$'s and -1's satisfies $-c \leqslant s \leqslant c$. These shifts arise in engineering applications and often go by the name "DC-free sequences." See Immink [Imm2, Chapter 6]. □

Example 1.2.8. Let $\mathcal{A} = \{e, f, g\}$, and X be the set of points in the full $\mathcal{A}$-shift for which e can be followed only by e or f, f can be followed only by g, and g can be followed only by e or f. A point in this space is then just a bi-infinite path on the graph shown in Figure 1.2.1 This is an example of a *shift of finite type*. These shifts are the focus of the next chapter. □

Example 1.2.9. X is the set of points in the full shift $\{a, b, c\}^{\mathbb{Z}}$ so that a block of the form $ab^m c^k a$ may occur in the point only if $m = k$. We will refer to this example as the *context-free shift*. □

You can make up infinitely many shift spaces by using different forbidden collections $\mathcal{F}$. Indeed, there are uncountably many shift spaces possible (see Exercise 1.2.12). As subsets of full shifts, these spaces share a common feature called *shift invariance*. This amounts to the observation that the constraints on points are given in terms of forbidden blocks alone, and do not involve the coordinate at which a block might be forbidden. It follows that if x is in $X_{\mathcal{F}}$, then so are its shifts $\sigma(x)$ and $\sigma^{-1}(x)$. This can be neatly expressed as $\sigma(X_{\mathcal{F}}) = X_{\mathcal{F}}$. The *shift map* σ_X on X is the restriction to X of the shift map σ on the full shift.

This shift invariance property allows us to find subsets of a full shift that are not shift spaces. One simple example is the subset X of $\{0, 1\}^{\mathbb{Z}}$

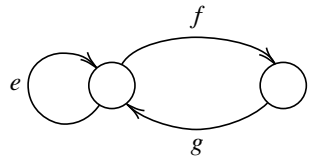


FIGURE 1.2.1. A graph defining a shift space.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

consisting of the single point

$$x = \ldots 0101.0101 \ldots = (01)^\infty.$$

Since $\sigma(x) = (10)^\infty \notin X$, we see that X is not shift invariant, so it is not a shift space.

However, shift invariance alone is not enough to have a shift space. What is missing is a sort of "closure" (see Corollary 1.3.5 and Theorem 6.1.21). This is illustrated by the following example.

Example 1.2.10. Let $X \subseteq \{0,1\}^{\mathbb{Z}}$ be the set of points each of which contains exactly one symbol 1 and the rest 0's. Clearly X is shift invariant. If X were a shift space, then no block of 0's could be forbidden. But then the point $0^\infty = \ldots 000.000 \ldots$ would necessarily belong to X, whereas it does not. The set X lacks the "closure" necessary for a shift space. □

Since a shift space X is contained in a full shift, Definition 1.1.3 serves to define what it means for $x \in X$ to be fixed or periodic under σ_X. However, unlike full shifts and many of the examples we have introduced, there are shift spaces that contain no periodic points at all (Exercise 1.2.13).

EXERCISES

1.2.1. Find a collection $\mathcal{F}$ of blocks over $\{0,1\}$ so that $\mathbf{X}_\mathcal{F} = \varnothing$.

1.2.2. For Examples 1.2.5 through 1.2.9 find a set of forbidden blocks describing the shift space.

1.2.3. Let X be the subset of $\{0,1\}^{\mathbb{Z}}$ described in Example 1.2.10. Show that $X \cup \{0^\infty\}$ is a shift space.

1.2.4. Find two collections $\mathcal{F}_1$ and $\mathcal{F}_2$ over $\mathcal{A} = \{0,1\}$ with $\mathbf{X}_{\mathcal{F}_1} = \mathbf{X}_{\mathcal{F}_2} \neq \varnothing$, where $\mathcal{F}_1$ is finite and $\mathcal{F}_2$ is infinite.

1.2.5. Show that $\mathbf{X}_{\mathcal{F}_1} \cap \mathbf{X}_{\mathcal{F}_2} = \mathbf{X}_{\mathcal{F}_1 \cup \mathcal{F}_2}$. Use this to prove that the intersection of two shift spaces over the same alphabet is also a shift space. Extend this to arbitrary intersections.

1.2.6. Show that if $\mathcal{F}_1 \subseteq \mathcal{F}_2$, then $\mathbf{X}_{\mathcal{F}_1} \supseteq \mathbf{X}_{\mathcal{F}_2}$. What is the relationship between $\mathbf{X}_{\mathcal{F}_1} \cup \mathbf{X}_{\mathcal{F}_2}$ and $\mathbf{X}_{\mathcal{F}_1 \cap \mathcal{F}_2}$?

1.2.7. Let X be the full $\mathcal{A}$-shift.

(a) Show that if X_1 and X_2 are shift spaces such that $X_1 \cup X_2 = X$, then $X_1 = X$ or $X_2 = X$ (or both).

(b) Extend your argument to show that if X is the union of any collection $\{X_\alpha\}$ of shift spaces, then there is an α such that $X = X_\alpha$.

(c) Explain why these statements no longer hold if we merely assume that X is a shift space.

1.2.8. If S is an infinite subset of $\{0,1,2,\ldots\}$, show that the collection of all binary sequences of the form

$$x = \ldots 1\,0^{n_{-1}}\,1\,0^{n_0}\,1\,0^{n_1}\,1 \ldots,$$

where each $n_j \in S$, is not a shift space.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

1.2.9. Let X_i be a shift over $\mathcal{A}_i$ for $i = 1, 2$. The *product shift* $X = X_1 \times X_2$ consists of all pairs $(x^{(1)}, x^{(2)})$ with $x^{(i)} \in X_i$. If we identify a pair (x, y) of sequences with the sequence $(\dots (x_{-1}, y_{-1}), (x_0, y_0), (x_1, y_1), \dots)$ of pairs, we can regard $X_1 \times X_2$ as a subset of $(\mathcal{A}_1 \times \mathcal{A}_2)^{\mathbb{Z}}$. With this convention, show that $X_1 \times X_2$ is a shift space over the alphabet $\mathcal{A}_1 \times \mathcal{A}_2$.

1.2.10. Let X be a shift space, and $N \geqslant 1$. Show that there is a collection $\mathcal{F}$ of blocks, all of which have length at least N, so that $X = \mathsf{X}_{\mathcal{F}}$.

1.2.11. For which sets S does the S-gap shift have infinitely many periodic points?

1.2.12. Show there are uncountably many shift spaces contained in the full 2-shift. [*Hint*: Consider S-gap shifts.]

***1.2.13.** Construct a nonempty shift space that does not contain any periodic points.

***1.2.14.** For a given alphabet $\mathcal{A}$, let

$$X = \{x \in \mathcal{A}^{\mathbb{Z}} : x_{i+n^2} \neq x_i \text{ for all } i \in \mathbb{Z} \text{ and } n \geqslant 1\}.$$

(a) If $|\mathcal{A}| = 2$, prove that $X = \varnothing$.
(b) If $|\mathcal{A}| = 3$, show that $X = \varnothing$. [*Hint*: $3^2 + 4^2 = 5^2$.]

§1.3. Languages

It is sometimes easier to describe a shift space by specifying which blocks are allowed, rather than which are forbidden. This leads naturally to the notion of the language of a shift.

Definition 1.3.1. Let X be a subset of a full shift, and let $\mathcal{B}_n(X)$ denote the set of all n-blocks that occur in points in X. The *language of* X is the collection

$$\mathcal{B}(X) = \bigcup_{n=0}^{\infty} \mathcal{B}_n(X).$$

Example 1.3.2. The full 2-shift has language

$$\{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, \dots\}. \qquad \square$$

Example 1.3.3. The golden mean shift (Example 1.2.3) has language

$$\{\varepsilon, 0, 1, 00, 01, 10, 000, 001, 010, 100, 101, 0000, \dots\}. \qquad \square$$

The term "language" comes from the theory of automata and formal languages. See [HopU] for a lucid introduction to these topics. Think of the language $\mathcal{B}(X)$ as the collection of "allowed" blocks in X. For a block $u \in \mathcal{B}(X)$, we sometimes use alternative terminology such as saying that u *occurs in* X or *appears in* X or *is in* X or *is allowed in* X.

Not every collection of blocks is the language of a shift space. The following proposition characterizes those which are, and shows that they provide an alternative description of a shift space. In what follows we will denote the complement of a collection $\mathcal{C}$ of blocks over $\mathcal{A}$ relative to the collection of all blocks over $\mathcal{A}$ by $\mathcal{C}^c$.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

Proposition 1.3.4.

(1) *Let X be a shift space, and $\mathcal{L} = \mathcal{B}(X)$ be its language. If $w \in \mathcal{L}$, then*
   (a) *every subblock of w belongs to $\mathcal{L}$, and*
   (b) *there are nonempty blocks u and v in $\mathcal{L}$ so that $uwv \in \mathcal{L}$.*

(2) *The languages of shift spaces are characterized by* (1)*. That is, if $\mathcal{L}$ is a collection of blocks over $\mathcal{A}$, then $\mathcal{L} = \mathcal{B}(X)$ for some shift space X if and only if $\mathcal{L}$ satisfies condition* (1)*.*

(3) *The language of a shift space determines the shift space. In fact, for any shift space, $X = \mathsf{X}_{\mathcal{B}(X)^c}$. Thus two shift spaces are equal if and only if they have the same language.*

PROOF: (1) If $w \in \mathcal{L} = \mathcal{B}(X)$, then w occurs in some point x in X. But then every subblock of w also occurs in x, so is in $\mathcal{L}$. Furthermore, clearly there are nonempty blocks u and v such that uwv occurs in x, so that $u, v \in \mathcal{L}$ and $uwv \in \mathcal{L}$.

(2) Let $\mathcal{L}$ be a collection of blocks satisfying (1), and X denote the shift space $\mathsf{X}_{\mathcal{L}^c}$. We will show that $\mathcal{L} = \mathcal{B}(X)$. For if $w \in \mathcal{B}(X)$, then w occurs in some point of $\mathsf{X}_{\mathcal{L}^c}$, so that $w \notin \mathcal{L}^c$, or $w \in \mathcal{L}$. Thus $\mathcal{B}(X) \subseteq \mathcal{L}$. Conversely, suppose that $w = x_0 x_1 \dots x_m \in \mathcal{L}$. Then by repeatedly applying (1b), we can find symbols x_j with $j > m$ and x_i with $i < 0$ so that by (1a) every subblock of $x = (x_i)_{i \in \mathbb{Z}}$ lies in $\mathcal{L}$. This means that $x \in \mathsf{X}_{\mathcal{L}^c}$. Since w occurs in x, we have that $w \in \mathcal{B}(\mathsf{X}_{\mathcal{L}^c}) = \mathcal{B}(X)$, proving that $\mathcal{L} \subseteq \mathcal{B}(X)$.

(3) If $x \in X$, no block occurring in x is in $\mathcal{B}(X)^c$ since $\mathcal{B}(X)$ contains all blocks occurring in all points of X. Hence $x \in \mathsf{X}_{\mathcal{B}(X)^c}$, showing that $X \subseteq \mathsf{X}_{\mathcal{B}(X)^c}$. Conversely, since X is a shift there is a collection $\mathcal{F}$ for which $X = \mathsf{X}_{\mathcal{F}}$. If $x \in \mathsf{X}_{\mathcal{B}(X)^c}$, then every block in x must be in $\mathcal{B}(X) = \mathcal{B}(\mathsf{X}_{\mathcal{F}})$, and so cannot be in $\mathcal{F}$. Hence $x \in \mathsf{X}_{\mathcal{F}}$, proving that $X = \mathsf{X}_{\mathcal{F}} \supseteq \mathsf{X}_{\mathcal{B}(X)^c}$. □

This result shows that although a shift X can be described by different collections of forbidden blocks, there is a largest collection $\mathcal{B}(X)^c$, the complement of the language of X. This is the largest possible forbidden collection that describes X. For a minimal forbidden collection, see Exercise 1.3.8. The proposition also gives a one-to-one correspondence between shifts X and languages $\mathcal{L}$ that satisfy (1). This correspondence can be summarized by the equations

$$\mathcal{L} = \mathcal{B}(\mathsf{X}_{\mathcal{L}^c}), \qquad X = \mathsf{X}_{\mathcal{B}(X)^c}. \tag{1–3–1}$$

A useful consequence of part (3) above is that to verify that a point x is in a given shift space X, you only need to show that each subblock $x_{[i,j]}$ is in $\mathcal{B}(X)$. In fact, this gives a characterization of shift spaces in terms of "allowed" blocks.

Corollary 1.3.5. *Let X be a subset of the full $\mathcal{A}$-shift. Then X is a shift space if and only if whenever $x \in \mathcal{A}^{\mathbb{Z}}$ and each $x_{[i,j]} \in \mathcal{B}(X)$ then $x \in X$.*

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

PROOF: The "whenever" condition is equivalent to the condition that $X = X_{\mathcal{B}(X)^c}$. Thus the corollary follows from Proposition 1.3.4(3). □

If X is a shift space, the first part of Proposition 1.3.4 shows that every block $w \in \mathcal{B}(X)$ can be extended on both sides to another block $uwv \in \mathcal{B}(X)$. However, given two blocks u and v in $\mathcal{B}(X)$, it may not be possible to find a block w so that $uwv \in \mathcal{B}(X)$. For example, let $X = \{0^\infty, 1^\infty\} \subseteq \{0,1\}^{\mathbb{Z}}$, and $u = 0$, $v = 1$. Shift spaces for which two blocks can always be "joined" by a third play a special and important role.

Definition 1.3.6. A shift space X is *irreducible* if for every ordered pair of blocks $u, v \in \mathcal{B}(X)$ there is a $w \in \mathcal{B}(X)$ so that $uwv \in \mathcal{B}(X)$.

Note that if u, v is an ordered pair of blocks in $\mathcal{B}(X)$, then so is v, u. Thus to verify that X is irreducible, we must be able to find blocks w_1 and w_2 so that both uw_1v and vw_2u are in $\mathcal{B}(X)$.

The reader should verify that Examples 1.2.2 through 1.2.9 are irreducible. Indeed, most shift spaces we encounter will be irreducible. Those which are not can usually be decomposed into irreducible "pieces," and the theory we develop for irreducible shifts can then be applied to each piece.

There are close connections between symbolic dynamics and the theory of formal languages. For example, special shift spaces called sofic shifts that we will explore in Chapter 3 correspond to regular languages, i.e., those languages accepted by a finite-state automaton. In addition, ideas and techniques from formal languages can sometimes be used in symbolic dynamics. For instance, the idea behind the Pumping Lemma for regular languages is used in Example 3.1.7 to prove that the context-free shift is not sofic. See the notes section of Chapter 3 for more on these ideas.

EXERCISES

1.3.1. Determine the language of the full shift, the even shift (Example 1.2.4), the $(1,3)$ run-length limited shift (Example 1.2.5), and the charge constrained shift (Example 1.2.7).

1.3.2. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are languages satisfying condition 1 of Proposition 1.3.4, show that $\mathcal{L}_1 \cup \mathcal{L}_2$ also satisfies the condition. Use this to prove that the union of two shift spaces is also a shift space. If $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \ldots$ are languages over the same alphabet, show that $\bigcup_{n=1}^{\infty} \mathcal{L}_n$ is also a language. Why can't you use this to prove that the union of an infinite number of shift spaces over the same alphabet is also a shift space?

1.3.3. Is the intersection of the languages of two shift spaces also the language of a shift space?

1.3.4. If X and Y are shift spaces, describe the languages of their intersection $X \cap Y$ and of their product $X \times Y$ (defined in Exercise 1.2.9).

1.3.5. Is the intersection of two irreducible shift spaces always irreducible? The product (defined in Exercise 1.2.9)?

1.3.6. Let $\mathcal{A} = \{0,1\}$ and $\mathcal{F} = \{01\}$. Is $X_{\mathcal{F}}$ irreducible?

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

1.3.7. Let X be an irreducible shift space. Show that for every ordered pair of blocks $u, v \in \mathcal{B}(X)$, there is a *nonempty* block $w \in \mathcal{B}(X)$ such that $uwv \in \mathcal{B}(X)$.

***1.3.8.** Let X be a shift space. Call a word w a "first offender" for X if $w \notin \mathcal{B}(X)$, but every proper subword of w is in $\mathcal{B}(X)$. Let $\mathcal{O}$ be the collection of all first offenders for X.

(a) Prove that $X = \mathsf{X}_{\mathcal{O}}$.

(b) If $X = \mathsf{X}_{\mathcal{F}}$, show that for every $w \in \mathcal{O}$ there is a $v \in \mathcal{F}$ such that w is a subword of v, but v contains no other first offenders.

(c) Use (b) to show that $\mathcal{O}$ is a *minimal* forbidden set, in the sense that if $\mathcal{F} \subseteq \mathcal{O}$ and $\mathsf{X}_{\mathcal{F}} = X$, then $\mathcal{F} = \mathcal{O}$.

§1.4. Higher Block Shifts and Higher Power Shifts

One of the basic constructions in symbolic dynamics involves widening our attention from a single symbol to a block of consecutive symbols, and considering such blocks as letters from a new, more elaborate alphabet. This process, which we will call "passing to a higher block shift," is a very convenient technical device, and we will be using it often. It provides an alternative description of the same shift space.

Let X be a shift space over the alphabet $\mathcal{A}$, and $\mathcal{A}_X^{[N]} = \mathcal{B}_N(X)$ be the collection of all allowed N-blocks in X. We can consider $\mathcal{A}_X^{[N]}$ as an alphabet in its own right, and form the full shift $(\mathcal{A}_X^{[N]})^{\mathbb{Z}}$. Define the *Nth higher block code* $\beta_N \colon X \to (\mathcal{A}_X^{[N]})^{\mathbb{Z}}$ by

$$(\beta_N(x))_{[i]} = x_{[i,i+N-1]}. \tag{1–4–1}$$

Thus β_N replaces the ith coordinate of x with the block of coordinates in x of length N starting at position i. This becomes clearer if we imagine the symbols in $\mathcal{A}_X^{[N]}$ as written vertically. Then the image of $x = (x_i)_{i \in \mathbb{Z}}$ under β_4 has the form

(1–4–2)

$$\beta_4(x) = \ldots \begin{bmatrix} x_0 \\ x_{-1} \\ x_{-2} \\ x_{-3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_0 \\ x_{-1} \\ x_{-2} \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \\ x_0 \\ x_{-1} \end{bmatrix} . \begin{bmatrix} x_3 \\ x_2 \\ x_1 \\ x_0 \end{bmatrix} \begin{bmatrix} x_4 \\ x_3 \\ x_2 \\ x_1 \end{bmatrix} \begin{bmatrix} x_5 \\ x_4 \\ x_3 \\ x_2 \end{bmatrix} \ldots \in (\mathcal{A}_X^{[4]})^{\mathbb{Z}}.$$

Definition 1.4.1. Let X be a shift space. Then the Nth *higher block shift* $X^{[N]}$ or *higher block presentation* of X is the image $X^{[N]} = \beta_N(X)$ in the full shift over $\mathcal{A}_X^{[N]}$.

Notice that in (1–4–2) consecutive symbols from $\mathcal{A}_X^{[N]}$ overlap. If $u = u_1 u_2 \ldots u_N$ and $v = v_1 v_2 \ldots v_N$ are N-blocks, let us say that u and v *overlap progressively* if $u_2 u_3 \ldots u_N = v_1 v_2 \ldots v_{N-1}$. If the 2-block uv over the alphabet $\mathcal{A}_X^{[N]}$ occurs in some image point $\beta_N(x)$, then a glance at (1–4–2) shows that u and v must overlap progressively. Also observe from

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

(1–4–2) that by knowing the bottom letter in each symbol of $\beta_N(x)$ we can reconstruct the entire image, as well as the original point x. In this sense $X^{[N]}$ is simply another description of the same shift space X.

Example 1.4.2. Let X be the golden mean shift of Example 1.2.3. Then

$$\mathcal{A}_X^{[2]} = \{a = 00, b = 01, c = 10\},$$

and $X^{[2]}$ is described by the constraints $\mathcal{F} = \{ac, ba, bb, cc\}$. Each of these 2-blocks is forbidden since they fail to overlap progressively. For example, the second symbol of $a = 00$ does not match the first of $c = 10$, so ac is forbidden. Naturally, the block 11 is also forbidden, since it is forbidden in the original shift. This is expressed by its absence from $\mathcal{A}_X^{[2]}$. □

The terminology "higher block shift" implies that it is a shift space. We can verify this as follows.

Proposition 1.4.3. *The higher block shifts of a shift space are also shift spaces.*

PROOF: Let X be a shift space over $\mathcal{A}$, and $N \geqslant 1$. Then there is a collection $\mathcal{F}$ of blocks over $\mathcal{A}$ so that $X = X_{\mathcal{F}}$. Create a new collection $\widetilde{\mathcal{F}}$ by replacing each block u in $\mathcal{F}$ such that $|u| < N$ by all N-blocks over $\mathcal{A}$ containing u. Then clearly $X = X_{\widetilde{\mathcal{F}}}$, and every block in $\widetilde{\mathcal{F}}$ has length $\geqslant N$. (See Exercise 1.2.10.)

For each $w = a_1 a_2 \dots a_m \in \widetilde{\mathcal{F}}$ let

$$w^{[N]} = (a_1 a_2 \dots a_N)(a_2 a_3 \dots a_{N+1}) \dots (a_{m-N+1} a_{m-N+2} \dots a_m)$$

be the corresponding $(m - N + 1)$-block over $\mathcal{A}^N$. Let $\mathcal{F}_1$ denote the set of all blocks over the alphabet $\mathcal{A}^N$ of the form $w^{[N]}$ for some $w \in \widetilde{\mathcal{F}}$. This represents one set of constraints on $X^{[N]}$, namely those coming from the constraints on the original shift. It follows that $X^{[N]} \subseteq X_{\mathcal{F}_1}$.

Points in $X^{[N]}$ also satisfy the overlap condition illustrated in (1–4–2). Thus we let

$$\mathcal{F}_2 = \{uv : u \in \mathcal{A}^N, v \in \mathcal{A}^N, \text{ and } u \text{ and } v \text{ do not overlap progressively}\}.$$

Then $X^{[N]} \subseteq X_{\mathcal{F}_2}$, so that by Exercise 1.2.5

$$X^{[N]} \subseteq X_{\mathcal{F}_1} \cap X_{\mathcal{F}_2} = X_{\mathcal{F}_1 \cup \mathcal{F}_2}.$$

Conversely, suppose that $y \in X_{\mathcal{F}_1 \cup \mathcal{F}_2}$, and let x be the point of $\mathcal{A}^{\mathbb{Z}}$ reconstructed from the "bottom" symbols as described after Definition 1.4.1. Then $x \in X = X_{\mathcal{F}}$ since y satisfies the constraints from $\mathcal{F}_1$, and $y = \beta_N(x)$ by the overlap constraints from $\mathcal{F}_2$. This proves that $X^{[N]} \supseteq X_{\mathcal{F}_1 \cup \mathcal{F}_2}$, so that $X^{[N]} = X_{\mathcal{F}_1 \cup \mathcal{F}_2}$ is a shift space. □

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

The Nth higher block shift of X uses overlapping blocks. The same sort of construction can be made with nonoverlapping blocks, and leads to the notion of the Nth higher power shift of X.

Using the same notation as at the beginning of this section, define the *Nth higher power code* $\gamma_N : X \to (\mathcal{A}_X^{[N]})^{\mathbb{Z}}$ by

$$(\gamma_N(x))_{[i]} = x_{[iN, iN+N-1]}.$$

Here γ_N chops up the coordinates of x into consecutive N-blocks and assembles the pieces into a point over $\mathcal{A}_X^{[N]}$. The image of $x = (x_i)_{i\in\mathbb{Z}}$ under γ_4 has the form

(1–4–3)

$$\gamma_4(x) = \ldots \begin{bmatrix} x_{-9} \\ x_{-10} \\ x_{-11} \\ x_{-12} \end{bmatrix} \begin{bmatrix} x_{-5} \\ x_{-6} \\ x_{-7} \\ x_{-8} \end{bmatrix} \begin{bmatrix} x_{-1} \\ x_{-2} \\ x_{-3} \\ x_{-4} \end{bmatrix} . \begin{bmatrix} x_{3} \\ x_{2} \\ x_{1} \\ x_{0} \end{bmatrix} \begin{bmatrix} x_{7} \\ x_{6} \\ x_{5} \\ x_{4} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{10} \\ x_{9} \\ x_{8} \end{bmatrix} \ldots \in (\mathcal{A}_X^{[4]})^{\mathbb{Z}}.$$

Compare this with (1–4–2). Note that the bottom symbols here will usually *not* determine the rest of the symbols since there is no overlapping.

Definition 1.4.4. Let X be a shift space. The Nth *higher power shift* X^N of X is the image $X^N = \gamma_N(X)$ of X in the full shift over $\mathcal{A}_X^{[N]}$

Example 1.4.5. Let X be the golden mean shift of Example 1.2.3, and $N = 2$. Then $\mathcal{A}_X^{[2]} = \{a = 00, b = 01, c = 10\}$. The 2nd higher power shift X^2 is described by $\mathcal{F} = \{bc\}$, since words containing $bc = 0110$ are the only ones containing the forbidden word 11 of the original shift. □

As before, a higher power shift is also a shift space.

Proposition 1.4.6. *The higher power shifts of a shift space are also shift spaces.*

PROOF: The proof is very similar to that of Proposition 1.4.3, and is left to the reader. □

EXERCISES

1.4.1. For Examples 1.2.3 and 1.2.4 describe explicitly the 3rd higher block shift $X^{[3]}$. To do this, you need to specify the alphabet $\mathcal{A}_X^{[3]}$, and then describe a collection $\mathcal{F}$ of blocks over this alphabet so that $X^{[3]} = X_{\mathcal{F}}$. [*Hint*: Use the proof of Proposition 1.4.3 as a guide.]

1.4.2. If X and Y are shift spaces over the same alphabet, show that

$$(X \cap Y)^{[N]} = X^{[N]} \cap Y^{[N]} \quad \text{and} \quad (X \cup Y)^{[N]} = X^{[N]} \cup Y^{[N]}.$$

1.4.3. If X and Y are shift spaces over possibly different alphabets, show that

$$(X \times Y)^{[N]} = X^{[N]} \times Y^{[N]}$$

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

1.4.4. If X is a shift space with shift map σ_X, and $X^{[N]}$ is its Nth higher block shift with corresponding shift map $\sigma_{X^{[N]}}$, prove that $\beta_N \circ \sigma_X = \sigma_{X^{[N]}} \circ \beta_N$. [*Hint*: Compute the ith coordinate of each image.]

1.4.5. If X^N is the Nth higher power shift of X, and σ_{X^N} is its shift map, prove that $\gamma_N \circ \sigma_X^N = \sigma_{X^N} \circ \gamma_N$. Here σ_X^N is the N-fold composition of σ_X with itself.

1.4.6. Find an example of a shift space for which the bottom symbols in (1–4–3) of $\gamma_4(x)$ do not determine the rest of the symbols. Find another example for which they do determine the rest.

§1.5. Sliding Block Codes

Suppose that $x = \ldots x_{-1}x_0x_1\ldots$ is a sequence of symbols in a shift space X over $\mathcal{A}$. We can transform x into a new sequence $y = \ldots y_{-1}y_0y_1\ldots$ over another alphabet $\mathfrak{A}$ as follows. Fix integers m and n with $-m \leqslant n$. To compute the ith coordinate y_i of the transformed sequence, we use a function Φ that depends on the "window" of coordinates of x from $i-m$ to $i+n$. Here $\Phi\colon \mathcal{B}_{m+n+1}(X) \to \mathfrak{A}$ is a fixed *block map* , called an $(m+n+1)$-*block map* from allowed $(m+n+1)$-blocks in X to symbols in $\mathfrak{A}$, and so

$$y_i = \Phi(x_{i-m}x_{i-m+1}\ldots x_{i+n}) = \Phi(x_{[i-m,i+n]}). \tag{1–5–1}$$

Definition 1.5.1. Let X be a shift space over $\mathcal{A}$, and $\Phi\colon \mathcal{B}_{m+n+1}(X) \to \mathfrak{A}$ be a block map. Then the map $\phi\colon X \to \mathfrak{A}^{\mathbb{Z}}$ defined by $y = \phi(x)$ with y_i given by (1–5–1) is called the *sliding block code* with *memory* m and *anticipation* n *induced by* Φ. We will denote the formation of ϕ from Φ by $\phi = \Phi_\infty^{[-m,n]}$, or more simply by $\phi = \Phi_\infty$ if the memory and anticipation of ϕ are understood. If not specified, the memory is taken to be 0. If Y is a shift space contained in $\mathfrak{A}^{\mathbb{Z}}$ and $\phi(X) \subseteq Y$, we write $\phi\colon X \to Y$.

Figure 1.5.1 illustrates the action of a sliding block code. The window is slid one coordinate to the right to compute the next coordinate of the image.

The simplest sliding block codes are those with no memory or anticipation, i.e., with $m = n = 0$. Here the ith coordinate of the image of x depends only on x_i. Such sliding block codes are called *1-block codes*. By our convention about memory in Definition 1.5.1, when Φ is a 1-block map, then $\phi = \Phi_\infty$ is taken to be a 1-block code if no memory is specified.

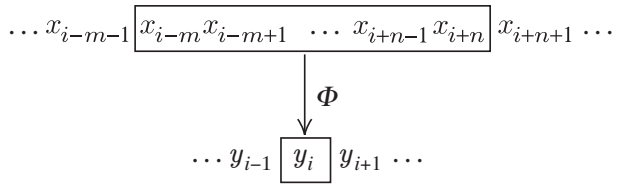


FIGURE 1.5.1. Sliding block code.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

Example 1.5.2. Let X be a shift space over an alphabet $\mathcal{A}$, $\mathfrak{A} = \mathcal{A}$, $m = 0$, $n = 1$, and $\Phi(a_0a_1) = a_1$. Then $\phi = \Phi_\infty^{[0,1]} = \Phi_\infty$ is the shift map σ_X. What happens if we let $\Phi(a_0a_1) = a_0$ instead?

Now let $m = 1$, $n = 0$, and $\Psi(a_{-1}a_0) = a_{-1}$. Then $\psi = \Psi_\infty^{[-1,0]} = \Psi_\infty$ is the inverse σ_X^{-1} of the shift map, so that $\psi(\phi(x)) = x = \phi(\psi(x))$ for all $x \in X$.

Notice that if $\Theta(a) = a$ for all $a \in \mathcal{A}$, then $\phi = \Theta_\infty^{[1,1]}$ and $\psi = \Theta_\infty^{[-1,-1]}$. So, there may be many ways of representing a mapping between shift spaces as a sliding block code. □

Example 1.5.3. Let $\mathcal{A} = \{0,1\} = \mathfrak{A}$, $X = \mathcal{A}^{\mathbb{Z}}$, $m = 0$, $n = 1$, and $\Phi(a_0a_1) = a_0 + a_1 \pmod 2$. Then $\phi = \Phi_\infty$ is the code ϕ discussed near the end of §1.1. □

Example 1.5.4. Let $\mathcal{A} = \{0,1\}$, $\mathfrak{A} = \{a,b\}$, $m = n = 0$, and $\Phi(0) = a$, $\Phi(1) = b$. Then $\phi = \Phi_\infty$ is a 1-block code from the full 2-shift to the full $\mathfrak{A}$-shift. If $\Psi(0) = \Psi(1) = a$, then $\psi = \Psi_\infty$ collapses the full 2-shift to the single point a^∞. □

Example 1.5.5. Let X be a shift space over $\mathcal{A}$, $\mathfrak{A} = \mathcal{A}_X^{[N]}$, $m = 0$, $n = N-1$, $Y = X^{[N]}$, and

$$\Phi(a_0a_1 \dots a_{N-1}) = a_0a_1 \dots a_{N-1} \in \mathcal{A}_X^{[N]}.$$

Then $\phi = \Phi_\infty : X \to Y$ is the Nth higher block code β_N from §1.4. □

Suppose that $\Phi : \mathcal{B}_{m+n+1}(X) \to \mathfrak{A}$ is a block map which induces a sliding block code with memory m and anticipation n. It will sometimes be convenient to imagine Φ as having a larger "window," and ignore the extra coordinates. Thus if $M \geqslant m$ and $N \geqslant n$, define $\widehat{\Phi} : \mathcal{B}_{M+N+1}(X) \to \mathfrak{A}$ by

$$\widehat{\Phi}(x_{[-M,N]}) = \Phi(x_{[-m,n]}).$$

Clearly $\widehat{\Phi}_\infty^{[-M,N]} = \Phi_\infty^{[-m,n]}$. The process of passing from Φ to $\widehat{\Phi}$ is called "increasing the window size of Φ," and shows we can assume that a sliding block code is induced by a block map with as large a window as we like.

Let $\Phi : \mathcal{B}_{m+n+1}(X) \to \mathfrak{A}$ be a block map. We can extend Φ so that it maps $(m+n+k)$-blocks in X to k-blocks over $\mathfrak{A}$ by sliding its window as follows. If $x_{[-m,n+k-1]}$ is in $\mathcal{B}_{m+n+k}(X)$, put

$$\Phi(x_{[-m,n+k-1]}) = \Phi(x_{[-m,n]})\Phi(x_{[-m+1,n+1]}) \dots \Phi(x_{[-m+k-1,n+k-1]}).$$

For example, if Φ is the 2-block map of Example 1.5.3, then $\Phi(011010001) = 10111001$.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

Example 1.5.6. Let $\mathcal{A} = \mathfrak{A} = \{0,1\}$, X be the golden mean shift of Example 1.2.3, and Y be the even shift of Example 1.2.4. Let Φ be the 2-block map defined by $\Phi(00) = 1$, $\Phi(01) = 0$, and $\Phi(10) = 0$. We do not need to define $\Phi(11)$ since the block 11 does not occur in X. Then we will show that the induced sliding block code $\phi = \Phi_\infty : X \to Y$ is onto.

If 10^k1 occurs in $\phi(x)$, it must be the image under Φ of the block $0(01)^r00$, so that $k = 2r$ is even. This shows that $\phi(X) \subseteq Y$. Since each point $y \in Y$ has 1's separated by an even number of 0's, this same observation shows how to construct an $x \in X$ with $\phi(x) = y$, so that ϕ is onto. □

If $\phi : X \to Y$ is a sliding block code and $x \in X$, then computing ϕ at the shifted sequence $\sigma_X(x)$ gives the same result as shifting the image $\phi(x)$ using σ_Y. The commuting property is a key feature of sliding block codes.

Proposition 1.5.7. *Let X and Y be shift spaces. If $\phi : X \to Y$ is a sliding block code, then $\phi \circ \sigma_X = \sigma_Y \circ \phi$; i.e., the following diagram commutes.*

$$\begin{array}{ccc} X & \xrightarrow{\sigma_X} & X \\ {\scriptstyle\phi}\downarrow & & \downarrow{\scriptstyle\phi} \\ Y & \xrightarrow{\sigma_Y} & Y \end{array}$$

PROOF: Let ϕ be induced by the block map $\Phi : \mathcal{B}_{m+n+1}(X) \to \mathfrak{A}$ and have memory m and anticipation n. For $x \in X$,

$$(\sigma_Y \circ \phi)(x)_{[i]} = \phi(x)_{[i+1]} = \Phi(x_{[i+1-m,i+1+n]}),$$

while

$$\begin{aligned} (\phi \circ \sigma_X)(x)_{[i]} &= \phi(\sigma_X(x))_{[i]} \\ &= \Phi(\sigma_X(x)_{[i-m,i+n]}) \\ &= \Phi(x_{[i-m+1,i+n+1]}). \end{aligned}$$

Hence the ith coordinates of the images agree for each i, so the images are equal. □

However, shift-commuting is not enough to have a sliding block code (the reader is asked to give a specific example in Exercise 1.5.14). One also needs to know that $\phi(x)_0$ depends only on a central block of x.

Proposition 1.5.8. *Let X and Y be shift spaces. A map $\phi : X \to Y$ is a sliding block code if and only if $\phi \circ \sigma_X = \sigma_Y \circ \phi$ and there exists $N \geqslant 0$ such that $\phi(x)_0$ is a function of $x_{[-N,N]}$.*

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

PROOF: The necessity of the condition is clear from the definition and Proposition 1.5.7. For sufficiency, define the $(2N+1)$-block map Φ by $\Phi(w) = \phi(x)_0$ where x is any point in X such that $x_{[-N,N]} = w$. It is straightforward to check that $\phi = \Phi_\infty^{[-N,N]}$. □

If a sliding block code $\phi: X \to Y$ is onto, then ϕ is called a *factor code from X onto Y* . A shift space Y is a *factor* of X if there is a factor code from X onto Y. The sliding block codes ϕ in Examples 1.5.2 through 1.5.6 are factor codes. Factor codes are often called "factor maps" in the literature.

If $\phi: X \to Y$ is one-to-one, then ϕ is called an *embedding of X into Y*. The sliding block codes ϕ in Examples 1.5.2 and 1.5.4 are embeddings, as is the higher block code $\beta_N: X \to (\mathcal{A}_X^{[N]})^{\mathbb{Z}}$. The code in Example 1.5.3 is not an embedding since it is two-to-one everywhere.

Sometimes a sliding block code $\phi: X \to Y$ has an *inverse*, i.e., a sliding block code $\psi: Y \to X$ such that $\psi(\phi(x)) = x$ for all $x \in X$ and $\phi(\psi(y)) = y$ for all $y \in Y$. This is the case in Example 1.5.2. If ϕ has an inverse, it is unique (see Exercise 1.5.4), so we can write $\psi = \phi^{-1}$, and we call ϕ *invertible*.

Definition 1.5.9. A sliding block code $\phi: X \to Y$ is a *conjugacy from X to Y*, if it is invertible. Two shift spaces X and Y are *conjugate* (written $X \cong Y$) if there is a conjugacy from X to Y.

If there is a conjugacy from X to Y, we can think of Y as being a "recoded" version of X, sharing all of its properties. Then X and Y are merely different views of the same underlying object. We will explore this idea in greater detail in Chapter 6. Conjugacies are often called "topological conjugacies" in the literature.

Example 1.5.10. Let X be a shift space over $\mathcal{A}$, and $X^{[N]}$ be its Nth higher block shift. According to Example 1.5.5, $\beta_N: X \to X^{[N]}$ is a sliding block code. Define the 1-block map $\Psi: \mathcal{A}_X^{[N]} \to \mathcal{A}$ by $\Psi(a_0 a_1 \dots a_{N-1}) = a_0$, and put $\psi = \Psi_\infty: X^{[N]} \to X$. It is easy to check that $\psi = \beta_N^{-1}$, so that β_N is a conjugacy, and thus $X \cong X^{[N]}$. In this sense, $X^{[N]}$ is a recoded version of X. □

The behavior of periodic points under sliding block codes is described in the following result.

Proposition 1.5.11. *Let $\phi: X \to Y$ be a sliding block code. If $x \in X$ has period n under σ_X, then $\phi(x)$ has period n under σ_Y, and the least period of $\phi(x)$ divides the least period of x. Embeddings, and hence conjugacies, preserve the least period of a point.*

PROOF: If x has period n, then $\sigma_X^n(x) = x$. Hence

$$\sigma_Y^n(\phi(x)) = \phi(\sigma_X^n(x)) = \phi(x),$$

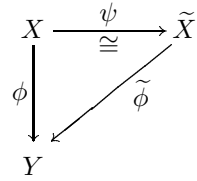
https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

so that $\phi(x)$ has period n. If x has least period n, then $\phi(x)$ has period n, and hence its least period divides n (see Exercise 1.1.5). If ϕ is one-to-one, then $\sigma_X^n(x) = x$ if and only if $\sigma_Y^n(\phi(x)) = \phi(x)$, so that x and $\phi(x)$ must have the same least period. □

Observe that this proposition shows that for each n, the number of points of period n is the same for all shifts conjugate to a given shift; i.e., it is an *invariant* of the shift. This gives a way to prove that some pairs of shift spaces cannot be conjugate, for example when one has a fixed point and the other doesn't. We shall be meeting several other kinds of invariants in the chapters ahead.

Let $\phi: X \to Y$ be a sliding block code. We show next that we can recode X to a conjugate shift $\widetilde{X}$ so that the corresponding sliding block code $\widetilde{\phi}: \widetilde{X} \to Y$ is a 1-block code. This process, called "recoding ϕ to a 1-block code," is often a starting point in proofs, since 1-block codes are much easier to think about. However, the penalty for making the map simpler is making the alphabet more complicated.

Proposition 1.5.12. *Let $\phi: X \to Y$ be a sliding block code. Then there exist a higher block shift $\widetilde{X}$ of X, a conjugacy $\psi: X \to \widetilde{X}$, and a 1-block code $\widetilde{\phi}: \widetilde{X} \to Y$ so that $\widetilde{\phi} \circ \psi = \phi$; i.e., the following diagram commutes.*



PROOF: Suppose that ϕ is induced by a block map Φ and has memory m and anticipation n. Let $\mathfrak{A} = \mathcal{B}_{m+n+1}(X)$, and define $\psi: X \to \mathfrak{A}^{\mathbb{Z}}$ by $\psi(x)_{[i]} = x_{[i-m,i+n]}$. Then $\psi = \sigma^{-m} \circ \beta_{m+n+1}$. Thus $\widetilde{X} = \psi(X) = X^{[m+n+1]}$ is a shift space, and since σ and β_{m+n+1} are conjugacies, so is ψ. Put $\widetilde{\phi} = \phi \circ \psi^{-1}$. Note that $\widetilde{\phi}$ is a 1-block code. □

We remark that if a sliding block code ϕ happens to be a conjugacy, then the recoding of ϕ to a 1-block code, given in Proposition 1.5.12, usually does not also recode ϕ^{-1} to a 1-block code (why?). Thus the cost of recoding to "simplify" a sliding block code in one direction is often to make it more "complicated" in the other.

We next show that for any sliding block code $\phi: X \to Y$, $\phi(X)$ is a shift space.

Theorem 1.5.13. *The image of a shift space under a sliding block code is a shift space.*

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

PROOF: Let X and Y be shift spaces, and $\phi: X \to Y$ be a sliding block code. By Proposition 1.5.12, we can assume that ϕ is a 1-block code. Let Φ be a 1-block map inducing ϕ. Put $\mathcal{L} = \{\Phi(w) : w \in \mathcal{B}(X)\}$. We will show that $\phi(X) = X_{\mathcal{L}^c}$, proving that the image of X is a shift space.

If $x \in X$, then every block in $\phi(x)$ is in $\mathcal{L}$, so that $\phi(x) \in X_{\mathcal{L}^c}$. This proves that $\phi(X) \subseteq X_{\mathcal{L}^c}$.

Suppose now that $y \in X_{\mathcal{L}^c}$. Then for each $n \geqslant 0$ the central $(2n+1)$-block of y is the image under Φ of the central $(2n+1)$-block of some point $x^{(n)}$ in X; i.e.,

$$\Phi(x^{(n)}_{[-n,n]}) = \phi(x^{(n)})_{[-n,n]} = y_{[-n,n]}. \tag{1–5–2}$$

We will use the $x^{(n)}$ to find a point $x \in X$ with $\phi(x) = y$.

First consider the 0th coordinates $x^{(n)}_{[0]}$ for $n \geqslant 1$. Since there are only finitely many symbols, there is an infinite set S_0 of integers for which $x^{(n)}_{[0]}$ is the same for all $n \in S_0$. Next, the central 3-blocks $x^{(n)}_{[-1,1]}$ for $n \in S_0$ all belong to the finite set of possible 3-blocks, so there is an infinite subset $S_1 \subseteq S_0$ so that $x^{(n)}_{[-1,1]}$ is the same for all $n \in S_1$. Continuing this way, we find for each $k \geqslant 1$ an infinite set $S_k \subseteq S_{k-1}$ so that all blocks $x^{(n)}_{[-k,k]}$ are equal for $n \in S_k$.

Define x to be the sequence with $x_{[-k,k]} = x^{(n)}_{[-k,k]}$ for all $n \in S_k$ (these blocks are all the same by our construction). Observe that since $S_k \subseteq S_{k-1}$, the central $(2k-1)$-block of $x_{[-k,k]}$ is $x_{[-k+1,k-1]}$, so that x is well-defined. Also observe that every block in x occurs in some $x_{[-k,k]} = x^{(n)}_{[-k,k]} \in \mathcal{B}(X)$, so that $x \in X$ since X is a shift space. Finally, for each $k \geqslant 0$ and $n \in S_k$ with $n \geqslant k$ we have, using (1–5–2), that

$$\Phi(x_{[-k,k]}) = \Phi(x^{(n)}_{[-k,k]}) = \phi(x^{(n)})_{[-k,k]} = y_{[-k,k]},$$

so that $\phi(x) = y$. This proves that $X_{\mathcal{L}^c} \subseteq \phi(X)$, completing the proof. □

This proof repays close study. It uses a version of the *Cantor diagonal argument*, one of the most important and subtle ideas in mathematics. This argument is used, for example, to show that the set of real numbers cannot be arranged in a sequence (i.e., the set of real numbers is uncountable). We will encounter it again in Chapter 6, when we discuss the notion of compactness.

Suppose that $\phi: X \to Y$ is an embedding. By the previous result, $\phi(X)$ is a shift space, and ϕ establishes a one-to-one correspondence between points in X and points in $\phi(X)$. Let $\psi: \phi(X) \to X$ be the reverse correspondence, so that $\psi(y) = x$ whenever $\phi(x) = y$. Then ψ is a mapping, but is it a sliding block code? Another application of the Cantor diagonal argument shows that it is.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

Theorem 1.5.14. *A sliding block code that is one-to-one and onto has a sliding block inverse, and is hence a conjugacy.*

PROOF: Let $\phi: X \to Y$ be a sliding block code that is one-to-one and onto. By recoding ϕ if necessary, we can assume that ϕ is a 1-block code. Let Φ be a 1-block map inducing ϕ. Let $\psi: Y \to X$ be the map on points inverse to ϕ, which we will show is a sliding block code.

First observe that if $y = \phi(x)$, then since $\phi(\sigma_X(x)) = \sigma_Y(\phi(x))$, we have that

$$\begin{aligned}\sigma_X(\psi(y)) &= \sigma_X(x) = \psi(\phi(\sigma_X(x))) \\ &= \psi(\sigma_Y(\phi(x))) = \psi(\sigma_Y(y)),\end{aligned}$$

so that $\psi \circ \sigma_Y = \sigma_X \circ \psi$. Hence by Proposition 1.5.8, to show that ψ is a sliding block code, it is enough to find an $n \geqslant 0$ such that the central $(2n+1)$-block of every y determines the 0th coordinate $\psi(y)_{[0]}$ of its image.

If this were not the case, then for every $n \geqslant 0$ there would be two points $y^{(n)}$ and $\tilde{y}^{(n)}$ in Y so that $y^{(n)}_{[-n,n]} = \tilde{y}^{(n)}_{[-n,n]}$ but $\psi(y^{(n)})_{[0]} \neq \psi(\tilde{y}^{(n)})_{[0]}$. Put $x^{(n)} = \psi(y^{(n)})$ and $\tilde{x}^{(n)} = \psi(\tilde{y}^{(n)})$.

Since there are only finitely many symbols in the alphabet of X, there would be distinct symbols $a \neq b$ and an infinite set S_0 of integers so that $x^{(n)}_{[0]} = \psi(y^{(n)})_{[0]} = a$ and $\tilde{x}^{(n)}_{[0]} = \psi(\tilde{y}^{(n)})_{[0]} = b$ for all $n \in S_0$.

Since the number of pairs of possible 3-blocks is finite, there would be an infinite subset $S_1 \subseteq S_0$ so that the $x^{(n)}_{[-1,1]}$ are all equal for $n \in S_1$ and the $\tilde{x}^{(n)}_{[-1,1]}$ are all equal for $n \in S_1$. Continuing this way, for each $k \geqslant 1$ we would find an infinite subset $S_k \subseteq S_{k-1}$ so that the $x^{(n)}_{[-k,k]}$ are all equal for $n \in S_k$, and the $\tilde{x}^{(n)}_{[-k,k]}$ are all equal for $n \in S_k$. As in the proof of Theorem 1.5.13, this would allow us to construct points x and $\tilde{x}$ in X defined by $x_{[-k,k]} = x^{(n)}_{[-k,k]}$ and $\tilde{x}_{[-k,k]} = \tilde{x}^{(n)}_{[-k,k]}$ for $n \in S_k$. Note that $x_{[0]} = a \neq b = \tilde{x}_{[0]}$, so that $x \neq \tilde{x}$. Now if $n \in S_k$ and $n \geqslant k$, then

$$\begin{aligned}\Phi(x_{[-k,k]}) &= \Phi(x^{(n)}_{[-k,k]}) = \phi(x^{(n)})_{[-k,k]} = y^{(n)}_{[-k,k]} \\ &= \tilde{y}^{(n)}_{[-k,k]} = \phi(\tilde{x}^{(n)})_{[-k,k]} = \Phi(\tilde{x}^{(n)}_{[-k,k]}) = \Phi(\tilde{x}_{[-k,k]}).\end{aligned}$$

But this would imply that $\phi(x) = \phi(\tilde{x})$. This contradiction shows that ψ must be a sliding block code. □

If we are given two shift spaces X and Y, it is natural to ask whether Y is conjugate to X, whether Y is a factor of X, or whether Y embeds into X. These questions are very difficult to settle for general shift spaces. Indeed, many of the ideas and results we will encounter originate in attempts to answer these fundamental questions for special classes of shift spaces.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

EXERCISES

1.5.1. Suppose that $\phi: X \to Y$ and $\psi: Y \to Z$ are sliding block codes. Show that $\psi \circ \phi: X \to Z$ is also a sliding block code. If ϕ and ψ are factor codes, show that $\psi \circ \phi$ is also a factor code, and similarly for embeddings and conjugacies.

1.5.2. Show that an invertible sliding block code must be one-to-one and onto, so it is simultaneously a factor code and an embedding.

1.5.3. Prove that conjugacy $\cong$ between shift spaces is an equivalence relation; that is, show that (a) $X \cong X$, (b) if $X \cong Y$ then $Y \cong X$, and (c) if $X \cong Y$ and $Y \cong Z$, then $X \cong Z$.

1.5.4. Prove that an invertible sliding block code can have only one inverse.

1.5.5. Does the sliding block code in Example 1.5.3 have an inverse? What about the sliding block codes in Example 1.5.4? Justify your answers.

1.5.6. Let X be a shift space.

(a) Show that $X^{[1]} = X$.
(b) Show that $(X^{[N]})^{[2]} \cong X^{[N+1]}$.

1.5.7. Let $X = \{0,1\}^{\mathbb{Z}}$, and $\Phi: \{0,1\} \to \{0,1\}$ be the 1-block map given by $\Phi(0) = 1$ and $\Phi(1) = 0$. Show that $\phi = \Phi_\infty: X \to X$ is a conjugacy of the full 2-shift to itself.

1.5.8. Let X be the full 2-shift. Define the block map Φ by

$$\Phi(abcd) = b + a(c+1)d \pmod 2,$$

and put $\phi = \Phi_\infty^{[-1,2]}$.

(a) Describe the action of ϕ on $x \in X$ in terms of the blocks 1001 and 1101 appearing in x.
(b) Show that $\phi^2(x) = x$ for all $x \in X$, and hence show that ϕ is a conjugacy of X to itself.
(c) Use this method to find other conjugacies of the full 2-shift to itself.

1.5.9. Recode Example 1.5.3 to a 1-block code.

1.5.10. Suppose that $X_1 \supseteq X_2 \supseteq X_3 \supseteq \dots$ are shift spaces whose intersection is X. For each $N \geqslant 1$, use the Cantor diagonal argument to prove that there is a $K \geqslant 1$ such that $\mathcal{B}_N(X_k) = \mathcal{B}_N(X)$ for all $k \geqslant K$.

1.5.11. (a) Is the full 2-shift conjugate to the full 3-shift?
(b) Find a factor code from the full 3-shift onto the full 2-shift. Can you find infinitely many such factor codes?
(c) Is there a factor code from the full 2-shift onto the full 3-shift?
(d) Is the golden mean shift conjugate to a full shift? To the even shift?

1.5.12. Let $\phi: X \to Y$ be a sliding block code, and Z be a shift space contained in Y. Show that $\phi^{-1}(Z) = \{x \in X : \phi(x) \in Z\}$ is a shift space.

1.5.13. (a) Let Z be the full k-shift, and $\phi: X \to Z$ be a sliding block code. If X is a subset of a shift space Y, show that ϕ can be extended to a sliding block code $\psi: Y \to Z$ such that $\psi(x) = \phi(x)$ for all $x \in X$.
(b) Find an example of shift spaces X, Y, and Z with $X \subset Y$, and a sliding block code $\phi: X \to Z$, such that there is no sliding block code from Y to Z extending ϕ.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

1.5.14. Find a point mapping from the full 2-shift to itself that commutes with the shift, but is *not* a sliding block code.

1.5.15. Show that for a forbidden list $\mathcal{F}$, $X_{\mathcal{F}} = \varnothing$ if and only if there exists N such that whenever u and v are blocks with $|u| = N$, then some subblock of uvu belongs to $\mathcal{F}$.

***1.5.16.** (a) Show that there is no 1-block or 2-block factor code from the even shift onto the golden mean shift.
(b) Find a 3-block factor code from the even shift onto the golden mean shift (compare with Example 1.5.6).

***1.5.17.** Show that the S-gap shift and the S'-gap shift are conjugate iff either $S = S'$ or for some n, $S = \{0, n\}$ and $S' = \{n, n+1, \dots\}$. Hence, there are uncountably many shifts no pair of which are conjugate.

§1.6. Convolutional Encoders

In symbolic dynamics the term "code" means a mapping from one shift space to another, or more loosely some sort of apparatus or procedure for constructing such a mapping. In the previous section we introduced sliding block codes. Later in this book we consider finite-state codes (Chapter 5), finite-to-one codes (Chapter 8), and almost invertible codes (Chapter 9).

However, in the subject of coding theory the term "code" means something different, namely a set $\mathcal{C}$ of sequences (often finite sequences, but sometimes right-infinite or bi-infinite sequences). The goal is to find "good" error-correcting codes. These are codes $\mathcal{C}$ for which any two distinct sequences in $\mathcal{C}$ differ in a relatively "large" number of coordinates. Thus if the sequences in $\mathcal{C}$ are regarded as messages and transmitted over a "noisy" channel that makes a relatively small number of errors, then these errors can be detected and corrected to recover the original message.

The two broad classes of error-correcting codes that have been studied over the past forty years are block codes and convolutional codes. A *block code* is defined as a finite set of sequences all of the same length over some finite alphabet. *Convolutional codes* are much closer in spirit to symbolic dynamics, and are used in various applications in communications and storage. Such a code is defined as the image of a mapping, called a convolutional encoder, defined below.

Recall that a *finite field* $\mathbb{F}$ is a finite set in which you can add, subtract, multiply and divide so that the basic associative, distributive and commutative laws of arithmetic hold. A good example to keep in mind (and the one that we are mostly concerned with) is the field $\mathbb{F}_2$ with just two elements. Thus $\mathbb{F}_2 = \{0, 1\}$ with the usual additive and multiplicative structure: $0+0 = 1+1 = 0$, $0+1 = 1+0 = 1$, $0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0$, and $1 \cdot 1 = 1$.

A *Laurent polynomial* over a field $\mathbb{F}$ is a polynomial $f(t)$ in the variable t and its inverse t^{-1} whose coefficients are in $\mathbb{F}$. A typical Laurent polynomial

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

looks like

$$f(t) = \sum_{j=-m}^{n} a_j t^j,$$

where the $a_j \in \mathbb{F}$. A *bi-infinite power series* over $\mathbb{F}$ is a series of the form

$$f(t) = \sum_{j=-\infty}^{\infty} a_j t^j. \tag{1–6–1}$$

Although these series resemble the infinite series studied in calculus, we are using them as a formal algebraic device and are not concerned with convergence. If the coefficients of two series lie in the same field $\mathbb{F}$, then we can add them coefficientwise. Note that the set of all bi-infinite power series with coefficients in $\mathbb{F}$ can be identified with the full shift over $\mathbb{F}$, where the series in (1–6–1) corresponds to the point $\ldots a_{-1}.a_0 a_1 \ldots \in \mathbb{F}^{\mathbb{Z}}$. We can also multiply a bi-infinite power series by a Laurent polynomial using the normal rules of algebra. A *Laurent polynomial matrix* is a (finite-dimensional) rectangular matrix whose entries are Laurent polynomials. For the k-dimensional vector space $\mathbb{F}^k$ over a field $\mathbb{F}$, we identify the full $\mathbb{F}^k$-shift with the set of all k-tuple row vectors of bi-infinite power series with coefficients in $\mathbb{F}$.

Definition 1.6.1. Let $G(t) = [g_{ij}(t)]$ be a $k \times n$ Laurent polynomial matrix. Use $G(t)$ to transform an *input vector* $I(t) = [I_1(t), \ldots, I_k(t)]$, whose components are bi-infinite power series, into an *output vector* $O(t) = [O_1(t), \ldots, O_n(t)]$ via the equation

$$O(t) = E(I(t)) = I(t)G(t). \tag{1–6–2}$$

A (k, n)-*convolutional encoder* is a mapping E from the full $\mathbb{F}^k$-shift to the full $\mathbb{F}^n$-shift of the form (1–6–2). A *convolutional code* is the image of a convolutional encoder.

The term "convolutional" is used because multiplying a power series by a polynomial is usually called a convolution. In coding theory, what we have defined as a convolutional encoder is often called a "Laurent polynomial convolutional encoder" to distinguish it from a more general class of encoders.

We illustrate these concepts with the following example.

Example 1.6.2. Let $\mathbb{F} = \mathbb{F}_2$ and

$$G(t) = \begin{bmatrix} 1 & 0 & 1+t \\ 0 & t & t \end{bmatrix}.$$

The image of the input vector $I(t) = [I_1(t), I_2(t)]$ under the corresponding convolutional encoder E is

$$E(I(t)) = I(t) \cdot G(t) = [I_1(t), tI_2(t), (1+t)I_1(t) + tI_2(t)].$$

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

To see how E represents a mapping from $X = (\mathbb{F}_2^2)^{\mathbb{Z}}$ to $Y = (\mathbb{F}_2^3)^{\mathbb{Z}}$, write $I_1(t) = \sum_j a_j t^j$ and $I_2(t) = \sum_j b_j t^j$. We are then identifying $I(t)$ with the point

$$\dots (a_{-1}, b_{-1}).(a_0, b_0)(a_1, b_1) \dots \in X,$$

and similarly for $O(t)$ and Y. With these identifications, the jth component of $O(t) = E(I(t))$ is $(a_j,\ b_{j-1},\ a_j + a_{j-1} + b_{j-1})$, whose entries are just the coefficients of t^j in $O_1(t)$, $O_2(t)$, and $O_3(t)$, respectively. Observe that E is actually a sliding block code. Specifically, $E = \Phi_\infty^{[-1,0]}$, where Φ is the 2-block map over $\mathbb{F}_2^2$ defined by

$$\Phi\big((a_{-1}, b_{-1})(a_0, b_0)\big) = (a_0,\ b_{-1},\ a_0 + a_{-1} + b_{-1}).$$

The corresponding convolutional code is

$$E(X) = \{\dots (a_j,\ b_{j-1},\ a_j + a_{j-1} + b_{j-1}) \dots \in (\mathbb{F}_2^3)^{\mathbb{Z}} : a_j, b_j \in \mathbb{F}_2\}.$$

To describe $E(X)$ more explicitly, let $\mathcal{F}$ be the finite collection of 2-blocks over $\mathbb{F}_2^3$ defined by

$$\mathcal{F} = \{(c, d, e)(c', d', e') : e' \neq c' + c + d'\}. \tag{1–6–3}$$

We leave it to the reader to check that $E(X)$ is the shift space defined using the set $\mathcal{F}$ of forbidden blocks, so that $E(X) = X_{\mathcal{F}}$. □

It is not an accident that the convolutional encoder in the previous example is a sliding block code, or that the corresponding convolutional code is a shift space. To prove that this holds generally, consider a convolutional encoder E defined by a Laurent polynomial matrix $G(t) = [g_{ij}(t)]$ over a finite field $\mathbb{F}$. Let M denote the largest power of t that occurs in any of the $g_{ij}(t)$, and N denote the smallest such power. Let g_{ij}^p be the coefficient of t^p in $g_{ij}(t)$. Similarly, if $I(t) = [I_1(t), \dots, I_k(t)]$ is an input vector of bi-infinite power series over $\mathbb{F}$, let I_i^p denote the coefficient of t^p in $I_i(t)$. Then $I(t)$ is identified with the point

$$\dots (I_1^{-1}, \dots, I_k^{-1}).(I_1^0, \dots, I_k^0)(I_1^1, \dots I_k^1) \dots \in (\mathbb{F}^k)^{\mathbb{Z}}.$$

It is then straightforward to check that

$$E = \Phi_\infty^{[-M,N]}, \tag{1–6–4}$$

where

$$\Phi\big((I_1^{-M}, \dots, I_k^{-M}) \dots (I_1^N, \dots, I_k^N)\big) = \left(\sum_{j=-M}^{N} \sum_{i=1}^{k} I_i^j g_{i,1}^{-j},\ \dots,\ \sum_{j=-M}^{N} \sum_{i=1}^{k} I_i^j g_{i,n}^{-j}\right).$$

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

Hence the convolutional code $E\big((\mathbb{F}^k)^{\mathbb{Z}}\big)$, being the image of a full shift under a sliding block code, is a shift space by Theorem 1.5.13.

Note that both $(\mathbb{F}^k)^{\mathbb{Z}}$ and $(\mathbb{F}^n)^{\mathbb{Z}}$ are (infinite-dimensional) vector spaces over $\mathbb{F}$. Also observe from what we just did that a convolutional encoder is a linear transformation. Since the image of a vector space under a linear transformation is again a vector space, it follows that the corresponding convolutional code $E\big((\mathbb{F}^k)^{\mathbb{Z}}\big)$ is a linear subspace of $(\mathbb{F}^n)^{\mathbb{Z}}$, i.e., is a *linear shift space*. Furthermore, it is easy to check that $E\big((\mathbb{F}^k)^{\mathbb{Z}}\big)$ is also irreducible (Exercise 1.6.4). This proves the following result

Theorem 1.6.3.

(1) *Every convolutional encoder is a linear sliding block code.*
(2) *Every convolutional code is a linear irreducible shift space.*

In fact, the converses hold: the convolutional encoders are precisely the linear sliding block codes, and the convolutional codes are precisely the linear irreducible shift spaces (see Exercises 1.6.3 and 1.6.6).

Convolutional encoders are usually viewed as operating on the set of all k-tuples of Laurent series, i.e., objects of the form $\sum_{j=j_0}^{\infty} a_j t^j$, where j_0 may be negative. We have focused on bi-infinite power series instead in order to view these encoders better within the framework of symbolic dynamics. Each formalism has its advantages.

EXERCISES

1.6.1. Verify that in Example 1.6.2, $E(X) = X_{\mathcal{F}}$ where $\mathcal{F}$ is defined by (1–6–3).

1.6.2. Verify (1–6–4).

1.6.3. Let $\mathbb{F}$ be a finite field and E be a mapping from the full shift over $\mathbb{F}^k$ into the full shift over $\mathbb{F}^n$. Show that the following are equivalent:
(a) E is a convolutional encoder.
(b) E is a linear sliding block code, i.e., a sliding block code which is linear as a mapping between the vector spaces $(\mathbb{F}^k)^{\mathbb{Z}}, (\mathbb{F}^n)^{\mathbb{Z}}$.
(c) E is a map from the full F^k-shift to the full F^n-shift of the form $E = \Phi_{\infty}^{[-M,N]}$ where Φ is a linear map $\Phi\colon (\mathbb{F}^k)^{M+N+1} \to \mathbb{F}^n$; here M, N are integers and we identify $(\mathbb{F}^k)^{M+N+1}$ with $\mathbb{F}^{k(M+N+1)}$.

1.6.4. Show that every convolutional code has a fixed point and is irreducible.

***1.6.5.** Let $G(t)$ be a Laurent polynomial matrix. Give an algorithm to construct a finite list $\mathcal{F}$ from $G(t)$ such that $X_{\mathcal{F}}$ is the convolutional code defined by $G(t)$.

***1.6.6.** Show that a subset of the full $\mathbb{F}^n$-shift is a convolutional code if and only if it is a linear irreducible shift space, i.e., an irreducible shift space that is a linear subspace of $(\mathbb{F}^n)^{\mathbb{Z}}$.

Notes

Symbolic dynamics goes back to Hadamard [Had] (1898) and Morse [Mor1, Mor2] (1921) in modeling of geodesics on surfaces of negative curvature. A notion of shift space described by spelling out an explicit list of restrictions on the

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press

allowed sequences was given by Morse and Hedlund [MorH1, pp. 822–824] (1938); see also [MorH2] and Hedlund [Hed4]. A space of sequences was often described by declaring the allowed blocks to be those that appear in a particular bi-infinite sequence; for instance, see Gottschalk and Hedlund [GotH, Chap. 12]. However, in the generality of our textbook, shift spaces were not formally defined until Smale [Sma] (1967); he called them *subshifts*, viewing them as closed, shift invariant subsets of full shifts (in his definition, he also assumed that periodic points are dense). We have chosen the term "shift space" to emphasize the basic nature of these spaces.

See §6.5 and §13.6 for discussions of how shift spaces can be used to model smooth dynamical systems. Shift spaces can also be used to model constraints that naturally occur in data recording; see §2.5. The *Scientific American* article [Mon] gives a lucid explanation of how compact audio disks make use of certain shifts in recording Beethoven symphonies.

Underlying the theory of shift spaces are some fundamental topological notions such as compactness and continuity. We will explain these thoroughly in Chapter 6. But for readers already familiar with these ideas, the following gives a brief account of how they connect with symbolic dynamics.

There is a metric on the full shift such that two points are "close" if and only if they agree in a "large" central block (such a metric is given in Example 6.1.10). With respect to this metric, the full shift is compact and the shift map is continuous. A subset of a full shift is a shift space precisely when it is compact and shift invariant. Sliding block codes are exactly those maps from one shift space to another that are continuous and commute with the shift map. Theorem 1.5.13 then follows from the result that the continuous image of a compact set is compact, while Theorem 1.5.14 follows from the general result that a continuous one-to-one map on a compact metric space has a continuous inverse. The Cantor diagonal argument replaces compactness in our proofs of these results. Other facts are easier to understand from the topological viewpoint. Exercise 1.2.5, for example, translates to the statement that the intersection of compact subsets is compact. We remark that the metric on the full shift mentioned above is compatible with the product topology on the full shift using the discrete topology on its alphabet.

There is a version of shift spaces in which the alphabet is allowed to be countably infinite (see §13.9). However, the requirement of a finite alphabet that we have imposed allows us to take advantage of two very important tools: the Cantor diagonalization argument and, as we shall see in later chapters, finite-dimensional linear algebra. There is also a version of shift spaces where the sequences are one-sided, i.e., indexed over the nonnegative integers rather than the integers (see §5.1 and §13.8). One reason we choose to focus on bi-infinite sequences is to allow for memory as well as anticipation in sliding block codes.

Convolutional encoders and convolutional codes are central objects in coding theory. We refer the reader to [McE], [LinC], [Pir] for further reading on this subject.

Exercise 1.2.14 is due to M. Keane; Exercise 1.3.8 to A. Khayrallah and D. Neuhoff [KhaN1]; Exercise 1.5.16 to E. Coven and M. Paul [CovP2]; Exercise 1.5.15 to V. DeAngelis; and Exercise 1.5.17 to D. A. Dastjerdi and S. Jangjoo [DasJ] and M. Hollander and J. Von Limbach.

https://doi.org/10.1017/9781108899727.003 Published online by Cambridge University Press