EXAMENSARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

# Shift Spaces and Two Notions of Entropy

av

**Gunnar Höglund**

2007 - No 11

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 10691 STOCKHOLM

# Shift Spaces and Two Notions of Entropy

Gunnar Höglund

Examensarbete i matematik 20 poäng, fördjupningskurs

Handledare: Yishao Zhou

2007

# Abstract

There are several different notions of entropy in mathematics. Chris Hillman discusses a couple of these in his paper “All Entropies Agree for an SFT”, and presents the result that they all agree numerically on shifts of finite type.

The focus in this report is on shift spaces in general and two notions of entropy in particular - topological and probabilistic - and their relationship will be shown more thoroughly. In order to do this we will begin with an overview of shift spaces and continue with basic graph theory. After presenting proof of some results for shift spaces and shift dynamical systems the report is concluded with one section on topological entropy and one on probabilistic entropy.

# Contents

# Notes

First and foremost I have relied on three different sources when writing this report. The sections on shift spaces and graph theory is mainly from [1], the section on topological entropy is mainly from [2] and the section on probabilistic entropy is mainly from [3] (for details, see the Reference Notes). When relying on a different source this will be noted in the text or in a footnote.

Regarding the proofs: I have chosen to move some of the proofs to the Appendix. This is done because I think they are tedious.

# 1 Fundamental Properties of Shift Spaces

Simplifying assumptions are frequently made in mathematics. For example, continuity is often assumed when dealing with a discrete case of some sort. In information theory a similar assumption is often made: Information is commonly represented as strings of ones and zeroes and even though these strings are finite it is in general useful to treat long strings as infinite in both directions.

Basically, a shift space is a compact space where every point is a bi-infinite sequence of symbols taken from a finite set. Associated with each shift space is a shift map that moves every sequence "one step to the left" and together, as will be seen later in the report, these form an elementary dynamical system.

## 1.1 Basic Definitions

**Definition** An alphabet $\mathcal{A}$ is a non-empty, finite set. The elements in $\mathcal{A}$ are called symbols or letters. A bi-infinite sequence of symbols $x_i \in \mathcal{A}, i \in \mathbb{Z}$ will be denoted by

$$x = \{x_i\}_{i\in\mathbb{Z}} \text{ or } x = \ldots x_{-1}\overset{*}{x}_0 x_1 \ldots .$$

and referred to as a point. The symbol $x_j$ is the $j$th coordinate of $x$ and the symbol "$*$" always occurs over the 0th coordinate of each point.

**Definition** The full $\mathcal{A}$-shift, $\mathcal{A}^{\mathbb{Z}} = \{(x_i)_{i\in\mathbb{Z}} : x_i \in \mathcal{A} \text{ for all } i \in \mathbb{Z}\}$ is the collection of all bi-infinite sequences of symbols from $\mathcal{A}$.

Finite strings of symbols, or blocks, will play an important part in characterizing the special subsets of the full shift, called shift spaces.

**Definition** A block, or word, over $\mathcal{A}$ is a finite sequence of symbols from $\mathcal{A}$. The sequence of no symbols, the empty block, is denoted by $\epsilon$. The lenght of a block $u$, denoted by $|u|$, is the number of symbols it contains. $|\epsilon| = 0$ and if $u$ is a non-empty block of $k$ symbols

$$u = a_1 a_2 \ldots a_k, \ \ |u| = k.$$

A k-block is a block of lenght $k$.

**Definition** For $x \in \mathcal{A}^{\mathbb{Z}}$ and $i \leqslant j$, define $x_{[i,j]} = x_i x_{i+1} \ldots x_j$. If $i{>}j$, define $x_{[i,j]} = \epsilon$. If $x_{[i,j]} = w$ we say that $w$ occurs in $x$.

**Remark** Two blocks $u, v$ can be put together forming new blocks $uv$ and $vu$, $|uv| = |vu| = |u| + |v|$. If $u$ is a letter or a block, the point $\ldots uuuuu \ldots$ is denoted by $u^{\infty}$.

Using this notation it is easy to define shift spaces.

**Definition** Let $\mathcal{F}$ be a collection of blocks over $\mathcal{A}$. For any such $\mathcal{F}$, define

$$\mathsf{X}_{\mathcal{F}} = \{x \in \mathcal{A}^{\mathbb{Z}} : \text{ no blocks in } \mathcal{F} \text{ occurs in } x\}.$$

A shift space or shift is a subset $X \subset \mathcal{A}^{\mathbb{Z}}$ such that $X = \mathsf{X}_{\mathcal{F}}$, for some $\mathcal{F}$.

**Example** Let $\mathcal{A} = \{0, 1\}$ and $X$ be all sequences in $\mathcal{A}^{\mathbb{Z}}$ with no two 1's next to each other. Then $X = \mathsf{X}_{\mathcal{F}}$, where $\mathcal{F} = \{11\}$. The shift $X$ is called the golden mean shift. The reason for the name will surface in section 5.3.

Up to this point shift spaces are static spaces. The shift map, now introduced, will add dynamics to the shift spaces and turn them into dynamical systems.

**Definition** Let $X$ be a shift space and $x = \ldots x_{-1}\overset{*}{x}_0 x_1 \ldots \in X$. The shift map or shift operator $\sigma_X : X \to X$ is defined by $(\sigma_X x)_i = x_{i+1}$ or

$$\sigma_X(\ldots x_{-1}\overset{*}{x}_0 x_1 \ldots) = \ldots x_{-1} x_0 \overset{*}{x}_1 x_2 \ldots, \text{ for all } x \in X.$$

In other words, $\sigma_X$ shifts each sequence one step to the left. There is also the inverse operation $\sigma_X^{-1}$ of shifting one step to the right. When $X$ is understood from the context we will denote $\sigma_X$ by $\sigma$.

Thinking of time as discrete and $x$ as a message emitted by some sort of permanent source[1], the $i$th coordinate of $x$ can then be thought of as the symbol emitted by the source at time $i$. Analogically, shifting the sequence one step to the left can be thought of as the passage of time by one time-unit.

**Definition** A point $x$ is periodic for $\sigma$ if $\sigma^n(x) = x$ for some $n \geqslant 1$. If $\sigma^n(x) = x$, $x$ is said to have period $n$ under $\sigma$.

The allowed blocks of a shift space are similar to the words of a language. When pursuing this analogy further the notion of languages of shift spaces is obtained; instead of describing a shift space by specifying which blocks are forbidden, the language specifies which blocks are allowed.

**Definition** Let $X \subset \mathcal{A}^{\mathbb{Z}}$ 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** The golden mean shift $X$ has language

$$\{\epsilon, 0, 1, 00, 01, 10, 000, 001, 010, 100, 101, 0000, \ldots\},$$

and if $f_n$ denotes the $n$th Fibonacci number $(f_1 = 0, f_2 = 1, f_3 = 1, \ldots)$ then the number of blocks of length $n$, $|\mathcal{B}_n(X)| = f_{n+3}$.

## 1.2 The Higher Block Presentation

Visualize an alphabet in which each symbol is a block of symbols from another alphabet. By using this "block"-alphabet it is possible to transform shift spaces into more complex ones.

This is the concept of higher block presentations of shift spaces and it provides a useful alternative description of a shift space.

[1] For a more comprehensive view on the subject, ct. Chapter II of [3].

**Definition** Let $u = u_1 u_2 \dots u_N$ and $v = v_1 v_2 \dots v_N$ be $N$-blocks ($N \geqslant 2$). We say that $u$ and $v$ overlap progressively if $u_2 u_3 \dots u_N = v_1 v_2 \dots v_{N-1}$. For the sake of simplicity we say that all 1-blocks overlap progressively.

**Definition** Let $X$ be a shift space over the alphabet $\mathcal{A}$. Then we can construct a new alphabet $\mathcal{A}_X^{[N]}$ by using blocks from $X$ such that $\mathcal{A}_X^{[N]} = \mathcal{B}_N(X)$. The "new" full shift is denoted $(\mathcal{A}_X^{[N]})^{\mathbb{Z}}$. Define the $N$th higher block code $\beta_N \colon X \to (\mathcal{A}_X^{[N]})^{\mathbb{Z}}$ by

$$(\beta_N(x))_i = x_{[i,i+N-1]}, \quad \text{for all } x \in X.$$

Basically, $\beta_N$ replaces the $i$th coordinate of $x$ with the $N$-block starting at position $i$. It might be easier to picture this if we imagine the symbols of $\mathcal{A}_X^{[N]}$ written vertically.

**Example**

$$\beta_4(x) = \dots \begin{bmatrix} x_1 \\ x_0 \\ x_{-1} \\ x_{-2} \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \\ x_0 \\ x_{-1} \end{bmatrix} \overset{*}{\begin{bmatrix} x_3 \\ x_2 \\ x_1 \\ x_0 \end{bmatrix}} \begin{bmatrix} x_4 \\ x_3 \\ x_2 \\ x_1 \end{bmatrix} \begin{bmatrix} x_5 \\ x_4 \\ x_3 \\ x_2 \end{bmatrix} \dots . \tag{1.1}$$

Having defined the higher block code, we use this to transform a whole shift space into a new one.

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

**Remark** Equation 2.1 emphasize two important characteristics of $X^{[N]}$:

1. The consecutive symbols of $x \in X^{[N]}$ overlap progressively.
2. The "bottom" letters in the symbols of $\beta_N(x)$ constitues the original point $x$. In this sense $X^{[N]}$ is simply another description of $X$.

Since the higher block code more or less just switches one symbol for another, the next proposition should not come as a surprise. The proof of it can be found in the appendix.

**Proposition 1.1** *Let $X$ be a shift space. The higher block presentations of $X$ are shift spaces.*

**Example** Let $X$ be the golden mean shift. Then

$$\mathcal{A}_X^{[2]} = \mathcal{B}_2(X) = \{a = 00, b = 01, c = 10\}$$

and $X^{[2]} = \mathsf{X}_{\mathcal{F}}$, where $\mathcal{F} = \{ac, ba, bb, cc\}$. The 2-blocks in $\mathcal{F}$ are forbidden since they fail to overlap progressively.

## 1.3 Sliding Block Codes

A sliding block code is a rule that transforms one string of symbols into another. The basic concept is that applying this rule to one sequence $x$ will transform it into another sequence $y$, where each coordinate $y_i$ is determined by a block surrounding the coordinate $x_i$. How big this surrounding block is, and exactly how the transformation presents itself depends on the sliding block code.

The main result of this section is a formal proof of that $X$ and $X^{[N]}$ are two equivalent representations of the same "structure".

**Definition** Let $X$ be a shift space, $\mathcal{A}'$ an alphabet and $\Phi\colon \mathcal{B}_{m+n+1}(X) \to \mathcal{A}'$. Then the map $\phi\colon X \to \mathcal{A}'^{\mathbb{Z}}$ defined by $y = \phi(x)$ with

$$y_i = \Phi(x_{[i-m,i+n]})$$

is called the sliding block code with memory $m$ and anticipation $n$ induced by $\Phi$. We write $\phi = \Phi_\infty^{[-m,n]}$ or $\phi = \Phi_\infty$ if the memory and anticiption of $\phi$ are understood.

**Remark** The shift map $\sigma$ is a sliding block code.

**Proof** Let $X$ be a shift space over the alphabet $\mathcal{A}$, $m = 0$, $n = 1$, $x \in X$ and

$$\Phi\colon \mathcal{B}_{m+n+1}(X) \to \mathcal{A} \text{ such that } \Phi(x_i x_{i+1}) = x_{i+1}.$$

Then $\phi = \Phi_\infty^{[0,1]}$ is the shift map $\sigma$.

**Definition** Let $\phi\colon X \to Y$ be a sliding block code. Then $\phi$ is a factor code from $X$ onto $Y$ if it is onto. A shift space $Y$ is a factor of $X$ if there is a factor code from $X$ onto $Y$. If $\phi$ is invertible, $\phi$ is said to be a conjugacy from $X$ to $Y$. Two shift spaces $X$ and $Y$ are conjugate, denoted $X \cong Y$, if there is a conjugacy from $X$ to $Y$.

**Remark** If $X \cong Y$, $X$ is simply a recoded version of $Y$ and vice versa. They share the same properties and are essentially the same.

Sometimes it is hard to tell two shift spaces from each other. As Proposition 1.3 will show and as was hinted in the remark on the previous page, even though $X$ and $X^{[N]}$ "look" quite different, they are merely recoded versions of each other. One way to find out if two shift spaces are not conjugate is to look at certain properties of them that do not change under conjugacies.

**Definition** A conjugacy invariant or invariant, assignes values to shift spaces in such a manner that conjugate shifts will obtain the same value.

Defined below is the notion of irreducibility and as shown in the following proposition, irreducibility is a conjugacy invariant.

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

**Proposition 1.2** *Let $X$ and $Y$ be two shift spaces such that $X \cong Y$. If $X$ is irreducible then so is $Y$.*

**Proof** Let $\phi = \Phi_{\infty}^{[-m,k]}$ be a sliding block code from $X$ onto $Y$. Then every block in $\mathcal{B}_n(Y)$ is the image of a block in $\mathcal{B}_{n+m+k}(X)$. I will denote this $u' = \Phi(u)$ for $u \in \mathcal{B}_{n+m+k}(X)$ and $u' \in \mathcal{B}_n(Y)$.

Now, assume that $X$ is irreducible and let $u', v' \in \mathcal{B}(Y)$ with $u' = \Phi(u),\ v' = \Phi(v)$ for some $u, v \in \mathcal{B}(X)$. Since $X$ is irreducible there exists a $w \in \mathcal{B}(X)$ such that $uwv \in \mathcal{B}(X)$.

$$\Phi(uwv) = \Phi(u)\Phi(u_{[|u|-m-k+1,|u|]}wv_{[1,k+m]})\Phi(v) =$$

$$= u'w'v',\ \text{for some } w' \in \mathcal{B}(Y).$$

Thus for all ordered pair of blocks $u', v' \in \mathcal{B}(Y)$ there exists a $w' \in \mathcal{B}(Y)$ such that $u'w'v' \in \mathcal{B}(Y)$ and $Y$ is irreducible.

Using the definition to determine whether a shift space is irreducible or not might seem a bit tiresome. Another, more functional way of determining this is presented in section 2.2.

The following proposition is essential for the report and will later on enable an extensive use of graphs to analyze the special kind of shift spaces covered in in 1.4, shifts of finite type.

**Proposition 1.3** *Let $X$ be a shift space over $\mathcal{A}$ and $N \in \mathbb{N}$. Then $X \cong X^{[N]}$.*

**Proof** Let $x \in X$ and define $\Phi\colon \mathcal{B}_N(X) \to \mathcal{A}_X^{[N]}$ by

$$\Phi(x_{[i,i+N-1]}) = x_{[i,i+N-1]}.$$

Then $\phi = \Phi_\infty\colon X \to X^{[N]}$ is the $N$th higher block code $\beta_N$. Now define $\Psi\colon \mathcal{A}_X^{[N]} \to \mathcal{A}$ by

$$\Psi(x_{[i,i+N-1]}) = x_i$$

then $\psi = \Psi_\infty\colon X^{[N]} \to X$ and $\psi = \beta_N^{-1}$. Thus $\beta_N\colon X \to X^{[N]}$ is a conjugacy and $X \cong X^{[N]}$. □

## 1.4 Shifts of Finite Type

Perhaps the most widely studied class of shift spaces are the shifts of finite type, or SFT for short. The SFT are the shifts that can be described by a finite set of forbidden blocks.

One reason why these shifts are so useful is that they are represented by finite directed graphs. Questions about a shift can be answered by examining the graph and its adjacency matrix. Because of this property they will be the main study for the rest of this report.

**Definition** Let $X = \mathsf{X}_{\mathcal{F}}$ be a shift space. If $\mathcal{F}$ is finite, $X$ is called a shift of finite type.

Note that it might be possible for a shift of finite type to be described by an infinite set of forbidden blocks.

**Definition** A shift of finite type is $M$-step (or has memory $M$) if it can be described by a collection of forbidden blocks all of which have length $M+1$.

The notion of memory for a shift of finite type can be thought of as the number of symbols necessary to keep in mind in order to determine whether a string is allowed or not. Consider a $M$-step shift of finite type and a block $a = a_1 a_2 ... a_n$, where $n$ much larger than $M$. One way of deciding whether $a$ is allowed or not is to check every $(M+1)$-block that occurs in $a$. To do this successfully one needs to remember the first $M$ symbols in every $(M+1)$-block.

**Example** The golden mean shift is a shift of finite type since it can be described by the forbidden set $\{11\}$. It is also a 1-step shift.

Not surprisingly, all shifts of finite type are $M$-step shifts, for some non-negative integer $M$.

**Remark** If $X$ is a shift of finite type, then there is an $M \geqslant 0$ such that $X$ is $M$-step.

**Proof** Let $X = \mathsf{X}_{\mathcal{F}}$, where $\mathcal{F}$ is finite. If $\mathcal{F} = \varnothing$ then $X$ is a full shift and $M = 0$ since the forbidden blocks are letters (which have length 1).

If $\mathcal{F} \neq \varnothing$: Let $M+1$ be the lenght of the longest block in $\mathcal{F}$ and create a new collection of forbidden blocks $\mathcal{F}'$ by replacing each $w \in \mathcal{F}$ with all blocks of lenght $M+1$ such that $w$ occurs in them. Then $\mathsf{X}_{\mathcal{F}'} = \mathsf{X}_{\mathcal{F}}$. □

# 2 Graphs and Their Shifts

The main results of this section is that every shift of finite type can be represented by a directed graph and is conjugate to a shift on this graph - the edge shift. Not only does this help to prove many important results but it also shows the connection between shifts of finite type and Markov chains.

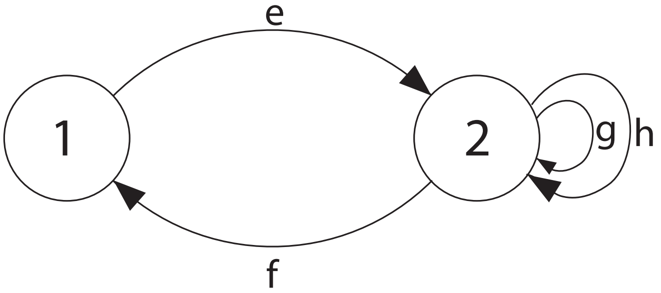
## 2.1 Basic Definitions

**Definition** A directed graph $G$ consists of a finite set $\mathcal{V}(G)$ of vertices together with a finite set $\mathcal{E}(G)$ of edges. Each edge $e \in \mathcal{E}(G)$ is an ordered set $(\mathsf{i}(e), \mathsf{t}(e))$ of two vertices called initial state and terminal state of $e$.

When $G$ is understood from the context we will denote $\mathcal{V}(G)$ by $\mathcal{V}$ and $\mathcal{E}(G)$ by $\mathcal{E}$.

Note that the definition does not allow one edge to start or end at multiple vertices.

**Example figure**



(2.1)

`Example graph.`

This graph has vertex set $\mathcal{V} = \{1, 2\}$ and edge set $\mathcal{E} = \{e, f, g, h\}$.

Another, perhaps more convenient way to describe a graph is in matrix form.

**Definition** Let $G$ be a graph with vertex set $\mathcal{V}$. For $i, j \in \mathcal{V}$, let $a_{ij}$ denote the number of edges in $G$ with initial state $i$ and terminal state $j$. The adjacency matrix of $G$, $\mathsf{A}_G = [a_{ij}]$.

Note that the adjacency matrix gives no information whatsoever of which one of the edges that start or end at a specific vertex.

**Example** The graph (2.1) has the adjacency matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}.$$

Similar to the construction of the adjacency matrix from a graph is the construction of a graph from a square matrix.

**Definition** Let $A = [a_{ij}]$ be an $r \times r$ matrix with non-negative integer entries. Then the graph of $A$, $G = \mathsf{G}_A$, is the graph with vertex set $\mathcal{V} = \{1, 2, \dots, r\}$, and with $a_{ij}$ distinct edges with initial state $i$ and terminal state $j$.

Note that if $G_1$ is a graph with adjacency matrix $A$ and $G_2$ is another graph constructed from $A$, $G_1$ and $G_2$ will only differ by the names of the vertices and edges.

## 2.2 The Edge Shift

By letting all the edges in a graph represent different letters, it is possible to turn a graph into a shift space over the alphabet composed of the edges of the graph. Every block belonging to a point in this shift space will correspond to a walk on the graph. Every point will correspond to a bi-infinite walk. Depending on the graph's appearance, some imaginable blocks might not be the product of any walk on the graph. These blocks correspond to the forbidden blocks.

In this paragraph it is shown that every shift of finite type is conjugate to a shift on a graph.

**Definition** Let $G$ be a graph with edge set $\mathcal{E}$ and adjacency matrix $A$. The edge shift $\mathsf{X}_G$ or $\mathsf{X}_A$ is the shift space over the alphabet $\mathcal{E}$ defined by

$$\mathsf{X}_G = \mathsf{X}_A = \{(e_i)_{i \in \mathbb{Z}} \colon \mathsf{t}(e_n) = \mathsf{i}(e_{n+1}), \text{ for all } n \in \mathbb{Z}\},$$

where $e_n$ are the ordered sets of edges of the graph $G$, for $n \in \mathbb{Z}$.

In other words, a bi-infinite sequence of edges is in $\mathsf{X}_G$ exactly when the terminal state of each edge is the initial state of the following one. We say that each sequence in $\mathsf{X}_G$ describes a bi-infinite walk on $G$

**Proposition 2.1** *If $G$ is a graph with adjacency matrix $A$, then the associated edge shift $\mathsf{X}_G = \mathsf{X}_A$ is a 1-step shift of finite type.*

**Proof** Let $\mathcal{A} = \mathcal{E}$ be the alphabet of $\mathsf{X}_G$. Define

$$\mathcal{F} = \{ef \colon e, f \in \mathcal{A}, \ \mathsf{t}(e) \neq \mathsf{i}(f)\}$$

Then $\mathsf{X}_G = \mathsf{X}_{\mathcal{F}}$ which is 1-step of finite type. □

So every edge shift is a 1-step shift of finite type. This result should not come as a surprise since a graph has "no memory". During a random walk on a graph the current position (vertex) is always known, but the path traveled up to that point could be impossible to tell.

**Example** The edge shift $\mathsf{X}_G$ constructed from the graph (2.1) is a 1-step shift of finite type:

$$\mathsf{X}_G = \mathsf{X}_{\mathcal{F}}, \ \mathcal{F} = \{fg, fh, ge, he\}.$$

A graph with vertices that are impossible to walk to or from, obviously generates the same edge shift as a graph without these "stranded" vertices. The following definitions and remark states this in a more formal way.

**Definition** A graph is essential if all vertices has at least one edge starting at it and at least one edge terminating at it.

**Definition** Let $G$ and $H$ be a graphs, $e_G \in \mathcal{E}(G)$ and $e_H \in \mathcal{E}(H)$. A graph $H$ is a subgraph of $G$ if

$$\mathcal{V}(H) \subset \mathcal{V}(G),\ \mathcal{E}(H) \subset \mathcal{E}(G),$$

and for each $e_H$ there exists $e_G$ such that

$$\mathsf{i}(e_H) = \mathsf{i}(e_G) \text{ and } \mathsf{t}(e_H) = \mathsf{t}(e_G).$$

**Remark** If $G$ is a graph, then there exists a unique subgraph $H$ of $G$ such that $H$ is essential and $\mathsf{X}_H = \mathsf{X}_G$.

**Proof** Let $\mathcal{E}(H)$ consist of the union of edges that appear in the bi-infinite walks on $G$ and let $\mathcal{V}(H)$ be the union of the vertices visited on such walks. Then $H$ is an essential subgraph of $G$ and any bi-infinite walk on $G$ is a walk on $H$ and vice versa. By definition $H$ is the largest essential subgraph of $G$ and since every edge in $H$ occurs in $\mathsf{X}_H$, $\mathsf{X}_H \neq \mathsf{X}_{H'}$, for any other essential subgraph $H'$. □

From now on, when dealing with edge shifts it can always be assumed that the graph generating the edge shift is essential.

**Definition** A path $\pi = e_1 e_2 \ldots e_m$ on a graph $G$ is a finite sequence of edges $e_j \in \mathcal{E}$ such that $\mathsf{t}(e_j) = \mathsf{i}(e_{j+1})$ for $1 \leqslant j \leqslant m-1$. The lenght of $\pi$, $|\pi|$ is the number of edges it traverses. The path $\pi$ starts at vertex $\mathsf{i}(\pi) = \mathsf{i}(e_1)$, terminates at vertex $\mathsf{t}(\pi) = \mathsf{t}(e_m)$ and is a path from $\mathsf{i}(\pi)$ to $\mathsf{t}(\pi)$. $\pi$ is a cycle if $\mathsf{i}(\pi) = \mathsf{t}(\pi)$. For each vertex $k$ there is an empty path $\varepsilon_k$ such that $|\varepsilon_k| = 0$ and $\mathsf{i}(\varepsilon_k) = \mathsf{t}(\varepsilon_k) = k$.

**Remark** There is a one-to-one correspondence between the paths on $G$ and the blocks in $\mathsf{X}_G$.

**Proposition 2.2** *Let $G$ be a graph with adjacency matrix $A$ and let $m \geqslant 0$. Then:*

1. *The number of paths of lenght $m$ from $i$ to $j$ is $(A^m)_{ij}$, the $(i,j)$th entry of $A^m$.*
2. *The number of cycles of lenght $m$ in $G$ equals the number of points in $\mathsf{X}_G$ with period $m$.*

**Proof** A proof is presented in the appendix. □

Recall that a shift space is irreducible if for all ordered pair of blocks $u, v \in \mathcal{B}(X)$ there exists a $w \in \mathcal{B}(X)$ such that $uwv \in \mathcal{B}(X)$. The corresponding property of a graph (and its matrix) is both less technical to define and easier to visualize.

**Definition** A graph $G$ is irreducible if for every ordered pair of vertices $i, j$ there is a path in $G$ from $i$ to $j$.

**Corollary 2.3** *An essential graph is irreducible if and only if its edge shift is irreducible.*

**Proof** This is a direct consequence of the previous remark. A more detailed proof is presented in the appendix. □

Note that an irreducible graph is essential.

**Definition** A non-negative, square matrix $A$ is irreducible if for each ordered pair of indices $i, j$ there exists $n \geqslant 0$ such that $(A^n)_{ij} > 0$. Since $A^0$ by definition equals $Id$ for any matrix $A$, the $1 \times 1$-matrix $[0]$ is irreducible. A non-negative matrix is essential if none of its rows or columns is zero.

Note that irreducibility of a graph is equivalent to irreducibility of its adjacency matrix. The same thing is true for essentiality.

The following theorem is the main theorem of this section. It shows that every shift of finite type is conjugate to an edge shift. This is a fundamental result and will be the core of many of the following ideas and proofs.

**Theorem 2.4** *If $X$ is an $M$-step shift of finite type, then there is a graph $G$ such that $X^{[M+1]} = \mathsf{X}_G$.*

**Proof** If $M = 0$, then $X$ is a full shift and we can take $G$ to have a single vertex and one edge for each symbol appearing in $X$. If $M \geqslant 1$, let $\mathcal{V}(G) = \mathcal{B}_M(X)$ and define $\mathcal{E}(G)$ as follows: Suppose that $a = a_1 a_2 \ldots a_M$ and $b = b_1 b_2 \ldots b_M$ are two vertices in $\mathcal{V}$. If

$$a \text{ and } b \text{ overlap progressively and if } a_1 a_2 \ldots a_M b_M \in \mathcal{B}(X)$$

then draw one edge in $G$ from $a$ to $b$ named $a_1 a_2 \ldots a_M b_M = a_1 b_1 b_2 \ldots b_M$. Otherwise, there is no edge from $a$ to $b$. A bi-infinite walk on $G$ is precisely a sequence of blocks in $\mathcal{B}_{M+1}(X)$ which overlap progressively. Whence $\mathsf{X}_G = X^{[M+1]}$. □

**Remark**

1. Let $M \geqslant 1$ and $G$ be a graph created using the method from the previous proof. If $A$ is the adjacency matrix of $G$ then $A = [a_{ij}]$ with $a_{ij} \in \{0, 1\}$.

2. If $X$ is an $(n-1)$-step shift of finite type, then $X \cong X^{[n]} = \mathsf{X}_G$ for some $G$. The matrix $\tilde{A}$ that Hillman defines on page 3 in [4] is the adjacency matrix $\mathsf{A}_G$.

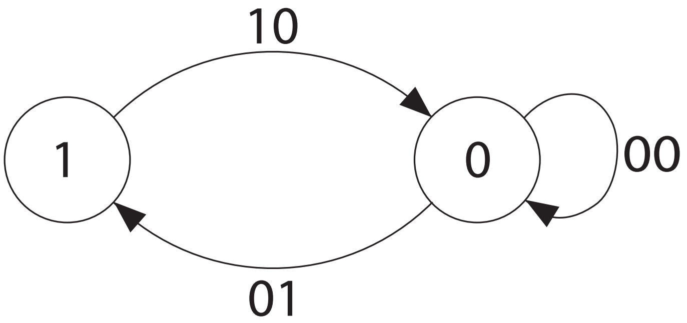
**Example** The golden mean shift is 1-step but is not itself an edge shift[2]. Carrying out the process described in the above proof shows that $X^{[2]} = \mathsf{X}_G$, where $G$ is the graph (2.2). The adjacency matrix of the recoded golden mean shift is[3]

$$\mathsf{A}_G = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

[2] A proof of this is presented on page 41 in [1].

[3] cf. [4] page 3.

**Example figure**



(2.2)

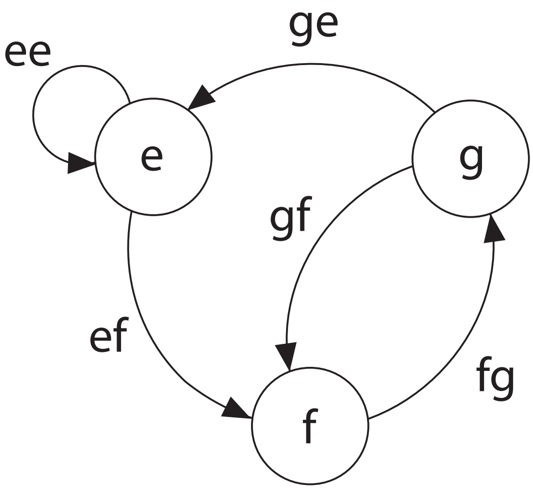
`The recoded golden mean shift.`

The following remark summarizes some important facts and illuminates the connection between irreducible shifts of finite type and irreducible graphs.

**Remark** Assume that $X$ is a shift of finite type. Then $X$ is conjugate to some edge shift (Proposition 1.3 and Theorem 2.4) and if $X$ is irreducible then so is the edge shift (Proposition 1.2). Then, according to Corollary 2.3, the essential graph that generates the edge shift is also irreducible. A similar revers argument is also true. In this sense there is a correspondence between the irreducible shifts of finite type and the irreducible graphs.

We conclude this paragraph by trying to use irreducibility to prove that two shift spaces are not conjugate.

**Example** Let $\mathcal{A} = \{e, f, g\}$ and let $X = \mathsf{X}_{\mathcal{F}}$ be the shift space over $\mathcal{A}$ described by the forbidden set $\mathcal{F} = \{eg, fe, ff, gg\}$ and the graph below



(2.3)

Obviously $X$ is a 1-step shift of finite type, just like the golden mean shift, but comparing the graphs it is not obvious whether this is a recoded version of the golden mean shift or not. Unfortunately, both graphs are irreducible so this will not help telling the shifts apart.

**Example** The full 2-shift is the full shift space over the alphabet $\mathcal{A} = \{0, 1\}$. Since there do not exist any forbidden strings, this shift too is irreducible. Even though we do not expect this shift to be conjugate the golden mean shift, we can not use this property to tell them apart.

The next section will show a correspondence between another type of shifts and graphs.

# 3 Further Properties of Shift Spaces

The focus in section 3.1 is on the adjacency matrix and periodicity. Also an important theorem is stated, which in section 3.2 leads to a special class of shift spaces called mixing shifts.

## 3.1 Periods of states, matrices and graphs

**Definition** Let $A$ be a non-negative matrix. The period of state $i$,

$$\text{per}(i) = \gcd\{n \geqslant 1\colon (A^n)_{ii} > 0\}.$$

If no such integers exist we define per($i$)= $\infty$. The period of a matrix $A$,

$$\text{per}(A) = \begin{cases} \gcd\{\text{finite per}(i), \text{ where } i \text{ are states of } A\}, \\ \infty, \text{ if per}(i) = \infty \text{ for all states } i \text{ of } A. \end{cases}$$

A matrix is aperiodic if it has period 1. The period of a graph, per($G$) is the period of its adjacency matrix.

**Definition** Let $X$ be a shift space and let $p_n(X)$ denote the number of points in $X$ with period $n$. The period of $X$,

$$\text{per}(X) = \begin{cases} \gcd\{n \geqslant 1\colon p_n(X) > 0\}, \\ \\ \infty, \text{ if } p_n(X) = 0 \text{ for all } n \in \mathbb{N}. \end{cases}$$

**Remark** Let $G$ be a graph. Every block in $\mathsf{X}_G$ with period $n$ corresponds to a cycle on $G$ of length $n$ and vice versa, thus

$$\text{per}(\mathsf{X}_G) = \text{per}(G).$$

**Definition** A matrix is primitive if it is irreducible and aperiodic. A graph is primitive if its adjacency matrix is primitive.

If $A$ is a matrix such that every element of $A$ is positive, we write $A > 0$.

The following theorem will be needed to gain a deeper understanding of the shifts presented in the next section.

**Theorem 3.1** *Let $A$ be a non-negative matrix. The following are equivalent:*

1. *$A$ is primitive.*
2. *$A^N > 0$ for some $N \geqslant 1$.*
3. *$A^N > 0$ for all sufficiently large $N$.*

**Proof** A proof is presented in the appendix.

## 3.2 Mixing Shifts

As seen earlier in 2.2, and as the name suggests, irreducible shifts of finite type correspond to irreducible graphs. This fact might raise the question: Which class of shifts of finite type correspond to the primitive graphs? The answer is the "mixing" shifts. The results of this section will be used in section 4.3 to prove some of the claims made by Hillman in [4].

**Definition** A shift space $X$ is mixing if for every ordered pair $u, v \in \mathcal{B}(X)$, there exists an $N$ such that for each $n \geqslant N$ there exists a $w \in \mathcal{B}_n(X)$ such that $uwv \in \mathcal{B}(X)$.

The next proposition is the "mixing analogue" of Proposition 1.2, Corollary 2.3 and the final remark on page 11.

**Proposition 3.2**

1. *Let $X$ and $Y$ be two shift spaces such that $X \cong Y$. If $X$ is mixing then so is $Y$.*
2. *If $G$ is an essential graph, then the edge shift $\mathsf{X}_G$ is mixing if and only if $G$ is primitive.*
3. *A shift of finite type is mixing if and only if it is conjugate to an edge shift $\mathsf{X}_G$ where $G$ is primitive.*

**Proof**

1. The proof of this is almost identical to the proof of Proposition 1.2 and is therefor omitted.
2. Suppose that $\mathsf{X}_G$ is mixing and let $\pi, \tau \in \mathcal{B}(\mathsf{X}_G)$ such that

   $$\mathsf{t}(\pi) = k,\ \mathsf{i}(\tau) = l, \text{ where } k, l \in \mathcal{V}.$$

   Let $\omega^{(n)} \in \mathcal{B}_n(\mathsf{X}_G)$ such that $\pi\omega^{(n)}\tau \in \mathcal{B}(\mathsf{X}_G)$, for all $n \geqslant N(k,l)$. Then there is a path from $k$ to $l$ of lenght $n$ for all $n \geqslant N(k,l)$. Let

   $$M = \max_{i,j \in \mathcal{V}} N(i,j),$$

   then $A^M > 0$ and $G$ is primitive.

   Conversely, suppose that $G = \mathsf{G}_A$ is primitive and let $\pi, \tau$ be paths on $G$ such that

   $$\mathsf{t}(\pi) = k, \mathsf{i}(\tau) = l, \text{ where } k, l \in \mathcal{V}.$$

   Since $A^n > 0$ for all $n \geqslant N$ there exists a path $\omega \in \mathcal{B}_n(\mathsf{X}_G)$, for all $n \geqslant N$ such that

   $$\mathsf{i}(\omega) = k,\ \mathsf{t}(\omega) = l.$$

   Then $\pi\omega\tau \in \mathcal{B}(\mathsf{X}_G)$ and $\mathsf{X}_G$ is mixing.
3. The result follows from *1*, *2* and Theorem 2.4.

Hence, analogue of the correspondence between irreducible shifts of finite type and irreducible graphs, there exists a correspondence between mixing shifts of finite type and primitive graphs. The reason why mixing shifts are called “mixing” will surface in the next section.

Since mixing is another conjugacy invariant, and a more “narrow” one than irreducibility, it is possible that it could help proving that the full 2-shift, the golden mean shift and the shift described by (2.3) are not conjugate. Unfortunately this can not be done since all three of them are mixing.

# 4 Dynamical Systems

By combining a shift space and its shift map with a metric, they form a dynamical system. Of course, the problem to distinguish different dynamical systems from each other persists and in 4.2 the topological analogues of "irreducibility" and "mixing" are presented. Later on their connection with the adjacency matrix are showed.

**Definition** A dynamical system $(M, \phi)$ consists of a compact metric space $M$ together with a continuous map $\phi\colon M \to M$. If $\phi$ is a homeomorphism[4] we call $(M, \phi)$ an invertible dynamical system.

## 4.1 Shift Dynamical Systems

Using a plausible metric, a shift space together with its shift map form an invertible dynamical system.

**Definition** Let $x, y \in \mathcal{A}^{\mathbb{Z}}$. A metric on $\mathcal{A}^{\mathbb{Z}}$ is given by

$$d(x,y) = \begin{cases} 2^{-n}, \text{ if } x \neq y, \text{ where } n = \min\{k\colon x_{[-k,k]} \neq y_{[-k,k]}\}, \\ 0, \text{ if } x = y. \end{cases} \tag{4.1}$$

We will use this as the default metric for all shift spaces. Note that this definition is equal to the one given on page 3 in [4][5] since $\inf(\varnothing) = \infty$.

**Proposition 4.1** *Let $X$ be a shift space and $\sigma$ the shift operator. Then $(X, \sigma)$ is an invertible dynamical system.*

**Proof** *$X$ is compact:* Let $x^{(n)}$ be a sequence in $X$ and let $a_j$ be arbitrary but fixed members of $\mathcal{A}$. Since $\mathcal{A}$ is finite it is possible to choose an $a_0$ such that

$$S_0 = \{i : x_0^{(i)} = a_0\}, \text{ is infinite.}$$

For the same reason it is possible to choose $a_{-k}, \ldots, a_k$ such that

$$S_k = \{i : x_{[-k,k]}^{(i)} = a_{-k}a_{-k+1}\ldots a_{-1}a_0a_1\ldots a_{k-1}a_k\}, \text{ is infinite.}$$

A sequence of sets such that $S_0 \supset S_1 \supset S_2 \supset \ldots$ is now obtained. Let

(i) $x_{[-k,k]} = x_{[-k,k]}^{(n)}$, for all $n \in S_k$,

(ii) $n_0 \in S_0$, and

(iii) $n_k \in S_k$ be the smallest element greater than $n_{k-1}$.

Then, $n_0 < n_1 < \ldots < n_k < \ldots$ and

$$x = \lim_{t\to\infty} x^{(n_t)} \in X.$$

Thus, every sequence in $X$ has a convergent subsequence and $X$ is compact.

[4] A function $\phi$ is a homeomorphism if it is continuous, one-to-one, onto and has a continuous inverse.

[5] $d(x,y) = 2^{-n},\ n = \inf\{|k|\colon x_k \neq y_k\}$

Moreover, $\sigma$ *is continuous*[6]: Let $x \in X$ and let $\epsilon > 0$. Choose $n$ such that $2^{-n} < \epsilon$ and let $\delta = 2^{-(n+1)}$. Then

$$d(\sigma(x), \sigma(y)) < 2^{-n} < \epsilon,$$

for all $y \in X$ such that $d(x, y) < \delta$. This is also true for $\sigma^{-1}$. Thus $(X, \sigma)$ is an invertible dynamical system. □

Note that all sliding block codes are continuous, since if two points are close to each other they agree on a large block centered around the 0th coordinate. This means that their images under a sliding block code will agree on a large (but smaller) central block.

If $X$ is a shift space and $\sigma$ is the shift operator, we call $(X, \sigma)$ shift dynamical system.

## 4.2 Invariants

In the same way that irreducibility and mixing divided shift spaces into two different classes each (irreducible and reducible, mixing and not mixing), there are invariants that divide dynamical system into similar classes. Here are three different invariants introduced: Topological transitive, topologically mixing and chaotic. The Curtis-Lyndon-Hedlund Theorem is also mentioned, which shows a link between sliding block codes and homomorphisms.

**Definition** Let $(M, \phi)$ and $(N, \psi)$ be dynamical systems. A homomorphism $\theta\colon (M, \phi) \to (N, \psi)$ is a continuous function $\theta\colon M \to N$ satisfying the commuting property that $\psi \circ \theta = \theta \circ \phi$.

**Remark** Let $\phi$ be a sliding block code between two shift dynamical systems, then $\phi$ is a homomorphism. In fact according to the Curtis-Lyndon-Hedlund Theorem[7] all homomorphisms between shift dynamical systems are sliding block codes.

The following two definitions introduce the "dynamical system"-notion of conjugacy and conjugacy invariant.

**Definition** Let $(M, \phi)$ and $(N, \psi)$ be dynamical systems and let $\theta\colon (M, \phi) \to (N, \psi)$ be a homomorphism. Then $\theta$ is called a topological conjugacy, denoted $\theta\colon (M, \phi) \cong (N, \psi)$, if it is one-to-one and onto. Two dynamical systems are topologically conjugate if there is a topological conjugacy between them.

**Remark** Two shift dynamical systems that are conjugate are topological conjugate and vice versa. This is a consequence of the Curtis-Lyndon-Hedlund Theorem.

**Definition** A conjugacy invariant or invariant, assignes values to dynamical systems in such a manner that topologically conjugate systems will obtain the same value.

[6] This proof is almost a direct copy of the one found on page 41 in [5]

[7] The complete theorem and proof can be found on page 186 in [1].

Three examples of conjugacy invariants are topologically transitive, topologically mixing and chaotic. Only two of these will be of interest for this report of reasons soon presented.

**Definition** Let $(M, \phi)$ be a dynamical system. Then $(M, \phi)$ is

1. topologically transitive if, for all ordered pair $U, V$ of non-empty, open sets in $M$ there exists an $n > 0$ such that $\phi^n(U) \cap V \neq \varnothing$.
2. topologically mixing if, for all ordered pair $U, V$ of non-empty, open sets in $M$ there exists an $n_0$ such that $\phi^n(U) \cap V \neq \varnothing$ for all $n \geqslant n_0$.
3. chaotic if it is topologically transitive and the periodic points of $\phi$ are dense in $M$.

The reason of the name "mixing" is that any two non-empty, open sets are eventually mixed up by $\phi$.

**Proposition 4.2** *Topological transitivity, topological mixing and chaos are invariants of topological conjugacy.*

**Proof** Let $(M, \phi)$ and $(N, \psi)$ be dynamical systems and let $\theta\colon (M, \phi) \cong (N, \psi)$. Then $\theta$ establishes a one-to-one correspondence between open sets of $M$ with those of $N$. Since $\theta$ is a homomorphism it also establishes a one-to-one correspondence between the periodic points of $\phi$ with those of $\psi$. Thus, topological transitivity, topological mixing and chaos are invariants of topological conjugacy. □

As a consequence of the definition, any two non-empty, open sets in a topologically transitive system will intersect each other infinitely many times.

**Remark** Let $(M, \phi)$ be topologically transitive and let $U, V \subset M$ be an ordered pair of non-empty, open sets. Since $\phi$ is continuous, onto and one-to-one, $\phi(U)$ is also open. Thus, there is an infinite number of $n_i \in \mathbb{N}$ such that $\phi^{n_i}(U) \cap V \neq \varnothing$.

## 4.3 Equivalences

There is a strong connection between the invariants of shift spaces introduced earlier and the ones introduced in the preceding section. This fact is stated in [4] but not proved. There is also another claim made in [4], that in at least one case does not seem to be true to me.

First of all, it is necessary to introduce the notion of cylinder sets.

**Definition** Let $X$ be a shift space, let $u \in \mathcal{B}(X)$ and $k \in \mathbb{Z}$. The cylinder set $C_k(u)$ is defined as

$$C_k(u) = \{x \in X\colon x_{[k,k+|u|-1]} = u\}.$$

In the following lemma and later on in section 5, $B(x; r)$ [$\bar{B}(x; r)$] will denote the open [closed] ball of radius $r$ centered around $x$.

**Lemma 4.3** *Cylinder sets are open.*

**Proof** First, observe that for any $x \in X$

$$C_{-n}(x_{[-n,n]}) = B(x; 2^{-n}).$$

Let $z \in C_k(u)$ and $n = \max\{|k|, |k + |u| - 1|\}$. Then

$$B(z; 2^{-n}) \subset C_k(u).$$

□

The following proposition shows that irreducibility [mixing] of a shift of finite type is equivalent to transitivity [mixing] of the corresponding shift dynamical system.

**Proposition 4.4** *Let $X \cong \mathsf{X}_G$ be a shift of finite type and let $(X, \sigma)$ be a shift dynamical system. Then the following three statements are equivalent:*

1. *$\mathsf{A}_G$ is irreducible,*
2. *$(X, \sigma)$ is topologically transitive,*
3. *$(X, \sigma)$ is chaotic.*

*The following two statements are also equivalent:*

4. *$\mathsf{A}_G$ is primitive,*
5. *$(X, \sigma)$ is topologically mixing.*

**Proof**

$1 \Rightarrow 2$ First note that irreducibility of $\mathsf{A}_G$ is equivalent to irreducibility of $X \cong \mathsf{X}_G$. Let $U, V \subset X$ be non-empty, open sets. Then there exists $k, l \in \mathbb{Z}$ such that

$$C_k(u) = \sigma^{-k}(C_0(u)) \subset U, u \in \mathcal{B}(X),$$

$$C_l(v) = \sigma^{-l}(C_0(v)) \subset V, v \in \mathcal{B}(X).$$

Since $X$ is irreducible it is possible to find a sequence $w^{(i)} \in \mathcal{B}(X)$ such that $uw^{(i)}v \in \mathcal{B}(X)$ and $|w^{(i)}| \to \infty$ as $i \to \infty$. Define $n_i = |uw^{(i)}| + k - l$. Since

$$z^{(i)} = \ldots uw^{(i)}v \ldots \in C_k(u), z_k^{(i)} = u_1,$$

it is easy to see that

$$\sigma^{n_i}(U) \cap V \neq \varnothing$$

where $n_i > 0$ for sufficiently large $i$. Thus $(X, \sigma)$ is topologically transitive.

$1 \Leftarrow 2$ Since $(X, \sigma)$ is topologically transitive, given an ordered pair $u, v \in \mathcal{B}(X)$ there exists $n \in \mathbb{N} > |u|$ such that[8]

$$\sigma^n(C_0(u)) \cap C_0(v) \neq \varnothing$$

$$\Leftrightarrow$$

$$C_0(u) \cap \sigma^{-n}(C_0(v)) \neq \varnothing.$$

For $z \in C_0(u) \cap \sigma^{-n}(C_0(v))$, we have

$$z_{[0,n+|v|-1]} = uz_{[|u|,n-1]}v = uwv \in \mathcal{B}(X).$$

Thus, $X$ is irreducible.

$2 \Leftrightarrow 3$ This follows from the definition of chaotic and the fact that the periodic points of $(X, \sigma)$ are dense in $X$[9].

$4 \Leftrightarrow 5$ The proof of this is very similar to the proof of $1 \Leftrightarrow 2$ and is therefore presented in the appendix.

A statement is made in [4] about a relationship between an adjacency matrix and its characteristic polynomial. In at least one case, it seems to me that this statement is not true.

**Remark** Let

(i) $X$ be an $M$-step shift of finite type,

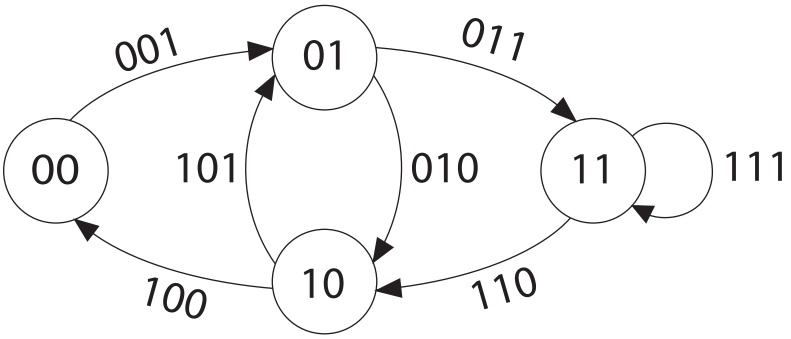
(ii) $G$ be a graph such that $X \cong X^{[M+1]} = \mathsf{X}_G$ and

(iii) $A = \mathsf{A}_G$.

The following are also claimed to be equivalent in [4]:

1. The characteristic polynomial $\chi_A(\lambda)$ is irreducible,
2. $A$ is irreducible.

**Counter example** Let $X = \mathsf{X}_\mathcal{F}$, where $\mathcal{F} = \{000\}$ and let $G$ be the graph below.



[8] This fact is noted in the Remark on page 18.

[9] For every $x \in X$ there exist a periodic point $y = (x_{[-n,n]})^\infty \in C_{-n}(x_{[-n,n]})$ that is arbitrarily close to $x$.

Then $X \cong X^{[3]} = \mathsf{X}_G$ and

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

$A$ is irreducible but $\chi_A(\lambda) = \lambda^4 - \lambda^3 - \lambda^2 - \lambda$ is not.

Even though (topological) transitivity and mixing are conjugacy invariants, we might suspect that they are not the best ones available. Especially since they only have two values each: “Yes” and “No”. In the following section, another more accurate conjugacy invariant will be defined: topological entropy.

# 5 Topological Entropy

Topological entropy was introduced in 1965 by Adler, Konheim and McAndrew as an invariant of topological conjugacy. The definition presented here is another but equivalent definition (due to Dinaburg and Bowen). The advantage of the Dinaburg and Bowen definition, as will be seen in section 6, is that it leads to proofs of the results connecting topological entropy with probabilistic. After defining topological entropy for metric spaces in general, this section is concluded with two paragraphs discussing topological entropy of shift dynamical systems and its properties.

## 5.1 Definition

In this section $(X, d)$ will be a metric space and $T$ a fixed, uniformly continuous function on $(X, d)$. We will define a new metric $d_n$ on $X$ by $d_n(x, y) = \max_{0 \leqslant i \leqslant n-1} d(T^i(x), T^i(y))$. Using the metric $d_n$, the ball of radius $r$ centered around $x$ equals

$$\bigcap_{i=0}^{n-1} T^{-i} B(T^i x; r).$$

**Definition** Let $K$ be a compact subset of $X$, $n \in \mathbb{N}$ and $\varepsilon > 0$. A subset $F \subset X$ $(n, \varepsilon)$-spans $K$ with respect to $T$ if for all $x \in K$ there exists $y \in F$ such that $d_n(x, y) \leqslant \varepsilon$. This condition is equivalent to

$$K \subset \bigcup_{y \in F} \bigcap_{i=0}^{n-1} T^{-i} \bar{B}(T^i y; \varepsilon).$$

**Definition** If $K$ is a compact subset of $X$, $n \in \mathbb{N}$ and $\varepsilon > 0$ let $r_n(\varepsilon, K)$ denote the smallest cardinality of any $(n, \varepsilon)$-spanning set for $K$ with respect to $T$. If we wish to emphasise the dependance on $T$ we will write $r_n(\varepsilon, K, T)$.

**Remark**

1. $r_n(\varepsilon, K) < \infty$ for all $n \in \mathbb{N}$ and $\varepsilon > 0$. According to the Heine-Borel Theorem, since $K$ is compact the covering of $K$ by open sets $\bigcap_{i=0}^{n-1} T^{-i} B(T^i x; r), x \in X$ has a finite subcover.
2. If $\varepsilon_1 < \varepsilon_2$ then $r_n(\varepsilon_1, K) \geqslant r_n(\varepsilon_2, K)$.

**Definition** If $K \subset X$ is compact and $\varepsilon > 0$ let

$$r(\varepsilon, K, T) = \limsup_{n \to \infty} \frac{1}{n} \log r_n(\varepsilon, K).$$

We will write $r(\varepsilon, K, T, d)$ whenever we wish to emphasise the dependance on $d$.

**Remark** $r(\varepsilon, K, T)$ is nondecreasing as $\varepsilon \to 0$. This follows from the previous remark.

**Definition** If $K$ is a compact subset of $X$ let $h(T; K) = \lim_{\varepsilon \to 0} r(\varepsilon, K, T)$. The topological entropy of $T$ is

$$h(T) = \sup_{K} h(T; K),$$

where the supremum is taken over the collection of all compact subsets $K \subset X$. If we wish to emphasise the dependance on the metric $d$ we will write $h_d(T)$.

Topological entropy is not dependent on the metric chosen as long as the metrics are uniformly equivalent.

**Definition** Two metrics $d$ and $d'$ are [uniformly] equivalent if

$$id.\colon (X, d) \to (X, d') \text{ and } id.\colon (X, d') \to (X, d)$$

are both [uniformly] continuous.

**Theorem 5.1** *If $d$ and $d'$ are uniformly equivalent then $h_d(T) = h_{d'}(T)$*

**Proof** Let $\varepsilon_1 > 0$ and choose $\varepsilon_2 > 0$ such that

$$d'(x, y) < \varepsilon_2 \Rightarrow d(x, y) < \varepsilon_1.$$

Furthermore, choose $\varepsilon_3 > 0$ such that

$$d(x, y) < \varepsilon_3 \Rightarrow d'(x, y) < \varepsilon_2.$$

Let $K$ be compact and assume that $F$ $(n, \varepsilon_2)$-spans $K$ in the metric $d'$. Then for all $x \in K$ there exists $y \in F$ such that

$$\max_i d'(T^i x, T^i y) \leqslant \varepsilon_2 \Rightarrow \max_i d(T^i x, T^i y) \leqslant \varepsilon_1.$$

Thus $F$ $(n, \varepsilon_1)$-spans $K$ in the metric $d$ and

$$r_n(\varepsilon_1, K, d) \leqslant r_n(\varepsilon_2, K, d').$$

Arguing in a similar fashion gives

$$r_n(\varepsilon_2, K, d') \leqslant r_n(\varepsilon_3, K, d).$$

Hence $r(\varepsilon_1, K, T, d) \leqslant r(\varepsilon_2, K, T, d') \leqslant r(\varepsilon_3, K, T, d)$. If $\varepsilon_1 \to 0$, then since the metrics are uniformly equivalent, $d'(x, y) \to 0$ and we can let $\varepsilon_2 \to 0$. For the same reason, when $\varepsilon_2 \to 0$ it is possible to let $\varepsilon_3 \to 0$ and

$$h_d(T; K) = h_{d'}(T; K).$$

□

**Remark** If $X$ is compact and the two metrics $d$ and $d'$ are equivalent, then they are uniformly equivalent. Therefore, given a compact metric space $X$ the entropy of $T$ is independent of the metric chosen (as long as they generate the same topology[10]).

[10] Two equivalent metrics generates the same topology.

The following theorem allows a simplification of the definition of $h(T)$ when $X$ is compact.

**Theorem 5.2** *If $K \subset K_1 \cup \ldots \cup K_m$ are all compact subsets of $X$ then*

$$h(T;K) \leqslant \max_{1 \leqslant i \leqslant m} h(T;K_i)$$

**Proof** To begin with, $r_n(\varepsilon, K) \leqslant r_n(\varepsilon, K_1) + \ldots + r_n(\varepsilon, K_m)$. Let $\varepsilon > 0$ and for all $n \in \mathbb{N}$ choose $K_{i(n,\varepsilon)}$ such that

$$r_n(\varepsilon, K_{i(n,\varepsilon)}) = \max_j r_n(\varepsilon, K_j).$$

Then $r_n(\varepsilon, K) \leqslant m \cdot r_n(\varepsilon, K_{i(n,\varepsilon)})$ and

$$\log r_n(\varepsilon, K) \leqslant \log m + \log r_n(\varepsilon, K_{i(n,\varepsilon)}).$$

Let $n_j$ be a subsequence to the natural numbers such that

$$\lim_{j\to\infty} \frac{1}{n_j} \log r_{n_j}(\varepsilon, K) = \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon, K). \tag{5.1}$$

Then there exists a subsequence $m_j$ to $n_j$ such that $K_{i(m_j,\varepsilon)}$ is independent of $j$ (i.e. $K_{i(m_j,\varepsilon)} = K_{i(\varepsilon)}$, $\forall j$). This is possible since there are only a finite number of $K_i$'s. Of course (6.1) holds with $m_j$ substituted for $n_j$. Thus

$$r(\varepsilon, K, T) \leqslant r(\varepsilon, K_{i(\varepsilon)}, T).$$

Choose $\varepsilon_q \to 0$ such that $K_{i(\varepsilon_q)}$ is constant ($K_{i(\varepsilon_q)} = K_{i(c)}$), then

$$h(T;K) \leqslant h(T;K_{i(c)}) \leqslant \max_j h(T;K_j)$$

□

**Corollary 5.3** *If $(X,d)$ is a compact metric space then $h_d(T) = h_d(T;X)$.*

**Proof** If $K \subset X$ is compact then $h_d(T;K) \leqslant h_d(T;X)$. □

Thus

$$h(T) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon, X).$$

## 5.2 Topological Entropy of Shift Spaces

As it turns out, the notion of topological entropy of a shift dynamical system is much less complicated.

In this section $X$ will be a shift space, $\sigma$ the shift operator and $(X,\sigma)$ a shift dynamical system. The metric used will be the one defined in (4.1).

First of all we need the following lemma.

**Lemma 5.4** *If $X$ is a shift space, then*

$$\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{B}_n(X)|$$

*exists, and equals*

$$\inf_{n \geqslant 1} \frac{1}{n} \log |\mathcal{B}_n(X)|$$

**Proof** [11] Let $m, n \geqslant 1$ then

$$|\mathcal{B}_n(X)| \leqslant |\mathcal{B}_{n+m}(X)| \leqslant |\mathcal{B}_n(X)| \cdot |\mathcal{B}_m(X)|$$

If $b_n = \log |\mathcal{B}_n(X)|$ then,

$$b_n \leqslant b_{n+m} \leqslant b_n + b_m.$$

Let $\varepsilon > 0$, $\inf_{n \geqslant 1} \frac{b_n}{n} = a$ and choose $q$ such that $\frac{b_q}{q} < a + \varepsilon$. For $n > q$ choose $k \in \mathbb{N}$ such that

$$(k-1)q < n \leqslant kq.$$

Then

$$\frac{b_n}{n} < \frac{b_{kq}}{(k-1)q} \leqslant \frac{k}{(k-1)} \frac{b_q}{q} < \frac{k}{(k-1)}(a+\varepsilon)$$

For large enough $n$ (and hence large $k$)

$$a \leqslant \frac{b_n}{n} < \frac{k}{(k-1)}(a+\varepsilon) < a + 2\varepsilon$$

$$\Longrightarrow \lim_{n\to\infty} \frac{1}{n} \log |\mathcal{B}_n(X)| = a$$

□

Since $X$ is compact and $\sigma$ is continuous, $\sigma$ is uniformly continuous. Thus the definition of topological entropy given in the former paragraph applies to $(X, \sigma)$.

**Proposition 5.5** *Let $(X, \sigma)$ be a shift dynamical system. Then the topological entropy*

$$h(\sigma) = \lim_{n\to\infty} \frac{1}{n} \log |\mathcal{B}_n(X)|.$$

**Proof** Let $\varepsilon = 2^{-(k+1)}$ and let $F \subset X$ $(n, \varepsilon)$-span $X$. Then for all $x \in X$ there exists $y \in F$ such that $\max_{0 \leqslant i \leqslant n-1} d(\sigma^i(x), \sigma^i(y)) \leqslant \varepsilon$. This is equivalent to:

$$x_{[-k,k]} = y_{[-k,k]}$$

$$x_{[-k+1,k+1]} = y_{[-k+1,k+1]}$$

$$\vdots$$

$$x_{[-k+n-1,k+n-1]} = y_{[-k+n-1,k+n-1]}.$$

This in turn is equivalent to $x_{[-k,k+n-1]} = y_{[-k,k+n-1]}$. Thus $F$ must have at least $|\mathcal{B}_{n+2k}(X)|$ number of elements and

$$r_n(2^{-(k+1)}, X, \sigma) = |\mathcal{B}_{n+2k}(X)|.$$

Using the result from the previous lemma we get:

$$h(\sigma) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon, X, \sigma) = \lim_{k\to\infty} \lim_{n\to\infty} \frac{1}{n} \log |\mathcal{B}_{n+2k}(X)| =$$

$$= \lim_{k\to\infty} \lim_{n\to\infty} \frac{n+2k}{n} \frac{1}{n+2k} \log |\mathcal{B}_{n+2k}(X)| = \lim_{m\to\infty} \frac{1}{m} \log |\mathcal{B}_m(X)|.$$

□

[11] This proof is almost a direct copy of the one found on pages 48-49 in [3].

From now on the topological entropy of a shift dynamical system $(X, \sigma)$ is denoted by $h(X)$ and since it solely (explicitly) depends on the number of blocks in $X$, this will sometimes be referred to as the entropy of a shift space.

Finding an upper limit for $h(X)$ is easy.

**Remark** If $X$ is a shift space over the alphabet $\mathcal{A}$ then $|\mathcal{B}_n(X)| \leqslant |\mathcal{A}|^n$ and

$$\frac{1}{n} \log |\mathcal{B}_n(X)| \leqslant \log |\mathcal{A}|.$$

Thus $h(X) \leqslant \log |\mathcal{A}|$. Also, if $X \neq \varnothing$, $h(X) \geqslant 0$. For $X = \varnothing$ we set $h(X) = -\infty$.

As stated earlier, topological entropy is a conjugacy invariant and despite that it is difficult utilizing the definition to calculate it, there is an easy way to compute it.

**Proposition 5.6** *If $Y$ is a factor of $X$, then $h(Y) \leqslant h(X)$.*

**Proof** Let $\phi = \Phi_\infty^{[-m,k]}$ be a sliding block code from $X$ onto $Y$. As stated before, every block in $\mathcal{B}_n(Y)$ is the image of a block in $\mathcal{B}_{n+m+k}(X)$. Hence $|\mathcal{B}_n(Y)| \leqslant |\mathcal{B}_{n+m+k}(X)|$ and

$$h(Y) = \lim_{n\to\infty} \frac{1}{n} \log |\mathcal{B}_n(Y)| \leqslant \lim_{n\to\infty} \frac{1}{n} |\log \mathcal{B}_{n+m+k}(X)| =$$

$$= \lim_{n\to\infty} \frac{n+m+k}{n} \frac{1}{n+m+k} \log |\mathcal{B}_{n+m+k}(X)| = h(X).$$

□

**Corollary 5.7** *Topological entropy of shift spaces is an invariant of topological conjugacy*

**Proof** Let $X$ and $Y$ be shift spaces and $X \cong Y$. Then $Y$ is a factor of $X$ and $X$ is a factor of $Y$. □

Actually, this result is not only true for shift spaces. A more general theorem and proof can be found on page 167 in [2].

## 5.3 Calculating Topological Entropy

Even if the appearance of entropy of a shift space is much less involved than for a general compact space, the definition do not give much assistance when computing this invariant. In this paragraph a simple method for computing the topological entropy of a shift of finite type is developed. This result will also help proving the relationship between topological and probabilistic entropy later in the report.

There is a fundamental connection between the adjacency matrix of a graph and the number of allowed blocks in the corresponding edge shift. As the following proposition and corollary shows, this fact enables an easy computation of the entropy of an edge shift.

**Proposition 5.8** *Let $A \neq 0$ be a non-negative matrix having an eigenvector $v > 0$ and a corresponding eigenvalue $\lambda > 0$.*[12] *Then there are constants $c_0,\ d_0 > 0$ such that*

$$c_0\lambda^n \leqslant \sum_{i,j}(A^n)_{ij} \leqslant d_0\lambda^n$$

**Proof** $A^n v = \lambda^n v$, for $n \geqslant 1$. Hence, for every non-negative integer $i$

$$\sum_j (A^n)_{ij} v_j = \lambda^n v_i.$$

Let

$$c = \min_i\{v_i\} \text{ and } d = \max_i\{v_i\}$$

Then

$$c\sum_j (A^n)_{ij} \leqslant \sum_j (A^n)_{ij} v_j = \lambda^n v_i \leqslant d\lambda^n. \tag{5.2}$$

To estimate $\sum_{i,j}(A^n)_{ij}$ from above, divide (5.2) by $c$ and sum over $i$,

$$\sum_{i,j}(A^n)_{ij} \leqslant \sum_i \frac{d}{c}\lambda^n = d_0\lambda^n,\ d_0 > 0.$$

To estimate $\sum_{i,j}(A^n)_{ij}$ from below, note that

$$c\lambda^n \leqslant \lambda^n v_i = \sum_j (A^n)_{ij} v_j \leqslant d\sum_j (A^n)_{ij} \leqslant d\sum_{i,j}(A^n)_{ij}$$

$$\Longrightarrow$$

$$c_0\lambda^n \leqslant \sum_{i,j}(A^n)_{ij},\ c_0 = \frac{c}{d} > 0.$$

Thus

$$c_0\lambda^n \leqslant \sum_{i,j}(A^n)_{ij} \leqslant d_0\lambda^n \tag{5.3}$$

□

**Corollary 5.9** *Let $A$ and $\lambda$ be the same as in Proposition 5.8. If $G$ is a graph with adjacency matrix $A$, then $h(\mathsf{X}_G) = \log\lambda$.*

**Proof** Note that the central term of (5.3)

$$\sum_{i,j}(A^n)_{ij} = |\mathcal{B}_n(\mathsf{X}_G)|.$$

Applying the result from the previous theorem we get

$$\frac{1}{n}\log c_0\lambda^n \leqslant \frac{1}{n}\log|\mathcal{B}_n(X)| \leqslant \frac{1}{n}\log d_0\lambda^n.$$

Letting $n \to \infty$ completes the proof. □

[12] Actually if $A \neq 0$ is non-negative matrix having an eigenvector $v > 0$ the corresponding eigenvalue will be positive.

Step by step, the restraints on what classes of shifts the above corollary applies to, will loosen. Next, the parts of the Perron-Frobenius Theorem important to this paper is presented. The proof is omitted but the interested reader can find the complete theorem and proof in the beginning of [6]. Thanks to this theorem, the result from Corollary 5.9 will be shown for all irreducible shifts of finite type.

**Theorem 5.10 (Perron-Frobenius Theorem)** *Let $A \neq 0$ be an irreducible matrix. Then there is a positive eigenvalue $\lambda_A$ such that if $\mu$ is a another eigenvalue of $A$, then $|\mu| \leqslant \lambda_A$. Corresponding to $\lambda_A$ is a left eigenvector $u_A > 0$ and a right eigenvector $v_A > 0$.*

For an irreducible matrix $A$, we call $\lambda_A$ the Perron eigenvalue of $A$.

**Corollary 5.11** *If $G$ is an irreducible graph with adjacency matrix $A$, then $h(\mathsf{X}_G) = \log \lambda_{\mathsf{A}}$.*

**Proof** Irreducibility of $G$ implies irreducibility of $A$. The Perron-Frobenius Theorem tells us that $A$ has a positive eigenvector with corresponding eigenvalue $\lambda_A > 0$. The result follows from Corollary 5.9. □

**Corollary 5.12** *If $X$ is an irreducible $M$-step shift of finite type and $G$ is the essential graph for which $X^{[M+1]} = \mathsf{X}_G$, then*

$$h(X) = \log \lambda_{\mathsf{A}_G}.$$

**Proof** Since topological entropy is a conjugacy invariant and $X \cong X^{[M+1]}$

$$h(X) = h(X^{[M+1]}) = h(\mathsf{X}_G).$$

Since irreducibility is also an invariant, irreducibility of $X$ implies irreducibility of $\mathsf{X}_G$ which in turn implies irreducibility of $G$ (Corollary 2.3). The result follows from the previous corollary. □

Computing the entropy of irreducible shifts thus breaks down to the computing of the eigenvalues of $A$. But how about the rest of the shifts, the reducible ones? As will be shown, by breaking down reducible graphs into irreducible subgraphs, a similar result applies to the reducible shifts.

The following passage states some important properties of graphs and matrices with irreducible components. I have chosen to omit the proofs since it would slow down the report considerably.

**Definition** Let $A$ be a square matrix and $n, m \in \mathbb{N}$. If we permute row $n$ and $m$ and column $n$ and $m$ we call it a simultaneous permutations of rows and columns.

A result from the theory of Markov chains is that every essential and reducible matrix $A$ - by simultaneous permutations of rows and columns - can assume a block triangular form[13]

$$A = \begin{bmatrix} A_1 & 0 & 0 & \dots & 0 \\ * & A_2 & 0 & \dots & 0 \\ * & * & A_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ * & * & * & \dots & A_n \end{bmatrix}$$

[13] ct. [8] or Example 4.4.1. in [1] for an idea how to do this.

where the $*$'s represent possibly nonzero matrices and the $A_i$'s are irreducible matrices. Let $G_i = G_{A_i}$. The $A_i$'s are called the irreducible components of $A$ and the $G_i$'s are called the irreducible components of $G$.

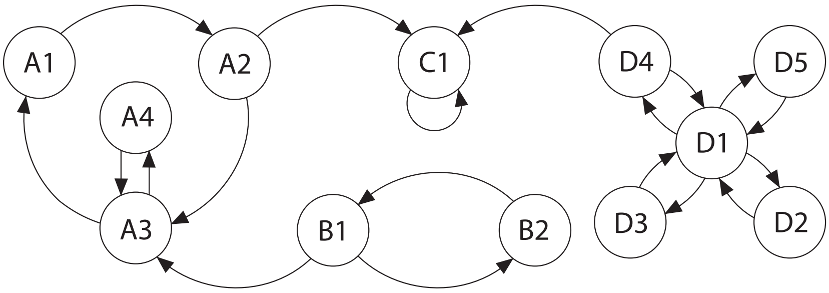
Note that it is only possible to "walk" from a vertex in $G_a$ to a vertex in $G_b$ when $a \geqslant b$.

**Remark** Since the characteristic polynomial $\chi_A(t)$ of $A$ is unchanged by simultaneous permutations of rows and columns it equals

$$\chi_A(t) = \chi_{A_1}(t)\chi_{A_2}(t)\dots\chi_{A_n}(t).$$

Thus the eigenvalues of $A$ are the eigenvalues of the $A_i$'s.

**Example figure**



Reducible graph with irreducible subgraphs.

The Perron eigenvalue of a *reducible* matrix is the largest Perron eigenvalue of its irreducible submatrices.

**Definition** Let $A$ be a non-negative matrix with irreducible components $A_1, \dots, A_n$. The Perron eigenvalue $\lambda_A$ of $A$ is

$$\lambda_A = \max_{1\leqslant i\leqslant n} \lambda_{A_i}.$$

**Corollary 5.13** *For an arbitrary non-negative, essential matrix A, its Perron eigenvalue $\lambda_A$ is the largest eigenvalue of A.*

**Proof** The proof follows from the Perron-Frobenius Theorem, the previous remark and the definition. □

This section is concluded with the final theorem which shows that all shifts of finite type have an entropy equal to $\log \lambda_A$.

**Theorem 5.14** *Let G be an essential graph with adjacency matrix A, then*

$$h(X_G) = \log \lambda_A$$

**Proof** Let $A$ have irreducible components $A_1, \dots, A_p$ and let

$$\lambda_A = \lambda_{A_q} = \max_{1 \leqslant i \leqslant p} \lambda_{A_i}.$$

If $\lambda_A = 0$, then each $A_i = [0]$. Thus $G_A$ has no bi-infinite walks and $\mathsf{X}_{G_A} = \varnothing$, which gives the result. Let $\lambda_A > 0$. Since $|\mathcal{B}_n(\mathsf{X}_G)| \geqslant |\mathcal{B}_n(\mathsf{X}_{G_q})|$,

$$h(\mathsf{X}_G) \geqslant h(\mathsf{X}_{G_q}) = \log \lambda_{A_q} = \log \lambda_A.$$

To prove that $h(\mathsf{X}_G) \leqslant \log \lambda_A$ we will estimate the number of paths of lenght $n$ in $G$. Such a path breaks down into subpaths on the irreducible components $G_i$ of $G$ and transitional edges between the $G_i$'s. Thus any $\pi \in \mathcal{B}_n(\mathsf{X}_G)$ has the form

$$\pi = \pi_1 e_1 \pi_2 e_2 \dots \pi_{j-1} e_{j-1} \pi_j, \tag{5.4}$$

where $\pi_i$ is a path in $G_{q(i)}$ of length $n_i$ and $e_i$ is a transitional edge from a vector in $G_{q(i)}$ to one in $G_{q(i+1)}$. Thus

$$q(1) > q(2) > \dots > q(j) \text{ where } j \leqslant p,$$

the number of irreducible submatrices of $A$. Let $T$ be the total number of transitional edges between subgraphs of $G$. Then there are at most $T$ possible choices for each $e_i$ in (5.4), and at most $n$ places where each could occur. Hence the number of arrangements of transitional edges in $\pi$ is bounded above by $(Tn)^p$. Also, the number of ways to choose $\pi_i$ is bounded above by $|\mathcal{B}_{n_i}(\mathsf{X}_{G_{q(i)}})|$. By the Perron-Frobenius Theorem and Proposition 5.8 there exists a $d > 0$, such that for every $G_{q(i)}$

$$|\mathcal{B}_{n_i}(\mathsf{X}_{G_{q(i)}})| \leqslant d\lambda_{A_{q(i)}}^{n_i} \leqslant d\lambda_A^{n_i}.$$

Hence the total number of ways to choose the $\pi_i$'s is bounded above by

$$\prod_{i=1}^{j} |\mathcal{B}_{n_i}(\mathsf{X}_{G_{q(i)}})| \leqslant d^j \lambda_A^{n_1 + \dots + n_j} \leqslant d^p \lambda_A^n.$$

Since we can choose the $e_i$'s in $\pi$ in less than $(Tn)^p$ different ways and the $\pi_i$'s in less than $d^p\lambda_A^n$ different ways, $\pi$ can be chosen in less than $(Tn)^p d^p \lambda_A^n$ different ways and

$$|\mathcal{B}_n(\mathsf{X}_G)| \leqslant (Tn)^p d^p \lambda_A^n.$$

Hence

$$h(\mathsf{X}_G) = \lim_{n\to\infty} \frac{1}{n} \log |\mathcal{B}_n(\mathsf{X}_G)|$$
$$\leqslant \lim_{n\to\infty} \left( \frac{1}{n} \log (Td)^p + \frac{p}{n} \log n + \log \lambda_A \right) = \log \lambda_A.$$

Which gives the result. □

This paragraph can be summed up in the following way: Let $X$ be a $M$-step shift of finite type. Since $X \cong X^{[M+1]} = \mathsf{X}_G$ for some essential graph $G$ with adjacency matrix $A$ and entropy is a conjugacy invariant:

$$h(X) = \log \lambda_A.$$

Following example shows that entropy somehow is a better conjugacy invariant than irreducibility and mixing.

**Example** Let $X$ be the golden mean shift and recall that $X \cong \mathsf{X}_G$, where $G$ is the graph (2.2) with adjacency matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then $\chi_A(\lambda) = \lambda^2 - \lambda - 1$, $\lambda_A = (1+\sqrt{5})/2$ and

$$h(X) = \log \frac{1+\sqrt{5}}{2}.$$

The origin of the name "the golden mean shift" is of course the fact that $\lambda_A$ equals "the golden mean". But the golden mean shift is far from the only shift with this property.

Let $Y$ be the shift described by the graph (2.3), then

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

which in turn gives $\chi_A(\lambda) = \lambda(\lambda^2 - \lambda - 1)$ and thus

$$h(Y) = \log \frac{1+\sqrt{5}}{2}.$$

It turns out that this shift too has the entropy $\log(1+\sqrt{5})/2$ and even though one could speculate about whether or not these shifts are merely two different representations of the same underlying object, we can not answer this question by looking at the entropy alone. However, by changing the name of $e$ to 00, $f$ to 01 and $g$ to 10 it is easy to see that these two shifts actually are conjugate.

The entropy of the full 2-shift (here denoted $Z$) is easiest calculated by using the definition. Since $|\mathcal{B}_n(Z)| = 2^n$ we see that

$$h(Z) = \log 2$$

and thus it is not conjugate to the shift above.

# 6 Probabilistic Entropy

In 1948 Shannon introduced the information entropy, his work was later on continued by McMillan and Feinstein. In 1953 Khinchin refined their work and the following section is based on his two papers gathered in [3]. The relationship with information theory will not be the focus, instead the focus will be on the relationship with topological entropy.

## 6.1 Definition and Basic Properties

Let $A$ be a finite probability space consisting of the elementary events $A_1, A_2, \ldots, A_n$ with probabilities $p_1, p_2, \ldots, p_n$ $(p_i \geqslant 0, \sum_i p_i = 1)$. We write

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}.$$

Sometimes we will refer to $A$ as a finite scheme or a system.

**Example** Let $A$ and $B$ be two finite schemes where

$$A = \begin{pmatrix} A_1 & A_2 \\ 0.99 & 0.01 \end{pmatrix}, \; B = \begin{pmatrix} B_1 & B_2 \\ 0.5 & 0.5 \end{pmatrix}.$$

Obviously the system $B$ represents more uncertainty than $A$. If we were to predict the outcome of a random trial in system $A$ we would without hesitation pick $A_1$, while if we were to do the same thing with $B$ we would be indifferent in choosing between $B_1$ and $B_2$.

So, what is expected of a measure of uncertainty? First of all, it is natural to expect the measure to be equal to 0 when there is no uncertainty. Second, the uncertainty of a system is expected to assume its largest value when all $n$ events has the same probability $1/n$. Another plausible property is that the measure somehow should increase with the number of events.

With these properties in thought, entropy of a finite probability space is now introduced. Not only does the probabilistic entropy fulfill the attributes mentioned above, but given a set of modest axioms it is the *only* reasonable way of measuring the uncertainty inherent in such a space. Note, since removing uncertainty can be considered equivalent to obtaining information we will also use entropy as a measure of information obtained.

**Definition** Let $A$ be a finite probability space composed of elementary events $A_1, A_2, \ldots, A_n$ with probabilities $p_1, p_2, \ldots, p_n \geqslant 0$; $\sum_{i=1}^{n} p_i = 1$. The entropy of $A$ is defined as

$$H(A) = H(p_1, p_2, \ldots, p_n) = -\sum_{k=1}^{n} p_k \log p_k,$$

where we take $p_i \log p_i = 0$ if $p_i = 0$. The logarithms are taken in an arbitrarily but fixed basis.

**Remark**

1. $H(A) = 0$ iff $p_j = 1$ for some $1 \leqslant j \leqslant n$. This is reasonable since if every "random" trial has the same outcome $A_j$ there is no uncertainty whatsoever what the outcome will be.

2. $H(A)$ assumes its largest value when $p_i = \frac{1}{n}$ for all $1 \leqslant i \leqslant n$. This too is reasonable since we expect the system with equally likely outcomes to be the one with most uncertainty.

**Proof** A result of Jensen's inequality[14] is that for every continuous, convex function $f(x)$ and non-negative numbers $a_i$ we have

$$f\left(\frac{1}{n}\sum_{k=1}^{n} a_k\right) \leqslant \frac{1}{n}\sum_{k=1}^{n} f(a_k).$$

Since $H(1/n, 1/n, \ldots, 1/n) = \log n$, then setting $f(x) = x \log x$, $a_i = p_i$ and bearing in mind that $p_i \geqslant 0; \sum_i p_i = 1$ we get

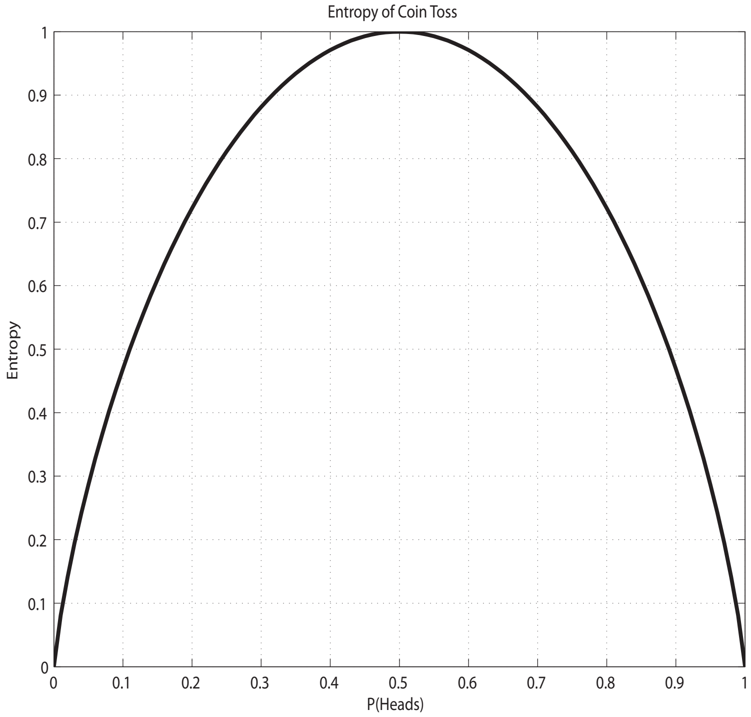
$$-\log n = nf(\frac{1}{n}\sum_k p_k) \leqslant \sum_k f(p_k) = -H(p_1, p_2, \ldots, p_n).$$

Thus

$$H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) = \log n \geqslant H(p_1, p_2, \ldots, p_n).$$

□

**Example** To illustrate the previous two remarks, consider flipping a coin with the probability of heads ranging between 0 and 1.



[14] ct. Theorem 1.1.14. in [7]

Here the entropy of the coin toss is plotted against the probability of heads. When computing the entropy, logarithms with base 2 have been used (for aesthetic reasons only). As expected, the maximum value of the entropy is obtained when the coin is fair. In the degenerated case with only one possible outcome the entropy is 0.

Having looked at the entropy of a single probability space, it is intuitive to proceed with the entropy of product spaces.

**Definition** Let $A$ and $B$ be two finite schemes such that

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_m \\ p_1 & p_2 & \cdots & p_m \end{pmatrix}, \ B|A_i = \begin{pmatrix} B_1 & B_2 & \cdots & B_n \\ q_{1|i} & q_{2|i} & \cdots & q_{n|i} \end{pmatrix}.$$

Where $B|A_i$ is the scheme $B$ when the event $A_i$ of $A$ has occurred and $q_{j|i} = P(B_j|A_i)$. The product of $A$ and $B$ is the system $AB$ composed of the elementary events $A_iB_j$ with probabilities $p_iq_{j|i}$.

For the rest of section 6, $A$ and $B$ will denote two finite schemes.

Obviously the entropy of a product space is closely related to the entropy of the individual spaces that make up the product system. The following remark shows the simple, expected relationship when the two systems are independent.

**Remark** If $A$ and $B$ are mutually independent, $H(AB) = H(A) + H(B)$. This is also something we would expect from a measure of uncertainty since a random trial in $AB$ is equivalent to one in $A$ and one in $B$.

**Proof** A straightforward calculation gives the result:

$$-H(AB) = \sum_{i,j} p_iq_j \log p_iq_j = \sum_{i,j} p_iq_j(\log p_i + \log q_j) =$$

$$= \sum_j q_j \sum_i p_i \log p_i + \sum_i p_i \sum_j q_j \log q_j = -H(A) - H(B).$$

□

**Definition** We will let $H_i(B)$ denote the conditional entropy of $B$ given the event $A_i$ of $A$ occurred. Thus

$$H_i(B) = H(B|A_i) = -\sum_j P(B_j|A_i) \log P(B_j|A_i).$$

Further we let $H_A(B)$ denote the expected value of $H_k(B)$:

$$H_A(B) = E(H_k(B)) = \sum_i P(A_i)H_i(B).$$

We can interpret $H_A(B)$ as the average additional amount of uncertainty removed (or information obtained) from a random trial in $B$ performed after a random trial in $A$.

In the independent case, the entropy of a product system equals the sum of the entropies, the general case is a straightforward extension of this result.

**Remark**

1. $H(AB) = H(A) + H_A(B)$.

**Proof** Let $p_i = P(A_i)$ and $q_{j|i} = P(B_j|A_i)$. Then $P(A_iB_j) = p_iq_{j|i}$ and

$$-H(AB) = \sum_{i,j} p_iq_{j|i}(\log p_i + \log q_{j|i}) = \sum_i p_i \log p_i \cdot \sum_j q_{j|i} +$$

$$+ \sum_i p_i \sum_j q_{j|i} \log q_{j|i} = -H(A) - \sum_i p_i H_i(B).$$

□

2. $H_A(B) \leqslant H(B)$. This is the same as saying that knowledge of the outcome of a random trial in $A$ can only decrease the uncertainty of $B$.

**Proof** For every continuous convex function $f(x)$ and non-negative numbers $a_i$; $\sum_i a_i = 1$ we have

$$\sum_k a_k f(x_k) \geqslant f\left(\sum_k a_k x_k\right).$$

Setting $f(x) = x \log x$, $a_i = p_i$ and $x_i = q_{j|i}$ we find for arbitrary $l$ that

$$\sum_k p_k q_{l|k} \log q_{l|k} \geqslant \sum_k p_k q_{l|k} \log \sum_k p_k q_{l|k} = q_l \log q_l.$$

Summing both sides over $l$ gives the result. □

## 6.2 The Uniqueness Theorem

We will now prove the statement made in the previous section that entropy is the only reasonable way to measure the uncertainty inherent in a finite probability space.

Remember that $A$ and $B$ denotes two finite schemes.

**Theorem 6.1** *Let $\Delta_n = \{(p_1, p_2, \ldots, p_n) \in \mathbb{R}^n : p_i \geqslant 0, \sum_{i=1}^n p_i = 1\}$ and suppose $H : \bigcup_{n=1}^{\infty} \Delta_n \to \mathbb{R}$ has the following properties:*

*(i)* *For each $n \geqslant 1$, $H|_{\Delta_n}$ is continuous.*

*(ii)* *$H(p_1, p_2, \ldots, p_n) = 0$ iff some $p_i = 1$.*

*(iii)* *$H(p_1, p_2, \ldots, p_n) = H(p_1, p_2, \ldots, p_n, 0)$.*

*(iv)* *For each $n \geqslant 1$, $H|_{\Delta_n}$ has its largest value at $(1/n, 1/n, \ldots, 1/n)$.*

*(v)* *$H(AB) = H(A) + H_A(B)$.*

*Then*

$$H(p_1, p_2, \ldots, p_n) = -\lambda \sum_{k=1}^n p_k \log p_k$$

*where $\lambda > 0$.*

**Remark** Property (*iii*) is just the statement that adding the impossible event (or any number of impossible events) does not change the uncertainty of the system.

**Proof of Theorem 6.1** Let $L(n) = H(1/n, 1/n, \dots, 1/n)$. Then

$$L(n) = H\left(\frac{1}{n}, \cdots, \frac{1}{n}, 0\right) \leqslant H\left(\frac{1}{(n+1)}, \cdots, \frac{1}{(n+1)}\right) = L(n+1)$$

by (*iii*) and (*iv*), so $L(n)$ is non-decreasing with respect to $n$. Let $m$ and $r$ be positive integers and let

$$S_i = \begin{pmatrix} S_{i_1} & S_{i_2} & \cdots & S_{i_r} \\ 1/r & 1/r & \cdots & 1/r \end{pmatrix}, \ (1 \leqslant i \leqslant m)$$

be mutually independent. Then $H(S_i) = L(r)$ and by (*v*) we have

$$H(S_1 S_2 \dots S_m) = \sum_{i=1}^{m} H(S_i) = mL(r).$$

The product space $S_1 S_2 \dots S_m$ obviously consists of $r^m$ equally likely events so $H(S_1 S_2 \dots S_m) = L(r^m)$ and

$$L(r^m) = mL(r), \text{ for all } m, r \in \mathbb{N}.$$

Now let $1 < r \leqslant s$, $n \in \mathbb{N}$ and $m = \max\{k \colon r^k \leqslant s^n\}$ so that

$$r^m \leqslant s^n < r^{m+1}.$$

Then

$$m \log r \leqslant n \log s < (m+1) \log r$$

and

$$\frac{m}{n} \leqslant \frac{\log s}{\log r} < \frac{m}{n} + \frac{1}{n}.$$

Since $L$ is non-decreasing we also have that

$$L(r^m) \leqslant L(s^n) \leqslant L(r^{m+1})$$

which is equivalent to

$$mL(r) \leqslant nL(s) \leqslant (m+1)L(r),$$

so

$$\frac{m}{n} \leqslant \frac{L(s)}{L(r)} \leqslant \frac{m}{n} + \frac{1}{n}.$$

Thus

$$\left| \frac{L(s)}{L(r)} - \frac{\log s}{\log r} \right| \leqslant \frac{1}{n},$$

and since the left side is independent of $m$, $n$ can be chosen arbitrarily large in the right side and

$$\frac{L(s)}{\log s} = \frac{L(r)}{\log r}.$$

Whence $L(n) = \lambda \log n$ and by the monotonicity of $L$ and $(ii)$ we have $\lambda > 0$. Note that this also holds for $n = 1$ since by $(ii)$ $L(1) = 0$.

To prove the general case consider the rational numbers $p_k,\ k = 1, 2, \ldots, n$:

$$p_k = \frac{g_k}{g};\ g_k > 0,\ \sum_{k=1}^{n} g_k = g$$

and let $A$ consist of $n$ events with probabilities $p_1, p_2, \ldots, p_n$. To define $H(A)$, consider the scheme $B$,

$$B = \begin{pmatrix} B_1 & B_2 & \ldots & B_g \\ q_1 & q_2 & \ldots & q_g \end{pmatrix}$$

and divide the $g$ events of $B$ into $n$ groups containing $g_1, g_2, \ldots, g_n$ events respectively.

Now let $B$ be dependent on $A$ in the following fashion: If the event $A_k$ of $A$ occurred we reduce $B$ to the $g_k$ events in group $k$, all with probability $1/g_k$. Then $B|A_k$ is a system containing $g_k$ equally likely events and

$$H_k(B) = H(B|A_k) = H(1/g_k, 1/g_k, \ldots, 1/g_k) = \lambda \log g_k.$$

Also

$$H_A(B) = \sum_{i=1}^{n} p_i H_i(B) = \lambda \sum_{i=1}^{n} p_i \log g_i = \lambda \sum_{i=1}^{n} p_i \log p_i + \lambda \log g. \qquad (6.1)$$

If we now consider the system $AB$ composed of the events $A_iB_j$, we see that such an event only is possible if $B_j$ belongs to the $i$th group. Thus $AB|A_i$ consists of $g_i$ events and the total number of possible events in $AB$ is $\sum g_i = g$. Furthermore, the probability of the event $A_iB_j$ is $p_i/g_i = 1/g$. Thus $AB$ consists of $g$ equally likely events and $H(AB) = \lambda \log g$. Using property $(v)$ and (6.1) we find

$$\lambda \log g = H(A) + \lambda \sum_{i=1}^{n} p_i \log p_i + \lambda \log g$$

$$\Updownarrow$$

$$H(A) = H(p_1, p_2, \ldots, p_n) = -\lambda \sum_{k=1}^{n} p_k \log p_k.$$

This result is also true for all real $p_i$ since $H$ is continuous by $(i)$. □

Having made it clear that the measure of uncertainty sought is in fact entropy, the section is continued with the entropy of Markov chains. As mentioned earlier, there is a strong connection between Markov chains and shifts of finite type, and this will be helpful when proving the relationship between topological and probabilistic entropy.

## 6.3 Markov Chains

Let $X$ be a Markov chain with a finite number of states $S_1, S_2, \ldots, S_n$ and transition probabilities $p_{ij}$. If the system is in state $S_k$, then the one step transitions of $X$ form the finite scheme

$$X|S_k = \begin{pmatrix} S_1 & S_2 & \ldots & S_n \\ p_{k1} & p_{k2} & \cdots & p_{kn} \end{pmatrix}.$$

The entropy of $X|S_k$

$$H_k(X) = H(X|S_k) = -\sum_{l=1}^{n} p_{kl} \log p_{kl}$$

is a measure of the amount of information obtained when the Markov chain moves one step ahead from $S_k$.

Averaging the entropies of the one step transitions gives the entropy of a Markov chain.

**Definition** Let $X$ be a irreducible Markov chain with a finite state space, transition probabilities $p_{ij}$ and initial probabilities $\pi_i$ $(1 \leqslant i, j \leqslant n)$. We define the entropy of $X$ as

$$H(X) = \sum_{k=1}^{n} \pi_k H_k(X) = -\sum_{k=1}^{n}\sum_{l=1}^{n} \pi_k p_{kl} \log p_{kl}.$$

**Remark** The entropy of a Markov chain is a measure of the average information obtained when moving one step ahead in the chain.

We will now define a Markov chain on a graph.

**Definition** A Markov chain on a graph $G$ is an assignment of probabilities $\pi_i \geqslant 0$ for $i \in \mathcal{V}(G)$ and conditional probabilities $P(e|i) \geqslant 0$ for $e \in \mathcal{E}(G)$ and $i \in \mathcal{V}(G)$ such that

$$\sum_{i \in \mathcal{V}} \pi_i = 1, \quad \sum_{e \in \mathcal{E}_i} P(e|i) = 1, \text{ for all } i \in \mathcal{V},$$

where $\mathcal{E}_i$ denotes the edges starting at vertex $i$.

**Remark** For a Markov chain on a graph, the probability of a path $\tau = e_1 e_2 \ldots e_n$, starting at vector $i$ is

$$P(\tau) = \pi_i P(e_1|i) P(e_2|\mathsf{t}(e_1)) \ldots P(e_n|\mathsf{t}(e_{n-1}))$$

After showing these properties for probabilistic entropy and Markov chains it is now time for the final theorem. The statement is easy but the theory underlying the proof is quite involved. Because of this I have chosen to prove just the main result and not the related facts stated in the remark below the theorem.

Remember from the Perron-Frobenius Theorem that for an irreducible matrix $A \neq [0]$ we let $\lambda_A$ denote the Perron eigenvalue and $u$, $v$ the corresponding left and right eigenvector.

**Theorem 6.2** *Let $G$ be an irreducible graph, $\mathsf{A}_G = [a_{ij}]$, $a_{ij} \in \{0,1\}$ and let $X$ be a Markov chain on $G$ defined by the following probabilities*

$$\pi_i = u_i v_i \text{ and } p_{ij} = \frac{a_{ij} v_j}{\lambda_A v_i}, \text{ where } \sum_i u_i v_i = 1.$$

*Then $H(X) = \log \lambda_A$.*

**Remark** This measure, the Perry measure is the unique measure with this property. It is also the measure that maximizes $H$. For a proof of this, see the pages 195-196 in [2].

**Proof of Theorem 6.2** Given the Perry measure

$$H(X) = -\sum_{i,j} \pi_i p_{ij} \log p_{ij} = -\sum_{i,j} u_i v_i \frac{a_{ij} v_j}{\lambda_A v_i} \log \frac{a_{ij} v_j}{\lambda_A v_i}$$

$$= -\sum_{i,j} u_i \frac{a_{ij} v_j}{\lambda_A} \log \frac{a_{ij} v_j}{\lambda_A v_i} = -\sum_{i,j} \frac{u_j a_{ij} v_i}{\lambda_A} (\log a_{ij} + \log v_j - \log \lambda_A - \log v_i)$$

$$= 0 - \sum_j u_j v_j \log v_j + \log \lambda_A + \sum_i u_i v_i \log v_i = \log \lambda_A$$

where the second to last equality follows since $a_{ij} \in \{0,1\}$ and $u$ and $v$ are eigenvectors. □

The result stated is that there exists a measure such that when applied to a Markov chain on a graph, the entropy of the Markov chain will equal the entropy of the shift space described by the same graph.

As an example of this result, consider the graph of the golden mean shift with the Perry measure on it.

**Example** If $G$ is the graph of the recoded golden mean shift then

$$\mathsf{A}_G = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \lambda = \lambda_A = \frac{1+\sqrt{5}}{2}, u_A = \begin{pmatrix} \lambda & 1 \end{pmatrix}, v_A = \begin{pmatrix} \frac{\lambda}{\lambda^2+1} \\ \frac{1}{\lambda^2+1} \end{pmatrix},$$

and we can define a Markov chain $X$ on $G$ by assigning probabilities to the edges. If we assign the Perry probabilities:

$$\pi = \begin{pmatrix} \frac{\lambda^2}{\lambda^2+1} \\ \frac{1}{\lambda^2+1} \end{pmatrix}, P = \begin{pmatrix} \frac{1}{\lambda} & \frac{1}{\lambda^2} \\ 1 & 0 \end{pmatrix}$$

then (since $\lambda^2 = \lambda + 1$)

$$H(X) = -\frac{\lambda^2}{\lambda^2+1} \left( \frac{1}{\lambda} \log \frac{1}{\lambda} + \frac{1}{\lambda^2} \log \frac{1}{\lambda^2} \right) = \frac{\lambda^2}{\lambda^2+1} \frac{\lambda+2}{\lambda^2} \log \lambda = \log \lambda.$$

# A Appendix

**Proof of Proposition 1.1** Let $X$ be a shift space over the alphabet $\mathcal{A}$ and $N \geqslant 1$. Then there is a collection $\mathcal{F}$ of blocks over $\mathcal{A}$ such that $X = \mathsf{X}_{\mathcal{F}}$. Create a new collection $\mathcal{F}'$ by replacing each block $u \in \mathcal{F}$ such that $|u|<N$ by all $N$-blocks over $\mathcal{A}$ containing $u$. Then $X = \mathsf{X}_{\mathcal{F}'}$ and $|v| \geqslant N$ for all $v \in \mathcal{F}'$. For each $w = a_1a_2 \dots a_m \in \mathcal{F}'$ let

$$w^{[N]} = (a_1a_2 \dots a_N)(a_2a_3 \dots a_{N+1}) \dots (a_{m-N+1}a_{m-N+2} \dots a_m)$$

be the corresponding $(m-N+1)$-block over $\mathcal{B}_N(\mathcal{A}^{\mathbb{Z}})$. Let $\mathcal{F}_1 = \{w^{[N]} : \ w \in \mathcal{F}'\}$ then $X^{[N]} \subset \mathsf{X}_{\mathcal{F}_1}$. Let

$$\mathcal{F}_2 = \{uv : u, v \in \mathcal{B}_N(\mathcal{A}^{\mathbb{Z}}), u \text{ and } v \text{ do not overlap progressively}\}.$$

Then it follows from part 1 of the remark in section 1.2 that $X^{[N]} \subset \mathsf{X}_{\mathcal{F}_2}$ and

$$X^{[N]} \subset \mathsf{X}_{\mathcal{F}_1} \cap \mathsf{X}_{\mathcal{F}_2} = \mathsf{X}_{\mathcal{F}_1 \cup \mathcal{F}_2}.$$

Conversely, suppose that $y \in \mathsf{X}_{\mathcal{F}_1 \cup \mathcal{F}_2}$ and let $x \in \mathcal{A}^{\mathbb{Z}}$ be the point reconstructed from the "bottom" letters of $y$ (as mentioned in part 2 of the remark in section 1.2). Then $x \in X = \mathsf{X}_{\mathcal{F}}$ since $y$ satisfies the constraints from $\mathcal{F}_1$ and $y = \beta_N(x)$ by the overlap constraints from $\mathcal{F}_2$. Whence $X^{[N]} \supset \mathsf{X}_{\mathcal{F}_1 \cup \mathcal{F}_2}$, and $X^{[N]} = \mathsf{X}_{\mathcal{F}_1 \cup \mathcal{F}_2}$ is a shift space. □

**Proof of Proposition 2.2**

1. If $m = 0$: The only paths of length zero are the empty paths (from $i$ to $i$). $A^0 = Id$, verifying the result in this case. According to the definition, it is also true for $m = 1$ . Suppose the result is true for $m = k$. Then

   $$(A^{k+1})_{ij} = \sum_l A_{il}(A^k)_{lj}$$

   equals the total number of paths of length $k+1$ from $i$ to $j$.

2. If $\pi$ is a cycle in $G$ of length $m$, then $\pi^\infty$ is a point of period $m$ in $\mathsf{X}_G$. Conversely, if $x \in \mathsf{X}_G$ has period $m$, then $x_{[0,m-1]}$ must be a cycle in $G$ of length $m$.

□

**Proof of Corollary 2.3.** Let $G$ be an irreducible graph, and $\pi, \tau \in \mathcal{B}(\mathsf{X}_G)$. Suppose that $\pi$ terminates at vertex $i$ and $\tau$ starts at vertex $j$. Since $G$ is irreducible there is a path $\omega \in \mathcal{B}(\mathsf{X}_G)$ from $i$ to $j$. Then $\pi\omega\tau$ is a path on $G$ and $\pi\omega\tau \in \mathcal{B}(\mathsf{X}_G)$. Conversely, suppose that $G$ is essential and $\mathsf{X}_G$ is irreducible. Let $j, k$ be vertices of $G$. Since $G$ is essential there are edges $e$ and $f$ such that $\mathsf{t}(e) = j$ and $\mathsf{i}(f) = k$. By irreducibility of $\mathsf{X}_G$ there is a block $h$ such that $ehf \in \mathcal{B}(\mathsf{X}_G)$. Then $h$ is a path in $G$ from $j$ to $k$. □

**Proof of Theorem 3.1** The proof of the theorem relies on the following lemma.

**Lemma A.1** *If $A$ is irreducible, then all states have the same period.*

**Proof** Let $i$ be a state, and let $p$ =per($i$). If $p = \infty$, then $A = [0]$ and the proof is completed. Assume $p < \infty$. Let $j$ be another state, then there exists $r, s \geqslant 1$ such that

$$(A^r)_{ij} > 0, \ (A^s)_{ji} > 0.$$

Let $(A^n)_{jj} > 0$, then

$$(A^{r+s})_{ii} \geqslant (A^r)_{ij}(A^s)_{ji} > 0,$$

and

$$(A^{r+n+s})_{ii} \geqslant (A^r)_{ij}(A^n)_{jj}(A^j)_{ji} > 0.$$

Thus $p$ divides both $r+s$ and $r+n+s$ which means that $p \mid n$. Hence $p$ divides all $n$ such that $(A^n)_{jj} > 0 \Rightarrow p \mid \text{per}(j)$. Reversing the roles of $i$ and $j$ shows that per($i$) =per($j$). □

We are now ready to prove the theorem:

$2 \Rightarrow 1$ Since $A^N > 0$ for some $N \geqslant 1$ there is a path of length $N$ between all states in $\mathsf{G}_A$. This establishes irreducibility of $A$. Also, since $A$ has no zero rows $A^{N+1} = A \cdot A^N > 0$. Thus per($A$) divides both $N$ and $N+1$ and per($A$) $= 1$.

$1 \Rightarrow 3$ Let $A$ be primitive. We first show that for each state $i$ there exists an $N_i$ such that $(A^n)_{ii} > 0$ for all $n \geqslant N_i$. Let $R_i = \{n \geqslant 1 \colon (A^n)_{ii} > 0\}$. By definition (and lemma A.1) gcd$\{n \in R_i\}$ =per($i$) =per($A$) = 1. Hence there are numbers $m_k, n_l \in R_i$ and $a_k, b_l \in \mathbb{N}$ such that

$$1 = \underbrace{\sum_{k=1}^{p} a_k m_k}_{\text{K}} - \underbrace{\sum_{l=1}^{q} b_l n_l}_{\text{L}} .$$

Let $N_i = L^2$. If $n \geqslant N_i$, then $n = cL + d$ where $c \geqslant L$ and $0 \leqslant d < L$. Hence

$$n = cL + d = cL + d(K - L) = (c-d)L + dK,$$

where $c - d \geqslant L - d > 0$ and $d \geqslant 0$. Thus

$$n = \sum_{l=1}^{q} [(c-d)b_l]n_l + \sum_{k=1}^{p} [da_k]m_k$$

is a linear combination of numbers in $R_i$, hence is in $R_i$. This shows that $R_i$ contains all $n \geqslant N_i$.

To complete the proof: Since $A$ is irreducible we can choose $M$ so that between every pair of states there is a path of lenght less than or equal to $M$. Let

$$N = M + \max_i N_i.$$

Then between any states $i, j$ there is a path of length $s \leqslant M$, and since $N - s \geqslant N_j$ there is a cycle of length $N - s$ starting at $j$. Hence there is a path of length $N$ from $i$ to $j$. This proves that $A^N > 0$, and since $A$ is primitive no row of $A$ can be zero so $A^n > 0$ for all $n \geqslant N$.

3 $\Rightarrow$ 2 Obvious.

□

**Proof of Proposition 4.4**

4 $\Rightarrow$ 5 First note that mixing of $X \cong \mathsf{X}_G$ is equivalent to primitiveness of $\mathsf{A}_G$ (this follows from Proposition 3.2). Let $U, V \subset X$ be non-empty, open sets. Then there exists $k, l \in \mathbb{Z}$ such that

$$C_k(u) = \sigma^{-k}(C_0(u)) \in U, u \in \mathcal{B}(X),$$

$$C_l(v) = \sigma^{-l}(C_0(v)) \in V, v \in \mathcal{B}(X).$$

Since $X$ is mixing, for every ordered pair $u, v \in \mathcal{B}(X)$ there exists an $N$ such that for all $i \geqslant N$ there exists a $w^{(i)} \in \mathcal{B}_i(X)$ such that $uw^{(i)}v \in \mathcal{B}(X)$. Once again, let $n_i = |uw^{(i)}| + k - l$. Since

$$z_i = \ldots uw^{(i)}v \ldots \in C_k(u) \text{ for all } i \geqslant N, z_k = u_1,$$

it is easy to see that

$$\sigma^{n_i}(U) \cap V \neq \varnothing \text{ for all } n_i \geqslant n_N.$$

Thus, $(X, \sigma)$ is topologically mixing.

4 $\Leftarrow$ 5 Since $(X, \sigma)$ is topologically mixing: given any ordered pair $u, v \in \mathcal{B}(X)$ there exists an $N > |u|$ such that

$$C_0(u) \cap \sigma^{-n}(C_0(v)) \neq \varnothing, \text{ for all } n \geq N.$$

For $z \in C_0(u) \cap \sigma^{-n}(C_0(v))$ we have

$$z_{[0,n+|v|-1]} = uz_{[|u|,n-1]}v \in \mathcal{B}(X), \text{ for all } n \geqslant N.$$

Thus, $X$ is mixing. □

# References

1. Douglas Lind and Brian Marcus - An Introduction to Symbolic Dynamics and Coding. Cambridge University Press 1995, ISBN: 0521559006.

2. Peter Walters - An Introduction to Ergodic Theory. Springer 2000. ISBN: 0387951520.

3. A.I. Kinchin - Mathematical Foundations of Information Theory. Dover Publications, Inc., New York 1957. ISBN: 0486604349.

4. Chris Hillman - All Entropies Agree for an SFT. February 18, 1998. http://www.math.uni-hamburg.de/home/gunesch/Entropy/PUB/aeasft.ps

5. Robert L. Devaney - An Introduction to Chaotic Dynamical Systems, Second Edition. Addison-Wesley Publishing Company 1989. ISBN: 0201130467.

6. Eugene Seneta - Non-Negative Matrices, An Introduction to Theory and Applications. George Allen & Unwin Ltd 1973. ISBN: 0045190119.

7. Lars Hörmander - Notions of Convexity. Birkhäser Boston 1994. ISBN: 0817637990.

8. URL: http://mathworld.wolfram.com/ReducibleMatrix.html

## Reference notes

The leftmost numbers refer to the pages in this report. The numbers to the right is the references for respective page.

### Fundamental Properties of Shift Spaces

2: 1-5 [1]
3: 3, 4, 6, 9 [1]
4: 12, 13 [1]
5: 11, 15, 16, 18 [1]
6: 18, 28, 29 [1]
7: 29, 30 [1]

### Graphs and Their Shifts

8: 33-35 [1]
9: 35, 36 [1]
10: 37, 38 [1]
11: 39, 41, 108 [1]
12: 42 [1]

### Further Properties of Shift Spaces

14: 125-128 [1]
15: 129 [1]

### Dynamical Systems

17: 177, 184 [1]
18: 185, 186, 188 [1]
19: 178, 189, 190 [1]
20: 178, 189 [1]
21: 189 [1]

### Topological Entropy

23: 168 [2]
24: 169, 171, 172 [2]
25: 172 [2], 104, [1]
26: 190 [1]
27: 100, 104 [1]
28: 106, 107 [1]
29: 112, 113, 119 [1]
30: 119, 120 [1]
31: 120, 121[1]

### Probabilistic Entropy

33: 2-4, 34 [3]
34: 4 [3]
35: 4-6 [3]
36: 5, 6, 9 [3], 77-78 [2]
37: 10, 11[3]
38: 12, 13 [3]
39: 14 [3], 329 [1]