Class Degree and Measures of Relative Maximal Entropy

by

Mahsa Allahbakhshi
B.Sc., University of Tehran, 2000
M.Sc., University of Tehran, 2003

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Mathematics and Statistics

Class Degree and Measures of Relative Maximal Entropy

by

Mahsa Allahbakhshi
B.Sc., University of Tehran, 2000
M.Sc., University of Tehran, 2003

Supervisory Committee

Dr. Anthony Quas, Supervisor
(Department of Mathematics and Statistics)

Dr. Chris Bose, Departmental Member
(Department of Mathematics and Statistics)

Dr. Ian F. Putnam, Departmental Member
(Department of Mathematics and Statistics)

Dr. Venkatesh Srinivasan, Outside Member
(Department of Computer Science)

Supervisory Committee

Dr. Anthony Quas, Supervisor
(Department of Mathematics and Statistics)

Dr. Chris Bose, Departmental Member
(Department of Mathematics and Statistics)

Dr. Ian F. Putnam, Departmental Member
(Department of Mathematics and Statistics)

Dr. Venkatesh Srinivasan, Outside Member
(Department of Computer Science)

ABSTRACT

Given a factor code π from a shift of finite type X onto an irreducible sofic shift Y, and a fully supported ergodic measure ν on Y, we give an explicit upper bound on the number of ergodic measures on X which project to ν and have maximal entropy among all measures in the fiber $\pi^{-1}\{\nu\}$. This bound is invariant under conjugacy. We relate this to an important construction for finite-to-one symbolic factor maps.

Contents

List of Figures

ACKNOWLEDGEMENTS

In the first place I would like to record my sincere gratitude to Anthony Quas for his supervision and guidance from the very earliest stage of this research, as well as for providing me with advice and support in various ways since I moved to Canada to start my program. His truly mathematical intuition and his logical way of thinking have always been of great value to me (and I hope I have inherited some of it!). I would like to thank him for his constant questions and his patience with my less frequent answers; for guiding me on how best to express my ideas, and for his financial support throughout these years.

It is a pleasure to thank Chris Bose for his good teaching, guidance and thoughtful insights since the first day I started my program.

I gratefully acknowledge Ian Putnam who was very kind to carefully read a draft of this thesis before submission, and provide outstanding advice on how it could be improved.

I also thank Brian Marcus and Venkatesh Srinivasan, who agreed to participate in the examining committee on such a short notice.

I have great regard for the many people who shared their experience and knowledge during the time of my studies. I am especially thankful to Mate Wierdl, Karl Petersen, Steven Kalikow, Hillel Furstenberg, and Jean-Paul Thouvenot.

Many friends not only have helped me stay sane through difficult moments of these years but also made very enjoyable times for me in Victoria. Mojtaba and Christianne, I am deeply grateful to you for your friendship and hospitality, for helping and treating me like family, and for all the dances and laughs at your house. Maryam, your care and your sense of beauty brought me wisdom and joy when I needed it the most. To Muhammad, thank you for making my life more cheerful, for being a source of energy at stressful times, and for sharing your peanut butter sandwiches with me every single time!

Goldis, you are part of my family. No matter when or where, you are always there for me, and I thank you. Thank you Faranak, for all the encouragement and your constant faith in me.

Words fail me to express my appreciation to my sister Mehrnoosh, and my mother, whose love and prayers have taken the loads off my shoulders. Mehrnoosh, you with your unfailing instincts, are my angel on earth. Thank you for all the chats and dreams we had together. Even when you are in your most diabolical mode, you are still the best sister ever!

Mom, you are my hero. Thank you for giving me courage at the times when I had none. Thank you for all the times you came to my rescue and lifted me above my fears. Alireza, to whom this dissertation is dedicated, you have been a constant source of strength and support; a strong backbone of the family whom we can always count on when times are rough. Dad, I always know how difficult it is for you knowing that I am away from home, where you can take my hands and feel that I am safe. Maryam, your house has always been my sanctuary, where I can clear my mind as well as enjoy all sorts of yummy snacks you always treat us with. Now I need to thank my parents in Persian.

مادر عزیزم، بغضی بی انتها گلویم را سخت میفشارد و فرو میمانم که کدامین کلمات در شاءن تواند. هرگز از یاد نخواهم برد تمامی لحضات طولانی را که با دل و جان به من گوش فرا میدهی و آنقدر نغمه عشق سر میدهی تا تمامی غمها و نگرانیها از تک تک یاختهای بدنم پر کشند و نور وجود تو جای گزینشان شود. همیشه به خوبی میدانم هر جا که باشم در سختترین لحضات زندگیم شنیدن صدای تو و مناجات آسمانی توست که نجاتم میدهد.

پدر نازنینم ، تشویش دستانت را آن هنگام که به آغوشم کشیدی و پذیرفتی که خانه را به قصد هدفم ترک میکنم به خوبی به یاد میاورم. ای کاش میتوانستم رنج نگاهت و صدای لرزانت را در تمامی لحضاتی که نگران سلامت و امنیت من بودی و هستی را پاسخی گویم.

پدر و مادر عزیزم، میخواهم دستان پر مهرتان را بوسه بارن کنم و سپاستان گویم برای همه عمری که بر من سپری کردید و خواهید کرد، برای همه مهربانیتان، برای لبخندهای شیرینتان و نگاه های زندگی بخشتان.

دوستتان دارم.

Chapter 1

Introduction

All of our work deals with special dynamical systems. A dynamical system is specified by a set X together with a map from X to X (which we shall call T). Intuitively we think of the points of X as describing the "state" of a physical system. If the physical system is in state x at some time n, then $T(x)$ describes the state of the system at time $n+1$. Note that it is crucial that the same map T is applied at every stage (this assumption ensures that we are dealing with a homogeneous dynamical system). In ergodic theory, the dynamical systems have additional structure, namely a measure. A measure is a function mapping subsets of X to $[0,1]$ which we think of as measuring the size of the subset. Of crucial importance are the invariant measures: those for which $T(A)$ and A always have the same measure.

Symbolic dynamics is a rapidly growing part of dynamical systems which allows us to study complex dynamical systems by means of discretizing space as well as time. Shift spaces are basic notions in symbolic dynamics. Each point in a shift space represents a bi-infinite walk on a directed graph and the passage of time T corresponds to shifting a sequence one place to the left.

One of the major concepts which is sensitive to the measure defined on X is the entropy of T. Entropy is a non-negative number measuring the average amount of information provided by the knowledge of the "present state" given the knowledge of an arbitrary long past. Thus, a system with zero entropy can be viewed as strongly deterministic in the sense that the knowledge of the entire past precisely determines the future behavior of the system.

In symbolic dynamics measures of maximal entropy have been thoroughly studied and many of their properties are known. It is a well-known result of Parry [21], generalizing an earlier result of Shannon [27], that every irreducible shift of finite type has a unique measure of maximal entropy which is given by an explicit formula. In contrast, we consider the relative case in which one is given a factor code $\pi : X \to Y$ from a shift of finite type X to a sofic shift Y, and a measure ν on Y. In this case measures on X in the fibre $\pi^{-1}\{\nu\}$ having maximal entropy in the fibre, so-called measures of relative maximal entropy, are not well understood.

Measures of relative maximal entropy appear frequently in different areas of Mathematics. One of the applications is providing some techniques to compute Hausdorff dimension. This reveals the connections of measures of relative maximal entropy with functions of Markov chains [3, 4, 5, 20], measures that maximize a weighted entropy functional [8, 28], the theory of pressure and equilibrium states [10, 13, 26], relative pressure and relative equilibrium states [17, 18, 30], and compensation functions [4, 30]. Other uses of such measures arise from their application in the mathematics of information transfer [23] and information-compressing channels [20].

The main goal of this thesis is finding an upper bound on the number of ergodic measures of relative maximal entropy. In 2003, Petersen, Quas, and Shin [24] found the following upper bound on the number of such measures.

Theorem 1.0.1. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let ν be an ergodic measure on Y and*

$$N_\nu(\pi) = \min\{|\pi^{-1}(w)| : w \textit{ in the alphabet of } Y,\, \nu[w] > 0\}.$$

The number of ergodic measures of maximal entropy over ν is at most $N_\nu(\pi)$.

This bound suffers from not being invariant under conjugacy and becomes arbitrarily large simply by recoding to a higher block presentation. A simple example is the map $\pi : X \to Y$ where X is the full 2-shift $\{0,1\}^{\mathbb{Z}}$, $Y = \{0\}^{\mathbb{Z}}$ and ν is the trivial measure on the full 1-shift. Then $N_\nu(\pi) = |\pi^{-1}(0)| = 2$. The 2nd-higher block presentation of X, $X^{[2]}$, is conjugate to X. However, $N(\bar{\pi}) = 4$ where $\bar{\pi} : X^{[2]} \to \{0\}^{\mathbb{Z}}$.

To avoid this issue one possibility is to take the minimum of this bound over all shifts of finite type $\tilde{X}$ which are conjugate to X. However, no algorithm for computing this minimum has been found. On the other hand, in this thesis we find a simpler conjugacy-invariant upper bound which can be strictly better and is never worse than the minimum bound mentioned above over the conjugates.

When a factor code is finite-to-one, there is a well-known quantity assigned to the code called the "degree" of the code. The degree of a code is invariant under conjugacy and can be easily computed. We show there is a quantity assigned to a general factor code, called the "class degree" of the code, analogous to the degree of a finite-to-one code. By defining a "transition equivalence relation" on X the class degree of the code is defined to be the number of transition classes in the preimage of a "typical" point of Y. It is shown that the class degree of a code is invariant under conjugacy. Given an ergodic measure ν on Y, we show that the number of transition classes above a typical point of ν is an upper bound on the number of measures of relative maximal entropy. This bound is equal to the class degree of the code when ν is ergodic and fully supported.

One of the ingredients we use to prove the above statement is a key property of measures of relative maximal entropy. Given a finite set $G \subseteq \mathbb{Z}$, the boundary of the complement of G is $\partial G^c = \{i \in G^c \colon \exists j \in G \text{ with } |i-j| = 1\}$. We prove if μ is a measure of maximal entropy in the fibre $\pi^{-1}\{\nu\}$ then the conditional distribution of μ on any finite set $G \subseteq \mathbb{Z}$ given the configuration on G^c is μ-a.s. uniform over all configurations on G which extend the configuration on ∂G^c and map to the same configuration in Y under the factor code π. The proof of this result follows techniques developed by Burton and Steif [6] where they show that if μ is a measure of maximal entropy for a d-dimensional shift of finite type then the conditional distribution of μ on any finite set $G \subseteq \mathbb{Z}^d$ given the configuration on G^c is μ-a.s. uniform over all configurations on G which extend the configuration on ∂G^c (this result is related to an earlier work of Lanford and Ruelle [16]).

The content of the thesis is as follows. We have divided the material into four main parts (Chapters 2, 3, 4, and 5) where the main goal is achieved in Chapter 5. In Chapter 2 we discuss the background material from dynamical systems. In this chapter we follow [29] and chapters 1, 2, 9 of [19]. If the reader is familiar with the topics of these two references, they may skip the second chapter and refer back to it for the statement of theorems as needed.

We start the third chapter with an introduction to conditional expectation and its properties which are needed later in the thesis. We prove the uniform conditional distribution property of measures of relative maximal entropy in Section 3.2. This property is used in Chapter 5.

The degree of a finite-to-one code is reviewed in the beginning of Chapter 4. We present an efficient algorithm for the computation of the degree in Section 4.1.2. A major result of this thesis is discussed in Section 4.2. Given a factor code $\pi : X \to Y$, we introduce an equivalence relation on X, called the transition equivalence relation. The minimum number of transition classes above points of Y is called the class degree of the code. Then in Section 4.3 we show that the number of transition classes above each typical point is the same and equal to the class degree.

Chapter 5 starts with reviewing the old upper bound on the number of ergodic measures of relative maximal entropy and finishes with presenting a new upper bound on such measures.

Chapter 2

Background

In this chapter we review some background materials from ergodic theory and symbolic dynamics. In many cases we refer the reader to other resources for further details. However, we do provide examples and proofs as well as giving an interpretation for topics which are key ingredients in a future chapter. To reach this goal, we describe the standard ideas of ergodic theory in the special case of shift spaces that is the principal context for our study.

The chapter starts with introducing shift spaces, in particular shifts of finite type and their connection with finite directed graphs. We develop methods and tools that are useful in the study of these spaces and measures supported on them in Sections 2.1.1 and 2.1.2. Entropy is defined in Section 2.2. We discuss this important topic from a few viewpoints and compute it in some cases. Section 2.3 studies sofic shifts which are factors of shifts of finite type.

2.1 Shift Spaces

Definition 2.1.1. *Let $\mathcal{A}$ be a finite set, called the alphabet. The* full $\mathcal{A}$-shift*, denoted by*

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

is the collection of all bi-infinite sequences of symbols from $\mathcal{A}$ equipped with the shift map $T : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ where $T(x)_n = x_{n+1}$ for each $n \in \mathbb{Z}$.

Each bi-infinite sequence $x \in \mathcal{A}^{\mathbb{Z}}$ is called a point of the full shift. We sometimes use the notation $\ldots x_{-2}x_{-1}\overset{*}{x}_0x_1x_2\ldots$ for a point in $\mathcal{A}^{\mathbb{Z}}$. The * indicates the 0-th position (or time 0) in the sequence. T can be written as

$$T(\ldots x_{-2}x_{-1}\overset{*}{x}_0x_1x_2\ldots) = \ldots x_{-2}x_{-1}x_0\overset{*}{x}_1x_2\ldots.$$

For example $\{0,1\}^{\mathbb{Z}}$ is the set of all binary sequences. The point $\ldots 00\overset{*}{1}01\ldots$ maps to $\ldots 001\overset{*}{0}1\ldots$ under the shift map T.

The full n-shift is the full shift over the alphabet $\{0, 1, \ldots, n-1\}$. If $\mathcal{A}$ has size n, then there is a natural correspondence between the full n-shift and the full $\mathcal{A}$-shift. For example it is convenient to refer to the full shift over $\{0,1\}$ or the full shift over $\{a,b\}$ as the full 2-shift.

A metric on $\mathcal{A}^{\mathbb{Z}}$ is given by $d(x,y) = 2^{-k}$, where k is the smallest integer so that $x_{[-k,k]} \neq y_{[-k,k]}$. In other words, the distance between x and y is 2^{-k} if k is the number of positions away from the 0th coordinate where the first disagreement between x and y occurs (with the convention that if no disagreement occurs, then $d(x,y) = 2^{-\infty} = 0$). So two points are "close" if they have many initial symbols in common. It is not hard to verify that d defines a metric on X.

Definition 2.1.2. *If X is a non-empty shift-invariant closed subset of $\mathcal{A}^{\mathbb{Z}}$ for some alphabet $\mathcal{A}$, then the system (X, T_X) is called a shift space where T_X is the restriction to X of the shift map T on the full shift.*

It is worth mentioning that some authors prefer to use the term "subshift" rather than "shift space".

Definition 2.1.3. *Let $\mathcal{A}$ be an alphabet. A block over $\mathcal{A}$ is a finite sequence of symbols from $\mathcal{A}$ including the sequence of no symbols, called the empty block.*

A $\mathbb{Z}$-interval is a finite contiguous subset of $\mathbb{Z}$. If there is a point x in X and a $\mathbb{Z}$-interval G such that $x_G = B$, then we say B is a block of X or B occurs in x.

Definition 2.1.4. *A shift space* (X, T_X) *is called a* shift of finite type (SFT) *if there exists a finite set of (forbidden) blocks* $\mathcal{F}$ *defining* X *as*

$$X = X_{\mathcal{F}} = \{x \in \mathcal{A}^{\mathbb{Z}} : \text{ no block } W \in \mathcal{F} \text{ occurs in } x\}.$$

If B is a block of X, we sometimes refer to B as an "allowable block" of X.

Let $\mathcal{A}$ be an alphabet. The length of a block B over $\mathcal{A}$ is the number of symbols it contains, and is denoted by $|B|$. A n-block is a block of length n.

Definition 2.1.5. *We say* X *is a SFT of step size at most* N *if there is a forbidden block set consisting of blocks of length at most* $N+1$*, and say* X *is an* N-step SFT *if it is of step size at most* N *but not of step size at most* $N-1$.

One can think of a 1-step SFT as being given by a collection of symbols together with a description of which symbols can follow which others. The following example describes a 1-step SFT.

Example 2.1.6. *Let* X *be the set of binary sequences with no two consecutive 1's. Then* $X = X_{\mathcal{F}}$*, where* $\mathcal{F} = \{11\}$*. The 1-step shift of finite type* X *is called the* golden mean shift*.*

Most of the time we denote a SFT (X, T_X) by only X, the shift map being understood implicitly. When dealing with several SFTs, their possibly different alphabets will be denoted by $\mathcal{A}(X), \mathcal{A}(Y)$.

Here we show an alternative description of 1-step shifts of finite type using directed graphs.

Definition 2.1.7. *A finite directed* graph G *consists of a finite set of vertices* $\mathcal{V}$ *together with a finite set of edges* $\mathcal{E}$*. Each edge* ξ *in* $\mathcal{E}$ *starts at an initial vertex of* $\mathcal{V}$ *denoted by* $i(\xi)$*, and terminates at a terminal vertex of* $\mathcal{V}$ *denoted by* $t(\xi)$*. Note that loops are allowed in a graph, meaning that there might be an edge* ξ *where* $t(\xi)$ *is the same as* $i(\xi)$*.*

Through this work we only consider graphs in which there is no more than one edge from a vertex to another.

Let G be a finite directed graph with vertex set $\mathcal{V}$ (where there is no more than one edge from a vertex to another). For vertices I and J in $\mathcal{V}$, let M_{IJ} denote the number of edges in G with initial vertex I and terminal vertex J. Note that M_{IJ} is 1 if there is an edge from I to J and it is 0 otherwise. The adjacency matrix of G is $M_G = [M_{IJ}]$. A walk is a sequence of graph vertices $v_i \dots v_j$ such that $M_{v_k v_{k+1}}$ is 1 for each $i \leq k < j$. The vertex shift X_M is the shift space over the alphabet $\mathcal{V}$ specified by

$$X_M = \{v \in \mathcal{V}^{\mathbb{Z}} \colon M_{v_i v_{i+1}} = 1 \text{ for each } i \in \mathbb{Z}\}.$$

Note that the vertex shift X_M is a 1-step SFT defined by the list

$$\mathcal{F} = \{v_i v_j \colon M_{v_i v_j} = 0\}$$

of forbidden 2-blocks. Conversely, every 1-step SFT is a vertex shift. Because if $X = X_{\mathcal{F}}$ where $\mathcal{F}$ consists of 2-blocks, then X may be regarded as the vertex shift X_M, where M is the 0-1 matrix indexed by the alphabet of X and $M_{ij} = 0$ if and only if $ij \in \mathcal{F}$. Later we show that any shift of finite type can be viewed as a vertex shift.

Example 2.1.8. *Let G be a finite directed graph with vertex set $\mathcal{V} = \{0, 1\}$ and*

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

shown in Figure 2.1. The vertex shift X_M is the golden mean shift described in Example 2.1.6.

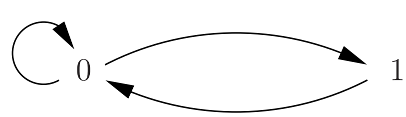

Figure 2.1: Graph for Example 2.1.8 representing the golden mean shift

2.1.1 Invariant Measures

Let X be a set. A σ-algebra $\mathscr{B}_X$ is a collection of subsets of X that is closed under countable unions and complements and also contains the set X itself. By a measure on X we mean a function $\mu : \mathscr{B}_X \to [0,1]$ satisfying $\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n)$ whenever $(B_n)_{n=1}^{\infty}$ is a sequence of pairwise disjoint members of $\mathscr{B}_X$. A probability measure is a measure μ with the added assumption that $\mu(X)$ is 1. For our purpose, measures will always be probability measures.

Let X be a shift space and x in X. The block of coordinates in x from time i to time j is denoted by $x_{[i,j]} = x_i x_{i+1} \dots x_j$. We denote the set of points in X in which a block $B = B_0 \dots B_{n-1}$ occurs starting at position k by

$$C_k(B) = {}_k[B_0 \dots B_{n-1}] = \{x \in X : x_{[k,\,k+n-1]} = B\}.$$

The set $C_k(B)$ is called the cylinder set defined by B with starting position k. The collection of all cylinder sets of X is denoted by $\mathscr{L}(X)$, and $\mathscr{B}_X$ denotes the smallest σ-algebra which contains the cylinder sets. It is worth noting that a subset of a shift space is both open and closed if and only if it is a finite union of cylinder sets.

To define a measure on X, one can assign probabilities to cylinder sets and then use standard measure theory to extend these assignment to all members of $\mathscr{B}_X$. We need the following definitions before stating Theorem 2.1.11 which allows us to extend a measure in this way.

Definition 2.1.9. *Let $\mathscr{L}$ be a collection of sets that contains the empty set. A function $\tau : \mathscr{L} \to \mathbb{R}^+$ is called a pre-measure if*

1. $\tau(\emptyset) = 0$.
2. *If $\{A_i\}$ is a countable collection of disjoint sets of $\mathscr{L}$ and $\bigcup_i A_i \in \mathscr{L}$, then*

$$\tau\left(\bigcup_i A_i\right) = \sum_i \tau(A_i).$$

Definition 2.1.10. *Let X be a set. A* semi-algebra $\mathscr{L}$ *is a collection of subsets of X such that*

1. $\emptyset \in \mathscr{L}$.
2. $\mathscr{L}$ *is closed under finite intersections.*
3. *If $A \in \mathscr{L}$ then A^c may be written as a finite union of disjoint elements of $\mathscr{L}$.*

The usefulness of semi-algebras comes from the following theorem. Note that when X is a shift space then $\mathscr{L}(X)$, the collection of all cylinder sets of X, forms a semi-algebra.

Theorem 2.1.11. *Let X be a shift space. Let $\tau : \mathscr{L}(X) \to [0,1]$ be a probability pre-measure on $\mathscr{L}(X)$. Then τ uniquely extends to a probability measure on X.*

A proof of Theorem 2.1.11 can be found in [15]

Example 2.1.12. *Let $X = \{0, 1, \ldots, k-1\}^{\mathbb{Z}}$ and let $P = (p_0, p_1, \ldots, p_{k-1})$ be a probability vector (i.e., $p_i \geq 0$ for each i and $\sum_{i=0}^{k-1} p_i = 1$). Define the function τ on the set of cylinder sets by*

$$\tau({}_r[a_0, a_1, \ldots, a_n]) = \prod_{i=0}^{n} p_{a_i}.$$

Then one can extend τ to a measure μ on X (for more details, see Theorem 0.2 and 0.4 of [15]). The measure μ on X is called the Bernoulli-$(p_0, p_1, \ldots, p_{k-1})$ measure.

We shall denote by $M(X)$ the collection of all (probability) measures defined on X.

Definition 2.1.13. *Let X be a shift space. A measure μ on X is called* T-invariant *if for each $B \in \mathscr{B}_X$ we have $\mu(T^{-1}B) = \mu(B)$.*

We most of the time refer to μ as an invariant measure rather than a T-invariant measure when T is being understood implicitly.

Theorem 2.1.14. *[19, Theorem 1.1] Let X be a shift space and μ be a measure on X. If for each cylinder set C we have $\mu(T^{-1}C) = \mu(C)$ then μ is invariant.*

In simple words μ is invariant if and only if the starting position of any cylinder set does not affect its measure.

The subset of $M(X)$ consisting of the invariant measures is denoted by $M(X, T)$. By the above Theorem, it is clear that the Bernoulli-$(p_0, p_1, \ldots, p_{k-1})$ measure defined in Example 2.1.12 is invariant and therefore belongs to $M(X, T)$.

To investigate the topological properties of $M(X)$ we need to define continuous and integrable functions on X. Let X be a shift space and μ be a measure on X. A function $f: X \to \mathbb{R}$ is $\mathscr{B}_X$-measurable if $f^{-1}(c, \infty) \in \mathscr{B}_X$ for all $c \in \mathbb{R}$. A function $f: X \to \mathbb{R}$ is called continuous if $f^{-1}(c, \infty)$ is open for all $c \in \mathbb{R}$. The set of all continuous functions $f: X \to \mathbb{R}$ is denoted by $C(X)$. If f is in $C(X)$, we define the uniform norm [1] of f to be

$$\|f\|_u = \sup\{|f(x)| : x \in X\}.$$

Note that the shift space X is a compact space and since any continuous function on a compact space is bounded we have $\|f\|_u < \infty$ for every f in $C(X)$. (To see a justification that shift spaces are compact see [19] page 177.) The function $d(f, g) = \|f - g\|_u$ is easily seen to be a metric on $C(X)$. We say $f: X \to \mathbb{R}$ is μ-integrable if f is $\mathscr{B}_X$-measurable and $\int f \, d\mu < \infty$ (a full description of integration may be found in [7]). We say $f = g$ μ-a.e. if $\mu(\{x \in X : f(x) \neq g(x)\}) = 0$. The space of all μ-integrable functions $f: X \to \mathbb{R}$ where two such functions are identical if they are equal μ-a.e. is denoted by $L^1(X, \mathscr{B}_X, \mu)$. In speaking of $\mathscr{B}_X$-measurable functions or μ-integrable functions we shall usually omit the $\mathscr{B}_X$ and the μ, and simply write measurable functions or integrable functions. We may also write L^1 rather than $L^1(X, \mathscr{B}_X, \mu)$ or a.e. rather than μ-a.e. if the σ-algebra and the measure defined on X is understood from the context.

[1] A norm on a real vector space V (for example the vector space $C(X)$) is a function $v \to \|v\|$ from V to $[0, \infty)$ such that (a)$\|v + w\| \leq \|v\| + \|w\|$ for all v, w in V, (b)$\|\lambda v\| = |\lambda| \|v\|$ for all v in V and λ in $\mathbb{R}$, and (c)$\|v\| = 0$ only when $v = 0$.

We need the following theorem to describe a topology on the set of invariant measure on a shift space X.

Theorem 2.1.15. *[29] Let X be a compact metric space. The set $C(X)$ of continuous functions on X has a countable dense subset.*

There is a standard topology on the set of invariant measures on X, $M(X)$, called the weak* topology [1] which is generated by the following metric on $M(X)$. Fix a countable dense subset of $C(X)$, $(f_n)_{n=1}^{\infty}$. For μ and ν in $M(X)$ define

$$d(\mu, \nu) = \sum_{n=1}^{\infty} \frac{|\int f_n \, d\mu - \int f_n \, d\nu|}{2^n \|f_n\|_u}.$$

It is straightforward to see $d\colon M(X) \times M(X) \to \mathbb{R}$ is a metric.

Theorem 2.1.16. *[29, Theorem 6.5] Let X be a shift space. The space $M(X)$ is compact in the weak* topology.*

Theorem 2.1.17. *[29, Theorem 6.9] Let X be a shift space and x in X. Form a sequence $(\mu_N)_{N=1}^{\infty}$ of measures on X by $\mu_N = \dfrac{1}{N}\sum_{i=0}^{N-1} \delta_x \circ T^{-i}$ where δ_x stands for the point mass measure at x. Then any limit point μ of $(\mu_N)_{N=1}^{\infty}$ is an invariant measure; i.e., μ is a member of $M(X,T)$.* (Note that such limit points exist by the compactness of $M(X)$.)

Definition 2.1.18. *Let X be a shift space. A measure $\mu \in M(X,T)$ is called* ergodic *if the only members $B \in \mathscr{B}_X$ with $T^{-1}B = B$ satisfy $\mu(B) = 0$ or $\mu(B) = 1$.*

The idea is working with spaces in which there are no portion that remains isolated from the rest if we allow the system to evolve long enough.

Theorem 2.1.17 shows that $M(X,T)$ is non-empty. We, moreover, have the following properties of $M(X,T)$ which immediately implies that the set of ergodic measures of X is also non-empty.

[1]The most common way to define this topology on $M(X)$ is by applying Riesz Representation Theorem to relate elements of $M(X)$ to linear functionals on $C(X)$ and get a topology on $M(X)$ from the weak* topology on $C(X)^*$. For a detailed discussion on Riesz Representation Theorem and weak* topology see [7] Chapter 5 and [29] page 148.

Theorem 2.1.19. *[29, Theorem 6.10] Let X be a shift space. Then*

(a) $M(X,T)$ is convex and is a compact subset of $M(X)$.

(b) μ is an ergodic measure if and only if μ is an extreme point of $M(X,T)$.

The Ergodic Decomposition theorem shows that any invariant measure on X can be decomposed to ergodic measures as follows.

Theorem 2.1.20. (Ergodic Decomposition Theorem) Let $E(X,T)$ denote the set of all ergodic measures on X. Then for each μ in $M(X,T)$ there is a unique measure τ on $M(X,T)$ such that $\tau(E(X,T)) = 1$ and for each f in $C(X)$,

$$\int_X f(x)\,d\mu(x) = \int_{E(X,T)} \left(\int_X f(x)\,d\rho(x) \right) d\tau(\rho).$$

We write $\mu = \displaystyle\int_{E(X,T)} \rho\, d\tau(\rho)$ and call it the ergodic decomposition of μ. From Theorem 2.1.20 we see that τ represents the "weights" of the ergodic measures that make up μ. A proof of the Ergodic Decomposition Theorem may be found in [25]

Determining if a measure μ on X is ergodic is not always straightforward by only using Definition 2.1.18. There are several other ways of stating the ergodicity condition. We state two of them in the following theorem. The first criterion gives a way of checking the ergodicity by reducing the computations to a class of sets we can manipulate, and the second result expresses the ergodicity in functional form. We apply (b) to check ergodicity in Example 2.1.22 and use (c) in Theorem 2.1.26.

Theorem 2.1.21. *Let μ be an invariant measure on a shift space X. Then the following are equivalent:*

(a) μ is ergodic.

(b) For any cylinder sets U and V of X,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \mu(T^{-i}U \cap V) = \mu(U)\mu(V).$$

(c) For any continuous function f and for any integrable function g,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \int f(T^i x) g(x)\, d\mu = \int f\, d\mu \int g\, d\mu.$$

A proof of this Theorem may be found in [29]. In particular, the reader is referred to Theorem 1.17 and Lemma 6.11 of [29].

Example 2.1.22. *The Bernoulli-$(p_0, \ldots, p_{k-1})$ measure is ergodic.*

Proof. To see this, let $U = {}_a[i_0 \ldots i_r]$ and $V = {}_b[j_0 \ldots j_s]$ be two cylinder sets of X. For $n > b + s - a$ we have

$$\mu(T^{-n}U \cap V) = p_{j_0} \ldots p_{j_s} p_{i_0} \ldots p_{i_r} = \mu(U)\mu(V).$$

Therefore,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \mu(T^{-i}U \cap V) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=b+s-a+1}^{N-1} \mu(T^{-i}U \cap V) = \mu(U)\mu(V).$$

□

The orbit of a point x in X is the sequence $(T^i x)_{i=-\infty}^{\infty}$. The Birkhoff Ergodic Theorem shows that if a function f has a finite integral with respect to an ergodic measure μ then the integral of the function is its average value along the orbit of μ-almost every point.

Theorem 2.1.23. *(*Birkhoff Ergodic Theorem) Let X be a shift space. Let μ be an ergodic measure on X and f be an integrable function. Then

$$\lim_{N\to\infty} S_N f(x) = \int f\, d\mu$$

for μ-almost every point x in X, where

$$S_N f(x) = \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x).$$

A short proof of the Birkhoff Ergodic Theorem may be found in [12]. For a proof based on nonstandard-analysis see [2].

Definition 2.1.24. *Let μ be an invariant measure on a shift space X. A point x in X is called* **generic** *for μ if $\lim_{N\to\infty} S_N f(x) = \int f\, d\mu$ for every f in $C(X)$.*

Theorem 2.1.26 shows that given the ergodicity property, almost every point of a shift space is generic. To prove this, we need the following basic theorem on integrating a sequence of functions.

Theorem 2.1.25. (Dominated Convergence Theorem) Let $g : X \to \mathbb{R}$ be μ-integrable and $(f_n)_{n=1}^{\infty}$ be a sequence of measurable real-valued functions with $|f_n| \leq g$ μ-almost everywhere. Suppose that for μ-almost every x the sequence $f_n(x)$ is convergent, and let $f(x) = \lim_{n\to\infty} f_n(x)$ μ-a.e. Then f is integrable and $\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu$.

For a proof of the dominated convergence theorem see [7, Theorem 2.24].

Theorem 2.1.26. *Let μ be an invariant measure on a shift space X. Then μ is ergodic if and only if μ-almost all points of X are generic.*

Proof. Suppose μ is ergodic. Let $(f_i)_{i=1}^{\infty}$ be a dense subset of $C(X)$. By the Birkhoff Ergodic Theorem for any function f_i we have $\lim_{N\to\infty} S_N f_i(x) = \int f_i\, d\mu$ for μ-almost every point x in X. Let

$$E_i = \{x \in X \colon \lim_{N\to\infty} S_N f_i(x) = \int f_i\, d\mu\}.$$

Since each E_i has full μ-measure, their countable intersection

$$E = \bigcap_{i=1}^{\infty} E_i = \{x \in X \colon \lim_{N\to\infty} S_N f_i(x) = \int f_i\, d\mu, \text{ for each } i\}$$

has full μ-measure as well. Now let f be in $C(X)$. There is a subsequence $(f_{i_j})_{j=1}^{\infty}$ which converges uniformly to f; i.e, for a given $\epsilon > 0$ there is J such that for any $j > J$ we have $\|f_{i_j} - f\|_u = \sup\{|f_{i_j}(x) - f(x)| \colon x \in X\} < \epsilon$. Since for each N we have

$$\|S_N f_{i_j} - S_N f\|_u \leq \|f_{i_j} - f\|_u < \epsilon,$$

and for each integer j and each x in E

$$\lim_{N\to\infty} S_N f_{i_j}(x) = \int f_{i_j}\, d\mu$$

it follows that for $x \in E$

$$\lim_{N\to\infty} S_N f(x) = \int f\, d\mu.$$

Thus E consists of generic points.

Conversely, suppose μ-almost all points of X are generic; i.e, there is $E \subseteq X$, with $\mu(E) = 1$ such that for each $x \in E$, and $f \in C(X)$ we have $\lim_{N\to\infty} S_N f(x) = \int f\, d\mu$.

Let x be a point in E, f be a function in $C(X)$, and g be an integrable function. Then

$$\lim_{N\to\infty} S_N f(x) g(x) = g(x) \int f\, d\mu.$$

Since $S_N f(x) g(x) \leq \|f\|_u g(x)$, then applying the dominated convergence theorem implies the ergodicity condition of Theorem 2.1.21(c). $\square$

2.1.2 Typical Points

In this section we show certain points of a shift space X may be more "typical" than others, meaning that the set consisting of these points is the complement of a set which is negligible in a probabilistic sense.

Definition 2.1.27. *Let X be a shift space. A point x in X is* right transitive *if the right orbit of $x = \{x, T(x), T^2(x), \dots\}$ is dense in X. In other words for any cylinder set $C_k(U)$ of X there is n in $\mathbb{N}$ such that $T^n(x) \cap C_k(U)$ is nonempty. A point x in X is* transitive *if both the right orbit and the left orbit of x are dense in X.*

Definition 2.1.28. *A shift space X is irreducible if for every ordered pair of blocks U, V of X there exists a block W of X so that UWV is an allowable block of X.*

A directed graph G is irreducible (strongly connected) if for every ordered pair of vertices I and J there is a path in G starting at I and terminating at J. It is clear that G is irreducible if and only if the vertex shift X_M where M is the adjacency matrix of G is irreducible.

Example 2.1.29. *The full 2-shift and the golden mean shift are irreducible.*

If X has a transitive point, it is not hard to see that X is irreducible. Below in Theorem 2.1.31, we show that the converse is also true. To prove this we need the following theorem.

Theorem 2.1.30. (Baire Category Theorem) Let X be a shift space, and suppose that $A_1, A_2, \ldots$ are each open and dense in X. Then $\bigcap_{n=1}^{\infty} A_n$ is dense in X.

A proof of the Baire Category Theorem in the general case (when X is a complete metric space) may be found in [7].

Theorem 2.1.31. *Any irreducible shift space possesses a transitive point.*

Proof. Let X be an irreducible shift space. Recall that $\mathscr{L}(X)$ is a countable set containing cylinder sets of X. For a cylinder set $C_0(U)$, the set $A = \bigcup_{n\geq 0} T^{-n}(C_0(U))$ is open by continuity of T. This set is also dense in X. To see this, let B be an open nonempty set. Then B contains a cylinder set

$$C_k(V) = T^{-k}(C_0(V))$$

of X. Since X is irreducible there are blocks $W^{(i)}$ so that $VW^{(i)}U$ are allowable blocks of X and $\lim_{i\to\infty} |W^{(i)}| = \infty$. Let $n_i = |VW^{(i)}|$. Then

$$T^{n_i}(C_0(V)) \cap C_0(U) \neq \emptyset,$$

or equivalently

$$C_0(V) \cap T^{-n_i}(C_0(U)) \neq \emptyset.$$

It follows that $T^{-n_i-k}(C_0(U)) \cap B$ is nonempty which confirms the density of A by noting that for large enough i we have $n_i + k > 0$.

By the Baire Category Theorem, the set

$$D_1 = \bigcap_{C_0(U) \in \mathscr{L}(X)} \bigcup_{n \geq 0} T^{-n}(C_0(U)),$$

which is a G_δ-set (a countable intersection of open sets) is dense in X. The left orbit of any point $x \in D_1$ is dense. Since T is invertible we deduce that the G_δ-set

$$D_2 = \bigcap_{C_0(U) \in \mathscr{L}(X)} \bigcup_{n \geq 0} T^{n}(C_0(U))$$

is also dense in X. The right orbit of any point in D_2 is dense. Let $D = D_1 \cap D_2$. Then D which is the intersection of two G_δ-sets is dense in X and every point of D is transitive. □

Remark 2.1.32. *The proof of Theorem 2.1.31 implies that not only there is a transitive point in an irreducible shift space, but such points are dense.*

Theorem 2.1.34 shows that transitive points are actually typical points of a shift space.

Definition 2.1.33. *Let X be a shift space. A measure μ on X is called fully supported if it assigns positive measure to each nonempty open set of X.*

Theorem 2.1.34. *Let X be a shift space and μ be a fully supported ergodic measure on X. Then the set of transitive points has μ-measure 1.*

Proof. Let k be a finite positive integer and let B be a block of X. Define

$$M_k(B) = \{x \colon B \text{ does not occur in } x \text{ more than } k \text{ times}\}.$$

Then $M_k(B)$ is a T-invariant member of $\mathscr{B}(X)$. Since μ is ergodic we have $\mu(M_k(B)) = 0$ or $\mu(M_k(B)) = 1$. Moreover, we show $M_k(B)$ is a proper closed subset of X. Then from the assumption that μ is a fully supported measure it follows that the μ-measure of $M_k(B)$ must be zero.

To see $M_k(B)$ is closed, let x be a point in the X-complement of $M_k(B)$. Then B occurs in x at least $k+1$ times. Suppose $i_0, \ldots, i_k$ are some starting positions where B occurs in x. Let $|i_j| = \max\{|i_t| \colon 0 \leq t \leq k\}$. Then any point y with the distance less than $2^{-|i_j|-|B|}$ from x lies in the complement of $M_k(B)$. It follows that the complement of $M_k(B)$ is open, or equivalently, $M_k(B)$ is closed.

Let

$$E = \{x : \text{there is a block of } X \text{ occurring finitely many times in } x\}.$$

We have

$$E = \bigcup_{B} \bigcup_{k \in \mathbb{N}} M_k(B),$$

where B is a block of X. Since for each block B of X and each k in $\mathbb{N}$ we have $\mu\left(M_k(B)\right) = 0$ it follows that $\mu(E) = 0$.

The X-complement of E is a subset of the set of transitive points of X with μ-measure 1. It follows that the set of transitive points of X has μ-measure 1. $\square$

2.2 Entropy

Entropy is an important quantity which, in some sense, measures the complexity of a system. In this section we formally define entropy for shift spaces and compute it for a Bernoulli measure. In order to define entropy of a measure on a shift space we take three steps. First, the entropy of a partition is defined, then the entropy of a measure with respect to a partition is discussed for taking the final step which is defining the entropy of a measure. We present an important characteristic of entropy which results in an equivalent definition of it due to Katok [11].

Definition 2.2.1. *A finite partition of a shift space X is a finite pairwise disjoint* *collection of elements of $\mathscr{B}_X$ whose union is X. Let $\xi = \{A_1, \ldots, A_m\}$ and $\zeta = \{B_1, \ldots, B_r\}$ be two partitions of X. Their* **join** *is the partition*

$$\xi \vee \zeta = \{A_i \cap B_j \colon 1 \le i \le m,\, 1 \le j \le r\},$$

omitting the terms $A_i \cap B_j$ when $A_i \cap B_j = \emptyset$. If $n \in \mathbb{Z}$ then $T^{-n}\xi$ denotes the partition $\{T^{-n}A_1, \ldots, T^{-n}A_m\}$.

Example 2.2.2. *Let X be the golden mean shift and $\xi = \{{}_0[00]\,, {}_0[01]\,, {}_0[10]\}$ be a partition of X. Then $T^{-1}\xi = \{{}_1[00]\,, {}_1[01]\,, {}_1[10]\}$ and*

$$\xi \vee T^{-1}\xi = \{{}_0[000]\,, {}_0[001]\,, {}_0[010]\,, {}_0[100]\,, {}_0[101]\}.$$

As in probability theory we can consider a partition $\xi = \{A_1, \ldots, A_m\}$ as a list of the possible outcomes of an experiment, where the probability of the outcome A_i is $\mu(A_i)$. Then the entropy of the partition ξ, denoted by $H_\mu(\xi)$, is a number assigned to ξ which measures the knowledge gained by performing the experiment represented by ξ. Requiring $H_\mu(\xi)$ to satisfy a handful of reasonable properties may be shown to force $H_\mu(\xi)$ to be as in Definition 2.2.3. These properties can be found in [14].

Definition 2.2.3. *Let X be a shift space and μ be a measure on X. Let $\xi = \{A_1, \ldots, A_m\}$ be a finite partition of X. The entropy of ξ is the number*

$$H_\mu(\xi) = -\sum_{i=1}^{m} \mu(A_i) \log \mu(A_i),$$

with $0 \log 0$ defined to be 0.

Example 2.2.4. *Let μ be the Bernoulli-$(1/2, 1/2)$ measure on full 2-shift $X = \{0,1\}^{\mathbb{Z}}$. Let $\xi = \{{}_0[00]\,, {}_0[01]\,, {}_0[10]\,, {}_0[11]\}$ be a partition of X. Then*

$$H_\mu(\xi) = -\sum_{i=1}^{4} \frac{1}{4} \log \frac{1}{4} = \log 4.$$

Let $\zeta = \{{}_0[0]\,, {}_0[10]\,, {}_0[11]\}$ *be another partition of* X*. Then we have*

$$H_\mu(\zeta) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{4}\right) < H_\mu(\xi).$$

If $\bar{\mu}$ *is the Bernoulli-*$(1/3, 2/3)$ *measure on* X *then*

$$H_{\bar{\mu}}(\xi) = -\left(\frac{1}{9}\log\frac{1}{9} + \frac{4}{9}\log\frac{2}{9} + \frac{4}{9}\log\frac{4}{9}\right) = 2\log 3 - \frac{4}{3}\log 2 < H_\mu(\xi).$$

If we think of the shift T as an evolution by one unit of time, then $\bigvee_{i=0}^{n-1} T^{-i}\xi$ represents the combined experiment of performing the original experiment ξ on n consecutive units of time. Therefore $\frac{1}{n}H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right)$ measures the average information per unit of time that we get from performing the experiment ξ on n consecutive units of time.

Definition 2.2.5. *Let* μ *be an invariant measure on a shift space* X*. If* ξ *is a finite partition of* X *then*

$$h_\mu(T, \xi) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right)$$

is called the **entropy of μ with respect to ξ***. (See [29, Theorem 4.9.1] for a justification that the above limit always exists.)*

Now by taking the supremum of $h_\mu(T, \xi)$ over all finite partitions of X we get the entropy of the measure μ, denoted by $h_\mu(T)$.

Definition 2.2.6. *Let* μ *be an invariant measure on a shift space* X*. Then*

$$h_\mu(T) = \sup\{h_\mu(T, \xi) \colon \xi \text{ is a finite partition of } X\}.$$

The definition of entropy is more readily computed than it seems. The Kolmogorov-Sinai theorem shows that in many cases $h_\mu(T) = h_\mu(T, \xi)$ for an appropriately chosen ξ.

Before stating the Kolmogorov-Sinai theorem, we need to know the following facts about partitions and σ-algebras. If ξ is a finite partition of X then the collection of all elements of $\mathscr{B}_X$ which are unions of elements of ξ is a finite sub-σ-algebra of $\mathscr{B}_X$. We denote it by $\sigma(\xi)$ and call it the σ-algebra generated by partition ξ. Conversely if $\mathcal{C} = \{C_1, \ldots, C_m\}$ is a finite sub-σ-algebra of X, then the non-empty sets of the form $B_1 \cap \cdots \cap B_m$, where $B_i = C_i$ or $X - C_i$, form a finite partition of X. We denote it by $\mathcal{P}(\mathcal{C})$. We have $\sigma(\mathcal{P}(\mathcal{C})) = \mathcal{C}$ and $\mathcal{P}(\sigma(\xi)) = \xi$ which gives a one-to-one correspondence between finite partitions and finite sub-σ-algebras of $\mathscr{B}_X$.

If $(\mathcal{A}_n)_n$ is a sequence of σ-algebras of $\mathscr{B}_X$ then $\bigvee_n \mathcal{A}_n$ denotes the sub-σ-algebra of $\mathscr{B}_X$ generated by $(\mathcal{A}_n)_n$; i.e., $\bigvee_n \mathcal{A}_n$ is the intersection of all sub-σ-algebras of $\mathscr{B}_X$ that contain every $\mathcal{A}_n$. A countable partition ξ of X is called a generator of $\mathscr{B}_X$ if

$$\bigvee_{n=-\infty}^{\infty} T^n \sigma(\xi) = \mathscr{B}_X.$$

Theorem 2.2.7. (Kolmogorov-Sinai Theorem) Let ξ be a finite partition of a shift space X. Suppose ξ is a generator of the σ-algebra $\mathscr{B}_X$. Then

$$h(T) = h(T, \xi).$$

For a proof of the Kolmogorov-Sinai Theorem see [29, Theorem 4.17].

Example 2.2.8. *Let μ be the Bernoulli-(p_0, p_1) measure on the full 2-shift $\{0,1\}^{\mathbb{Z}}$. Then $h_\mu(T) = -(p_0 \log p_0 + p_1 \log p_1)$.*

Proof. Let ξ be the partition $\{{}_0[0]\,, {}_0[1]\}$ of X. By the definition of σ-algebra $\mathscr{B}_X$, we have

$$\bigvee_{i=-\infty}^{\infty} T^i \sigma(\xi) = \mathscr{B}_X.$$

By the Kolmogorov-Sinai Theorem,

$$h_\mu(T) = h_\mu(T, \xi) = \lim_{n\to\infty} \frac{1}{n} H(\xi \vee \cdots \vee T^{-(n-1)}\xi).$$

Note that $\xi \vee \cdots \vee T^{-(n-1)}\xi = \{{}_0[i_0, i_1, \ldots, i_{n-1}] : i_j \in \{0,1\}\}$ where

$$\mu({}_0[i_0, i_1, \ldots, i_{n-1}]) = p_{i_0} p_{i_1} \ldots p_{i_{n-1}}.$$

It follows that

$$\begin{aligned} &H(\xi \vee \cdots \vee T^{-(n-1)}\xi) \\ &\quad = -\sum p_{i_0} \ldots p_{i_{n-1}} \log(p_{i_0} \ldots p_{i_{n-1}}) \\ &\quad = -\sum_{i_0,\ldots,i_{n-1}=0}^{1} p_{i_0} \ldots p_{i_{n-1}} (\log p_{i_0} + \cdots + \log p_{i_{n-1}}) \\ &\quad = -n \sum_{i_0,\ldots,i_{n-1}=0}^{1} p_{i_0} \ldots p_{i_{n-1}} \log p_{i_0} \\ &\quad = -n(p_0 \log p_0 + p_1 \log p_1). \end{aligned}$$

Therefore, $h_\mu(T) = -(p_0 \log p_0 + p_1 \log p_1)$. $\square$

Example 2.2.9. *Let μ be the Bernoulli-$(1/2, 1/2)$ measure on the full 2-shift. Then $h_\mu(T) = \log 2$. The Bernoulli-$(1/3, 2/3)$ measure on the full 2-shift has entropy $\log 3 - \frac{2}{3}\log 2 < h_\mu(T)$.*

Definition 2.2.10. *Let X be a shift space. A measure $\mu \in M(X, T)$ is called a measure of maximal entropy if*

$$h_\mu(T) = \sup\{h_\rho(T) \colon \rho \in M(X, T)\}.$$

There is a well-known result of Parry which shows that every irreducible shift of finite type has a unique measure of maximal entropy.

Theorem 2.2.11. *[29, Theorem 8.10] Every irreducible shift of finite type has a unique measure of maximal entropy. Moreover, such a measure is ergodic.*

It is worth mentioning that there is an explicit formula for computing the measure of maximal entropy on an irreducible SFT. A full description of such measure may be found in [29, page 194].

Example 2.2.12. *[29, Theorem 8.9] The Bernoulli-*$(1/2, 1/2)$ *measure is the unique measure of maximal entropy on the full* 2*-shift.*

One of the important theorems about entropy is the Shannon-McMillan-Breiman theorem stated below.

Theorem 2.2.13. (Shannon-McMillan-Breiman Theorem) Let μ be an ergodic measure on a shift space X. Then

$$\frac{1}{n}\log \mu(x_{[1,n]}) \to -h_{\mu}(T)$$

for μ-almost every x in X.

A proof of the Shannon-McMillan-Breiman theorem may be found in [22].

Here we discuss the application of this theorem in counting the number of "typical" cylinder sets of X. Since almost everywhere convergence in Theorem 2.2.13 implies convergence in measure, given $\epsilon > 0$, there is $N(\epsilon) > 0$ such that if $n \geq N(\epsilon)$, then

$$\mu(\{x \colon |\frac{1}{n}\log \mu(x_{[1,n]}) + h_{\mu}(T)| < \epsilon\}) > 1 - \epsilon,$$

or equivalently

$$\mu(\{x \colon e^{-n(h_{\mu}(T)+\epsilon)} < \mu(x_{[1,n]}) < e^{-n(h_{\mu}(T)-\epsilon)}\}) > 1 - \epsilon.$$

It follows that, given $\epsilon > 0$ there is an integer $N(\epsilon)$ such that for any $n > N(\epsilon)$ one can separate the set of n-cylinder sets of X into two disjoint subsets $\mathscr{B}_n$ and $\mathscr{G}_n$, where

$$\mu\left(\bigcup_{B \in \mathscr{B}_n} B\right) \leq \epsilon,$$

and for any $B \in \mathscr{G}_n$

$$e^{-n(h_{\mu}(T)+\epsilon)} < \mu(x_{[1,n]}) < e^{-n(h_{\mu}(T)-\epsilon)}.$$

For any n-cylinder set B in $\mathscr{G}_n$ we have

$$1/n \log \mu(B) = -h_\mu(T) \pm \epsilon.$$

Such cylinder sets are called "good" cylinder sets. The remaining cylinder sets (those in $\mathscr{B}_n$) are called "bad" cylinders sets, but their total measure is small. An immediate consequence is that $\mathscr{G}_n$ has at least $e^{nh_\mu(T)}$ elements. Using this idea, an alternative description of entropy is given in the following theorem.

Theorem 2.2.14. *Let μ be an ergodic measure on a shift space X with $h_\mu(T) > 0$. For a given $\epsilon \in (0,1)$ let $r_n(\epsilon, \mu)$ be the minimum number of n-cylinder sets ${}_1[i_1, \ldots, i_n]$ which their union has μ-measure more than or equal to $1 - \epsilon$. Then*

$$h_\mu(T) = \lim_{n\to\infty} \frac{1}{n} \log r_n(\epsilon, \mu).$$

A detailed proof of Theorem 2.2.14 in a more general case (X a compact metric space), due to A. B. Katok, can be found in [11].

Recall that the entropy of the Bernoulli $(1/3, 2/3)$-measure is $-1/3 \log 1/3 - 2/3 \log 2/3$ (Example 2.2.8). We re-compute the entropy of this measure by counting the number of typical cylinder sets in $\{0,1\}^{\mathbb{Z}}$ and using Theorem 2.2.14.

Example 2.2.15. *Let μ be the Bernoulli-$(1/3,\, 2/3)$ measure on the full 2-shift. Then $h_\mu(T) = -1/3 \log 1/3 - 2/3 \log 2/3$.*

Proof. Let $\epsilon = 1/2$. We use Theorem 2.2.14 to show that

$$\lim_{n\to\infty} \frac{1}{n} \log r_n(1/2, \mu) = -1/3 \log 1/3 - 2/3 \log 2/3.$$

Let D be the set of all n-cylinder sets in X starting at position 1. Then D is the union of n pairwise disjoint subsets D_i where D_i contains cylinder sets of D in which 1 occurs exactly i times within coordinates 1 and n. It follows that

$$|D_i| = \binom{n}{i},$$

and each cylinder set C of D_i has measure

$$\mu(C) = (2/3)^i (1/3)^{n-i}.$$

For convenience, we introduce the following notation. When $A = \{A_1, \ldots, A_j\}$ is a set of pairwise disjoint cylinder sets, we denote $\mu^*(A)$ to be

$$\mu^*(A) = \mu\left(\bigcup_{A_i \in A} A_i\right) = \sum_{A_i \in A} \mu(A_i).$$

Let A be a set containing the minimum number of n-cylinder sets of D such that $\mu^*(A) \geq 1/2$; i.e., $|A| = r_n(1/2, \mu)$. Let $k = \dfrac{2n - 2\sqrt{n}}{3}$, and let $E = \bigcup_{i=k}^{n} D_i$. First we show that

$$\mu^*(E) \geq 1/2.$$

Then since for each C in E and C' in $D - E$ we have $\mu(C) \geq \mu(C')$ it follows that A is a subset of E.

Claim. *Using above notations, we have* $\mu^*(E) \geq 1/2$.

Proof. Define the collection $(f_i)_{i=1}^{\infty}$ of functions $f_i : X \to \{0, 1\}$ where $f_i(x) = x_i$. For a given $n \in \mathbb{N}$ form the function $S_n = f_1 + \cdots + f_n$. Then $S_n(x)$ counts the number of 1's occurring in the n-block $x_1, \ldots, x_n$. We show

$$\mu\left(\bigcup_{i=0}^{k} D_i\right) = \mu\left(\{x \colon S_n(x) \leq \frac{2n - 2\sqrt{n}}{3}\}\right) \leq 1/2, \tag{2.2.1}$$

which directly leads to the result $\mu^*(E) \geq 1/2$.

To show Equation (2.2.1) note that

$$\begin{aligned}
\mu\left(\{x\colon |S_n(x) - 2n/3| \geq 2\sqrt{n}/3\}\right) &= \mu\left(\{x\colon (S_n(x) - 2n/3)^2 \geq 4n/9\}\right) \\
&\leq \frac{\int_X \left(S_n - 2n/3\right)^2 \, d\mu}{4n/9} \qquad (2.2.2) \\
&= \frac{9}{4n}\int_X (S_n)^2 \, d\mu + n - 3\int_X S_n \, d\mu.
\end{aligned}$$

For each i we have

$$\int_X f_i \, d\mu = \int_X \mathbf{1}_{\,i[1]} \, d\mu = \mu([1]) = 2/3,$$

and for each $j \neq i$

$$\int_X f_i f_j \, d\mu = \int_X f_i \, d\mu \int_X f_j \, d\mu.$$

Moreover, $\displaystyle\int_X S_n \, d\mu = 2n/3$. It follows that

$$\begin{aligned}
\int_X (S_n)^2 \, d\mu &= \int_X \left(f_1 + \cdots + f_n\right)^2 \, d\mu \\
&= n\int_X (f_1)^2 \, d\mu + 2\binom{n}{2}\int_X f_1 f_2 \, d\mu \\
&= 2n/3 + n(n-1)4/9 \\
&= 4n^2/9 + 2n/9.
\end{aligned}$$

Recall Equation(2.2.2) to obtain

$$\mu\left(\{x\colon |S_n(x) - 2n/3| \geq 2\sqrt{n}/3\}\right) \leq 1/2. \qquad (2.2.3)$$

Consequently we have

$$\mu\left(\{x\colon S_n(x) \leq \frac{2n - 2\sqrt{n}}{3}\}\right) \leq 1/2.$$

$\square$

Proof of Example 2.2.15 (continued). Let $r = \frac{2n+2\sqrt{n}}{3}$. Note that Equation (2.2.3), above, implies

$$\mu\left(\{x\colon S_n(x) \geq r\}\right) \leq 1/2,$$

which means

$$\sum_{i=r}^{n} \mu^*(D_i) \leq 1/2.$$

Since $|D_r| = \binom{n}{r}$, it follows that

$$|A| \geq \binom{n}{r}. \tag{2.2.4}$$

Recall that A is a subset of E. Moreover,

$$|E| = \sum_{i=k}^{n} \binom{n}{i}.$$

Since $\displaystyle\binom{n}{i} > \binom{n}{i+1}$ for every $i \geq \dfrac{n}{2}$, it follows that

$$|A| \leq |E| \leq n\binom{n}{k}.$$

From this and Equation (2.2.4) we conclude that

$$\binom{n}{r} \leq |A| \leq n\binom{n}{k}.$$

Therefore

$$\lim_{n\to\infty} \frac{1}{n} \log \binom{n}{r} \leq \lim_{n\to\infty} \frac{1}{n} \log |A| \leq \lim_{n\to\infty} \frac{1}{n} \log n\binom{n}{k}. \tag{2.2.5}$$

Note that from Stirling's approximation,

$$\log m! = m \log m - m + o(m),$$

for any integer s in $(n/4, 3n/4)$ we have

$$\begin{aligned}
\log\binom{n}{s} &= \log n! - \log s! - \log(n-s)! \\
&= n\log n - n - s\log s + s - (n-s)\log(n-s) + (n-s) + o(n) \\
&= n\log n - s\left(\log n + \log\left(\frac{s}{n}\right)\right) - (n-s)\left(\log n + \log\left(\frac{n-s}{n}\right)\right) \\
&+ o(n) \\
&= -s\log(\frac{s}{n}) - (n-s)\log(\frac{n-s}{n}) + o(n).
\end{aligned}$$

It follows that

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\log\binom{n}{r} &= -\lim_{n\to\infty}\frac{2n+2\sqrt{n}}{3n}\log\frac{2n+2\sqrt{n}}{3n} \\
&- \lim_{n\to\infty}\frac{n-2\sqrt{n}}{3n}\log\frac{n-2\sqrt{n}}{3n} \\
&= -2/3\log 2/3 - 1/3\log 1/3.
\end{aligned}$$

Similarly we have

$$\lim_{n\to\infty}\frac{1}{n}\log\binom{n}{k} = -2/3\log 2/3 - 1/3\log 1/3.$$

Thus, by Equation (2.2.5), we have

$$\lim_{n\to\infty}\frac{1}{n}\log|A| = -2/3\log 2/3 - 1/3\log 1/3.$$

Then theorem 2.2.14 implies that

$$h_\mu(T) = -1/3\log 1/3 - 2/3\log 2/3.$$

$\square$

2.3 Sofic Shifts

In this section we present a method to construct new shift spaces from old ones by a special class of mapping called sliding block codes. It is worth mentioning that sofic shifts and regular languages in automata theory are closely related. See [9] to read on automata theory and for more details on the connection between sofic shifts and regular languages see [1].

Recall that a n-block is a block of length n. The set of all n-blocks of a shift space X is denoted by $\mathcal{B}_n(X)$.

Definition 2.3.1. *Let X be a shift space and $\mathfrak{A}$ be an alphabet. For fixed integers m and n with $-m \leq n$ a map $\pi_b : \mathcal{B}_{m+n+1}(X) \to \mathfrak{A}$ is called a **block map**. The map $\pi : X \to \mathfrak{A}^{\mathbb{Z}}$ defined by $\pi(x)_i = \pi_b(x_{[i-m,i+n]})$, and denoted by $|\mathsf{m+n+1}|$-**block code**, is called the **sliding block code** with memory m and anticipation n induced by π_b.*

Example 2.3.2. *Let X be the full 2-shift $\{0,1\}^2$. Define the block map $\pi_b : \mathcal{B}_2(X) \to \{0,1\}$ by $\pi_b([x_i x_{i+1}]) = x_i + x_{i+1}$ mod 2. Then the sliding block code $\pi : X \to \{0,1\}^2$ where $\pi(x)_i = x_i + x_{i+1}$ mod 2 is a 2-block code with memory 0 and anticipation 1. Figure 2.2 illustrates the action of π on the point $\ldots 01\overset{*}{1}00\ldots$.*

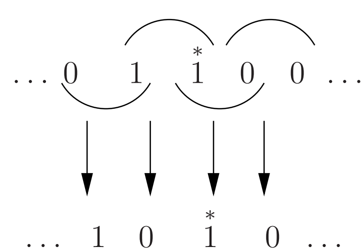

Figure 2.2: $\ldots 01\overset{*}{1}00\ldots$ projects to $\ldots 10\overset{*}{1}0\ldots$ under the sliding block code defined in Example 2.3.2

The simplest sliding block codes are those with no memory or anticipation; i.e., with $m = n = 0$. Such sliding block codes are called 1-block codes. Using standard tools which we shall describe later, it will be sufficient to establish most of theorems in this work in the case of 1-block codes. When π is a 1-block code, then π_b is acting on the alphabet of X. We can naturally extend this map to act on any block of X. For our purpose in this work, whenever $\pi : X \to Y$ is a 1-block code then the block code π_b is considered to be a function from the set of blocks of X to the set of blocks of Y.

Proposition 2.3.3. *[19, Proposition 1.5.8] Let X and Y be shift spaces. A map $\pi : X \to Y$ is a sliding block code if and only if $\pi \circ T_X = T_Y \circ \pi$ and there exists $N \geq 0$ such that $\pi(x)_0$ is a function of $x_{[-N,N]}$.*

A **homomorphism** from a shift space X to a shift space Y is a continuous function $\phi : X \to Y$ satisfying $T_Y \circ \phi = \phi \circ T_X$ (the shift-commuting property). In other words ϕ is a continuous function so that the diagram in Figure 2.3 commutes. The

$$\begin{tikzcd} X \arrow[r, "T_X"] \arrow[d, "\phi"'] & X \arrow[d, "\phi"] \\ Y \arrow[r, "T_Y"] & Y \end{tikzcd}$$

Figure 2.3: ϕ is a continuous shift-commuting map

Curtis-Lyndon-Hedlund Theorem characterizes the homomorphisms from one shift space to another as being exactly the sliding block codes.

Theorem 2.3.4. (Curtis-Lyndon-Hedlund Theorem) *Let X and Y be shift spaces. A map $\pi : X \to Y$ is a sliding block code if and only if it is a homomorphism.*

A proof of the Curtis-Lyndon-Hedlund Theorem may be found in [19, Theorem 6.2.9].

Definition 2.3.5. *A sliding block code $\pi : X \to Y$ is a **conjugacy** from X to Y, if it is invertible. Two shift spaces X and Y are **conjugate**, written by $X \cong Y$, if there is a conjugacy from X to Y.*

Let X be a shift space over the alphabet $\mathcal{A}$. Let $\mathcal{A}_X^{[N]} = \mathcal{B}_N(X)$ be the collection of all allowed N-blocks of X. Form the full shift over the new alphabet $\mathcal{A}_X^{[N]}$ and define the Nth higher block code $\gamma_N : X \to \left(\mathcal{A}_X^{[N]}\right)^{\mathbb{Z}}$ by $(\gamma_N(x))_i = x_{[i,i+N-1]}$. This code becomes more clear if we imagine the symbols in $\mathcal{A}_X^{[N]}$ as written vertically. Then, for example, the image of $x = (x_i)_{i\in\mathbb{Z}}$ under γ_2 is

$$\gamma_2(x) = \ldots \begin{bmatrix} x_{-1} \\ x_{-2} \end{bmatrix} \begin{bmatrix} x_0 \\ x_{-1} \end{bmatrix} . \begin{bmatrix} x_1 \\ x_0 \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} \begin{bmatrix} x_3 \\ x_2 \end{bmatrix} \cdots \in \left(\mathcal{A}_X^{[2]}\right)^{\mathbb{Z}} \tag{2.3.1}$$

Definition 2.3.6. *Let X be a shift space. The Nth higher block presentation of X, denoted by $X^{[N]}$, is the image of X under γ_N in the full shift over $\mathcal{A}_X^{[N]}$.*

Notice that every point in $X^{[2]}$ has to be of the form shown in (2.3.1). If $u = u_1u_2$ and $v = v_1v_2$ are 2-blocks of X, then uv is a block of $X^{[2]}$ if and only if $u_2 = v_1$.

Proposition 2.3.7. *The higher block presentations of a shift space X are shift spaces conjugate to X.*

A proof of this proposition may be found in [19]. In particular, the reader is referred to Proposition 1.4.3 and Example 1.5.10 of [19].

Recall that a SFT is a shift space with a finite set of forbidden blocks, and we call it a N-step SFT if it is a SFT of step size at most N but not of step size at most $N-1$. Any N-step SFT is conjugate to a 1-step SFT according to the following result.

Proposition 2.3.8. *[19, Proposition 2.3.9] If X is a N-step SFT, then $X^{[N]}$ is a 1-step SFT, equivalently a vertex shift.*

Example 2.3.9. *Let $\mathcal{A} = \{0, 1\}$ and $\mathcal{F} = \{00, 111\}$. The SFT $X_{\mathcal{F}}$ is a 2-step SFT which can not be described by a directed graph. Below, we construct $X^{[2]}$ which is a 1-step SFT conjugate to X and furthermore, it can be described by a directed graph. Form the collection of all allowed 2-blocks of X.*

$$\mathcal{A}_X^{[2]} = \{a = 01,\, b = 11,\, c = 10\}.$$

Then $X^{[2]}$ *is described by the set of forbidden blocks* $\{bb, aa, ba, cc, cb\}$*. Note that the block* bb *is a forbidden block of* $X^{[2]}$ *since* 111 *is a forbidden block of the original shift* X*. The rest of the forbidden blocks of* $X^{[2]}$ *come from the construction of a 2nd-higher block presentation as they fail to overlap progressively. For example, the second symbol of* $a = 01$ *does not match the first symbol of* $a = 01$*, so* aa *is forbidden. Figure 2.4 represents the vertex shift* $X^{[2]}$*. Note that if* $u = u_1u_2$ *and* $v = v_1v_2$ *are* 2*-blocks of* X *then* $uv \in X^{[2]}$ *if and only if* $u_2 = v_1$*.*

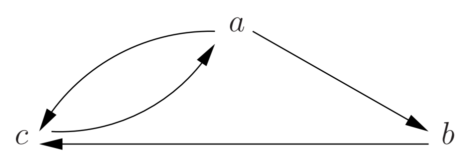

Figure 2.4: Graph for Example 2.3.9

Given a sliding block code $\pi : X \to Y$, next proposition shows that there is a shift space $\tilde{X}$ conjugate to X so that the corresponding sliding block code $\tilde{\pi} : \tilde{X} \to Y$ is a 1-block code.

Proposition 2.3.10. *[19, Proposition 1.5.12] Let* $\pi : X \to Y$ *be a sliding block code. There exists a higher block shift space* $\tilde{X}$ *conjugate to* X*, a conjugacy* $\psi : X \to \tilde{X}$*, and a 1-block code* $\tilde{\pi} : \tilde{X} \to Y$ *so that* $\tilde{\pi} \circ \psi = \pi$*.*

Combining Proposition 2.3.8 and Proposition 2.3.10 imply the following corollary.

Corollary 2.3.11. *Let* $\pi : X \to Y$ *be a sliding block code from a SFT* X *to a shift space* Y*. There exists a 1-step SFT* $\tilde{X}$ *conjugate to* X *under a conjugacy* $\psi : X \to \tilde{X}$*, and a 1-block code* $\tilde{\pi} : \tilde{X} \to Y$ *so that* $\tilde{\pi} \circ \psi = \pi$*.*

The process in Corollary 2.3.11, called “recoding”, is often starting point in proofs, since 1-step SFTs and 1-block codes are much easier to work with.

Images of shifts of finite type under sliding block codes form an important class of shift spaces, called sofic shifts.

Definition 2.3.12. *A subset Y of a full shift is called a* **sofic shift** *if there exists a SFT X and a block code $\pi : X \to Y$ such that $Y = \pi(X)$. In this case $\pi : X \to Y$ is called a* **factor code** *and the triple (X, Y, π) is called a* **factor triple**.

When $\pi : X \to Y$ is a 1-block factor code from a 1-step SFT X to a sofic shift Y, then Y can be presented by a finite directed graph where several vertices may carry the same label.

Definition 2.3.13. *Let G be a directed graph with vertex set $\mathcal{V}$. Let $\mathcal{A}$ be an alphabet. A* **labeling** *of G is a function $\mathcal{L} : \mathcal{V} \to \mathcal{A}$. For each vertex $v \in \mathcal{V}$ the element $\mathcal{L}(v)$ is called the* **label** *of v. A* **labeled graph** *$\mathcal{G}$ is a pair $(G, \mathcal{L})$, where G is a directed graph, and $\mathcal{L}$ is a labeling of G.*

If $\mathcal{G} = (G, \mathcal{L})$ is a labeled graph, then the labeling $\mathcal{L} : \mathcal{V} \to \mathcal{A}$ can be extended to a map from X_G to $\mathcal{A}^{\mathbb{Z}}$. The set of all bi-infinite walks on $\mathcal{G}$ is denoted by $Y_{\mathcal{G}} = \{y \in \mathcal{A}^{\mathbb{Z}} : y = \mathcal{L}(x) \text{ for some } x \in X_G\}$. A **graph presentation** of a sofic shift Y is a labeled graph $\mathcal{G}$ for which $Y_{\mathcal{G}} = Y$.

Example 2.3.14. *The labeled graph shown in Figure 2.5 represents a factor code π from the golden mean shift to the sofic shift $\{0\}^{\mathbb{Z}}$ where each vertex maps to 0.*

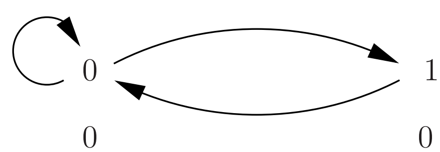

Figure 2.5: Graph for Example 2.3.14

The following theorem shows that entropy is a conjugacy invariant. However, it is worth mentioning that entropy is not a complete invariant, meaning that two spaces which are not conjugate may have equal entropy. For an example, the reader is referred to see [29, page 104].

Theorem 2.3.15. *[29, Theorem 4.11] Let X be a shift space conjugate to Y and μ be an invariant measure on X. Then*

$$h_{\mu}(T_X) = h_{\mu \circ \pi^{-1}}(T_Y).$$

The following Proposition shows that when $\pi : X \to Y$ is a factor code then the entropy of the SFT can not be less than the entropy of the sofic shift.

Proposition 2.3.16. *Given a factor triple (X, Y, π) and an invariant measure μ on X, we have*

$$h_\mu(T_X) \geq h_{\mu\circ\pi^{-1}}(T_Y).$$

For a proof of Proposition 2.3.16, see [29, page 89].

Definition 2.3.17. *A factor code $\pi : X \to Y$ is called* finite-to-one *if for every y in Y, the preimage $\pi^{-1}(y)$ is finite. π is called* bounded-to-one *if there is an integer M such for every y in Y, $\pi^{-1}(y)$ contains at most M elements .*

We need to develop the following definition and theorems to prove Theorem 2.3.20, below, which shows that a finite-to-one factor code preserves entropy.

Definition 2.3.18. *Let $\pi : X \to Y$ be a factor code from a shift space X to a sofic shift Y. A* diamond *for π is a pair of distinct points in X differing in only finitely many coordinates, with the same image under π.*

Theorem 2.3.19. *[19, Theorem 8.1.16] Let $\pi : X \to Y$ be a factor code from an irreducible SFT X to a sofic shift Y. Then the following are equivalent.*

(a) π is finite-to-one.

(b) π is bounded-to-one.

(c) π has no diamonds.

Theorem 2.3.20. *Let $\pi : X \to Y$ be a finite-to-one factor code from an irreducible SFT X to a sofic shift Y. Let μ be an invariant measure on X. Then*

$$h_\mu(T_X) = h_{\mu\circ\pi^{-1}}(T_Y).$$

Proof. Denote the measure $\mu \circ \pi^{-1}$ on Y by ν. Proposition 2.3.16 shows that

$$h_\mu(T_X) \geq h_\nu(T_Y).$$

Here we show $h_\mu(T_X) \leq h_\nu(T_Y)$, by using the alternative definition of entropy given in Theorem 2.2.14:

$$h_\nu(T_Y) = \lim_{n\to\infty} \frac{1}{n} \log r_n(\epsilon, \nu),$$

where $r_n(\epsilon, \nu)$ is the minimum number of n-cylinder sets ${}_1[i_1, \ldots, i_n]$ in Y which their union has ν-measure more than or equal to $1 - \epsilon$ for each $\epsilon > 0$.

By theorem 2.3.19, a finite-to-one factor code does not have any diamonds. Thus, for any a and a' in the alphabet of X, and each block W of Y there is at most one block of X which maps to W under π, begins at a, and ends at a'. Let

$$B = \{B_1, \ldots, B_{r_n(\epsilon,\nu)}\}$$

be a set of n-cylinder sets of Y whose union has ν-measure more than or equal to $1 - \epsilon$. Consider the set A containing the preimages of each B_i in B. Then we have

$$|A| \leq |\mathcal{A}(X)|^2 r_n(\epsilon, \nu). \tag{2.3.2}$$

Moreover, A is a set of some m-cylinder sets in X, for some $m \geq n$ (if π is a 1-block code then $m = n$), which their union has μ-measure more than or equal to $1 - \epsilon$. It follows that

$$r_n(\epsilon, \mu) \leq r_m(\epsilon, \mu) \leq |A|.$$

From this and Equation (2.3.2) we conclude that

$$\begin{aligned} h_\mu(T_X) &= \lim_{n\to\infty} \frac{1}{n} \log r_n(\epsilon, \mu) \\ &\leq \lim_{n\to\infty} \frac{1}{n} \log r_n(\epsilon, \nu) \\ &= h_\nu(T_Y). \end{aligned}$$

□

Definition 2.3.21. *Let (X, Y, π) be a factor triple. Let ν be an invariant measure on Y. The* fibre *of ν, denoted by $\pi^{-1}\{\nu\}$, is the set of all invariant measures μ on X such that for each $B \in \mathscr{B}_Y$ we have $\nu(B) = \mu \circ \pi^{-1}(B)$.*

Given an ergodic measure ν on Y there can be many measures in the fibre of ν. For example given $\pi : \{0,1\}^{\mathbb{Z}} \to \{0\}^{\mathbb{Z}}$ and the trivial measure ν on $\{0\}^{\mathbb{Z}}$, any Bernoulli-$(p, 1-p)$ measure on the full-2 shift is in the fibre $\pi^{-1}\{\nu\}$. The following result shows that the fibre of any fully supported ergodic measure is non-empty.

Proposition 2.3.22. *Let (X, Y, π) be a factor triple and ν be an ergodic measure on Y. Then $\pi^{-1}\{\nu\}$ contains at least one ergodic measure.*

Proof. Let $y \in Y$ be a generic point of ν; i.e., $\lim_{N\to\infty} S_N f(y) = \int f \, d\nu$ for all f in $C(Y)$. Let N in $\mathbb{N}$, x in $\pi^{-1}(y)$, and

$$\mu_N = 1/N \sum_{i=0}^{N-1} \delta_x \circ T_X^{-i},$$

where δ_x stands for the point mass measure at x. Then μ_N is a measure on X. Moreover,

$$\mu_N \circ \pi^{-1} = 1/N \sum_{i=0}^{N-1} \delta_y \circ T_Y^{-i},$$

is a measure on Y. Choose a subsequence (N_j) so that (μ_{N_j}) converges (such a subsequence exists by the compactness of $M(X)$). Let $\mu_{N_j} \to \mu$. Measure μ is T_X-invariant by Theorem 2.1.17. Moreover, for each cylinder set C of Y we have

$$\begin{aligned}
\mu \circ \pi^{-1}(C) &= \lim_{N_j\to\infty} 1/N_j \sum_{i=0}^{N_j-1} \delta_y \circ T_Y^{-i}(C) \\
&= \lim_{N_j\to\infty} 1/N_j \sum_{i=0}^{N_j-1} \mathbf{1}_C(T_Y^i(y)) \\
&= \nu(C).
\end{aligned}$$

Theorem 2.1.11 implies that $\mu \circ \pi^{-1}(B) = \nu(B)$ for each B in $\mathscr{B}_Y$, or equivalently $\mu \in \pi^{-1}\{\nu\}$. We now use Theorem 2.1.20 to show there is such a measure which is also ergodic. Suppose the measure μ, obtained above, is not ergodic. Let

$\mu = \displaystyle\int_{E(X,T)} \rho \, d\tau(\rho)$ be the ergodic decomposition of μ. Then

$$\nu = \pi\mu = \int_{E(X,T)} \pi\rho \, d\tau(\rho).$$

Since ν is ergodic, $\pi\rho = \nu$ for τ-almost every measure ρ in $E(X,T)$, which implies $\pi^{-1}\{\nu\}$ contains at least one ergodic measure. $\square$

Definition 2.3.23. *Let (X, Y, π) be a factor triple and ν be an ergodic measure on Y. An ergodic measure on X that projects under π to ν and has maximal entropy in the fibre $\pi^{-1}\{\nu\}$ is called a measure of relative maximal entropy.*

The following example shows that there can be more than one such ergodic relatively maximal measure over a given ν.

Example 2.3.24. *Let $X = Y$ be the full 2-shift $\{0,1\}^{\mathbb{Z}}$, and π be a 2-block factor code defined by $\pi(x)_0 = x_0 + x_1$ mod 2. Let $p \neq 1/2$. Let μ_1 be the Bernoulli-$(p, 1-p)$ measure and μ_2 be the Bernoulli-$(1-p, p)$ measure on X. Both measures μ_1 and μ_2 map to the same measure ν on Y. Moreover, measure ν inherits the ergodicity and being fully supported from μ_i. Since above every point of $y \in Y$ there are only finitely many (two) points in X which are mapped to y by π we have $h_\nu(T_Y) = h_{\mu_1}(T_X) = h_{\mu_2}(T_X)$ by Theorem 2.3.20.*

Chapter 3

Uniform Conditional Distribution

It is a well-known result of Parry (Theorem 2.2.11), generalizing an earlier result of Shannon [27], that in one dimension every irreducible shift of finite type has a unique measure of maximal entropy. Burton and Steif [6] give a counterexample to this statement in higher dimensions. However, they show that such measures all have the uniform conditional distribution property stated in Theorem 3.0.25 (which is related to an earlier work of Lanford and Ruelle [16]). Given a finite set $G \subseteq \mathbb{Z}^d$, the boundary of the complement of G is $\partial G^c = \{i \in G^c \colon \exists j \in G \text{ with } \|i - j\| = 1\}$.

Theorem 3.0.25. *[6, Proposition 1.19] Let μ be a measure of maximal entropy for a SFT in d dimensions. Then the conditional distribution of μ on any finite set $G \subseteq \mathbb{Z}^d$ given the configuration on G^c is μ-a.s. uniform over all configurations on G which extend the configuration on ∂G^c.*

At the end of Chapter 2 we showed that given a factor triple (X, Y, π) and an ergodic measure ν on Y, there can exist more than one ergodic measure of relative maximal entropy; i.e., there can be more than one ergodic measure on X which projects to ν and has maximal entropy among all measures in the fibre $\pi^{-1}\{\nu\}$. In this chapter we show the uniform conditional distribution property for measures of relative maximal entropy. In the first section we define conditional entropy and develop some of its properties which are used later in the chapter. In Section 3.2 we follow techniques developed by Burton and Steif in the proof of Theorem 3.0.25 to show the uniform conditional distribution property for measures of relative maximal entropy.

3.1 Conditional Entropy

Let μ be an invariant measure on a shift space X. Let $\xi = \{A_1, \ldots, A_m\}$ and $\zeta = \{B_1, \ldots, B_n\}$ be two partitions of X. If we think of ξ and ζ as lists of possible outcomes of two experiments, then the entropy of ξ given ζ measures the information gained by performing the experiment ξ given that we will be told the outcome of ζ.

Definition 3.1.1. *Let μ be a measure on a shift space X. Let $\xi = \{A_1, \ldots, A_m\}$ and $\zeta = \{B_1, \ldots, B_n\}$ be two partitions of X. The* entropy of ξ given ζ *is the number*

$$H(\xi|\zeta) = -\sum_{i,j} \mu(A_i \cap B_j) \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)},$$

omitting the j-terms when $\mu(B_j) = 0$.

Note that if $\zeta = \{X\}$ is the trivial partition of X, then $H(\xi|\zeta) = H(\xi)$ for any partition ξ of X.

Theorem 3.1.2. *Let μ be a measure on a shift space X. Let $\mathcal{C}$ be a sub-σ-algebra of $\mathscr{B}_X$. For each $f \in L^1(X, \mathscr{B}_X)$ there is a unique function $\mathbb{E}(f|\mathcal{C}) \in L^1(X, \mathcal{C})$ such that $\int_C \mathbb{E}(f|\mathcal{C})\, d\mu = \int_C f\, d\mu$ for each $C \in \mathcal{C}$.*

Theorem 3.1.2 is a direct corollary of the Radon-Nikodym Theorem which may be found in [29].

The operator $\mathbb{E}(.|\mathcal{C}) : L^1(X, \mathscr{B}_X) \to L^1(X, \mathcal{C})$ is called the conditional expectation operator. The following property of this operator will be used in a future chapter.

Proposition 3.1.3. (Law of Iterated Expectations) If $\mathcal{C}$ and $\mathcal{D}$ are sub-σ algebras of $\mathscr{B}_X$ then

$$\mathbb{E}\left(\mathbb{E}(f|\mathcal{C} \vee \mathcal{D})|\mathcal{C}\right) = \mathbb{E}(f|\mathcal{C}).$$

A proof of Proposition 3.1.3 may be found in [15].

If $\mathcal{C}$ is a finite sub-σ-algebra of $\mathscr{B}_X$ with $\mathcal{P}(\mathcal{C}) = \{C_1, \ldots, C_m\}$ then one can verify that

$$\mathbb{E}(f|\mathcal{C})(x) = \sum_{j=1}^{m} \mathbf{1}_{C_j}(x) \frac{1}{\mu(C_j)} \int_{C_j} f \, d\mu.$$

If $\mathcal{A}$ also is a finite sub-σ-algebra of $\mathscr{B}_X$ with $\mathcal{P}(\mathcal{A}) = \{A_1, \ldots, A_m\}$ then

$$\begin{aligned} H(\mathcal{A}|\mathcal{C}) &= -\sum_{i,j} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \\ &= -\sum_{i=1}^{m} \int \mathbf{1}_{A_i} \log \mathbb{E}(\mathbf{1}_{A_i}|\mathcal{C}) \, d\mu \\ &= -\int \sum_{i=1}^{m} \mathbb{E}(\mathbf{1}_{A_i}|\mathcal{C}) \log \mathbb{E}(\mathbf{1}_{A_i}|\mathcal{C}) \, d\mu. \end{aligned}$$

This leads us to define conditional entropy $H(\mathcal{A}|\mathcal{C})$ when $\mathcal{A}$ is a finite sub-σ-algebra of $\mathscr{B}_X$ and $\mathcal{C}$ is an arbitrary sub-σ-algebra of $\mathscr{B}_X$.

Definition 3.1.4. *Let X be a shift space. Let $\mathcal{A}$ be a finite sub-σ-algebra of $\mathscr{B}_X$ with $\mathcal{P}(\mathcal{A}) = \{A_1, \ldots, A_m\}$ and let $\mathcal{C}$ be an arbitrary sub-σ-algebra of $\mathscr{B}_X$. The entropy of $\mathcal{A}$ given $\mathcal{C}$ is the quantity*

$$H(\mathcal{A}|\mathcal{C}) = -\int \sum_{i=1}^{m} \mathbb{E}(\mathbf{1}_{A_i}|\mathcal{C}) \log \mathbb{E}(\mathbf{1}_{A_i}|\mathcal{C}) \, d\mu.$$

The following Theorem provides an applicable tool in computing conditional entropy of a finite sub-σ-algebra given an arbitrary sub-σ-algebra.

Theorem 3.1.5. *[29, Theorem 4.7] Let μ be a measure on a shift space X. Let $\mathcal{A}$ be a finite sub-algebra of $\mathscr{B}_X$ and let $(\mathcal{C}_n)_{n=1}^{\infty}$ be an increasing sequence of sub-σ-algebras of $\mathscr{B}_X$ with $\bigvee_{n=1}^{\infty} \mathcal{C}_n = \mathcal{C}$. Then $H(\mathcal{A}|\mathcal{C}_n) \to H(\mathcal{A}|\mathcal{C})$.*

Here we state some well-known properties of entropy which will be used later in this chapter. See [29] for justifications.

Proposition 3.1.6. *Let X be a shift space and $\mathcal{J}$ be a finite sub-σ-algebra of $\mathscr{B}_X$ then:*

(a) If $\mathscr{B}_X = \bigvee_{n=-\infty}^{\infty} T^n\mathcal{J}$ then $h_\mu(T) = H(\mathcal{J}| \bigvee_{n=1}^{\infty} T^n\mathcal{J})$.

(b) For $k > 0$, $h_\mu(T) = kh_\mu(T^k)$

Proposition 3.1.7. *Let X be a shift space. Let μ, $\mu' \in M(X,T)$ and $p \in [0,1]$. Then*

$$h_{p\mu+(1-p)\mu'}(T) = ph_\mu(T) + (1-p)h_{\mu'}(T).$$

Proposition 3.1.8. *Let X be a shift space. Let $\mathcal{A}$ and $\mathcal{J}$ be finite sub-σ-algebras of $\mathscr{B}_X$ and $\mathcal{D}$ be an arbitrary sub-σ-algebra of $\mathscr{B}_X$. Then*

$$H(\mathcal{J}|\mathcal{D}) = H(\mathcal{A}|\mathcal{D}) + H(\mathcal{J}|\mathcal{D} \vee \mathcal{A}).$$

3.2 Uniform Conditional Distribution

Given a 1-step SFT, a point x in X may be regarded as a function $x\colon \mathbb{Z} \to \mathcal{A}(X)$ such that $x(i)x(i+1)$ is an allowable block. For $G \subseteq \mathbb{Z}$, let x_G denote the restriction of x to G. A function $\alpha\colon G \to \mathcal{A}(X)$ is called a configuration on G if there is a point x in X with $x_G = \alpha$; i.e., $x(i) = \alpha(i)$ for each i in G. Let α be a configuration on G. We identify α with the set $\{x \in X\colon x_G = \alpha\}$. Then given a measure μ on X, $\mu(\alpha)$ denote $\mu(\{x \in X\colon x_G = \alpha\})$. Let η be a configuration on ∂G^c. We say α extends η if there is x in X with $x_G = \alpha$ and $x_{\partial G^c} = \eta$.

Theorem 3.2.1. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y, ν an invariant measure on Y, and μ an invariant measure of relative maximal entropy over ν. Then the conditional distribution of μ on any finite set $G \subseteq \mathbb{Z}$ given the configuration on G^c is μ-a.s. uniform over all configurations on G which extend the configuration on ∂G^c and map to the same configuration in Y under the factor code π.*

Proof. Let G be a finite subset of $\mathbb{Z}$. Let Δ be a configuration of Y on G. Pick a configuration η of X on ∂G^c such that $\mu(\eta \cap \pi^{-1}(\Delta)) > 0$. Starting from μ, we define a measure $\tilde{\gamma}$ on X by uniformizing over preimages of Δ that have η on the boundary. We then show that if μ does not have the required uniform conditional distribution property then $\tilde{\gamma}$ has greater entropy than μ, but still is an element of the fibre $\pi^{-1}\{\nu\}$. This is a contradiction and will therefore establish the required uniform conditional distribution property of μ.

Let $D = \{\alpha_1, \ldots, \alpha_L\}$ be the set of all configurations on G which extend η and map to Δ under the factor code π. Let $R = \{-m, \ldots, m-1\}$ be large enough so that $G \cup \partial G^c \subseteq R$. Let $(R_n)_{n\in\mathbb{Z}}$ be the partition of $\mathbb{Z}$ by translates of R ($R_0 = R$), i.e.

$$R_n = \{(2n-1)m, \ldots, (2n+1)m - 1\}.$$

Let G_n and ∂G_n^c be the corresponding translates of G and ∂G^c in R_n.

We obtain the measure γ from μ and η as follows. Define $\Phi : X \times D^{\mathbb{Z}} \to X$ by

$$\Phi(x, \zeta)_{G_n} = \begin{cases} \zeta_n & \text{if } x_{\partial G_n^c} = \eta \text{ and } x_{G_n} \in D \\ x_{G_n} & \text{otherwise,} \end{cases}$$

and $\Phi(x, \zeta)_{G_n^c \cap R_n} = x_{G_n^c \cap R_n}$ for each $n \in \mathbb{Z}$. In simple words what Φ does to a pair (x, ζ) is to check R_n's one by one: if $x_{\partial G_n^c} = \eta$ and x_{G_n} in D then Φ replaces x_{G_n} by ζ_n. Since $\zeta_n \in D$ and each element of D extends η, the assumption that X is a 1-step SFT implies $\Phi(x, \zeta) \in X$. For each $\zeta \in D^{\mathbb{Z}}$ we have

$$\pi(x) = \pi(\Phi(x, \zeta)),$$

since $\Phi(x, \zeta)$ and x are the same except maybe having alternative α_i's in the same positions. Let $A \in \mathscr{B}_X$. Define

$$\gamma(A) = (\mu \times \lambda)\Phi^{-1}A,$$

where λ is the Bernoulli-$(1/L, \ldots, 1/L)$ measure on $D^{\mathbb{Z}}$. The measure γ is not necessarily invariant under T. However, we claim that γ is T^{2m}-invariant; i.e., for each $A \in \mathscr{B}_X$ we have

$$\gamma(A) = \gamma(T^{-2m}A).$$

Proof. By the definition of γ, since μ and λ are T-invariant it suffices to show that, given $\Phi^{-1}A = E \times K$ for some E in $\mathscr{B}_X$ and K in $\mathscr{B}_{D^{\mathbb{Z}}}$ we have

$$\Phi^{-1} \circ T^{-2m}A = T^{-2m}E \times T^{-1}K.$$

This is equivalent to showing that for each x in X and every ζ in $D^{\mathbb{Z}}$ we have

$$\Phi(x, \zeta) = T^{2m} \circ \Phi\left(T^{-2m}x,\, T^{-1}\zeta\right). \tag{3.2.1}$$

Let x be in X and ζ be a point in $D^{\mathbb{Z}}$. Then for every integer n we have

$$T^{2m} \circ \Phi\left(T^{-2m}x,\, T^{-1}\zeta\right)_{G_n} = \Phi\left(T^{-2m}x,\, T^{-1}\zeta\right)_{G_{n+1}}. \tag{3.2.2}$$

Moreover, by the definition of Φ,

$$\begin{aligned}
&\Phi\left(T^{-2m}x,\, T^{-1}\zeta\right)_{G_{n+1}} \\
&\quad = \begin{cases} (T^{-1}\zeta)_{n+1} & \text{if } (T^{-2m}x)_{\partial G_{n+1}^c} = \eta,\ (T^{-2m}x)_{G_{n+1}} \in D \\ (T^{-2m}x)_{G_{n+1}} & \text{otherwise} \end{cases} \\
&\quad = \begin{cases} \zeta_n & \text{if } x_{\partial G_n^c} = \eta,\ x_{G_n} \in D \\ x_{G_n} & \text{otherwise} \end{cases} \\
&\quad = \Phi(x, \zeta)_{G_n}.
\end{aligned}$$

From this and Equation (3.2.2) it follows that for each integer n,

$$T^{2m} \circ \Phi\left(T^{-2m}x,\, T^{-1}\zeta\right)_{G_n} = \Phi(x, \zeta)_{G_n}. \tag{3.2.3}$$

Furthermore, we have

$$\begin{aligned} T^{2m} \circ \Phi\left(T^{-2m}x, T^{-1}\zeta\right)_{G_n^c \cap R_n} &= \Phi\left(T^{-2m}x, T^{-1}\zeta\right)_{G_{n+1}^c \cap R_{n+1}} \\ &= \left(T^{-2m}x\right)_{G_{n+1}^c \cap R_{n+1}} \qquad (3.2.4) \\ &= x_{G_n^c \cap R_n}. \end{aligned}$$

Equation (3.2.1) follows from the combination of Equation (3.2.3) and Equation (3.2.4). □

Define a new measure $\tilde{\gamma}$ on X by

$$\tilde{\gamma}(C) = \frac{1}{2m}\left(\gamma(C) + \gamma(T^{-1}C) + \cdots + \gamma(T^{-2m+2}C) + \gamma(T^{-2m+1}C)\right).$$

Then

$$\begin{aligned} \tilde{\gamma}\left(T^{-1}C\right) &= \frac{1}{2m}\left(\gamma(T^{-1}C) + \gamma(T^{-2}C) + \cdots + \gamma(T^{-2m+1}C) + \gamma(T^{-2m}C)\right) \\ &= \frac{1}{2m}\left(\gamma(T^{-1}C) + \gamma(T^{-2}C) + \cdots + \gamma(T^{-2m+1}C) + \gamma(C)\right) \\ &= \tilde{\gamma}(C). \end{aligned}$$

Thus $\tilde{\gamma}$ is a T-invariant measure on X.

Claim. *Both measures γ and $\tilde{\gamma}$ map to ν under π.*

Proof. Let $E \in \mathscr{B}_Y$. To show $\gamma \circ \pi^{-1}(E) = \nu(E)$, it is enough to show that

$$\Phi^{-1}\left(\pi^{-1}E\right) = \pi^{-1}E \times D^{\mathbb{Z}}. \qquad (3.2.5)$$

Then we have

$$\gamma(\pi^{-1}E) = (\mu \times \lambda)\Phi^{-1}(\pi^{-1}E) = (\mu \times \lambda)(\pi^{-1}E \times D^{\mathbb{Z}}) = \mu(\pi^{-1}E) = \nu(E).$$

To show Equation (3.2.5) recall that for each $x \in X$ and $\zeta \in D^{\mathbb{Z}}$ we have $\pi(x) =$

$\pi(\Phi(x,\zeta))$. It follows that

$$\begin{aligned}(x,\zeta)\in\Phi^{-1}\left(\pi^{-1}E\right) &\iff \Phi(x,\zeta)\in\pi^{-1}E\\ &\iff \pi(x)=\pi(\Phi(x,\zeta))\in E\\ &\iff x\in\pi^{-1}E\\ &\iff (x,\zeta)\in\pi^{-1}E\times D^{\mathbb{Z}}.\end{aligned}$$

Now we show $\tilde{\gamma}$ also maps to ν under π. For each $i\in\mathbb{Z}$,

$$\gamma\left(T_X^i(\pi^{-1}E)\right)=\gamma\left(\pi^{-1}(T_Y^iE)\right)=\nu(T_Y^iE)=\nu(E).$$

Therefore for each $E\in\mathscr{B}_Y$ we have

$$\begin{aligned}\tilde{\gamma}(\pi^{-1}E)&=\frac{1}{2m}\left(\gamma\left(\pi^{-1}E\right)+\cdots+\gamma\left(T_X^{-2m-1}(\pi^{-1}E)\right)\right)\\ &=\frac{1}{2m}(\nu(E)+\cdots+\nu(E))\\ &=\nu(E).\end{aligned}$$

□

We aim to show that $h_{\tilde{\gamma}}(T)\geq h_\mu(T)$. First we need to introduce some notations and σ-algebras. If S is a subset of $\mathbb{Z}$ then $\mathcal{P}(X_S)$ is the partition of X generated by the configurations of X on S, and $\sigma(X_S)$ is the σ-algebra generated by $\mathcal{P}(X_S)$. When $S=\{a,\dots,b\}$ is a $\mathbb{Z}$-interval, we sometimes denote X_S by X_a^b. Considering $R=\{-m,\dots,m-1\}$ as before, $\sigma\left(X_{-m}^{m-1}\right)$ is the finite σ-algebra generated by the partition

$$\mathcal{P}\left(X_{-m}^{m-1}\right)=\{{}_{-m}[x_{-m}x_{-m+1}\dots x_{m-1}]_{m-1}:x_{-m}x_{-m+1}\dots x_{m-1}\in\mathcal{B}_m(X)\}.$$

Denote

$$\sigma\left(X_{-\infty}^{-m-1}\right)=\sigma\left(X_{-3m}^{-m-1}\right)\vee\sigma\left(X_{-5m}^{-3m-1}\right)\vee\dots.$$

Define an equivalence relation on X as follows; suppose $x,x'\in X$, say $x\sim_0 x'$ if either $x_R = x'_R$ or else $x_{R\cap G^c} = x'_{R\cap G^c}$, $x_{\partial G^c} = x'_{\partial G^c} = \eta$, and x_G, x'_G in D. Denote the equivalence class containing x by $C_0(x)$. Such equivalence classes form a sub-partition of $\mathcal{P}(X_{-m}^{m-1})$. Let $\mathcal{A}$ be the σ-algebra generated by these equivalence classes. We show

$$h_{\tilde{\gamma}}(T) \geq h_\mu(T)$$

as follows, using Lemma 3.2.2 which appears below; In the following equation, "Prop" stands for "by Proposition", and "Lem" stands for "by Lemma".

$$\begin{aligned}
& h_{\tilde{\gamma}}(T) && (3.2.6)\\
&= \frac{1}{2m} h_{\tilde{\gamma}}\left(T^{2m}\right) && \text{Prop 3.1.6}(b)\\
&= \frac{1}{2m} h_{\gamma}\left(T^{2m}\right) && \text{Prop 3.1.7}\\
&= \frac{1}{2m} H_{\gamma}\left(\sigma\left(X_{-m}^{m-1}\right)|\sigma\left(X_{-\infty}^{-m-1}\right)\right) && \text{Prop 3.1.6}(a)\\
&= \frac{1}{2m} H_{\gamma}\left(\mathcal{A}|\sigma\left(X_{-\infty}^{-m-1}\right)\right) + \frac{1}{2m} H_{\gamma}\left(\sigma\left(X_{-m}^{m-1}\right)|\sigma\left(X_{-\infty}^{-m-1}\right)\vee\mathcal{A}\right) && \text{Prop 3.1.8}\\
&= \frac{1}{2m} H_{\gamma}\left(\mathcal{A}|\sigma\left(X_{-\infty}^{-m-1}\right)\right) + \frac{1}{2m} H_{\gamma}\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) && \text{Lem 3.2.2}(a)\\
&\geq \frac{1}{2m} H_{\gamma}\left(\mathcal{A}|\sigma\left(X_{-\infty}^{-m-1}\right)\right) + \frac{1}{2m} H_{\mu}\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) && \text{Lem 3.2.2}(b)\\
&\geq \frac{1}{2m} H_{\mu}\left(\mathcal{A}|\sigma\left(X_{-\infty}^{-m-1}\right)\right) + \frac{1}{2m} H_{\mu}\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) && \text{Lem 3.2.2}(c)\\
&\geq \frac{1}{2m} H_{\mu}\left(\mathcal{A}|\sigma\left(X_{-\infty}^{-m-1}\right)\right) + \frac{1}{2m} H_{\mu}\left(\sigma(X_{-m}^{m-1})|\sigma\left(X_{-\infty}^{-m-1}\right)\vee\mathcal{A}\right) && \text{Prop 3.1.8}\\
&= \frac{1}{2m} H_{\mu}\left(\sigma\left(X_{-m}^{m-1}\right)|\sigma\left(X_{-\infty}^{-m-1}\right)\right)\\
&= \frac{1}{2m} h_{\mu}\left(T^{2m}\right)\\
&= h_{\mu}(T).
\end{aligned}$$

Lemma 3.2.2. *Reusing previous notations, we have*

(a) $H_\gamma\left(\sigma(X_{-m}^{m-1})|\sigma(X_{-\infty}^{-m-1})\vee\mathcal{A}\right) = H_\gamma\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right)$.

(b) $H_\gamma\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right) \geq H_\mu\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right)$. *Equality occurs if and only if*

$$\frac{\mu(\alpha \cap \bar{A})}{\mu(\bar{A})} = 1/L,$$

for each α *in* D *and* $\bar{A}$ *in* $\mathcal{P}(\mathcal{A})$ *such that for each* x *in* $\bar{A}$ *we have* $x_{\partial G^c} = \eta$ *and* x_G *in* D.

(c) $H_\gamma\left(\mathcal{A}|\sigma(X_{-\infty}^{-m-1})\right) \geq H_\mu\left(\mathcal{A}|\sigma(X_{-\infty}^{-m-1})\right)$.

Proof. Notice that by definition, for $\rho \in \{\mu, \gamma\}$ we have

$$H_\rho\left(\sigma(X_{-m}^{m-1})|\mathcal{A}\right) = -\sum_{i,k} \rho(O_i \cap A_k) \log \frac{\rho(O_i \cap A_k)}{\rho(A_k)}$$

where O_i is in the partition $\mathcal{P}\left(X_{-m}^{m-1}\right)$ and A_k in the partition $\mathcal{P}(\mathcal{A})$; with the convention that if $\rho(O_i \cap A_k) = 0$ then $\rho(O_i \cap A_k) \log \frac{\rho(O_i \cap A_k)}{\rho(A_k)} = 0$. Let x be a point in the cylinder set A_k. If $x_{\partial G^c} \neq \eta$ or x_G is not a configuration in D then for every cylinder set O_i in $\mathcal{P}\left(X_{-m}^{m-1}\right)$ we have either $O_i \cap A_k = \emptyset$ or $O_i \cap A_k = A_k$ which both imply $\rho(O_i \cap A_k) \log \frac{\rho(O_i \cap A_k)}{\rho(A_k)} = 0$. Let $\{\bar{A}_1, \ldots, \bar{A}_M\}$ be the set consisting of elements of the partition $\mathcal{P}(\mathcal{A})$ for which for any point x in $\bar{A}_k$ we have $x_{\partial G^c} = \eta$ and x_G in D. Then

$$H_\rho(\sigma(X_{-m}^{m-1})|\mathcal{A}) = -\sum_{i,k} \rho(O_i \cap \bar{A}_k) \log \frac{\rho(O_i \cap \bar{A}_k)}{\rho(\bar{A}_k)}.$$

There are exactly L pairwise disjoint cylinder sets O_i in the partition $\mathcal{P}\left(X_{-m}^{m-1}\right)$ defined by blocks which agree everywhere except on G, and form a partition of $\bar{A}_k$. Let these cylinder sets be denoted by $O_{k,1}, \ldots, O_{k,L}$ where $O_{k,i} = \alpha_i \cap \bar{A}_k$ for each $1 \leq i \leq L$. It follows that

$$\begin{aligned} H_\rho(\sigma(X_{-m}^{m-1})|\mathcal{A}) &= -\sum_{k,i} \rho(O_{k,i} \cap \bar{A}_k) \log \frac{\rho(O_{k,i} \cap \bar{A}_k)}{\rho(\bar{A}_k)} \\ &= -\sum_{k,i} \rho(\alpha_i \cap \bar{A}_k) \log \frac{\rho(\alpha_i \cap \bar{A}_k)}{\rho(\bar{A}_k)}. \end{aligned} \tag{3.2.7}$$

Let $n \geq 1$. By definition of γ, for each configuration α in D, each cylinder set P in the partition $\mathcal{P}\left(X_{-(2n+1)m}^{-m-1}\right)$, and $1 \leq k \leq M$ we have

$$\frac{\gamma(\alpha \cap P \cap \bar{A}_k)}{\gamma(P \cap \bar{A}_k)} = \frac{\gamma(\alpha \cap \bar{A}_k)}{\gamma(\bar{A}_k)} = \frac{1}{L}.$$

It follows that

$$\begin{aligned} H_\gamma\left(\sigma\left(X_{-m}^{m-1}\right)|\sigma\left(X_{-(2n+1)m}^{-m-1}\right) \vee \mathcal{A}\right) &= -\sum_{i,j,k} \gamma(\alpha_i \cap P_j \cap \bar{A}_k) \log \frac{\gamma(\alpha_i \cap P_j \cap \bar{A}_k)}{\gamma(P_j \cap \bar{A}_k)} \\ &= \log L \sum_{i,k} \gamma(\alpha_i \cap \bar{A}_k) \\ &= H_\gamma\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) \end{aligned}$$

(use the same argument we had before Equation (3.2.7) to get the first equality above). Then (a) follows from Theorem 3.1.5 and the fact that $\left(\sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right)_{n=1}^{\infty}$ is an increasing sequence of σ-algebras with

$$\bigvee_{n=1}^{\infty} \sigma\left(X_{-(2n+1)m}^{-m-1}\right) = \sigma\left(X_{-\infty}^{-m-1}\right).$$

To show (b) note that

$$\begin{aligned} H_\mu\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) &= -\sum_{i,k} \mu(\alpha_i \cap \bar{A}_k) \log \frac{\mu(\alpha_i \cap \bar{A}_k)}{\mu(\bar{A}_k)} \\ &= -\sum_k \mu(\bar{A}_k) \sum_{i=1}^{L} \frac{\mu(\alpha_i \cap \bar{A}_k)}{\mu(\bar{A}_k)} \log \frac{\mu(\alpha_i \cap \bar{A}_k)}{\mu(\bar{A}_k)} \\ &= \sum_k \mu(\bar{A}_k) \sum_{i=1}^{L} \psi\left(\frac{\mu(\alpha_i \cap \bar{A}_k)}{\mu(\bar{A}_k)}\right) \end{aligned}$$

where $\psi(x) = -x \log x$ ($\psi(0) = 0$) on the interval $[0, 1]$. Since $\psi(x)$ is a strictly concave function it follows that $H_\mu(\sigma(X_{-m}^{m-1})|\mathcal{A})$ attains its maximum if and only if for each $1 \leq k \leq M$ and α in D we have

$$\frac{\mu(\alpha \cap \bar{A}_k)}{\mu(\bar{A}_k)} = 1/L.$$

Therefore

$$\begin{aligned} H_\mu\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) &\leq \log L \sum_{k,i} \mu(\alpha_i \cap \bar{A}_k) \\ &= \log L \sum_{k,i} \gamma(\alpha_i \cap \bar{A}_k) \\ &= H_\gamma\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right), \end{aligned}$$

with equality if and only if $\frac{\mu(\alpha \cap \bar{A}_k)}{\mu(\bar{A}_k)} = 1/L$ for each $\alpha \in D$ and $1 \leq k \leq M$.

We prove (c) by showing that

$$H_\gamma(\mathcal{A}|\sigma\left(X_{-\infty}^{-m-1}\right)) = H_\mu(\mathcal{A}|\mathcal{A}_{-\infty}^{-m-1}) \tag{3.2.8}$$

where $\mathcal{A}_{-\infty}^{-m-1} = \bigvee_{n=1}^{\infty} T^{-2nm}\mathcal{A}$ (the choice of notation is intended to remind the reader that $\mathcal{A}_{-\infty}^{-m-1}$ is a sub-σ-algebra of $\sigma(X_{-\infty}^{-m-1})$). The fact that $\mathcal{A}_{-\infty}^{-m-1}$ is a sub-σ-algebra of $\sigma(X_{-\infty}^{-m-1})$ implies

$$H_\mu(\mathcal{A}|\mathcal{A}_{-\infty}^{-m-1}) \geq H_\mu(\mathcal{A}|\sigma(X_{-\infty}^{-m-1}))$$

which will complete the proof.

Define an equivalence relation $\sim_n$ on X as follows; say $x \sim_n x'$ provided that for all $-n \leq i \leq -1$ we either have $x_{R_i} = x'_{R_i}$ or else $x_{R_i \cap G_i^c} = x'_{R_i \cap G_i^c}$, $x_{\partial G_i^c} = x'_{\partial G_i^c} = \eta$, and x_{G_i}, x'_{G_i} in D. Denote the equivalence class containing x by $C_n(x)$. Such equivalence classes form a sub-partition of $\mathcal{P}\left(X_{-(2n+1)m}^{-m-1}\right)$ and these equivalence classes generate the σ-algebra

$$\mathcal{A}_{-(2n+1)m}^{-m-1} = T^{-2m}\mathcal{A} \vee \cdots \vee T^{-2nm}\mathcal{A}.$$

Let x in X. Then x is in E for some E in $\mathcal{P}\left(\mathcal{A}_{-(2n+1)m}^{-m-1}\right)$. Set

$$K_x = \{-n \le k \le -1 \colon x_{\partial G_k^c} = \eta,\ x_{G_k} \in D\}.$$

There are exactly $L^{|K_x|}$ pairwise disjoint cylinder sets $P_1, \dots, P_{L^{|K_x|}}$ in the partition $\mathcal{P}\left(X_{-(2n+1)m}^{-m-1}\right)$ which are the same except having alternative α_i's in the same positions and they form a partition of E. Let A in $\mathcal{A}$. By definition of γ for each $1 \le j \le L^{|K_x|}$ we have

$$\frac{\gamma(A \cap P_j)}{\gamma(P_j)} = \frac{L^{-|K_x|}\gamma(A \cap E)}{L^{-|K_x|}\gamma(E)} = \frac{\gamma(A \cap E)}{\gamma(E)} = \frac{\mu(A \cap E)}{\mu(E)}.$$

It follows that

$$\mathbb{E}_\gamma\left(\mathbf{1}_A | \sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right) = \mathbb{E}_\mu\left(\mathbf{1}_A | \mathcal{A}_{-(2n+1)m}^{-m-1}\right). \tag{3.2.9}$$

Let x be a point in X, F be in the σ-algebra $\mathcal{A}_{-\infty}^{-m-1}$, and ζ be a point in $D^{\mathbb{Z}}$. Then the two sided implication

$$\Phi(x, \zeta) \in F \iff x \in F,$$

implies that

$$\mathbf{1}_F \circ \Phi(x, \zeta) = \mathbf{1}_F(x).$$

Let g be a bounded $\mathcal{A}_{-\infty}^{-m-1}$-measurable function. Since there is a sequence of $\mathcal{A}_{-\infty}^{-m-1}$-measurable simple functions converging uniformly to g we deduce that

$$g \circ \Phi(x, \zeta) = g(x).$$

Moreover,

$$\int_F g\, d\gamma = \int_F g\, d(\mu \times \lambda)(\Phi^{-1}) = \int_{F \times D^{\mathbb{Z}}} g \circ \Phi\, d(\mu \times \lambda) = \int_F g\ d\mu, \tag{3.2.10}$$

where the second equality follows from Equation (3.2.11) and recalling that there is a sequence of $\mathcal{A}_{-\infty}^{-m-1}$-measurable simple functions converging uniformly to g. Given a characteristic $\mathcal{A}_{-\infty}^{-m-1}$-measurable function $\mathbf{1}_S$, we have

$$\begin{aligned}\int_F \mathbf{1}_S \, d(\mu\times\lambda)(\Phi^{-1}) &= (\mu\times\lambda)(\Phi^{-1}(S\cap F)) \\ &= \int_{\Phi^{-1}(F)} \mathbf{1}_{\Phi^{-1}(S)} \, d(\mu\times\lambda) \\ &= \int_{\Phi^{-1}(F)} \mathbf{1}_S \circ \Phi \, d(\mu\times\lambda) \\ &= \int_{F\times D^{\mathbb{Z}}} \mathbf{1}_S \circ \Phi \, d(\mu\times\lambda).\end{aligned} \tag{3.2.11}$$

Let $g = \mathbb{E}_\mu\left(\mathbf{1}_A|\mathcal{A}_{-\infty}^{-m-1}\right)$ for some A in $\mathcal{P}(\mathcal{A})$ in Equation (3.2.10). Then we have

$$\begin{aligned}H_\gamma\left(\mathcal{A}|\sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right) &= \int_X \sum_{A\in\mathcal{P}(\mathcal{A})} \psi\left(\mathbb{E}_\gamma\left(\mathbf{1}_A|\sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right)\right)\, d\gamma \\ &= \int_X \sum_{A\in\mathcal{P}(\mathcal{A})} \psi\left(\mathbb{E}_\mu\left(\mathbf{1}_A|\mathcal{A}_{-(2n+1)m}^{-m-1}\right)\right)\, d\gamma \qquad \text{by (3.2.9)} \\ &= \int_X \sum_{A\in\mathcal{P}(\mathcal{A})} \psi\left(\mathbb{E}_\mu\left(\mathbf{1}_A|\mathcal{A}_{-(2n+1)m}^{-m-1}\right)\right)\, d\mu \qquad \text{by (3.2.10)} \\ &= H_\mu\left(\mathcal{A}|\mathcal{A}_{-(2n+1)m}^{-m-1}\right).\end{aligned}$$

Then Lemma 3.2.2(c) follows directly from Theorem 3.1.5 and the fact that $\left(\sigma\left(X_{-(2n+1)m}^{-m-1}\right)\right)_{n=1}^{\infty}$ and $\left(\mathcal{A}_{-(2n+1)m}^{-m-1}\right)_{n=1}^{\infty}$ are both increasing sequences with

$$\bigvee_{n=1}^{\infty} \sigma\left(X_{-(2n+1)m}^{-m-1}\right) = \sigma\left(X_{-\infty}^{-m-1}\right),$$

and

$$\bigvee_{n=1}^{\infty} \mathcal{A}_{-(2n+1)m}^{-m-1} = \mathcal{A}_{-\infty}^{-m-1}.$$

$\square$

Proof of Theorem 3.2.1 (continued). Since μ is an invariant measure of relative maximal entropy it follows that all of the inequalities in Equation (3.2.6) are forced to be equalities. In particular, we have

$$H_\mu\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right) = H_\gamma\left(\sigma\left(X_{-m}^{m-1}\right)|\mathcal{A}\right).$$

Then Lemma 3.2.2(b) implies that

$$\frac{\mu\left(\alpha \cap \bar{A}\right)}{\mu(\bar{A})} = 1/L, \tag{3.2.12}$$

for each configuration α in D and $\bar{A}$ in the partition $\mathcal{P}(\mathcal{A})$ in which for each x in $\bar{A}$ we have $x_{\partial G^c} = \eta$ and x_G in D. Note that given a finite subset $G \subseteq \mathbb{Z}$, both configurations $\Delta \in \mathscr{B}_Y$ on G and $\eta \in \mathscr{B}_X$ on ∂G^c with $\mu(\eta \cap \pi^{-1}(\Delta)) > 0$ are chosen arbitrarily. By choosing different configurations and noting that

$$\bar{A} \in \mathcal{P}\left(\pi^{-1}(\sigma(Y_G)) \bigvee \sigma(X_{R\cap G^c})\right) = \mathcal{P}\left(\pi^{-1}(\sigma(Y_R)) \bigvee \sigma(X_{R\cap G^c})\right),$$

Equation (3.2.12) implies that for any configuration α on G we have

$$\mathbb{E}\left(\mathbf{1}_\alpha|\pi^{-1}(\sigma(Y_R)) \vee \sigma(X_{R\cap G^c})\right)(x) = \begin{cases} \frac{1}{L_{(x,G)}} & \text{if } \alpha \text{ extends } x_{\partial G^c},\ x_G \in \pi_{\mathrm{b}}^{-1}(\pi_{\mathrm{b}}(\alpha)) \\ 0 & \text{otherwise.} \end{cases}$$

where $L_{(x,G)}$ is the number of configurations of X occurring at G which extend $x_{\partial G^c}$ and project to $\pi_{\mathrm{b}}(x_G)$. Now for $t \in \mathbb{N}$ let

$$R^{(t)} = \{-(m+t), \ldots, m+t-1\},$$

and $(R_n^{(t)})_{n\in\mathbb{Z}}$ be the partition of $\mathbb{Z}$ by translates of $R^{(t)}$ where $R_0^{(t)} = R^{(t)}$, i.e.

$$R_n^{(t)} = \{(2n-1)(m+t), \ldots, (2n+1)(m+t)-1\}.$$

Let $R^{(0)} = R$. Since $\left(\sigma\left(Y_{R^{(t)}}\right)\right)_{t=0}^{\infty}$ and $\left(\sigma\left(X_{R^{(t)} \cap G^c}\right)\right)_{t=0}^{\infty}$ are increasing sequences with

$$\bigvee_{t=0}^{\infty} \sigma(Y_{R^{(t)}}) = \mathscr{B}_Y,$$

and

$$\bigvee_{t=0}^{\infty} \sigma(X_{R^{(t)} \cap G^c}) = \sigma(X_{G^c}),$$

it follows from Proposition 3.1.5 that

$$\mathbb{E}\big(1_\alpha | \pi^{-1}(\mathscr{B}_Y) \vee \sigma(X_{G^c})\big)(x) = \begin{cases} \frac{1}{L_{(x,G)}} & \text{if } \alpha \text{ extends } x_{\partial G^c},\ x_G \in \pi_{\mathrm{b}}^{-1}(\pi_{\mathrm{b}}(\alpha)) \\ 0 & \text{otherwise.} \end{cases}$$

$\square$

Chapter 4

Class Degree

When π is a finite-to-one factor code from a SFT X to a sofic shift Y, there is a uniform upper bound on the number of preimages of points in Y. The minimal number of π-preimages of points in Y is called the degree of the code. In this chapter we find a quantity analogous to the degree when π is an infinite-to-one factor code. This will be done by introducing an equivalence relation on X. We start the chapter by discussing some well-known properties of the degree of a finite-to-one code. In Section 4.1.2 we present an algorithm to compute the degree of a finite-to-one code. Section 4.2 describes an equivalence relation on X where the minimum number of equivalence classes over points of Y is called the class degree of a code. We show that this number agrees with the degree of a code in the case of having a finite-to-one code. In section 4.3 we justify that there are the same number of such equivalence classes over each typical point of an irreducible sofic shift.

4.1 Degree and an Algorithm for Computation of the Degree

We divide this section into two parts. The first part is devoted to the description of the degree of a finite-to-one factor code which is mainly adapted from [19, Chapter 9]. Then we present an algorithm in the second part of this section which computes the degree of a code.

4.1.1 Degree

Let $\pi : X \to Y$ be a 1-block factor code. Above every Y-block W of length n there is a set of X-blocks W' of length n which are sent to W by π_{b}; i.e., $\pi_{\mathrm{b}}(W') = W$. Given $0 \le i \le n-1$, set

$$d(W,i) = |\{a \in \mathcal{A}(X) \colon \exists W' \text{ with } \pi_{\mathrm{b}}(W') = W,\, W'_i = a\}|,$$

and let

$$d^*_\pi = \min\{d(W,i) : W \text{ a block of } Y,\, 0 \le i \le |W|-1\}.$$

A magic block is a block W such that $d(W,i) = d^*_\pi$ for some $0 \le i \le |W|-1$. Such an index i is called a magic coordinate of W. A factor code π has a magic symbol if there is a magic block of π of length 1.

Example 4.1.1. *Figure 4.1 illustrates an example where $d^*_\pi = 2$. $\{a,b,c,d\}$ is the alphabet of the original SFT and $\{0,1\}$ is the alphabet of the sofic shift.*

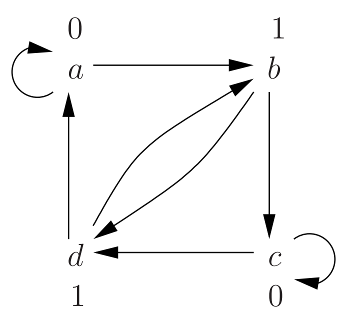

Figure 4.1: Graph for Example 4.1.1

Above each binary block starting with 0 there is a unique block starting with a or c. Similarly if a binary block starts with 1 then it is presented by a unique block starting with b or d.

Example 4.1.2. *Figure 4.2 represents an example where* $d^*_\pi = 1$. *The alphabet of the SFT is* $\{a, b, c, d\}$ *and the alphabet of the sofic shift is* $\{0, 1\}$. *Note that the shortest magic block has length two:* 00 *is such a magic block.*

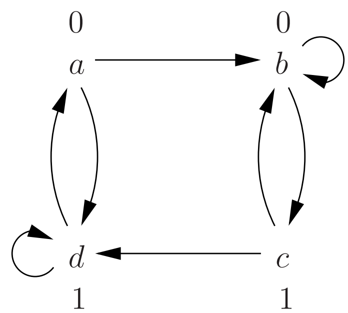

Figure 4.2: Graph for Example 4.1.2

Definition 4.1.3. *Let* (X, Y, π) *be a factor triple. The minimal number of* π*-preimages of points in* Y *is called the* **degree** *of the code and denoted by* d_π.

Theorem 4.1.4. *Let* $\pi : X \to Y$ *be a finite-to-one factor code from a SFT* X *to an irreducible sofic shift* Y. *Then every doubly transitive point of* Y *has exactly* d_π *preimages.*

Theorem 9.1.11 in [19] shows this result when X is irreducible. However, the proof only depends on the irreducibility of Y.

We say two factor triples (X, Y, π) and $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ are **conjugate**, and denote it by $(X, Y, \pi) \cong (\tilde{X}, \tilde{Y}, \tilde{\pi})$, if X is conjugate to $\tilde{X}$ under a conjugacy ϕ, Y is conjugate to $\tilde{Y}$ under a conjugacy ψ, and $\tilde{\pi} \circ \phi^{-1} = \psi^{-1} \circ \pi$. See Figure 4.3.

Theorem 4.1.5. *Let* (X, Y, π) *and* $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ *be conjugate factor triples. Then we have* $d_\pi = d_{\tilde{\pi}}$.

A proof of this Corollary can be found in [19]. In particular, the reader is referred to Theorem 9.1.16 of [19].

Let (X, Y, π) be a factor triple. Since d_π is invariant under conjugacy it would be ideal to work with the simplest possible factor triples. The following result which is adapted from a theorem in [19] allows us to do this.

$$\begin{tikzcd} X \arrow[r, "\phi"] \arrow[d, "\pi"'] & \tilde{X} \arrow[d, "\tilde{\pi}"] \\ Y \arrow[r, "\psi"] & \tilde{Y} \end{tikzcd}$$

Figure 4.3: Conjugate factor triples, $(X, Y, \pi) \cong (\tilde{X}, \tilde{Y}, \tilde{\pi})$

Theorem 4.1.6. *Let (X, Y, π) be a factor triple. There is a factor triple $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ conjugate to (X, Y, π) such that $\tilde{X}$ is a 1-step SFT and $\tilde{\pi}$ is a 1-block code with a magic symbol.*

Proof. By recoding, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code (see Corollary 2.3.11). Let W be a magic block of π of length n with a magic coordinate t. Define two n-blocks U and V in X to be equivalent if $\pi(U) = \pi(V)$ and $U_t = V_t$. This obviously defines an equivalence relation on $\mathcal{B}_n(X)$. Denote the equivalence class containing U by $C(U)$. Let $\tilde{X}$ be the new 1-step SFT whose alphabet consists of equivalence classes of blocks in $\mathcal{B}_n(X)$, and the legal transitions between the equivalence classes be defined by saying $C(U)$ can be followed by $C(V)$ if and only if there are $U' \in C(U)$ and $V' \in C(V)$ such that $U'_{i+1} = V'_i$ for $i = 0, \ldots, n-2$. Let $\phi : \tilde{X} \to X$ be the 1-block code induced by the map ϕ_{b} which takes $C(U)$ to U_t. Then ϕ does map $\tilde{X}$ into X, since whenever $C(U)C(V)$ is a 2-block in $\tilde{X}$, then U_tV_t is a 2-block in X. Moreover, ϕ is a conjugacy with the inverse n-block code induced by the map which takes U to $C(U)$. Therefore, $\tilde{X}$ is a 1-step SFT conjugate to X.

Let $\tilde{Y}$ be the nth higher block presentation of Y and let $\psi : \tilde{Y} \to Y$ be the n-block code induced by the map ψ_{b} which takes $y_0 \ldots y_{n-1}$ to y_t. Define

$$\tilde{\pi} = \psi^{-1} \circ \pi \circ \phi : \tilde{X} \to \tilde{Y}.$$

Observe that $\tilde{\pi}$ is a 1-block code induced by the well-defined map which projects a symbol $C(U)$ in $\mathcal{A}(\tilde{X})$ to $\pi_{\mathrm{b}}(U)$ (regarding $\pi_{\mathrm{b}}(U)$ as a symbol of $\tilde{Y}$).

Note that if B is a m-block of $\tilde{Y}$ then it is of the form $E^{(0)}E^{(1)}\dots E^{(m-1)}$ where $E^{(i)}$ is a n-block of Y. Then B corresponds to the $(n+m-1)$-block

$$E_0^{(0)}E_1^{(0)}\dots E_{n-1}^{(0)}E_{n-1}^{(1)}E_{n-1}^{(2)}\dots E_{n-1}^{(m-1)}$$

of Y. For each $0 \leq i \leq m-1$ we have

$$\begin{aligned} d_{\tilde{\pi}}(B,i) &= d_{\tilde{\pi}}(E^{(0)}E^{(1)}\dots E^{(m-1)},\, i) \\ &= |\{C(U) \in \mathcal{A}(\tilde{X})\colon \exists \text{ block } F \text{ of } \tilde{X} \text{ with } \tilde{\pi}_{\mathrm{b}}(F) = B,\, F_i = C(U)\}| \\ &= |\{a \in \mathcal{A}(X)\colon \exists \text{ block } K \text{ of } X \text{ with } \pi_{\mathrm{b}}(K) = B,\, K_{i+t} = a\}| \\ &= d_{\pi}(E_0^{(0)}E_1^{(0)}\dots E_{n-1}^{(0)}E_{n-1}^{(1)}E_{n-1}^{(2)}\dots E_{n-1}^{(m-1)},\, i+t). \end{aligned}$$

Hence $d^*_{\tilde{\pi}} \geq d^*_{\pi}$. Since $d_{\tilde{\pi}}(W,0) = d_{\pi}(W,t)$ it follows that W is a magic symbol of $\tilde{\pi}$. □

Definition 4.1.7. *Bi-infinite sequences $x^{(1)},\dots,x^{(k)}$ are* mutually separated *if, for each integer i, $x_i^{(1)},\dots,x_i^{(k)}$ are all distinct.*

Proposition 4.1.8. *Let $\pi : X \to Y$ be a finite-to-one 1-block factor code from a SFT to an irreducible sofic shift Y. The preimages of a transitive point are mutually separated. If π has a magic symbol w then $d_\pi = |\pi_b^{-1}(w)|$.*

A proof of this result may be found in [19]. In particular the reader is referred to Proposition 9.1.9 and Exercise 9.1.3 of [19].

4.1.2 Degree Algorithm

Let X be a shift of finite type with alphabet $\mathcal{A}(X) = \{a_1,\dots,a_i\}$ and adjacency matrix I. Let $\pi : X \to Y$ be a finite factor code from X to a sofic shift Y with the alphabet $\mathcal{A}(Y) = \{b_1,\dots,b_j\}$. Let d_π be the degree of the factor code π. Make two graphs G and G' as follows. Let G and G' both have the same vertex set

$\mathcal{V} = \bigcup_{b\in\mathcal{A}(Y)} \mathscr{P}\{\pi_{\mathrm{b}}^{-1}(b)\} = \{A_1, \ldots, A_m\}$ where $\mathscr{P}\{\pi_{\mathrm{b}}^{-1}(b)\}$ stands for the power set of $\{\pi_{\mathrm{b}}^{-1}(b)\}$. For what we need later, divide the vertex set $\mathcal{V}$ into two parts $\mathcal{U}$ and $\mathcal{V} - \mathcal{U}$ where $\mathcal{U} = \{\pi_{\mathrm{b}}^{-1}(b) \colon b \in \mathcal{A}(Y)\}$. Form the adjacency matrix M of G and the adjacency matrix M' of G' as follows. Let $A, A' \in \mathcal{V}$, then $A \subseteq \pi_{\mathrm{b}}^{-1}(b)$ and $A' \subseteq \pi_{\mathrm{b}}^{-1}(b')$ for some b, b' in $\mathcal{A}(Y)$. Say $M_{AA'} = 1$ if the following conditions hold.

1. bb' is a block of Y,
2. A' is exactly the set of all symbols a' in $\pi_{\mathrm{b}}^{-1}(b')$ such that aa' is a block of X for some a in A.

Otherwise $M_{AA'} = 0$. Say $M'_{AA'} = 1$ if we have

1. bb' is a block of Y,
2. A is exactly the set of all symbols a in $\pi_{\mathrm{b}}^{-1}(b)$ such that aa' is a block of X for some a' in A'.

Otherwise $M'_{AA'} = 0$. Consider the following subsets of the vertex set $\mathcal{V}$,

$$S = \{A \in \mathcal{V} \colon \text{ there is a finite path in } G \text{ from } B \text{ to } A \text{ for some } B \in \mathcal{U}\},$$

and

$$S' = \{A \in \mathcal{V} \colon \text{ there is a finite path in } G' \text{ from } A \text{ to } B \text{ for some } B \in \mathcal{U}\}.$$

Claim. *Using above notations, we have*

$$d_\pi = \min_{\substack{A\in S \\ A'\in S'}} \{|A \cap A'| \colon A \cap A' \neq \emptyset\}.$$

Proof. Note that G and G' are finite directed graphs. Let X_G and $X_{G'}$ be the shift spaces represented by G and G' accordingly. Let $\bar{\pi}_{\mathrm{b}}$ be the map from $\mathcal{V}$ to Y taking $A \in \mathcal{V}$ to $\pi(a)$ for some $a \in A$ (note that $\bar{\pi}_{\mathrm{b}}(A)$ is independent of $a \in A$). Then $\bar{\pi}_{\mathrm{b}}$

induces a 1-block code $\bar{\pi}_G : X_G \to Y$ and a 1-block code $\bar{\pi}_{G'} : X_{G'} \to Y$.

The key feature of these two graphs is the following. For any block

$$W = W_0 \dots W_k$$

of Y there is a unique walk

$$U = U_0 \dots U_k$$

in G with the following properties.

1. $U_0 = \pi_{\mathrm{b}}^{-1}(W_0) = \bar{\pi}_{\mathrm{b}}^{-1}(W_0)$.
2. $\bar{\pi}_G(U) = W$.
3. $U_k = \{a \in \mathcal{A}(X) \colon \exists B \in \pi^{-1}(W) \text{ and } B_k = a\}$.

Similarly there is a unique walk

$$V = V_0 \dots V_k$$

in G' with the following properties.

1. $V_k = \pi_{\mathrm{b}}^{-1}(W_k) = \bar{\pi}_{\mathrm{b}}^{-1}(W_k)$.
2. $\bar{\pi}_{G'}(V) = W$.
3. $V_0 = \{a \in \mathcal{A}(X) \colon \exists\, B \in \pi^{-1}(W) \text{ and } B_0 = a\}$.

Let $A \in S$, $A' \in S'$, and $A \cap A' \neq \emptyset$. Let $U = U_0 \dots U_c$ be a walk in G where $U_0 \in \mathcal{U}$ and $U_c = A$. Let $V = V_0 \dots V_k$ be a walk in G' where $V_0 = A'$ and $V_k \in \mathcal{U}$. Since $A \cap A' \neq \emptyset$ we have $\bar{\pi}_G(A) = \bar{\pi}_{G'}(A')$, and consequently $\bar{\pi}_G(U)\bar{\pi}_{G'}(V) = W_0 \dots W_{c+k} = W$ is a block of Y. By the key feature discussed in the above paragraph, we have

$$A \cap A' = \{a \in \mathcal{A}(X) \colon W' \in \pi_{\mathrm{b}}^{-1}(W), \text{ and } W'_c = a\}.$$

Since by definition

$$d_\pi = \min_{\substack{W \text{ block of } Y \\ 0 \le i \le |W|}} |\{a \in \mathcal{A}(X) \colon \exists\, W' \in \pi^{-1}(W), W'_i = a\}|,$$

we have

$$d_\pi \le \min_{\substack{A \in S \\ A' \in S'}} \{|A \cap A'| \colon A \cap A' \ne \emptyset\}.$$

Let $W = W_0 \dots W_k$ be a magic block of π with a magic coordinate $0 \le c \le k$. Let

$$D = \{a \in \mathcal{A}(X) \colon W' \in \pi_{\mathrm{b}}^{-1}(W),\, W'_c = a\}.$$

Consider $U = W_0 \dots W_c$ and $V = W_c \dots W_k$. Let $\bar{U} = \bar{U}_0 \dots \bar{U}_c$ be the unique path in G which maps to U and $\bar{U}_0 = \pi_{\mathrm{b}}^{-1}(W_0)$. Then

$$\bar{U}_C = \{a \in \mathcal{A}(X) \colon F \in \pi_{\mathrm{b}}^{-1}(U),\, F_c = a\}.$$

Let $\bar{V} = \bar{V}_c \dots \bar{V}_k$ be a unique path in G' which maps to V and $\bar{V}_k = \pi_{\mathrm{b}}^{-1}(W_k)$. Then

$$\bar{V}_c = \{a \in \mathcal{A}(X) \colon G \in \pi_{\mathrm{b}}^{-1}(V),\, G_0 = a\}.$$

It follows that

$$\bar{U}_c \cap \bar{V}_c = \{a \in \mathcal{A}(X) \colon W' \in \pi_{\mathrm{b}}^{-1}(W),\, W'_c = a\} = D,$$

and therefore

$$d \ge \min_{\substack{A \in S \\ A' \in S'}} \{|A \cap A'| \colon A \cap A' \ne \emptyset\}.$$

$\square$

4.2 Class Degree

We will find a quantity analogous to the degree when π is an infinite-to-one factor code. This will be done by developing the following equivalence relation on X. Figure 4.4 illustrates Definition 4.2.1.

Definition 4.2.1. *Suppose (X, Y, π) is a factor triple and x, x' be two points in X. We say there is a* transition *from x to x' and denote it by $x \to x'$ if, for each integer n, there exists a point v in X so that*

1. *$\pi(v) = \pi(x) = \pi(x')$, and*
2. *$v_{-\infty}^{n} = x_{-\infty}^{n}$, $v_i^{\infty} = x_i'^{\infty}$ for some $i \geq n$.*

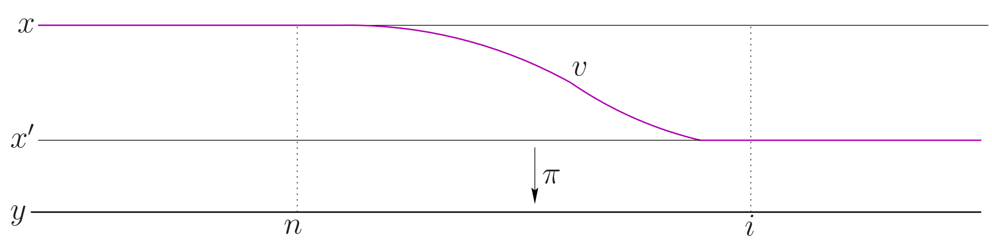

Figure 4.4: Transition from point x to x'

Write $x \nrightarrow x'$ if the above conditions do not hold. We write $x \sim x'$, and say x and x' are in the same transition class if $x \to x'$ and $x' \to x$.

To see that the relation $\sim$ is, indeed, an equivalence relation notice that reflexivity and symmetry are immediate from the definition. Here we show the transitivity property.

Claim. *The relation $\sim$, defined above, is transitive.*

Proof. Let a, b, $c \in X$. Suppose $a \sim b$ and $b \sim c$. To show $a \sim c$ we should show for a given integer n there is a point w in X so that $\pi(w) = \pi(a) = \pi(c)$, and moreover $w_{-\infty}^{n} = a_{-\infty}^{n}$, $w_j^{\infty} = c_j^{\infty}$ for some $j \geq n$.

Since $a \sim b$, we have $\pi(a) = \pi(b)$. Furthermore there is a point u in $\pi^{-1}(a)$ so that $u_{-\infty}^{n} = a_{-\infty}^{n}$, $u_i^{\infty} = b_i^{\infty}$ for some $i \geq n$. Having $b \sim c$ implies $\pi(b) = \pi(c)$ in addition to the existence of a point v in $\pi^{-1}(c)$ so that $v_{-\infty}^i = b_{-\infty}^i$, $v_j^\infty = c_j^\infty$ for some $j \geq i$. By now we have $\pi(a) = \pi(b) = \pi(c)$. Define the new point w as follows;

$$w_t = \begin{cases} u_t & t \leq i \\ v_t & t \geq i. \end{cases}$$

Note that since $u_i = v_i = b_i$, the point w lies in X. The fact that w is a concatenation of u and v and having $\pi(u) = \pi(a) = \pi(c) = \pi(v)$ imply that $\pi(w) = \pi(a) = \pi(c)$. Moreover, since $n \leq i \leq j$ we have $w_{-\infty}^n = u_{-\infty}^n = a_{-\infty}^n$. Also, $w_j^\infty = v_j^\infty = c_j^\infty$ (see Figure 4.5). It follows that $a \sim c$ and consequently $\sim$ is an equivalence relation on

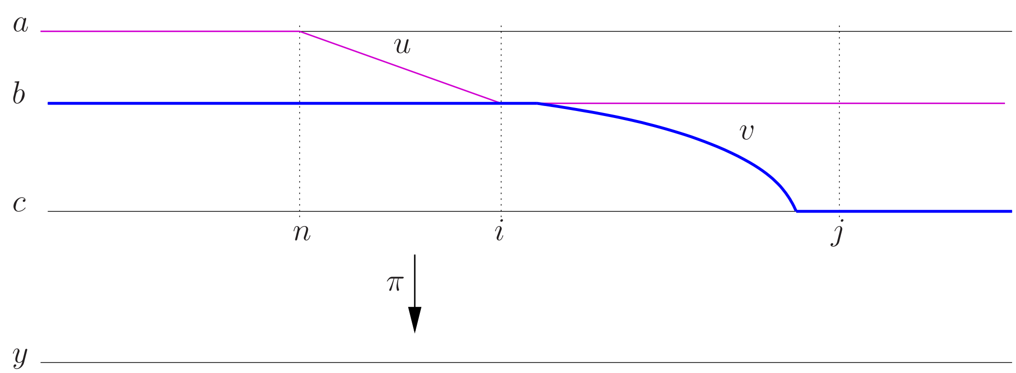

Figure 4.5: $a \sim b$ and $b \sim c$ imply $a \sim c$

X. $\square$

Denote the set of transition classes in X over $y \in Y$ by $\mathscr{C}_\pi(y)$. Sometimes we denote $\mathscr{C}_\pi(y)$ by only $\mathscr{C}(y)$ when there is no ambiguity in understanding π. We say $[x] \to [x']$ if $x \to x'$ (note that the relation is well-defined); use the notation $[x] \nrightarrow [x']$ otherwise.

The following fact is derived from Definition 4.2.1 immediately.

Fact 4.2.2. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let x, x' project to the same piont in Y, and $x_{a_i} = x'_{a_i}$ where $(a_i)_{i\in\mathbb{N}}$ is a strictly increasing sequence in $\mathbb{Z}$. Then we have $x \sim x'$.*

We mention that Fact 4.2.2 gives an obvious case when two points lie in the same transition class. Example 4.2.3 illustrates a more complicated case when two points are equivalent without having a common symbol at the same time.

Example 4.2.3. *Let $X_{\mathscr{F}} \subseteq \{a, b\}^{\mathbb{Z}}$ be a SFT with $\mathscr{F} = \{bb\}$. Let $\pi : X_{\mathscr{F}} \to \{0\}^{\mathbb{Z}}$, as shown by the labeled graph in Figure 4.6.*

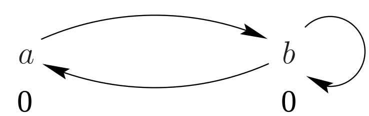

Figure 4.6: Graph for Example 4.2.3

Observe that points $x = \dots aba\overset{*}{b}aba \dots$ and $x' = \dots bab\overset{*}{a}bab \dots$ (where * indicates the 0-th position) have no common symbol at the same time but are equivalent. To see there is a transition from x to x', let n in $\mathbb{Z}$ and consider the point v where

$$v_i = \begin{cases} x_i & i \leq n \\ b & i = n+1 \\ x'_i & i > n+1. \end{cases}$$

It is clear that v holds the conditions in Definition 4.2.1. That $x' \to x$ is shown similarly.

The following two theorems give an upper bound on the number of transition classes over points of Y. Later in Theorem 4.3.12 we find the exact number of such classes over transitive points of Y.

Theorem 4.2.4. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let y in Y and let w be a symbol in Y occurring infinitely often in positive coordinates of y. Then $|\mathscr{C}(y)| \leq |\pi_b^{-1}(w)|$.*

Proof. Let $(a_j)_{j \in \mathbb{N}}$ be a strictly increasing sequence of integers such that $y_{a_j} = w$ for each $j \in \mathbb{N}$. Let $|\pi_{\mathrm{b}}^{-1}(w)| = d$ and suppose $\pi^{-1}(y)$ contains $d+1$ distinct transition classes $C_1, \dots, C_{d+1}$. Form a set $A = \{x_1, \dots, x_{d+1}\}$ where x_i is an arbitrary point of C_i. Since $\pi_{\mathrm{b}}^{-1}(w)$ contains exactly d symbols, it follows that for each $j \in \mathbb{N}$ there are at least two points in A with the same a_jth coordinate. The Pigeonhole Principle implies that there is a subsequence $(b_k)_{k\in\mathbb{N}}$ of $(a_j)_{j\in\mathbb{N}}$ and at least two points x and x' in A with $x_{b_k} = x'_{b_k}$ for each $k \in \mathbb{N}$. Fact 4.2.2 implies that $x \sim x'$. This contradicts the assumption that x and x' are in different transition classes. $\square$

A similar proof to the above implies the following theorem.

Theorem 4.2.5. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Then $|\mathscr{C}(y)| < \infty$ for each $y \in Y$. Moreover, $min_{w\in\mathcal{A}(Y)}\{|\pi_b^{-1}(w)|\}$ is the uniform upper bound on the number transition classes over each right transitive point of Y.*

Definition 4.2.6. *Let (X, Y, π) be a factor triple. The minimal number of transition classes over points of Y is called the* **class degree** *of π and denoted by c_π.*

Theorem 4.2.7 shows that conjugate factor triples have the same class degree.

Theorem 4.2.7. *Let (X, Y, π) and $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ be conjugate factor triples. Then we have $c_\pi = c_{\tilde{\pi}}$.*

We prove a more general theorem below.

Theorem 4.2.8. *Let (X, Y, π) and $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ be conjugate factor triples under conjugacies $\phi : X \to \tilde{X}$ and $\psi : Y \to \tilde{Y}$. Let y in Y. Then we have*

$$|\mathscr{C}_\pi(y)| = |\mathscr{C}_{\tilde{\pi}}(\psi(y))|.$$

Proof. For now assume X is a 1-step SFT and π is a 1-block factor code. Let $\phi : X \to \tilde{X}$ be a conjugacy induced by ϕ_{b} so that $\phi(x)_i = \phi_{\mathrm{b}}(x_{[i-m,i+t]})$ for some m, t in $\mathbb{N}$. Let y in Y and u, v in $\pi^{-1}(y)$. Since $\tilde{\pi} \circ \phi = \psi \circ \pi$ then $\phi(u), \phi(v)$ are in $\tilde{\pi}^{-1}(\psi(y))$. We show that there is a transition from u to v if and only if there is a transition from $\phi(u)$ to $\phi(v)$. Let n be an integer, then $u \to v$ implies that there is x in $\pi^{-1}(y)$ where

$$x_j = \begin{cases} u_j & \text{if } -\infty < j \leq n + t \\ v_j & \text{if } i \leq j < \infty \end{cases}$$

for some $i \geq t+n$. Clearly $\phi(x)$ is in $\tilde{\pi}^{-1}(\psi(y))$. Moreover, we have

$$\phi(x)_j = \begin{cases} \phi(u)_j & \text{if } -\infty < j < n \\ \phi(v)_j & \text{if } i+m \leq j < \infty. \end{cases}$$

Since n is arbitrary it follows that $\phi(u) \to \phi(v)$, and since ϕ is invertible the other implication follows similarly. This shows that C is a transition class over y if and only if $\phi(C)$ is a transition class over $\psi(y)$. It also shows that for any classes C_1 and C_2 in $\mathscr{C}_\pi(y)$ we have $C_1 \to C_2$ if and only if $\phi(C_1) \to \phi(C_2)$ which implies the equality of $|\mathscr{C}_\pi(y)|$ and $|\mathscr{C}_{\tilde{\pi}}(\psi(y))|$. This completes the proof of the result when X is a 1-step SFT and π is a 1-block factor code.

Now let (X, Y, π) be a general factor triple. By recoding, let $(\bar{X}, \bar{Y}, \bar{\pi})$ be a factor triple conjugate to (X, Y, π), such that $\bar{X}$ is a 1-step SFT and $\bar{\pi}$ is a 1-block factor code. Since conjugacy has the transitive property

$$(X, Y, \pi) \cong (\bar{X}, \bar{Y}, \bar{\pi}) \cong (\tilde{X}, \tilde{Y}, \tilde{\pi}),$$

the result follows immediately from the discussion above. □

4.3 Typical Points and Class Degree

In this section we show when $\pi : X \to Y$ is a factor code from a SFT X to an irreducible sofic shift Y, there are exactly c_π transition classes over a transitive point of Y. In order to show this, we introduce another quantity c^*_π, defined concretely in terms of blocks. Since the class degree is invariant under conjugacy of factor triples, by recoding we may focus only on 1-step SFTs and 1-block factor codes.

Definition 4.3.1. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let W be a Y-block of length $p+1$. Let n be an integer in $[0, p]$, and M be a subset of $\pi_b^{-1}(W_n)$. We say $U \in \pi_b^{-1}(W)$ **is routable through** $a \in M$ **at time** n if there is a block $U' \in \pi_b^{-1}(W)$ with $U'_0 = U_0$, $U'_n = a$, and $U'_p = U_p$. A triple (W, n, M) is called a **transition block** of π if every block in $\pi_b^{-1}(W)$ is routable through a symbol* *of M at time n. The cardinality of the set M is called the* **depth** *of the transition block (W, n, M).*

While Definition 4.3.1 seems complicated, Figure 4.7 illustrates it more clearly.

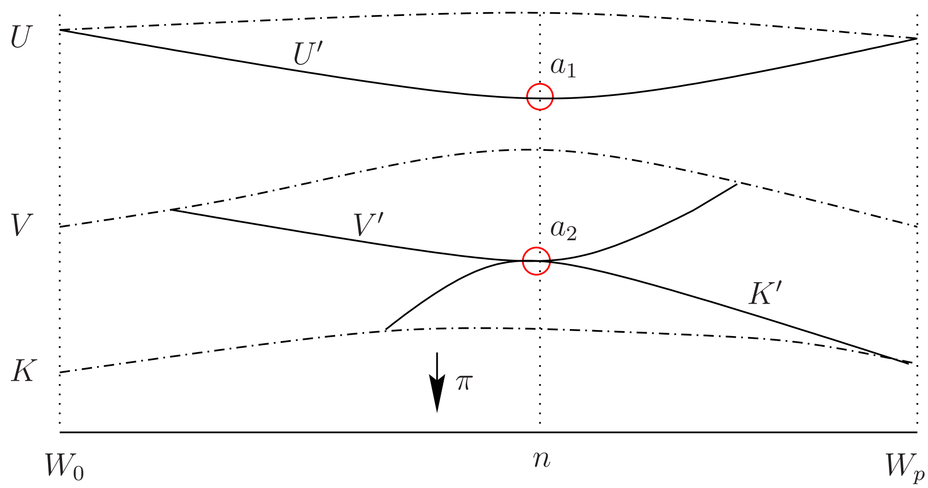

Figure 4.7: (W, n, M) is a transition block with $M = \{a_1, a_2\}$. The blocks U, V, $K \in \pi_b^{-1}(W)$ are routable through members of M at time n via blocks U', V', $K' \in \pi_b^{-1}(W)$.

Definition 4.3.2. *Let*

$$c_\pi^* = \min\{|M| \colon (W, n, M) \text{ is a transition block of } \pi\}.$$

A ***minimal transition block*** *of π is a transition block of depth c_π^*.*

Example 4.3.3. *In figure 4.8, we display an example of a labeled graph which defines an infinite-to-one 1-block factor code π. We see that $(001,\ 1,\ \{b\})$ is a minimal transition block of π of depth 1. For example, observe that block $U = aac$ is routable through b at time 1 by considering $U' = abc$.*

We need to develop some definitions and lemmas before proving Theorem 4.3.12 below. Given a factor triple (X, Y, π), a point y in Y and a block V occuring in y, not always all preimages of V can be extended to a point in X which projects to y

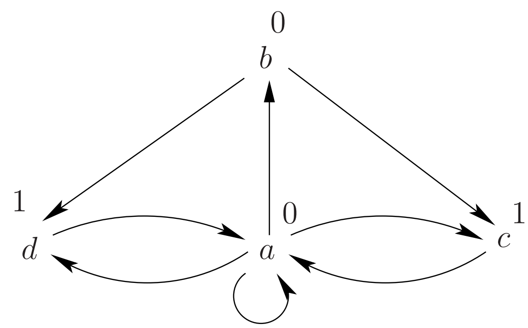

Figure 4.8: Graph for Example 4.3.3

under π. Here we show that in order to recognize those non-extendable blocks we do not need to deal with a bi-infinite sequence. A finite extension of a block is sufficient to determine if the block can be extended to a point above y.

Definition 4.3.4. *Let (X, Y, π) be a factor triple and y in Y. Let V be a block of Y occuring in y at a $\mathbb{Z}$-interval $[m, n]$. A Block W of Y is called a y-synchronizing extension of V if W occurs in y at $[m-l, n+l]$ for some integer $l \geq 0$ and*

$$\{U \in \pi_b^{-1}(V) \colon \exists x \in \pi^{-1}(y),\, x_{[m,n]} = U\}$$
$$= \{U \in \pi_b^{-1}(V) \colon \exists\, B \in \pi_b^{-1}(W),\, B_{[l,l+n-m]} = U\}.$$

Lemma 4.3.5. *Let (X, Y, π) be a factor triple and y in Y. Let V be a block occuring in y at a $\mathbb{Z}$-interval $[m, n]$. There is a y-synchronizing extension of V.*

Proof. Let

$$S = \{U \in \pi_{\mathrm{b}}^{-1}(V) \colon \exists\, x \in \pi^{-1}(y),\, x_{[m,n]} = U\}.$$

For $l \geq 0$ let

$$S^l = \{U \in \pi_{\mathrm{b}}^{-1}(V) \colon \exists\, B^l \in \pi_{\mathrm{b}}^{-1}(y_{[m-l,n+l]}),\, B^l_{[l,l+n-m]} = U\}.$$

We show for some l we have $S = S^l$. Then it will imply that $y_{[m-l,n+l]}$ is a y-synchronizing extension of V.

Clearly for all $l \geq 0$ we have $S \subseteq S^l$. We show $S^l \subseteq S$ for some l as follows. First note that since π is a factor code, S is not empty. It follows that the sequence $(S^l)_{l=0}^{\infty}$ is a non-increasing sequence of non-empty finite sets. Thus there is l such that for each $r \geq l$ we have $S^l = S^r$. We show $S^l \subseteq S$.

Let $U \in S^l$. Then U is in S^r for each $r \geq l$. It follows that for each $r \geq l$ there is B^r in $\pi_{\mathrm{b}}^{-1}(y_{[m-r,n+r]})$ with $B^r_{[r,r+n-m]} = U$. For each $r \geq l$ consider the cylinder set $C_{m-r}(B_r)$. The sequence $(C_{m-r}(B_r))_{r\geq l}$ is a non-increasing sequence of non-empty compact subsets of X. It follows that $\bigcap_{r\geq l} C_{m-r}(B_r)$ is not empty. Let $x \in \bigcap_{r\geq l} C_{m-r}(B_r)$. Then $x_{[m,n]} = U$ and $\pi(x) = y$ which implies that U is in S. $\square$

Here we introduce a relation between symbols and transition classes.

Definition 4.3.6. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let y be in Y. Let n be in $\mathbb{Z}$, i a symbol in $\pi_b^{-1}(y_n)$, and C be a transition class in $\mathscr{C}(y)$. Say i **belongs to C at time** n, and denote it by $i \in S_n(C)$, if*

$$\{D \in \mathscr{C}(y)\colon \text{there is } x \text{ in } D \text{ with } x_n = i\} = \{D \in \mathscr{C}(y) : C \to D\}.$$

In other words, the set of transition classes in $\mathscr{C}(y)$ containing a point with nth coordinate being i, is the same as the set of transition classes to which there is a transition from the class C (including C itself).

*Say i is **transient at time** n if there is no transition class C in $\mathscr{C}(y)$ for which $i \in S_n(C)$.*

Example 4.3.7 illustrates the definition above.

Example 4.3.7. *Consider the directed labeled graph in Figure 4.9 which presents a 1-block factor code $\pi : X \to Y$ where $X \subseteq \{a, b, c, d, e, f, g\}^{\mathbb{Z}}$ and $Y \subseteq \{0, 1\}^{\mathbb{Z}}$. Let y be the point $\cdots 01\overset{*}{0}1010\cdots$ in Y. By Definition 4.2.1, there are 3 distinct transition classes in $\mathscr{C}(y)$ as follows:*

1. *Transition Class $C_1 = [x_1] = \{x_1\}$ where $x_1 = \cdots ba\overset{*}{b}abab\cdots$.*

2. *Transition Class $C_2 = [x_2]$ where x_2 is a point in $\pi^{-1}(y)$ which does not contain symbols e or g but contains only d's and f's from some time onwards, for example: $x_2 = \cdots ba\overset{*}{b}cdfd\cdots$.*

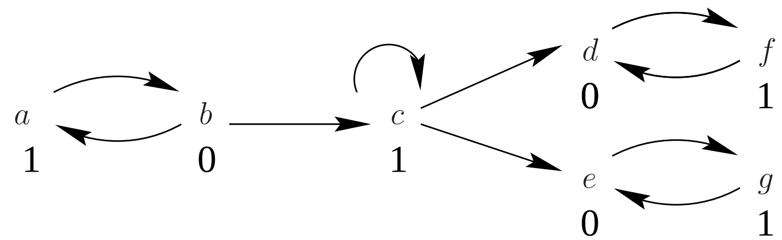

Figure 4.9: Graph for Example 4.3.7

3. *Transition Class* $C_3 = [x_3]$ *where* x_3 *is a point in* $\pi^{-1}(y)$ *which does not contain symbols* d *or* f *but contains only* e*'s and* g*'s from some time onwards, for example:* $x_3 = \cdots eg\overset{*}{e}gege\cdots$.

Clearly $C_1 \to C_2$ *and* $C_1 \to C_3$*, but not vice versa.* $C_2 \nrightarrow C_3$*,* $C_3 \nrightarrow C_2$*, and symbol* c *is transient at any time.*

Lemma 4.3.8. *Let* $\pi : X \to Y$ *be a* 1*-block factor code from a 1-step SFT* X *to a sofic shift* Y*. Let* x *be a point in* X*. There is* $m < \infty$ *such that for each* $n \geq m$ *the symbol* x_n *is not transient. In fact,* m *can be found in such a way that for each* $n \geq m$*,* x_n *belongs to the class* $[x]$ *at time* n*; i.e.,* $x_n \in S_n[x]$*.*

Proof. Let $\pi(x) = y$. Suppose that the transition class C in $\mathscr{C}(y)$ satisfies $[x] \nrightarrow C$. Then there exists $i < \infty$ such that for any z in $\pi^{-1}(y)$ with $z_i = x_i$ we have $z \notin C$. Denote the smallest such i by i^C. Note that i^C may be $-\infty$ (but not $+\infty$). Let $m = \max\{i^C : C \in \mathscr{C}(y), [x] \nrightarrow C\}$. Let $n \geq m$. We show

$$\{C \in \mathscr{C}(y) : [x] \to C\} = \{C \in \mathscr{C}(y) \colon \text{there is } z \in C \text{ with } z_n = x_n\}, \tag{4.3.1}$$

which will imply the result $x_n \in S_n[x]$.

Let C be a transition class in $\mathscr{C}(y)$ with $[x] \to C$. It follows that $x \to z$ for some z in C. Therefore, there is a point u in $\pi^{-1}(y)$ as follows.

$$u_t = \begin{cases} x_t & \text{if } -\infty < t \leq n \\ z_t & \text{if } l \leq t < \infty, \end{cases}$$

for some $l \geq n$. Since $u_t = z_t$ from time l onwards, it follows that u is in the class C. Moreover, we have $u_n = x_n$ by the construction. It follows that C belongs to the set in the right hand side of Equation (4.3.1).

Now let C be an transition class in $\mathscr{C}(y)$ with $[x] \nrightarrow C$. Suppose, for a contradiction, C contains a point z with $z_n = x_n$. Form the point u as follows.

$$u_t = \begin{cases} x_t & \text{if } -\infty < t \leq n \\ z_t & \text{if } n < t < \infty. \end{cases}$$

Note that since X is a 1-step SFT and $z_n = x_n$, the point u is in X. Moreover u maps to y by the construction. Noticing that $u_t = z_t$ from the time n onwards, shows that u lies in the class C. Recall that $n \geq i^C$. By the argument in the first paragraph of the proof, any point in $\pi^{-1}(y)$ with i^Cth coordinate equal to x_{i^C} does not belong to C. This contradicts that u is a point in C. $\square$

Lemma 4.3.9. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let $y \in Y$. There is an integer $m < \infty$ such that for each $n \geq m$ and C in $\mathscr{C}(y)$ there is a symbol i in $\pi_b^{-1}(y_n)$ which belongs to the class C at time n; i.e., $i \in S_n(C)$.*

Proof. Let $\mathscr{C}(y) = \{C_1, \dots, C_d\}$ for some finite d. Let $A = \{x^{(1)}, \dots, x^{(d)}\}$ be a set containing an arbitrary point $x^{(i)}$ of C_i for each C_i. By Lemma 4.3.8 there is a finite integer n_{C_i} such that for each $n \geq n_{C_i}$ the symbol $x_n^{(i)}$ belongs to C_i) at time n. Let $m = \sup\{n_{C_i} \colon 1 \leq i \leq d\}$. Note that $m < \infty$. Moreover, for each $n \geq m$ and C_i the symbol $x_n^{(i)}$ belongs to the class C_i at time n. $\square$

Proposition 4.3.10. *Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let ν be an invariant measure on Y. Then*

$$\nu\left(\{y \in Y \colon \forall C \in \mathscr{C}(y), \forall n \in \mathbb{Z} \text{ there is } i \in \pi_b^{-1}(y_n) \text{ with } i \in S_n(C)\}\right) = 1.$$

Proof. Let y in Y and let $m(y)$ be the infimum of the set of m's with the properties given in the statement of Lemma 4.3.9. We show

$$\nu\left(\{y \in Y : m(y) = -\infty\}\right) = 1$$

which will imply the result immediately. Note that $C \in \mathscr{C}(y)$ if and only if $T(C) \in \mathscr{C}(T(y))$. This shows $m(T(y)) = m(y) - 1$. For an integer $k < \infty$ let $A_k = \{y \in Y : m(y) = k\}$. We have $T(A_k) = A_{k-1}$. Since ν is T-invariant it follows that $\nu(A_k) = 0$. Therefore $m(y) = -\infty$ for ν-a.e. y in Y. □

Definition 4.3.11. *Let X be a shift space. We say the point x in X is measure recurrent if for each block B which occurs in x, B occurs infinitely often in the positive coordinates of x.*

Note that the set of measure recurrent point of X has measure one with respect to any ergodic measure on X.

Theorem 4.3.12. *Let π be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let y be a measure recurrent point of Y. Let d be the minimal depth of a transition block which occurs infinitely often in the positive coordinates of y. Then $|\mathscr{C}(y)| = d$.*

Proof. First we apply the same method we used in the proof of Theorem 4.2.4 to show that $|\mathscr{C}(y)| \leq d$. Let (W, n, M) with $|W| = p + 1$ be a transition block of depth d occuring infinitely often in the positive coordinates of y. Let $(t_m)_{m\in\mathbb{N}}$ with $t_{m+1} > p + t_m$ be a sequence of integers such that

$$y_{t_m} \ldots y_{p+t_m} = W.$$

Suppose $\mathscr{C}(y)$ contains more than d transition classes. Let A be a set containing one point from each of these transition classes. By the definition of a transition block, for each u in A there exists a point $x^{(u,m)} \in \pi^{-1}({}_{t_m}[W])$ with $x_{t_m} = u_{t_m}$, $x_{n+t_m} \in M$, and $x_{p+t_m} = u_{p+t_m}$. Construct a new point u' which agrees with u everywhere except at positions $t_m, \ldots, s + t_m$ where it agrees with $x^{(u,m)}$ for each m in $\mathbb{N}$. In other words we have

$$u'_i = \begin{cases} u_i & \text{if } -\infty < i < t_1,\ s + t_m < i < t_{m+1} \\ x_i^{(u,m)} & \text{if } t_m \leq i \leq s + t_m. \end{cases}$$

The assumption that X is a 1-step SFT guarantees that u' belongs to X, and Fact 4.2.2 implies that $u' \sim u$. For each element u of A construct such a point u' and collect them in a set denoted by A'. Since we assumed $d < |A'|$, for each m in $\mathbb{N}$ there must be at least two points in A' that agree in the $(n + t_m)$th position. It follows that there exists a subsequence $(k_l)_{l\in\mathbb{N}}$ of $(t_m)_{m\in\mathbb{N}}$ and two distinct points $u', v' \in A'$ such that for each l in $\mathbb{N}$ we have $u'_{n+k_l} = v'_{n+k_l}$. Then by Fact 4.2.2 we have $u' \sim v'$ and consequently $u \sim v$, contrary to the fact that the points u and v are chosen from distinct transition classes.

Now we show that $|\mathscr{C}(y)| \geq d$ by finding a transition block of depth $|\mathscr{C}(y)|$ occuring in y. We construct such a transition block in the following 4 stages. Let $h = |\mathscr{C}(y)|$ and suppose $\mathscr{C}(y) = \{C_1, \ldots, C_h\}$. Choose a finite integer n_0 satisfying the properties given in the statement of Lemma 4.3.9.

Stage *1*. We claim there is $n_1 \in [n_0, \infty)$ such that for each $x \in \pi^{-1}(y)$, there is $n_0 \leq t \leq n_1$ so that x_t is not transient.

Proof. Suppose there is no such n_1. It follows that for each $j \geq n_0$ there is a point $x^{(j)}$ in $\pi^{-1}(y)$ such that $x_l^{(j)}$ is transient for all $n_0 \leq l \leq j$. Consider the sequence $(x^{(j)})_{j\in\mathbb{Z}}$, and let x be the limit of a convergent subsequence of it. Clearly $x \in \pi^{-1}(y)$. However, x_l is transient for each $l \geq n_0$, contradicting Lemma 4.3.8. $\square$

Stage *2*. We claim there is $n_2 \in [n_1, \infty)$ and a set of symbols

$$M' = \{a_1, \ldots, a_h\}$$

with $a_e \in S_{n_2}(C_e)$ such that for each i in $S_{n_1}(C_e)$ there is a block

$$U \in \pi_{\mathrm{b}}^{-1}(y_{n_1} y_{n_1+1} \ldots y_{n_2-1} y_{n_2})$$

which begins with i and ends with a_e. See Figure 4.10.

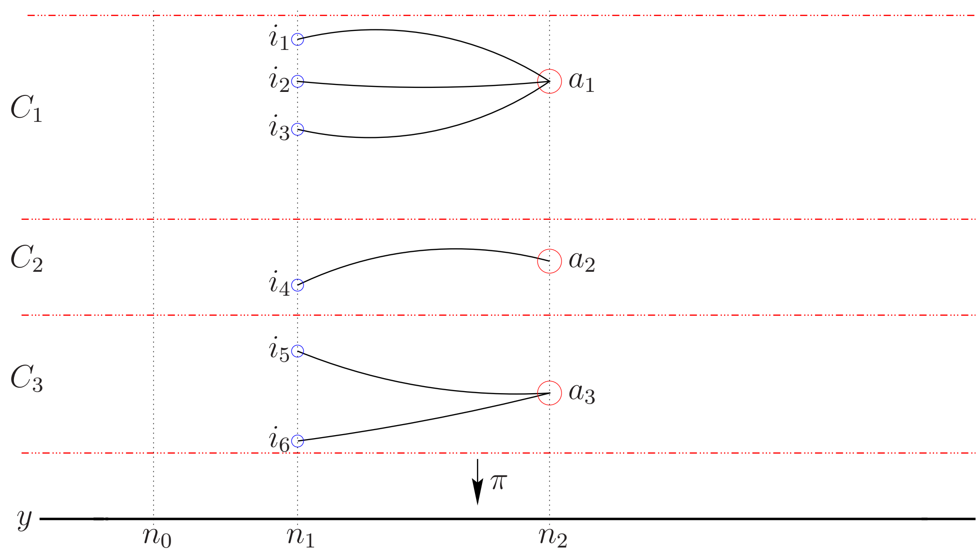

Figure 4.10: An example illustrating Stage 2. $\mathscr{C}(y) = \{C_1,\ C_2,\ C_3\}$, $i_1, i_2, i_3 \in S_{n_1}(C_1)$, $i_4 \in S_{n_1}(C_2)$, $i_5, i_6 \in S_{n_1}(C_3)$, and $M' = \{a_1,\ a_2,\ a_3\}$.

Proof. Let C_e be a transition class in $\mathscr{C}(y)$ and x^e be a point in C_e. For each symbol i in $S_{n_1}(C_e)$ (non-empty by the choice of n_0), there is a bi-infinite sequence z^i in the class C_e with $z^i_{n_1} = i$ such that z^i matches x^e from some time $k^i \in [n_1, \infty)$ onwards. Let

$$k^e = \max\{k^i : i \in S_{n_1}(C_e)\},$$

and

$$n_2 = \max\{k^e : C_e \in \mathscr{C}(y)\}.$$

Note that $n_2 < \infty$. Rename $x^e_{n_2}$ for each $C_e \in \mathscr{C}(y)$ as a_e, and let $M' = \{a_1, \ldots, a_h\}$. $\square$

Stage *3*. We claim there is $n_3 \in [n_2, \infty)$ such that for each point x in $\pi^{-1}(y)$ there is a point x' in $\pi^{-1}(y)$ so that $x'_r = x_r$ for every $r \in (-\infty, n_0] \cup [n_3, \infty)$, and $x'_{n_2} \in M'$. See Figure 4.11.

Proof. Let x be a point in $\pi^{-1}(y)$. By stage 1, there is $n_0 \leq t \leq n_1$ such that x_t is not transient; i.e., x_t belong to some transition class C_e at time t. It follows that there is

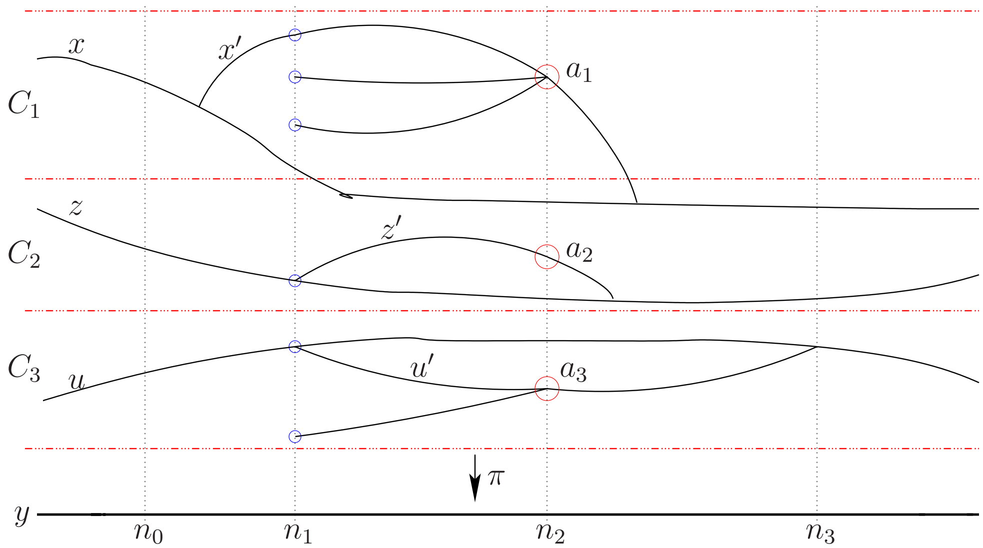

Figure 4.11: Graph for Stage 3. $M' = \{a_1, a_2, a_3\}$. $x, x' \in \pi^{-1}(y)$, $x'_r = x_r$ for each $r \in (-\infty, n_0] \cup [n_3, \infty)$, and $x'_{n_2} \in M'$. Same for z, z' and u, u'.

a point u in the class C_e with $u_t = x_t$. The new block

$$\ldots x_{n_0} \ldots x_t u_{t+1} \ldots u_{n_1} \tag{4.3.2}$$

is an allowable block of X (because X is a 1-step SFT). Moreover, this block maps to the block

$$\ldots y_t \ldots y_{n_1}.$$

Note that the symbol u_{n_1} belongs to the class C_e at time n_1. Then by Stage 2, there is a block starting at u_{n_1} ending at a_e which maps to

$$y_{n_1} \ldots y_{n_2}.$$

Denote this block by

$$u_{n_1} v_{n_1+1} \ldots v_{n_2-1} a_e. \tag{4.3.3}$$

Connect blocks in Equations (4.3.2) and (4.3.3) at u_{n_1} to get the following allowable block of X

$$\dots x_{n_0} \dots x_t u_{t+1} \dots u_{n_1} v_{n_1+1} \dots v_{n_2-1} a_e$$

Observe that having x_t in $S_t(C_e)$ implies that x must belong to a transition class C_f with $C_e \to C_f$. On the other hand having a_e in $S_{n_2}(C_e)$ implies that there is a point b in C_e with

$$b_{(-\infty, n_2]} = \dots x_{n_0} \dots x_t u_{t+1} \dots u_{n_1} v_{n_1+1} \dots v_{n_2-1} a_e$$

which matches x from some time j onwards for some $n_2 \leq j < \infty$. Let

$$n_x = \min_{n_2 \leq j < \infty} \{j : \exists x' \in \pi^{-1}(y) \text{ with } x'_r = x_r \forall r \in (-\infty, n_0] \cup [j, \infty),\ \ x'_{n_2} = a_e\}.$$

We claim that there is $n^e < \infty$ such that for each $x \in \pi^{-1}(y)$ with $x_t \in S_t(C_e)$ for some $n_0 \leq t \leq n_1$, there exists a point $x' \in \pi^{-1}(y)$ with $x'_r = x_r$ for each $r \in (\infty, n_0] \cup [n^e, \infty)$, and $x'_{n_2} = a_e$. Then letting

$$n_3 = \max\{n^e : C_e \in \mathscr{C}(y)\} < \infty$$

will complete the proof of Stage 3.

Suppose, for a contradiction, that there does not exist such an n^e. It follows that there is a sequence $x^{(l)}$ with $x^{(l)}_{t_l} \in S_{t_l}(C_e)$ for some $n_0 \leq t_l \leq n_1$, such that $\lim_{l\to\infty} n_{x^{(l)}} = \infty$. Let x^* be the limit of a convergent subsequence of $x^{(l)}$. Clearly x^* is in $\pi^{-1}(y)$. Assume $n_{x^*} = g < \infty$ then for every point k in the chosen subsequence of $x^{(l)}$ with $d(x^*, k) < 1/2^g$ we have $n_k = g$. This contradicts the assumption $\lim_{l\to\infty} n_{x^{(l)}} = \infty$. Therefore we must have $n_{x^*} = \infty$ which is a contradiction of the fact that for each x in $\pi^{-1}(y)$, $n_x < \infty$. $\square$

Stage *4*. Let W be a y-synchronizing extension of the block

$$V = y_{n_0} \dots y_{n_3}.$$

We claim W is a transition block of depth d.

Proof. Let $U \in \pi_{\mathrm{b}}^{-1}(W)$. By the definition of y-synchronizing extension, there is x in

$\pi^{-1}(y)$ with $x_{[n_0, n_3]} = U$. Then by Stages 1, 2, and 3 there is a point x' in $\pi^{-1}(y)$ with $x'_r = x_r$ for each $r \in (-\infty, n_0] \cup [n_3, \infty)$, and $x'_{n_2} \in M'$. Having $x'_{n_0} = x_{n_0}$, $x'_{n_2} \in M'$, $x'_{n_3} = x_{n_3}$ and $\pi_{\mathrm{b}}(x'_{[n_0, n_3]}) = V$ simply means that U is routable through a symbol of M' at time $n_2 + l$. Since U in $\pi_{\mathrm{b}}^{-1}(W)$ is arbitrary it follows that W is a transition block. Because M' has h elements it follows that W is a transition block of depth h. □

□

Corollary 4.3.13. *Let $\pi : X \to Y$ be a factor code from a SFT X to an irreducible sofic shift Y. Then*

1. $c_\pi = c^*_\pi$.
2. *There are exactly c_π transition classes over every right transitive point of Y.*
3. *Given an ergodic measure ν on Y, the number of transition classes over every ν-generic point of Y is the same.*

We now show that for a finite-to-one factor code, the degree and the class degree of the code are the same.

Theorem 4.3.14. *Let $\pi : X \to Y$ be a finite-to-one factor code from a SFT to an irreducible sofic shift Y. Then $c_\pi = d_\pi$.*

Proof. Theorem 4.1.5 and Theorem 4.2.7 imply that the degree of a code and the class degree of a code are both invariant under conjugacy of factor triples. So by recoding, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code. Let y be a transitive point of Y. By Proposition 4.1.8 the preimages of y are mutually seperated. It follows that each transition class above y has only one point. Therefore, $c_\pi = |\pi^{-1}(y)| = d_\pi$. □

Chapter 5

Bounding the Number of Ergodic Measures of relative maximal entropy

In Chapter 2 we showed that although every 1-dimensional irreducible SFT has a unique ergodic measure of maximal entropy (Parry measure), there can be more than one ergodic measure of relative maximal entropy over an ergodic measure; i.e., given a factor triple (X, Y, π) and a fully supported ergodic measure ν on Y, there can exist more than one ergodic measure on X that projects to ν under π and has maximal entropy among measures in the fibre $\pi^{-1}\{\nu\}$. In this chapter we show that the number of such measures can be no more than the class degree of π. In the first section we review a known bound for the number of measures of relative maximal entropy and discuss an unsatisfactory fact about it which is improved by the new bound. In section 5.2 we prove the main theorem which is showing that c_π is an upper bound on the number of such measures.

5.1 Old Bound on the Number of Ergodic Measures of Relative Maximal Entropy

In 2003, Petersen, Quas, and Shin [24] found an upper bound on the number of measures of relative maximal entropy.

Theorem 5.1.1. *[24, Corollary 1] Let $\pi : X \to Y$ be a 1-block code from a 1-step SFT X to a sofic shift Y. Let ν be an ergodic measure on Y and $N_\nu(\pi) = \min|\{\pi_b^{-1}(w) \colon w \in \mathcal{A}(Y),\, \nu[w] > 0\}|$. The number of ergodic measures of maximal entropy over ν is at most $N(\pi)$.*

We need the following theorem and Corollary to explain why we are not fully satisfied with this bound.

Theorem 5.1.2. *[29, Theorem 4.11] Let X and $\tilde{X}$ be conjugate shift spaces under a conjugacy $\phi : X \to \tilde{X}$. Let μ be an invariant measure on X. Then $\tilde{\mu} = \mu \circ \phi^{-1}$ is an invariant measure on $\tilde{X}$ and*

$$h_\mu(T_X) = h_{\tilde{\mu}}(T_{\tilde{X}})$$

Corollary 5.1.3 follows from Theorem 5.1.2 immediately.

Corollary 5.1.3. *Let (X, Y, π) and $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ be conjugate factor triples. Let ν be an ergodic measure on Y and $\tilde{\nu}$ be the corresponding ergodic measure on $\tilde{Y}$. The number of ergodic measures of relative maximal entropy over ν is the same as the number of ergodic measures of relative maximal entropy over $\tilde{\nu}$.*

This bound suffers from not being invariant under conjugacy. For example, the full 2-shift and any higher block presentation of it give different bounds on the number of ergodic measures of maximal entropy which map to the trivial measure on the full 1-shift.

One possibility to avoid having a non-invariant upper bound is to take the minimum of the bound in [24] over all conjugate factor triples with 1-block factor codes. This even improves the original bound. However, there is no known algorithm for computing this minimum.

Here we show that given an ergodic measure ν on Y, the number of transition classes over a ν-generic point of Y is an upper bound on the number of measures of relative maximal entropy over ν. Theorem 4.2.8 verifies that this bound is invariant under conjugacy of factor triples and Proposition 5.1.4 shows that it beats the bound mentioned above obtained by minimizing the bound in [24] over conjugate factor triples.

Proposition 5.1.4. *Let $\pi : X \to Y$ be a 1-block factor code from a SFT X to a sofic shift Y. Let ν be an ergodic measure on Y. Then we have $|\mathscr{C}_\pi(y)| \leq \min\{N_\nu(\tilde{\pi})\colon (X,Y,\pi) \cong (\tilde{X},\tilde{Y},\tilde{\pi}),\ \tilde{\pi}$ is 1-block$\}$ where y is a ν-generic point of Y. Equality holds if π is finite-to-one and ν is fully supported.*

Proof. Let $\min\{N_\nu(\tilde{\pi})\colon (X,Y,\pi) \cong (\tilde{X},\tilde{Y},\tilde{\pi}),\ \tilde{\pi}$ is 1-block$\}$ occur at a factor triple $(\tilde{X},\tilde{Y},\tilde{\pi})$ where X is conjugate to $\tilde{X}$ under a conjugacy $\phi : X \to \tilde{X}$, Y is conjugate to $\tilde{Y}$ under a conjugacy $\psi : Y \to \tilde{Y}$, and $\tilde{\pi} \circ \phi = \psi \circ \pi$. Denote the measure $\nu \circ \psi^{-1}$ by $\tilde{\nu}$. Let $N_{\tilde{\nu}}(\tilde{\pi}) = \min\{|\tilde{\pi}_{\mathrm{b}}^{-1}(w)|\colon w \in \mathcal{A}(\tilde{Y}), \tilde{\nu}[w] > 0\}$ occur at a symbol $\tilde{w}$ in $\mathcal{A}(\tilde{Y})$. Let $\tilde{y}$ be a ν-generic point of $\tilde{Y}$. There is a strictly increasing sequence of integers $(a_i)_{i\in\mathbb{N}}$ with $y_{a_i} = w$. Theorem 4.2.5 implies that $|\mathscr{C}_{\tilde{\pi}}(\tilde{y})| \leq |\tilde{\pi}_{\mathrm{b}}^{-1}(w)|$. Let $y = \psi^{-1}(y)$. Then by Theorem 4.2.8 we have $|\mathscr{C}_\pi(y)| = |\mathscr{C}_{\tilde{\pi}}(\tilde{y})| \leq |\tilde{\pi}_{\mathrm{b}}^{-1}(w)|$ which completes the proof of the first part of the result.

Now suppose π is finite-to-one and ν is fully supported. Since ν is fully supported we have

$$N_\nu(\pi) = \min\{|\tilde{\pi}_{\mathrm{b}}^{-1}(w)|\colon w \in \mathcal{A}(\tilde{Y})\}.$$

Therefore $\min\{N_\nu(\tilde{\pi})\colon (X,Y,\pi) \cong (\tilde{X},\tilde{Y},\tilde{\pi}),\ \tilde{\pi}$ is 1-block$\}$ occurs at a factor triple $(\tilde{X},\tilde{Y},\tilde{\pi})$ where $\tilde{\pi}$ has a magic symbol $\tilde{w}$ (such factor triple exists by Theorem 4.1.6). By Proposition 4.1.8 we have $|\tilde{\pi}_{\mathrm{b}}^{-1}(\tilde{w})| = d_{\tilde{\pi}}$ which implies

$$\min\{N_\nu(\tilde{\pi})\colon (X,Y,\pi) \cong (\tilde{X},\tilde{Y},\tilde{\pi}),\ \tilde{\pi} \text{ is 1-block}\} = d_{\tilde{\pi}}.$$

Since the degree of a code is invariant under conjugacy of factor triples we deduce that

$$\min\{N_\nu(\tilde{\pi})\colon (X,Y,\pi) \cong (\tilde{X},\tilde{Y},\tilde{\pi}),\ \tilde{\pi} \text{ is 1-block}\} = d_\pi.$$

Recalling the agreement of the degree and the class degree of π when π is finite-to-one, completes the proof. $\square$

Note that equality in Proposition 5.1.4 does not always hold. For example, consider the trivial factor code $\pi : \{0,1\}^{\mathbb{Z}} \to \{0\}^{\mathbb{Z}}$. Then $c_\pi = 1$; however, if $\left(\tilde{X}, \{0\}^{\mathbb{Z}}, \tilde{\pi}\right)$ is a factor triple conjugate to $\left(\{0,1\}^{\mathbb{Z}}, \{0\}^{\mathbb{Z}}, \pi\right)$ then $\mathcal{A}(\tilde{X})$ must be strictly greater than 1 and therefore $N_\nu(\tilde{\pi}) > 1$ for any ergodic measure ν on Y.

5.2 New Bound on the Number of Ergodic Measures of Relative Maximal Entropy

5.2.1 Relatively Independent Joining

Definition 5.2.1. *Let $\pi : X \to Y$ be a 1-block factor code from a SFT X to a sofic shift Y and ν be an ergodic measure on Y. Let $\mu_1, \ldots, \mu_n$ be invariant measures in the fibre $\pi^{-1}\{\nu\}$. The* relatively independent joining *of $\mu_1, \ldots, \mu_n$ over ν, denoted by $\tilde{\mu} = \mu_1 \otimes \cdots \otimes_\nu \mu_n$, is defined as follows; If $A_1, \ldots, A_n$ are measurable subsets of X then*

$$\hat{\mu}(A_1 \times \cdots \times A_n) = \int_Y \prod_{i=1}^{n} \mathbb{E}_{\mu_i}(\mathbf{1}_{A_i} | \pi^{-1}\mathscr{B}_Y) \circ \pi^{-1} \, d\nu.$$

Claim. *Writing p_i for the projection $X^n \to X$ onto the i'th coordinate, for $\hat{\mu}$-almost every $\hat{x} \in X^n$, $\pi(p_i(\hat{x}))$ is the same for each i.*

Proof. Let

$$A = \bigcup_{i \neq j} \{(x_1, \ldots, x_n) \in X^n \colon \pi(x_i) \neq \pi(x_j)\}.$$

We show for different i and j we have

$$\hat{\mu}\left(\{(x_1, \ldots, x_n) \in X^n \colon \pi(x_i) \neq \pi(x_j)\}\right) = 0.$$

Then it follows that $\hat{\mu}(A) = 0$.

Without loss of generality assume $i = 1$ and $j = 2$. We have

$$\{(x_1, \ldots, x_n) \in X^n \colon \pi(x_1) \neq \pi(x_2)\} = \bigcup_{r \in \mathbb{Z}} \bigcup_{\substack{a,b \in \mathcal{A}(Y) \\ a \neq b}} \pi^{-1}\left({}_r[a]\right) \times \pi^{-1}\left({}_r[b]\right) \times X \times \cdots \times X.$$

Therefore, it suffices to show that for $a \neq b$

$$\hat{\mu}\left(\pi^{-1}({}_r[a]) \times \pi^{-1}({}_r[b]) \times X \times \cdots \times X\right) = 0.$$

Let $a \neq b$. The characteristic function $\mathbf{1}_{\pi^{-1}({}_r[a])}$ is zero on the set

$$X - \pi^{-1}\left({}_r[a]\right) = \pi^{-1}\left(Y - {}_r[a]\right),$$

which is denoted by

$$\pi^{-1}({}_r[a]^c).$$

By the definition of conditional expectation, since $\pi^{-1}\left({}_r[a]^c\right)$ belongs to the σ-algebra $\pi^{-1}\mathscr{B}_Y$ we have

$$\begin{aligned} \int_{\pi^{-1}({}_r[a]^c)} \mathbb{E}_{\mu_1}(\mathbf{1}_{\pi^{-1}({}_r[a])} | \pi^{-1}\mathscr{B}_Y)\, d\mu_1 &= \int_{\pi^{-1}({}_r[a]^c)} \mathbf{1}_{\pi^{-1}({}_r[a])}\, d\mu_1 \\ &= 0. \end{aligned} \tag{5.2.1}$$

On the other hand, we have

$$\int_{{}_r[a]^c} \mathbb{E}_{\mu_1}(\mathbf{1}_{\pi^{-1}({}_r[a])} | \pi^{-1}\mathscr{B}_Y) \circ \pi^{-1}\, d\nu = \int_{\pi^{-1}({}_r[a]^c)} \mathbb{E}_{\mu_1}(\mathbf{1}_{\pi^{-1}({}_r[a])} | \pi^{-1}\mathscr{B}_Y)\, d\mu_1.$$

The combination of Equation (5.2.1) and the above equation implies that for ν-almost every y in ${}_r[a]^c$,

$$\mathbb{E}_{\mu_1}(\mathbf{1}_{\pi^{-1}({}_r[a])} | \pi^{-1}\mathscr{B}_Y) \circ \pi^{-1}(y) = 0.$$

A similar argument shows that for ν-almost every y in ${}_r[a]$,

$$\mathbb{E}_{\mu_2}(\mathbf{1}_{\pi^{-1}({}_r[b])}|\pi^{-1}\mathscr{B}_Y) \circ \pi^{-1}(y) = 0.$$

It follows that for ν-almost $y \in Y$ we have

$$\mathbb{E}_{\mu_1}(\mathbf{1}_{\pi^{-1}({}_r[a])}|\pi^{-1}\mathscr{B}_Y) \circ \pi^{-1}(y) \times \mathbb{E}_{\mu_2}(\mathbf{1}_{\pi^{-1}({}_r[b])}|\pi^{-1}\mathscr{B}_Y) \circ \pi^{-1}(y) = 0.$$

Thus we have,

$$\begin{aligned}
&\hat{\mu}\left(\{\pi^{-1}\left({}_r[a]\right) \times \pi^{-1}\left({}_r[b]\right) \times X \times \cdots \times X\}\right) \\
&= \int_Y \mathbb{E}_{\mu_1}(\mathbf{1}_{\pi^{-1}({}_r[a])}|\pi^{-1}\mathscr{B}_Y) \circ \pi^{-1} \times \mathbb{E}_{\mu_2}(\mathbf{1}_{\pi^{-1}({}_r[a])}|\pi^{-1}\mathscr{B}_Y) \circ \pi^{-1} \, d\nu. \\
&= 0.
\end{aligned}$$

□

5.2.2 New Bound

In this section we want to present a proof of the following theorem.

Theorem 5.2.2. *Let (X, Y, π) be a factor triple and ν be a fully supported ergodic measure on Y. The number of ergodic measures of relative maximal entropy over ν is at most c_π.*

We will use Theorem 5.2.3 below which is the main theorem from [24] to prove a stronger theorem (Theorem 5.2.4).

Theorem 5.2.3. *[24, Theorem 1] Let $\pi : X \to Y$ be a 1-block factor code from a 1-step SFT X to a sofic shift Y. Let ν be an ergodic measure on Y, and two distinct ergodic measures μ_1 and μ_2 be measures of relative maximal entropy over ν. Then $(\mu_1 \otimes \mu_2)\{(u, v) \in X \times X \colon u_0 = v_0\} = 0$.*

Theorem 5.2.4. *Let (X, Y, π) be a factor triple. Let ν be an ergodic measure on Y, and two distinct measures μ_1 and μ_2 be ergodic measures of relative maximal entropy over ν. Then $(\mu_1 \otimes \mu_2)\{(u, v) \in X \times X \colon u \sim v\} = 0$.*

Proof. First we show that, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code with a magic symbol. Suppose $(\tilde{X}, \tilde{Y}, \tilde{\pi})$ is a factor triple conjugate to (X, Y, π) and $\phi : X \to \tilde{X}$ is a conjugacy from X to $\tilde{X}$. By Theorem 4.2.7 we have

$$(\phi \times \phi)\{(u, v) \in X \times X \colon u \sim v\} = \{(\phi(u), \phi(v)) \in \tilde{X} \times \tilde{X} \colon \phi(u) \sim \phi(v)\}.$$

Moreover, the corresponding measure to $\mu_1 \otimes \mu_2$ under the conjugacy

$$\phi \times \phi : X \times X \to \tilde{X} \times \tilde{X},$$

is $\tilde{\mu}_1 \otimes \tilde{\mu}_2$ where $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are corresponding measures to μ_1 and μ_2 under ϕ. It follows that

$$(\mu_1 \otimes \mu_2)\{(u, v) \in X \times X \colon u \sim v\} = (\tilde{\mu}_1 \otimes \tilde{\mu}_2)\{(\phi(u), \phi(v)) \in \tilde{X} \times \tilde{X} \colon \phi(u) \sim \phi(v)\}.$$

Therefore by recoding, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code.

Let u, v be in X, m be in $\mathbb{Z}$, and a be a symbol in $\mathcal{A}(X)$. Write $u \overset{a}{\approx}_n v$ if $u \sim v$ and there are integers $m \leq n$ and $p \geq n$ such that

$$v_i = \begin{cases} u_i & -\infty < i < m,\ p < i < \infty \\ a & i = n \end{cases}$$

It is worth mentioning that the relation $\overset{a}{\approx}_n$ is not in general an equivalence relation on X. We have

$$\{(u, v) : u \sim v\} = \bigcup_{a \in \mathcal{A}(X)} \bigcup_{n \in \mathbb{Z}} \{(u, v) \colon u \overset{a}{\approx}_n v\}.$$

Now suppose, for a contradiction, that $(\mu_1 \otimes \mu_2)\{(u, v) : u \sim v\} > 0$. There must be an integer n and a symbol a in $\mathcal{A}(X)$ such that

$$(\mu_1 \otimes \mu_2)\{(u,v)\colon u \overset{a}{\approx}_n v\} > 0.$$

By applying $(T \times T)^{-n}$, without loss of generality, we may assume $n = 0$. Since we have

$$\{(u,v)\colon u \overset{a}{\approx}_0 v\} = \bigcup_{k\in\mathbb{Z}}\{(u,v)\colon u \overset{a}{\approx}_0 v,\, u_{-k}^{k}, v_{-k}^{k} \text{ are routable through } a \text{ at time } 0\ \},$$

it follows that there is an integer k so that

$$(\mu_1 \otimes \mu_2)\{(u,v)\colon u \overset{a}{\approx}_0 v,\, u_{-k}^{k}, v_{-k}^{k} \text{ are routable through } a \text{ at time } 0\ \} > 0.$$

Considering only blocks of length $2k+1$, there must be blocks A, B in X such that

$$(\mu_1 \otimes \mu_2)\big\{(u,v)\colon u \overset{a}{\approx}_0 v,\, u_{[-k,k]} = A,\, v_{[-k,k]} = B\big\} > 0. \tag{5.2.2}$$

Both A and B are routable through a at time n; i.e., there are blocks A', $B' \in \pi_{\mathrm{b}}^{-1}(W)$ with $A'_{-k} = A_{-k}$, $B'_{-k} = B_{-k}$, $A'_0 = B'_0 = a$, $A'_k = A_k$, and $B'_k = B_k$. Let $G \subseteq \mathbb{Z}$ be the set $\{-k,\ldots,k\}$. Basic properties of conditional expectation imply that

$$\begin{aligned}
\mathbb{E}_{\mu_1}\left(\mathbf{1}_{[A]}|\pi^{-1}(\mathscr{B}_Y)\right) &= \mathbb{E}\left(\mathbb{E}_{\mu_1}\left(\mathbf{1}_{\,_{-n}[A]}|\pi^{-1}(\mathscr{B}_Y) \vee \sigma(X_{G^c})\right)|\pi^{-1}(\mathscr{B}_Y)\right)\\
&= \mathbb{E}\left(\mathbb{E}_{\mu_1}\left(\mathbf{1}_{\,_{-n}[A']}|\pi^{-1}(\mathscr{B}_Y) \vee \sigma(X_{G^c})\right)|\pi^{-1}(\mathscr{B}_Y)\right)\\
&= \mathbb{E}_{\mu_1}\left(\mathbf{1}_{[A']}|\pi^{-1}(\mathscr{B}_Y)\right),
\end{aligned} \tag{5.2.3}$$

where the second equality follows from Theorem 3.2.1. Similarly we have

$$\mathbb{E}_{\mu_2}(\mathbf{1}_{[B]}|\pi^{-1}\mathscr{B}_Y) = \mathbb{E}_{\mu_2}(\mathbf{1}_{[B']}|\pi^{-1}\mathscr{B}_Y). \tag{5.2.4}$$

Let $D = \{(u, v)\colon u \overset{a}{\approx}_0 v,\ u_{[-k,k]} = A,\ v_{[k,k]} = B\}$. Since $D \subseteq {}_{-k}[A] \times {}_{-k}[B]$ we have

$$\begin{aligned}
\mu(D) &\leq (\mu_1 \otimes \mu_2)([A] \times [B]) \\
&= \int_Y \mathbb{E}_{\mu_1}(\mathbf{1}_{[A]}|\pi^{-1}\mathscr{B}_Y)\mathbb{E}_{\mu_2}(\mathbf{1}_{[B]}|\pi^{-1}\mathscr{B}_Y) \circ \pi^{-1}\, d\nu \\
&= \int_Y \mathbb{E}_{\mu_1}(\mathbf{1}_{[A']}|\pi^{-1}\mathscr{B}_Y)\mathbb{E}_{\mu_2}(\mathbf{1}_{[B']}|\pi^{-1}\mathscr{B}_Y) \circ \pi^{-1}\, d\nu \quad \text{using(5.2.3) and (5.2.4)} \\
&= (\mu_1 \otimes \mu_2)([A'] \times [B']) \\
&= 0
\end{aligned}$$

where the last equality follows from Theorem 5.2.3 since $A'_0 = B'_0 = a$. This contradicts Equation (5.2.2). $\square$

The following theorem is a more general case of Theorem 5.2.2.

Theorem 5.2.5. *Let (X, Y, π) be a factor triple and ν be an ergodic measure on Y. The number of ergodic measures of relative maximal entropy over ν is at most $|\mathscr{C}(y)|$ for a ν-generic point y in Y.*

Proof. Let $(\tilde{X}, \tilde{Y}, \pi)$ be a factor triple conjugate to (X, Y, π) under conjugacies $\phi : X \to \tilde{X}$ and $\psi : Y \to \tilde{Y}$. By Corollary 5.1.3, the number of ergodic measures of relative maximal entropy over ν and $\nu \circ \psi^{-1}$ is the same. Moreover, by Corollary 4.3.13(c), for each y in Y we have $|\mathscr{C}_\pi(y)| = |\mathscr{C}_{\tilde{\pi}}(\psi(y))|$. Therefore by recoding, without loss of generality, we may assume X is a 1-step SFT and π is a 1-block factor code.

Let y be a ν-generic point of Y. Suppose, for a contradiction, that there are $n > \mathscr{C}_\pi(y)$ ergodic measures $\mu_1, \ldots, \mu_n$ on X of relative maximal entropy over ν. Form the relatively independent joining $\hat{\mu}$ on X^n of the measures $\mu_1, \ldots, \mu_n$. Note that for $\hat{\mu}$-a.e. $\hat{x} \in X^n$, $\pi(p_i(\hat{x}))$ is a ν-generic point of Y with the same number of transition classes. The assumption $n > \mathscr{C}_\pi(y)$ implies that for $\hat{\mu}$-a.e. $\hat{x} = (x_1, \ldots, x_n) \in X^n$ there are distinct i, j such that $p_i(\hat{x}) \sim p_j(\hat{x})$; i.e.,

$$\hat{\mu}\left(\bigcup_{1 \leq i < j \leq n} \{\hat{x} = (x_1, \ldots, x_n)\colon x_i \sim x_j\}\right) = 1.$$

At least one of the sets $S_{i,j} = \{(x_1, \ldots, x_n)\colon x_i \sim x_j\}$ must have positive $\hat{\mu}$-measure. It follows that

$$\begin{aligned} 0 &< \hat{\mu}(S_{i,j}) \\ &= \hat{\mu}(\{(x_1, \ldots, x_n),\ x_i \sim x_j\}) \\ &= (\mu_i \otimes \mu_j)\{(u, v)\colon u \sim v\}. \end{aligned}$$

This contradicts Theorem 5.2.4. $\square$

Chapter 6

Outlook

Let (X, T) be a one-sided topologically mixing shift of finite type. An invariant measure μ on X is a Gibbs measure corresponding to a function f in $C(X)$ if there are constants $C_1,\ C_2 > 0$ and $P > 0$ such that

$$C_1 \leq \frac{\mu([x_0 x_1 \dots x_{n-1}])}{\exp\left(-Pn + (S_n f)(x)\right)} \leq C_2$$

for every point x in X and $n \geq 1$, where $(S_n f)(x) = \sum_{k=0}^{n-1} f(T^k(x))$.

Walters [31, 32] introduces a class $\mathrm{Bow}(X, T)$ of functions that contains the functions with summable variation, all of which have unique equilibrium states. Let

$$\mathrm{var}_n(f) = \sup\{|f(x) - f(y)| \colon x,\ y \in X,\ x_i = y_i \text{ for all } 0 \leq i \leq n-1\}.$$

Then $\mathrm{Bow}(X, T) = \{f \in C(X) \colon \sup_{n \geq 1} \mathrm{var}_n(S_n f) < \infty\}$.

Theorem 6.0.6. *[31, Theorem 2.16] Let $f \in Bow(X, T)$. Then f has a unique equilibrium state μ which is a Gibbs measure.*

The relative version of this result is an open question. We make the following conjecture;

Conjecture 6.0.7. *Let (X, Y, π) be a factor triple and ν be a fully supported ergodic measure on Y. The number of ergodic measures of maximal pressure in the fibre*

$\pi^{-1}\{\nu\}$ *is at most* c_π.

It would be desirable to develop a structural theory of transition classes. Finding a systematic way to have an order on a set of transition classes and understanding whether there is a transition between them is an interesting problem. In Section 4.1.2, we presented an algorithm to compute the degree of a finite-to-one code. It should be possible to generalize this algorithm to computing the class degree. We expect that knowing better the structure of transition classes can help in finding such algorithm.

Bibliography

[1] M.-P. Béal. *Codage Symbolique*. Masson, Paris, 1993.

[2] Tim Bedford, Michael Keane, and Caroline Series. *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces*. Oxford University Press, Oxford, 1991.

[3] D. Blackwell. The entropy of functions of finite-state Markov chains. *Transactions of the first Prague Conference on information theory, Statistical decision functions, random processes*, pages 13–20, 1956.

[4] M. Boyle and S. Tuncel. Infinite-to-one codes and Markov measures. *Trans. Amer. Math. Soc.*, 285:657–683, 1984.

[5] C. Burke and M. Rosenblatt. A Markovian function of a Markov chain. *Ann. Math. Statist.*, 29:1112–1122, 1958.

[6] R. Burton and J. Steif. Non-uniqueness of measures of maximal entropy for subshift of finite type. *Ergodic Theory Dynam. Systems.*, 14:213–235, 1994.

[7] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, New York, 1999.

[8] D. Gatzouras and Y. Peres. The variational principle for Hausdorff dimension. In *a survey. Ergodic Theory of* Z^d*-actions (Warwick, 1993–1994)*, volume 228, pages 113–125. London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, 1996.

[9] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading, 1979.

[10] R. Israel. *Convexity in the theory of lattice gases*. Princeton University Press, Princeton, 1979.

[11] A. B. Katok. Lyapunov exponents, entropy and periodic points for diffeomorphisms. *Inst. Hautes tudes Sci. Publ. Math.*, 51:137–173, 1980.

[12] Michael Keane and Karl Petersen. Nearly simultaneous proofs of the ergodic theorem and maximal ergodic theorem. In *Lecture Notes Monograph Series*, volume 0, pages 1–4. IMS, Berlin, 2005.

[13] G. Keller. *Equilibrium states in ergodic theory*. Cambridge Univ. Press, Cambridge, 1998.

[14] A. I. Khinchine. *Mathematical foundations of information theory*. Dover Publications, New York, 1957.

[15] J. F. C. Kingman and S. J. Taylor. *Introduction to measure and probability*. Cambridge University Press, Cambridge, 1966.

[16] O. E. Lanford and D. Ruelle. Observables at infinity and states with short range correlations in statistical mechanics. *Comm. Math. Phys.*, 13:194–215, 1969.

[17] F. Ledrappier and P. Walters. A relativised variational principle for continuous transformations. *J. London Math. Soc*, 16:568–576, 1977.

[18] F. Ledrappier and L. Young. The metric entropy of diffeomorphisms ii. *Ann. of Math.*, 122:540–574, 1985.

[19] D. Lind and B. Marcus. *Symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.

[20] B. Marcus, K. Petersen, and S. Williams. Transmission rates and factors of Markov chains. In *Conference in modern analysis and probability, Contemp. Math.*, volume 26, pages 279–293. Amer. Math. Soc., 1984.

[21] W. Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.*, 112:55–66, 1964.

[22] W. Parry. *Entropy and Generators in Ergodic Theory.* Benjamin, New York, 1969.

[23] K. Petersen. Information compression and retention in dynamical processes. In *Dynamics and randomness (Santiago, 2000)*. Kluwer Acad. Publ.

[24] K. Petersen, A. Quas, and S. Shin. Measures of maximal relative entropy. *Ergodic Theory Dynam. Systems.*, 23(1):207–223, 2003.

[25] A. Quas. Ergodicity and mixing. In *Encyclopedia of Complex Systems Science*, pages 2918–2933. Springer.

[26] D. Ruelle. Statistical mechanics on a compact set with Z^ν action satisfying expansiveness and specification. *Trans. Amer. Math. Soc.*, 187:237–251, 1973.

[27] C. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[28] S. Shin. Measures that maximize weighted entropy for factor maps between subshifts of finite type. *Ergodic Theory Dynam. Systems.*, 21:1249–1272, 2001.

[29] P. Walters. *An introduction to ergodic theory.* Springer Verlag New York, 1982.

[30] P. Walters. Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts. *Trans. Amer. Math. Soc.*, 296:1–31, 1986.

[31] P. Walters. Convergence of the Ruelle operator for a function satisfying bowen's condition. *Trans. Amer. Math. Soc.*, 353:327–347, 2001.

[32] P. Walters. Regularity conditions and Bernoulli properties of equilibrium states and g-measures. *J. London Math. Soc.*, 71(2):379–396, 2005.