MEASURING COMPLEXITY IN DYNAMICAL SYSTEMS

Benjamin Wilson

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Mathematics in the College of Arts and Sciences.

Chapel Hill
2015

Approved by:
Karl Petersen
Sue Goodman
Jane Hawkins
Jeremy Marzuola
Peter Mucha
Measuring the complexity of dynamical systems is important in order to classify them and better understand them. In 1958 Kolmogorov introduced to ergodic theory an analogue of Shannon’s information-theoretic entropy as a measure of disorder or uncertainty in a system. Based on this concept and ideas from neuroscience and information theory, we define the intricacy and average sample complexity of a topological dynamical system and a measure-preserving dynamical system. We examine these new complexity measurements in both the topological and measure-theoretic settings, including analysis of symbolic dynamical systems and Markov shifts. We compare these measurements to the usual measure-theoretic and topological entropies, give some properties of these quantities, and look at some questions that they raise.
ACKNOWLEDGMENTS

I would most of all like to thank my adviser, Karl Petersen, for his guidance, patience, and encouragement over the years. This work would have been impossible without his intelligence and creativity. I am lucky to have had such a helpful adviser, always willing to answer my questions and lead me in the right direction as well as put so much time and work into our project.

I would also like to thank the mathematicians that inspired our work including Jerome Buzzi and Lorenzo Zambotti. Other mathematicians who gave useful suggestions and thought about this project include Mike Boyle, Tomasz Downarowicz, Kevin McGoff, and Jean-Paul Thouvenot.

Additionally, I am grateful for the mathematicians that have supported me by coming to many of my talks, asking good questions, and giving good suggestions. These include my committee members Patrick Eberlein, Sue Goodman, Jane Hawkins, Jeremy Marzuola, and Peter Mucha, as well as Rachel Bayless, Joanna Furno, and Kathleen Carroll. I am also thankful to Michael Abel, Justin Allman, and Merrick Brown for discussions about mathematics and for helping me learn Mathematica.

I could not have completed this dissertation without my other friends in the department including Austin Baird, Namdi Brandon, Nate Bushek, Mayukh Mukherjee, and especially my officemates Sam Miller and Joe Seaborn.

I am grateful for my loving and supportive parents, sister, and brother. Successfully completing my dissertation would have been impossible without the encouragement of my family. Finally, I would like to thank Sydney Gross for her constant care and love. She has been there to keep me calm and focused throughout most of my graduate program.
# TABLE OF CONTENTS

| LIST OF FIGURES | .............................................................. viii |
| LIST OF TABLES | .............................................................. ix |
| INTRODUCTION | ............................................................... 1 |

## CHAPTER 1: BACKGROUND .......................................................... 5

1.1 Basic setup and notation .................................................. 5

1.2 Entropy ........................................................................... 6
   1.2.1 Topological entropy ............................................. 6
   1.2.2 Measure-theoretic entropy ................................... 9
   1.2.3 Variational principle .......................................... 12
   1.2.4 Topological pressure ........................................... 13

1.3 The isomorphism problem .................................................. 16
   1.3.1 Isomorphism of measure-preserving transformations ... 16
   1.3.2 Ergodicity and mixing ......................................... 17
   1.3.3 Spectral isomorphism ......................................... 17
   1.3.4 Topological conjugacy and discrete spectrum .......... 18

1.4 Neural complexity and intricacy ........................................... 19
   1.4.1 Entropy and mutual information in probability ...... 19
   1.4.2 Neural complexity ............................................. 20
   1.4.3 Intricacy in probability ...................................... 21

1.5 Examples of Systems ....................................................... 23
   1.5.1 Subshifts and shifts of finite type ....................... 23
   1.5.2 Other subshifts ............................................... 32
   1.5.3 Bernoulli shifts .............................................. 34
1.5.4 Markov shifts .................................................. 34
1.6 Other concepts of complexity ........................................ 36
  1.6.1 Sequence entropy ............................................ 36
  1.6.2 Maximal pattern complexity .................................. 38
  1.6.3 Eulerian entropy ............................................. 38
  1.6.4 Further readings on concepts of complexity ................. 39

CHAPTER 2: TOPOLOGICAL INTRICACY AND AVERAGE
SAMPLE COMPLEXITY ................................................. 40
  2.1 Definitions ..................................................... 40
  2.2 Calculations for the golden mean shift of finite type ........... 43
  2.3 Preliminary results .............................................. 44
  2.4 Definitions based on Bowen’s definition of entropy ............. 47
  2.5 Theorem relating topological average sample complexity and topological
    intricacy to topological entropy ................................ 49
  2.6 Average sample pressure ....................................... 56

CHAPTER 3: COMPLEXITY CALCULATIONS FOR SUBSHIFTS ...... 65
  3.1 Shifts of finite type ........................................... 65
  3.2 Average sample complexity and intricacy of other subshifts .... 80

CHAPTER 4: MEASURE-THEORETIC INTRICACY AND AVERAGE
SAMPLE COMPLEXITY ................................................. 83
  4.1 Definitions and preliminary results .............................. 83
  4.2 Main results for measure-theoretic systems ...................... 85
  4.3 Analysis of Markov shifts ..................................... 93

CHAPTER 5: FUTURE DIRECTIONS .................................... 101
  5.1 Improved computational methods and formulas .................... 101
  5.2 General weights ............................................... 102
  5.3 Further analysis of shifts of finite type ....................... 102
  5.4 Higher-dimensional shifts ..................................... 103
  5.5 Maximizing subsets $S \subset n^*$ ................................ 104
5.6 Analysis of more examples ................................................................. 104
5.7 Entropy is the only finitely observable invariant ................................. 105
5.8 Alternate definition of the average sample complexity function analogous
to the complexity function of a sequence ........................................... 105
5.9 Complexity of finite words ................................................................. 106
5.10 Partition \( n^* \) into \( m \) subsets ......................................................... 107
5.11 Definition based on Rokhlin entropy ................................................ 107
5.12 Application of topological average sample pressure to coding sequence density ................................................................. 108
5.13 Maximal measures, variational principle, and equilibrium states ........... 108

APPENDIX A: TABLES ................................................................. 110

A.1 Calculations for shifts of finite type using the uniform system of coefficients ................................................................. 110
A.2 Shifts of finite type with positive square adjacency matrices ............... 116

REFERENCES ................................................................. 120
LIST OF FIGURES

2.1 Visualization of sets $K_i$. ......................................................... 50
2.2 Visualization of sets $\tilde{E}_j$ and $G_j$. ............................................. 53

4.1 $\text{Asc}_\mu$ and $\text{Int}_\mu$ for 1-step Markov measures on the full 2-shift ................. 96
4.2 $\text{Int}_\mu$ for 1-step Markov measures on the full 2-shift with $P_{11} = 0$ ..................... 97
4.3 $h_\mu$ for 1-step Markov measures on the full 2-shift .................................................. 97
4.4 1-step Markov measures on the golden mean shift ..................................................... 99
4.5 $\text{Asc}_\mu$ and $\text{Int}_\mu$ for 2-step Markov measures on the golden mean shift ............. 100
4.6 $h_\mu$ for 2-step Markov measures on the golden mean shift ........................................... 100
# LIST OF TABLES

1.1 Golden mean shift ......................................................... 28
1.2 2nd higher block presentation of the golden mean shift .............. 30

2.1 $N(S)$ for the golden mean shift for all $S \subset 3^*$ ...................... 44
2.2 Calculations for the golden mean shift ..................................... 44

3.1 Two shifts of finite type with the same entropy .......................... 72
3.2 Comparison of $N(S)$ for two shifts of finite type ......................... 73
3.3 Two shifts of finite type with the same entropy .......................... 74
3.4 Table with calculations for SFTs with the same entropy .................. 75
3.5 Two shifts that have the same entropy, Asc, and Int. ..................... 78
3.6 Calculations of average sample pressure ................................... 79
3.7 Calculations for systems formed from substitution sequences .......... 80
3.8 Calculations for systems formed from Sturmian sequences ............. 81

4.1 1-step Markov measures on the full 2-shift ............................... 95
4.2 1-step Markov measures on the golden mean shift ......................... 98
4.3 2-step Markov measures on the golden mean shift ......................... 99
INTRODUCTION

In their study of high-level neural networks [TSE], G. Edelman, O. Sporns, and G. Tononi introduce a quantitative measure that they call neural complexity ($C_N$) that captures the interplay between two fundamental aspects of brain organization: the functional segregation of local areas and their global integration. $C_N$ is shown to be high when functional segregation coexists with integration and to be low when the components of a system are either completely independent (segregated) or completely dependent (integrated).

J. Buzzi and L. Zambotti [BZ] provide a mathematical foundation for neural complexity which belongs to a natural class of functionals: the averages of mutual information satisfying exchangeability and weak additivity. The former property means that the functional is invariant under permutations of the system, the latter that it is additive when independent systems are combined. They give a unified probabilistic representation of these functionals, which they call intricacies.

Our goal is to define and then study intricacy in dynamical systems based on the classical definition of topological entropy in dynamical systems and intricacy as defined by Buzzi and Zambotti. We will define topological intricacy and the closely related topological average sample complexity for a general topological dynamical system $(X, T)$ with respect to an open cover $\mathcal{U}$ of $X$. More specifically, denote by $n^*$ the set of integers $\{0, 1, \ldots, n-1\}$, let $S = \{s_0, s_1, \ldots, s_{|S|-1}\} \subset n^*$, let $S^c = n^* \setminus S$, let $c^n_S$ be a weighting function that depends on $S$ and $n$, let $\mathcal{U}_S = \bigvee_{i=0}^{|S|-1} T^{-s_i} \mathcal{U}$, and let $N(\mathcal{U})$ be the minimum cardinality of a subcover of $\mathcal{U}$. Then the topological intricacy of $(X, T)$ with respect to the open cover $\mathcal{U}$ is defined to be

$$\text{Int}(X, \mathcal{U}, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log \left( \frac{N(\mathcal{U}_S)N(\mathcal{U}_{S^c})}{N(\mathcal{U}_{n^*})} \right).$$

We will usually let $c^n_S = 2^{-n}$ for all $S$. Since we are averaging the quantity $\log(N(\mathcal{U}_S)N(\mathcal{U}_{S^c})/N(\mathcal{U}_{n^*}))$ over all subsets $S \subset n^*$, topological intricacy takes on high values for systems in which for most $S$ the product $N(\mathcal{U}_S)N(\mathcal{U}_{S^c})$ is large compared to $N(\mathcal{U}_{n^*})$, and this will only occur if $N(\mathcal{U}_S)$ and
$N(\mathcal{U}_S)$ are simultaneously large for most $S$. We will see that this happens in systems that are far from both total order and total disorder.

Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $\alpha = \{A_1, \ldots, A_k\}$ a finite measurable partition of $X$. Given $S \subset n^*$, $S^c$, and $c_n^S$ as above, let $\alpha_S = \bigwedge_{i=0}^{n-1} T^{-s_i} \alpha$ and $H_\mu(\alpha) = -\sum_{i=1}^k \mu(A_i) \log \mu(A_i)$. Then the measure-theoretic intricacy of $X$ and $T$ with respect to $\alpha$ is defined to be

$$
\text{Int}_\mu(X, \alpha, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_n^S \left[ H_\mu(\alpha_S) + H_\mu(\alpha_{S^c}) - H_\mu(\alpha_{n^*}) \right].
$$

For similar reasons as in the topological case, measure-theoretic intricacy takes high values for systems that are far from both order and disorder.

Two of the main results in this dissertation establish a relationship between topological intricacy and topological entropy as well as between measure-theoretic intricacy and measure-theoretic entropy. Entropy in dynamics classically is first defined with respect to either a specific cover of a topological space or a specific partition of a measure space. To find the entropy of a transformation, one then takes the supremum over all open covers or over all partitions. We define the intricacy with respect to a cover and with respect to a partition, but in a corollary of Theorem 2.5.1 we show, for $c_n^S = 2^{-n}$, that $\sup_\mathcal{U} \text{Int}(X, \mathcal{U}, T) = h_{\text{top}}(X, T)$, the usual topological entropy of the system. Similarly in the measure-theoretic setting, using Theorem 4.2.1 we show for $c_n^S = 2^{-n}$ that $\sup_{\alpha} \text{Int}_\mu(X, \alpha, T) = h_\mu(X, T)$, the usual measure-theoretic entropy. We still find interesting results by looking at these measurements for specific partitions and open covers. In particular, for certain covers of subshifts and partitions of Markov shifts we see that measuring the intricacy reveals interesting properties that are not found through other measurements of complexity.

The main setting we examine is that of subshifts, which are closed shift-invariant collections of infinite sequences of elements from a finite set called an alphabet. The topological entropy of a subshift is the exponential growth rate of the number of words of each length found in sequences in the subshift. To find the intricacy of a subshift, we do not just count all words of length $n$, but we count words seen at all places in a subset $S \subset n^*$ and take an average. This causes the measurement to be more sensitive to the structure of the sequences of a subshift than the entropy. Example 3.1.6 shows two subshifts that have the same entropy but different intricacies.
We show that intricacy is bounded above by entropy in both the topological and measure-theoretic settings. This tells us that systems of zero entropy also have zero intricacy. The fact that these completely dependent systems have zero intricacy shows that our generalization of neural complexity to dynamical systems accomplishes one of its goals: it takes on low values for integrated systems.

While we can approximate intricacy for subshifts, computing the actual quantity is very difficult in general. This is because for each \( n \) we have to make computations on all subsets of \( n^* \), and as \( n \) gets large doing \( 2^n \) computations is not feasible. Thus, Theorem 3.1.2 is an important result, as it gives a formula for the intricacy for particular covers of certain shifts of finite type. Using this formula we calculate the intricacies of some positive entropy systems. In particular, we find that for each \( r \geq 1 \), the intricacy with respect to the open cover by rank 0 cylinder sets of the full \( r \)-shift is zero. This is a system with maximum entropy and is completely independent. This shows that intricacy is also low for segregated systems, accomplishing the other goal of a generalization of neural complexity.

In the measure-theoretic setting Theorems 4.2.4 and 4.2.6 give a relationship between measure-theoretic average sample complexity with respect to a finite partition \( \alpha \) and a series involving the conditional entropies \( H_\mu(\alpha \mid \alpha_i^\infty) \). More specifically, we show

\[
\text{Asc}_\mu(X, \alpha, T) \geq \sum_{i=1}^{\infty} 2^{-i-1}H_\mu(\alpha \mid \alpha_i^\infty)
\]

with equality in certain cases. One of the cases where equality holds is for Markov shifts. In Section 4.3 we use this equation to compute the measure-theoretic average sample complexity and measure-theoretic intricacy for 1-step Markov measures on the full 2-shift and 1-step and 2-step Markov measures on the golden mean shift. Analysis of these data allows us to make conjectures about measures that maximize average sample complexity and measures that maximize intricacy.

This thesis is broken into 5 main chapters. Chapter 1 consists of background information including formal definitions and examples from ergodic theory. Chapter 2 contains definitions of topological intricacy and topological average sample complexity as well as the main results for topological dynamical systems. In Chapter 3 we look at calculations of topological intricacy and topological average sample complexity for subshifts. Chapter 4 includes definitions and results
pertaining to measure-theoretic intricacy and measure-theoretic average sample complexity as well
as an analysis of Markov shifts. In Chapter 5 we discuss some questions that arise from these
complexity measurements that may motivate future research. The appendix contains tables of
computations for many different shifts of finite type.
CHAPTER 1

Background

In this chapter we establish terminology and notation that will be used throughout the paper. Since entropy is the most well known and studied measurement of complexity in ergodic theory and our new measurements of complexity are based on entropy, we will give precise definitions of both topological entropy and measure-theoretic entropy. We will then give definitions of neural complexity and intricacy in the setting of probability. After describing some examples of dynamical systems we conclude the chapter by citing and describing several other measurements of complexity in ergodic theory. More details and definitions can be found in standard ergodic theory texts such as [Pet] and [Wal]. See [LM] for more information on symbolic dynamics.

1.1 Basic setup and notation

We will be discussing both topological and measure-theoretic dynamical systems. A topological dynamical system is defined as the combination of a compact Hausdorff (often metric) space $X$ with a continuous transformation $T : X \to X$. We denote a topological dynamical system by $(X, T)$. We will often denote an open cover of $X$ by $\mathcal{U}$ or $\mathcal{V}$ and elements of $\mathcal{U}$ and $\mathcal{V}$ by $U$ and $V$ respectively. If $X$ is a metric space with metric $d$ then we denote the open ball of radius $r$ around the point $x \in X$ by $B(x, r) = \{y \in X : d(x, y) < r\}$ and the closed ball of radius $r$ about $x \in X$ by $\overline{B}(x, r) = \{y \in X : d(x, y) \leq r\}$.

In the measure-theoretic case, we define a system $(X, \mathcal{B}, \mu, T)$ to consist of a complete probability space $X$, a $\sigma$-algebra $\mathcal{B}$ of measurable subsets of $X$, a complete measure $\mu$ on $\mathcal{B}$, and a one-to-one onto map $T : X \to X$, such that $T$ and $T^{-1}$ are both measurable. More precisely, $\mu$ is a non-negative, countably additive function on $\mathcal{B}$ such that $\mu(X) = 1$, $\mathcal{B}$ contains all sets of measure 0, and $T\mathcal{B} = T^{-1}\mathcal{B} = \mathcal{B}$. We also assume $\mu(T^{-1}E) = \mu(E)$ for all $E \in \mathcal{B}$, so that $T$ is a measure-preserving transformation. We will sometimes refer to just the probability space $(X, \mathcal{B}, \mu)$. 
We will often denote partitions of a measure space \( X \) by \( \alpha \) or \( \beta \). Recall, a partition of \( X \) is a family \( \alpha \) of nonempty disjoint subsets of \( X \) such that \( \bigcup_{\alpha} A = X \).

We denote by \( n^* \) the set of integers from 0 to \( n - 1 \). i.e.

\[
n^* = \{0, 1, \ldots, n - 1\}.
\]

(1.1.1)

Given a subset \( S \subset n^* \), denote its complement by \( S^c = n^* \setminus S \). We denote the number of elements in a set \( A \) by either \( \text{card}(A) \) or \( |A| \). For \( x \in \mathbb{R} \) we denote the floor of \( x \) by \( \lfloor x \rfloor = \max\{m \in \mathbb{Z} : m \leq x\} \) and the ceiling of \( x \) by \( \lceil x \rceil = \min\{m \in \mathbb{Z} : m \geq x\} \). Unless otherwise specified, logarithms will be taken base \( e \). We take the convention that \( 0 \log 0 = 0 \).

1.2 Entropy

This sections contains definitions of topological entropy and measure-theoretic entropy as well as some properties of these quantities.

1.2.1 Topological entropy

Let \( X \) be a a compact Hausdorff space and \( \mathcal{U} \) be an open cover of \( X \). We say \( \mathcal{V} \) is a refinement of \( \mathcal{U} \), and write \( \mathcal{V} \geq \mathcal{U} \), if every \( V \in \mathcal{V} \) is a subset of some \( U \in \mathcal{U} \). The least common refinement, or join, \( \mathcal{U} \lor \mathcal{V} \), of \( \mathcal{U} \) and \( \mathcal{V} \) consists of the nonempty members of \( \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\} \).

For each open cover \( \mathcal{U} \) of \( X \), let

\[
N(\mathcal{U}) = \text{the minimum of the cardinality of the subcovers of } \mathcal{U}
\]

and

\[
H(\mathcal{U}) = \log N(\mathcal{U}).
\]

Proposition 1.2.1. For open covers \( \mathcal{U} \) and \( \mathcal{V} \) of \( X \),

1. \( H(\mathcal{U} \lor \mathcal{V}) \leq H(\mathcal{U}) + H(\mathcal{V}) \).
2. If \( \mathcal{U} \leq \mathcal{V} \) then \( H(\mathcal{U}) \leq H(\mathcal{V}) \).
If $\mathcal{U}$ is an open cover of $X$ then $T^{-1}\mathcal{U}$ is the open cover consisting of all sets of the form $T^{-1}U$ where $U \in \mathcal{U}$. For any $-\infty < m \leq n < \infty$, define

$$\mathcal{U}^n_m = \bigcup_{k=m}^{n} T^{-k}\mathcal{U} = T^{-m}\mathcal{U} \cup T^{-(m+1)}\mathcal{U} \cup \cdots \cup T^{-n}\mathcal{U}.$$  

We will also use the notation

$$\mathcal{U}_k = \bigcup_{i=0}^{k-1} T^{-i}\mathcal{U}.$$  

In 1965, Adler, Konheim, and McAndrew [AKM] gave the following definition of topological entropy.

**Definition 1.2.2.** Let $T : X \to X$ be a continuous map on a compact Hausdorff space $X$ and let $\mathcal{U}$ be an open cover of $X$. Define the topological entropy of $T$ with respect to the open cover $\mathcal{U}$ to be

$$h_{\text{top}}(X, \mathcal{U}, T) = \lim_{n \to \infty} \frac{1}{n} H(\mathcal{U}^n_0) = \lim_{n \to \infty} \frac{1}{n} H(\mathcal{U} \cup T^{-1}\mathcal{U} \cup \cdots \cup T^{-n+1}\mathcal{U}).$$  

(1.2.1)

Then the topological entropy of the system $(X, T)$ is defined to be the supremum of $h_{\text{top}}(X, \mathcal{U}, T)$ over all possible covers $\mathcal{U}$ of $X$:

$$h_{\text{top}}(X, T) = \sup_{\mathcal{U}} h_{\text{top}}(X, \mathcal{U}, T).$$  

(1.2.2)

The limit in Equation 1.2.1 can be shown to exist by using the following lemma by Fekete [Fek] and the fact that the sequence $a_n = H(\mathcal{U}^n_0)$ is subadditive.

**Lemma 1.2.3 (Fekete’s Subadditive Lemma).** If $\{a_n, n \geq 1\}$ is a sequence of real numbers such that $a_{n+p} \leq a_n + a_p$ for all $n$ and $p$ then $\lim_{n \to \infty} \frac{a_n}{n}$ exists and equals $\inf_{n} \frac{a_n}{n}$.

The following notation will be useful later. Given a set of integers $S = \{s_0, s_2, \ldots, s_{|S|} - 1\}$, let

$$\mathcal{U}_S = T^{-s_0}\mathcal{U} \cup \cdots \cup T^{-s_{|S|} - 1}\mathcal{U}.$$  

Notice that $\mathcal{U}_{n^*} = \mathcal{U}^{n-1}$. For a point $x \in X$, $n \in \mathbb{N}$, and $S \subset n^*$ define the $S$-orbit of $x$ by $\{T^{-i}x : i \in S\}$. 


Bowen’s definition of topological entropy

Bowen defined topological entropy using the concepts of spanning sets and separated sets. It can be shown that the definition given in terms of spanning sets is equivalent to the definition given in terms of separated sets. We can also show that the definition of topological entropy given in terms of spanning sets or separated sets and the open cover definition by Adler, Konheim, and McAndrew are equivalent.

Let \((X, d)\) be a compact metric space and let \(T : X \to X\) be a continuous map. For each \(n \in \mathbb{N}\), given a subset \(S \subseteq n^*\), we define the metric

\[
d_S(x, y) = \max\{d(T^i x, T^i y) : i \in S\}.
\]

Given any \(\varepsilon > 0\) and \(n \in \mathbb{N}\), two points of \(X\) are \(\varepsilon\)-close with respect to this metric if their \(i \in S\) iterates under \(T\) are \(\varepsilon\)-close in the metric \(d\). This metric allows one to distinguish in a neighborhood of an \(S\)-orbit the points that move away from each other from the points that travel together. The open ball centered at \(x\) of radius \(r\) in the metric \(d_S\) is \(B_S(x, r) = \bigcap_{i \in S} T^{-i}B(T^i x, r)\), so

\[
B_S(x, r) = \{y \in X : d(T^i x, T^i y) < r \text{ for all } i \in S\}.
\]

**Definition 1.2.4.** Let \(n \in \mathbb{N}, S \subseteq n^*,\) and \(\varepsilon > 0\). A subset \(E \subseteq X\) is said to be an \((S, \varepsilon)\) spanning set of \(X\) with respect to \(T\) if for every \(x \in X\) there is \(y \in E\) with \(d_S(x, y) \leq \varepsilon\). In other words, for every \(x \in X\) there must be some point of \(E\) whose \(S\)-orbit stays within \(\varepsilon\) of the \(S\)-orbit of \(x\).

We let

\[
r(S, \varepsilon) = \min\{\text{card}(E) : E \subseteq X \text{ is an}(S, \varepsilon) \text{ spanning set}\}
\]

Notice that with our notation \((n^*, \varepsilon)\) spanning sets are the same as the usual \((n, \varepsilon)\) spanning sets.

**Definition 1.2.5.** If \(X\) is a compact metric space and \(T : X \to X\) is a continuous map, given \(\varepsilon > 0\) denote

\[
h_\varepsilon(X, T) = \limsup_{n \to \infty} \frac{1}{n} \log r(n^*, \varepsilon).
\]
Define the topological entropy of the system \((X, T)\) by

\[
h_{\text{top}}(X, T) = \lim_{\varepsilon \to 0^+} h_\varepsilon(X, T) = \lim_{\varepsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} \log r(n^*, \varepsilon). \tag{1.2.4}
\]

We also define the topological entropy in terms of \((n^*, \varepsilon)\) separated sets.

**Definition 1.2.6.** Let \(n \in \mathbb{N}, S \subset n^*, \text{and} \varepsilon > 0\). A subset \(E \subset X\) is said to be an \((S, \varepsilon)\) separated set of \(X\) with respect to \(T\) if for all \(x, y \in E\) with \(x \neq y\), \(d_S(x, y) > \varepsilon\). In other words, among the \(S\)-orbits of every distinct pair of points in \(E\) there is a pair of elements that are at least \(\varepsilon\) apart from one another. We let

\[
s(S, \varepsilon) = \max\{\text{card}(E) : E \subset X \text{ is an } (S, \varepsilon) \text{ separated set}\}
\]

Notice that with our notation \((n^*, \varepsilon)\) separated sets are the usual \((n, \varepsilon)\) separated sets.

**Definition 1.2.7.** If \(X\) is a compact metric space and \(T : X \to X\) is a continuous map, given \(\varepsilon > 0\) denote

\[
h'_\varepsilon(X, T) = \limsup_{n \to \infty} \frac{1}{n} \log s(n^*, \varepsilon). \tag{1.2.5}
\]

Define the topological entropy of the system \((X, T)\) by

\[
h_{\text{top}}(X, T) = \lim_{\varepsilon \to 0^+} h'_\varepsilon(X, T) = \lim_{\varepsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} \log s(n^*, \varepsilon). \tag{1.2.6}
\]

### 1.2.2 Measure-theoretic entropy

In the setting of ergodic theory, entropy was first defined by Kolmogorov [Kol2]. Given a measure-preserving system \((X, \mathcal{B}, \mu, T)\), we consider a finite partition

\[
\alpha = \{A_1, A_2, \ldots, A_n\}
\]
of \( X \) into finitely many pairwise disjoint measurable sets \( A_i \) of positive measure. Thus, up to measure 0, the sets of \( \alpha \) cover \( X \). Define the entropy of the partition \( \alpha \) to be

\[
H_\mu(\alpha) = -\sum_{i=1}^{n} \mu(A_i) \log \mu(A_i).
\]

Given partitions \( \alpha = \{A_1, \ldots, A_n\} \) and \( \beta = \{B_1, \ldots, B_m\} \) of \( X \), we define

\[
T^{-1} \alpha = \{T^{-1}A_1, \ldots, T^{-1}A_n\}
\]

and

\[
\alpha \vee \beta = \{A_i \cap B_j : i = 1, \ldots, n; j = 1, \ldots, m\}.
\]

**Definition 1.2.8.** Define the entropy of \( T \) with respect to \( \mu \) and \( \alpha \) as

\[
h_\mu(X, \alpha, T) = \lim_{n \to \infty} \frac{1}{n} H_\mu(\alpha \vee T^{-1} \alpha \vee \cdots \vee T^{-n+1} \alpha).
\]

The entropy of the transformation \( T \) is defined to be

\[
h_\mu(X, T) = \sup_\alpha h_\mu(X, \alpha, T). \tag{1.2.7}
\]

Given a partition \( \alpha \), for any \(-\infty < m \leq n < \infty\), define

\[
\alpha^n_m = \bigvee_{k=m}^{n} T^{-k} \alpha.
\]

We will also use the notation

\[
\alpha_k = \bigvee_{i=0}^{k-1} T^{-i} \alpha.
\]

Given a set of integers \( S = \{s_0, \ldots, s_{|S|-1}\} \), and a partition \( \alpha \), we write

\[
\alpha_S = T^{-s_0} \alpha \vee T^{-s_1} \alpha \vee \cdots \vee T^{-s_{|S|-1}} \alpha = \bigvee_{i \in S} T^{-i} \alpha.
\]

Notice that \( \alpha_n = \alpha_n^* = \alpha_0^{n-1} \).
Information and Conditioning

In this section we let \((X, \mathcal{B}, \mu)\) be a probability space. The characteristic function, \(\chi_A\), of a subset \(A \subset X\) is defined by

\[
\chi_A(x) = \begin{cases} 
  1, & \text{if } x \in A \\
  0, & \text{otherwise}
\end{cases}.
\]

**Definition 1.2.9.** Let \(\alpha = \{A_1, A_2, \ldots\}\) be a countable partition of \(X\). Then for each \(x \in X\) denote by \(\alpha(x)\) the element of \(\alpha\) to which \(x\) belongs. The information function associated to \(\alpha\) is defined to be

\[
I_\alpha(x) = -\log \mu(\alpha(x)) = -\sum_{A \in \alpha} \log \mu(A) \chi_A(x). 
\] (1.2.8)

Clearly

\[
H_\mu(\alpha) = -\sum_{A \in \alpha} \mu(A) \log \mu(A) = \int_X I_\alpha(x) d\mu(x).
\] (1.2.9)

**Definition 1.2.10.** Let \(\alpha = \{A_1, \ldots, A_n\}\) and \(\beta = \{B_1, \ldots, B_m\}\) be finite partitions of \(X\). For \(A_i \in \alpha\) and \(B_j \in \beta\), the conditional probability, \(\mu(A_i \mid B_j)\), of \(A_i\) given \(B_j\) is

\[
\mu(A_i \mid B_j) = \frac{\mu(A_i \cap B_j)}{\mu(B_j)}.
\]

It represents the likelihood that a point \(x \in X\) is in \(A_i\) given that it is in \(B_j\).

**Definition 1.2.11.** If \(\alpha\) and \(\beta\) are finite partitions of \(X\) then we define the conditional information function of \(\alpha\) given \(\beta\) by

\[
I_{\alpha \mid \beta}(x) = -\log \mu(\alpha(x) \mid \beta(x)) = -\sum_{A, \alpha} \sum_{B \in \beta} \log \frac{\mu(A \cap B)}{\mu(B)} \chi_{A \cap B}(x)
\] (1.2.10)
The conditional entropy of $\alpha$ given $\beta$ is defined by

$$H_{\mu}(\alpha \mid \beta) = - \sum_{A \in \alpha} \sum_{B \in \beta} \log \mu(A \mid B)\mu(A \cap B)$$

$$= \sum_{B \in \beta} \left( - \sum_{A \in \alpha} \log \mu(A \mid B)\mu(A \mid B) \right)\mu(B)$$

$$= - \int_X \log \mu(\alpha(x) \mid \beta(x)) d\mu(x)$$

$$= \int_X I_{\alpha|\beta}(x) d\mu(x)$$

$I_{\alpha|\beta}(x)$ represents the amount of information gained when we are told which cell of $\alpha$ the point $x$ is in, if we already know which cell of $\beta$ it is in.

1.2.3 Variational principle

The variational principle provides a relationship between topological entropy and measure-theoretic entropy. Let $T: X \to X$ be a continuous map on a compact metric space $(X, d)$. Denote by $M(X, T)$ the space of all $T$-invariant probability measures on the measurable space $(X, \mathcal{B})$, where $\mathcal{B}$ is the $\sigma$-algebra of Borel subsets of $X$.

**Definition 1.2.12.** The entropy map of the continuous map $T: X \to X$ is the map $\mu \mapsto h_{\mu}(X, T)$ which is defined on $M(X, T)$ and has values in $[0, \infty]$.

The variational principle is given by the following theorem.

**Theorem 1.2.13.** Let $T: X \to X$ be a continuous map on a compact metric space $X$. Then

$$h_{top}(X, T) = \sup_{\mu \in M(X, T)} h_{\mu}(X, T).$$

**Definition 1.2.14.** Let $T: X \to X$ be a continuous transformation on a compact metric space $X$. A measure $\mu \in M(X, T)$ is called a measure of maximal entropy for $T$ if $h_{\mu}(X, T) = h_{top}(X, T)$. Let $M_{\text{max}}(X, T)$ denote the collection of measures with maximal entropy.

The following theorem (see [Wal]) gives sufficient conditions for the existence of measures of maximal entropy.
Theorem 1.2.15. Let $T : X \to X$ be a continuous transformation on a compact metric space $X$.

(i) $M_{\text{max}}(X, T)$ is convex.

(ii) If $h_{\text{top}}(X, T) < \infty$ the extreme points of $M_{\text{max}}(X, T)$ are precisely the ergodic members of $M_{\text{max}}(X, T)$.

(iii) If $h_{\text{top}}(X, T) < \infty$ and $M_{\text{max}}(X, T) \neq \emptyset$ then $M_{\text{max}}(X, T)$ contains an ergodic measure.

(iv) If $h_{\text{top}}(X, T) = \infty$ then $M_{\text{max}} \neq \emptyset$.

(v) If the entropy map is upper semi-continuous then $M_{\text{max}}(X, T)$ is compact and nonempty.

1.2.4 Topological pressure

In this section we give a brief overview of topological pressure (see Chapter 9 of [Wal]). Topological pressure is a generalization of topological entropy. Let $(X, d)$ be a compact metric space, $C(X, \mathbb{R})$ the space of real-valued continuous functions on $X$, and $T : X \to X$ a continuous transformation. Just as we did with topological entropy, we may give the definition of topological pressure using spanning sets, separated sets, or open covers.

Definition 1.2.16. For $f \in C(X, \mathbb{R})$, $n \geq 1$ and $\varepsilon > 0$, let

$$Q_n(T, f, \varepsilon) = \inf \left\{ \sum_{x \in E} \exp \left( \sum_{i=0}^{n-1} f(T^i x) \right) : E \text{ is an } (n^*, \varepsilon) \text{ spanning set for } X \right\}.$$ 

We define the topological pressure of $T$ and $f$ to be

$$P(T, f) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log Q_n(T, f, \varepsilon).$$

The map $P(T, \cdot) : C(X, \mathbb{R}) \to \mathbb{R} \cup \{\infty\}$ is called the topological pressure function of $T$.

The function $f$ in Definition 1.2.16 is called the potential function. We see that $P(T, 0) = h_{\text{top}}(X, T)$. We also give the definition of topological pressure using separated sets.

Definition 1.2.17. For $f \in C(X, \mathbb{R})$, $n \geq 1$ and $\varepsilon > 0$ let

$$P_n(T, f, \varepsilon) = \sup \left\{ \sum_{x \in E} \exp \left( \sum_{i=0}^{n-1} f(T^i x) \right) : E \text{ is an } (n^*, \varepsilon) \text{ separated subset of } X \right\}.$$
It can be proven that $P(T, f)$ is also given by

$$P(T, f) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log P_n(T, f, \varepsilon).$$

We define topological pressure using open covers. The diameter of a subset $U$ of a metric space $(X, d)$ is the number $\text{diam}(U) = \inf \{r \in \mathbb{R}_{\geq 0} : d(x, y) \leq r \text{ for all } x, y \in U\}$. The Lebesgue covering lemma says that if $(X, d)$ is a compact metric space and $\mathcal{U}$ is an open cover of $X$ then there exists $\delta > 0$ such that each subset of $X$ of diameter less than or equal to $\delta$ lies in some member of $\mathcal{U}$. Such a $\delta$ is called a Lebesgue number for $\mathcal{U}$.

**Definition 1.2.18.** For $T : X \to X$ continuous, $f \in C(X, \mathbb{R})$, $n \geq 1$, and $\mathcal{U}$ an open cover of $X$, let

$$q_n(T, f, \mathcal{U}) = \inf \left\{ \sum_{V \in \mathcal{V}} \inf_{x \in V} \exp \left( \sum_{i=0}^{n-1} f(T^i x) \right) : \mathcal{V} \text{ is a finite subcover of } \mathcal{U}_n^* \right\}.$$

Also, for $T : X \to X$ continuous, $f \in C(X, \mathbb{R})$, $n \geq 1$, and $\mathcal{U}$ an open cover of $X$, let

$$p_n(T, f, \mathcal{U}) = \inf \left\{ \sum_{V \in \mathcal{V}} \sup_{x \in V} \exp \left( \sum_{i=0}^{n-1} f(T^i x) \right) : \mathcal{V} \text{ is a finite subcover of } \mathcal{U}_n^* \right\}.$$

The next result gives relationships between the different quantities used above to define the topological pressure.

**Proposition 1.2.19.** Let $T : X \to X$ be continuous and $f \in C(X, \mathbb{R})$.

(i) If $\mathcal{U}$ is an open cover of $X$ with Lebesgue number $\delta$ then $q_n(T, f, \mathcal{U}) \leq Q_n(T, f, \delta/2) \leq P_n(T, f, \delta/2)$.

(ii) If $\varepsilon > 0$ and $\mathcal{U}$ is an open cover with $\text{diam}(\mathcal{U}) \leq \varepsilon$ then $Q_n(T, f, \varepsilon) \leq P_n(T, f, \varepsilon) \leq p_n(T, f, \mathcal{U})$.

With this proposition in hand, one can show that the above definitions of topological pressure are all equivalent.

**Proposition 1.2.20** (see page 210 of [Wal]). If $T : X \to X$ is continuous and $f \in C(X, \mathbb{R})$ then each of the following equals the topological pressure $P(T, f)$:
\( \lim \delta \to 0 \left[ \sup_{\mathcal{U}} \{ \lim_{n \to \infty} (1/n) \log p_n(T, f, \mathcal{U}) : \text{diam}(\mathcal{U}) \leq \delta \} \right] . \)

(ii) \( \lim k \to \infty \left[ \lim_{n \to \infty} (1/n) \log p_n(T, f, \mathcal{U}_k) \right] , \text{if } \{ \mathcal{U}_k \} \text{ is a sequence of open covers of } X \text{ with } \text{diam}(\mathcal{U}_k) \to 0. \)

(iii) \( \lim \delta \to 0 \left[ \sup_{\mathcal{U}} \{ \lim \inf_{n \to \infty} (1/n) \log q_n(T, f, \mathcal{U}) : \text{diam}(\mathcal{U}) \leq \delta \} \right] . \)

(iv) \( \lim \delta \to 0 \left[ \sup_{\mathcal{U}} \{ \lim \sup_{n \to \infty} (1/n) \log q_n(T, f, \mathcal{U}) : \text{diam}(\mathcal{U}) \leq \delta \} \right] . \)

(v) \( \lim k \to \infty \left[ \lim \sup_{n \to \infty} (1/n) \log q_n(T, f, \mathcal{U}_k) \right] , \text{if } \{ \mathcal{U}_k \} \text{ is a sequence of open covers of } X \text{ with } \text{diam}(\mathcal{U}_k) \to 0. \)

(vi) \( \sup_{\mathcal{U}} \{ \lim \sup_{n \to \infty} (1/n) \log q_n(T, f, \mathcal{U}) : \mathcal{U} \text{ is an open cover of } X \} . \)

(vii) \( \lim \varepsilon \to 0 \lim \sup_{n \to \infty} \frac{1}{n} \log Q_n(T, f, \varepsilon) \)

(viii) \( \lim \varepsilon \to 0 \lim \sup_{n \to \infty} \frac{1}{n} \log P_n(T, f, \varepsilon) \)

(ix) \( \lim \varepsilon \to 0 \lim \inf_{n \to \infty} (1/n) \log Q_n(T, f, \varepsilon) . \)

(x) \( \lim \varepsilon \to 0 \lim \inf_{n \to \infty} (1/n) \log P_n(T, f, \varepsilon) . \)

**Extension of the variational principle**

We will now state the extension the variational principle of Section 1.2.3 to the case of topological pressure.

**Theorem 1.2.21** (See page 218 of [Wal]). Let \( T : X \to X \) be a continuous transformation on a compact metric space \( X \) and let \( f \in C(X, \mathbb{R}) \). Then

\[
P(T, f) = \sup_{\mu \in M(X, T)} \left\{ h_\mu(X, T) + \int f \, d\mu \right\} . \tag{1.2.11}
\]

**Equilibrium states**

One uses the concept of pressure and the variational principle to extend the idea of measures of maximal entropy. In other words, \( P(T, f) \) and Theorem 1.2.21 allow one to select members of \( M(X, T) \) in a natural way.
**Definition 1.2.22.** Let \( T : X \to X \) be a continuous map on a compact metric space \( X \) and let \( f \in C(X, \mathbb{R}) \). A measure \( \mu \in M(X, T) \) is called an equilibrium state for \( f \) if \( P(T, f) = h_\mu(X, T) + \int f \, d\mu \). Let \( M_f(X, T) \) denote the collection of all equilibrium states for \( f \). Note that if \( \mu \in M_0(X, T) \) then \( \mu \) is a measure of maximal entropy.

### 1.3 The isomorphism problem

One reason to study the complexity of dynamical systems is to decide when two measure-preserving dynamical systems are isomorphic or conjugate. The usual way of approaching the isomorphism problem is to search for isomorphism invariants. Entropy is one example of an isomorphism invariant, i.e. two transformation \( T_1 : X_1 \to X_1 \) and \( T_2 : X_2 \to X_2 \) are isomorphic if and only if \( h_\mu(X_1, T_1) = h_\mu(X_2, T_2) \). Another isomorphism invariant that is widely studied considers the group of eigenvalues of a measure-preserving transformation.

One reason this section is included is to add context to our motivation for studying measurements of complexity in dynamical systems. While the definitions and theorems presented in this section are not required in order to understand the main portion of the thesis, Chapters 2, 3, and 4, we present them for completeness. We also include this section since some of the its concepts will be used in Section 1.6 about other concepts of complexity.

#### 1.3.1 Isomorphism of measure-preserving transformations

**Definition 1.3.1.** Let \( (X_1, \mathcal{B}_1, \mu_1, T_1) \) and \( (X_2, \mathcal{B}_2, \mu_2, T_2) \) be two measure-preserving systems. We say \( T_1 \) is isomorphic to \( T_2 \) if there exist sets of measure zero \( Z_1 \subset X_1 \) and \( Z_2 \subset X_2 \) and a one-to-one onto map \( \phi : X_1 \setminus Z_1 \to X_2 \setminus Z_2 \) such that

\[
\phi T_1(x) = T_2 \phi(x) \quad \text{for all} \quad x \in X_1 \setminus Z_1
\]

and \( \mu_1(\phi^{-1}E) = \mu_2(E) \) for all measurable \( E \subset X_2 \setminus Z_2 \).

**Theorem 1.3.2.** Entropy is an isomorphism invariant, i.e. if \( (X_1, \mathcal{B}_1, \mu_1, T_1) \) and \( (X_2, \mathcal{B}_2, \mu_2, T_2) \) are isomorphic then \( h_{\mu_1}(X_1, T_1) = h_{\mu_2}(X_2, T_2) \).
1.3.2 Ergodicity and mixing

While entropy is a quantitative isomorphism invariant, there are some invariants which are qualitative properties of a system. Examples of qualitative isomorphism invariants include ergodicity, weak mixing, and strong mixing.

**Definition 1.3.3.** Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system. \(T\) is **ergodic** if the only members \(B \in \mathcal{B}\) with \(T^{-1}B = B\) satisfy \(\mu(B) = 0\) or \(\mu(B) = 1\).

**Definition 1.3.4.** Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system. \(T\) is **weak mixing** if for all \(A, B \in \mathcal{B}\)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0.
\]

(1.3.1)

\(T\) is **strong mixing** if for all \(A, B \in \mathcal{B}\),

\[
\lim_{n \to \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).
\]

(1.3.2)

It is clear that strong mixing transformation implies weak mixing which implies ergodic.

1.3.3 Spectral isomorphism

Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving dynamical system. Recall the space \(L^2(X, \mathcal{B}, \mu)\) is the equivalence class of functions \(f : X \to \mathbb{C}\) such that \(\int |f|^2 d\mu < \infty\) where \(f, g \in L^2(X, \mathcal{B}, \mu)\) are equivalent if \(f = g\) a.e. \((\mu\{x \in X : f(x) \neq g(x)\} = 0\). \(L^2(X, \mathcal{B}, \mu)\) is a Hilbert space with inner product \(\langle f, g \rangle = \int f\bar{g} \, d\mu\). A family of function \(\{f_i\}_{i=1}^\infty\) form an **orthonormal basis** of \(L^2(X, \mathcal{B}, \mu)\) if \(\langle f_i, f_j \rangle = 0\) for all \(i \neq j\), \(\langle f_i, f_i \rangle = 1\) for all \(i = 1, 2, \ldots\), and every \(f \in L^2(X, \mathcal{B}, \mu)\) can be written in the form \(f = \sum_{i=1}^\infty a_i f_i\) for \(a_i \in \mathbb{C}\). We define the linear operator induced by \(T\), \(U_T : L^2 \to L^2\) by \((U_Tf)(x) = f(Tx)\).

**Definition 1.3.5.** Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system. We say \(\lambda \in \mathbb{C}\) is an eigenvalue of \(T\) if it is an eigenvalue of the linear operator \(U_T\) induced by \(T\). In other words, if there exists a function \(f \in L^2(X, \mathcal{B}, \mu), f \neq 0\) with \(U_Tf = \lambda f\) (or \(f(Tx) = \lambda f(x)\) a.e.) In this case \(f\) is called an eigenfunction corresponding to \(\lambda\).
Definition 1.3.6. Given a measure-preserving system \((X, \mathcal{B}, \mu, T)\) we say \(T\) has continuous spectrum if 1 is the only eigenvalue of \(T\) and the only eigenfunctions are constants.

Definition 1.3.7. Let \((X_1, \mathcal{B}_1, \mu_1, T_1)\) and \((X_2, \mathcal{B}_2, \mu_2, T_2)\) be two measure-preserving systems. \(T_1\) and \(T_2\) are spectrally isomorphic if there is a linear operator \(W : L^2(X_2, \mathcal{B}_2, \mu_2) \rightarrow L^2(X_1, \mathcal{B}_1, \mu_1)\) such that \(W\) is invertible, \(\langle Wf, Wg \rangle = \langle f, g \rangle\) for all \(f, g \in L^2(X_2, \mathcal{B}_2, \mu_2)\) and \(U_{T_1} W = WU_{T_2}\).

Definition 1.3.8. Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system. If \(T\) is ergodic, then \(T\) has discrete spectrum if there exists an orthonormal basis for \(L^2(X, \mathcal{B}, \mu)\) which consists of eigenfunctions of \(T\).

Theorem 1.3.9. Let \((X_1, \mathcal{B}_1, \mu_1, T_1)\) and \((X_2, \mathcal{B}_2, \mu_2, T_2)\) be two measure-preserving systems. If \(T_1\) and \(T_2\) are ergodic and have discrete spectrum then \(T_1\) and \(T_2\) are spectrally isomorphic if and only if \(T_1\) and \(T_2\) have the same eigenvalues.

1.3.4 Topological conjugacy and discrete spectrum

In this section we give the topological analogue of the previous section.

Definition 1.3.10. Let \(T : X \rightarrow X\) and \(S : Y \rightarrow Y\) be homeomorphisms of compact metric spaces. We say \(T\) is topologically conjugate to \(S\) if there is a homomorphism \(\phi : X \rightarrow Y\) such that \(\phi T = S \phi\). \(\phi\) is called a conjugacy

Definition 1.3.11. Let \(X\) be a compact metric space, \(T : X \rightarrow X\) a homeomorphism, and \(f\) a complex-valued continuous function \(X\) which is not identically zero. We say \(f\) is an eigenfunction for \(T\) if there exists a \(\lambda \in \mathbb{C}\) such that \(f(Tx) = \lambda f(x)\) for all \(x \in X\). \(\lambda\) is called the eigenvalue for \(T\) corresponding to \(f\).

Definition 1.3.12. Let \(T\) be a homomorphism on a compact metric space \(X\). We say \(T\) has topological discrete spectrum if the eigenfunctions of \(T\) span \(C(X, \mathbb{C})\) the space of continuous complex-valued functions on \(X\). In other words, every function in \(C(X, \mathbb{C})\) can be written as a linear combination of eigenfunctions of \(T\).
1.4 Neural complexity and intricacy

In this section we describe the concept of neural complexity and its generalization to intricacy functionals in probability. Along with entropy, these concepts will be the basis for our new complexity measurements.

1.4.1 Entropy and mutual information in probability

The introduction of neural complexity by Edelman, Sporns, and Tononi [TSE] involves the concepts of entropy and mutual information in the setting of probability. While we have already defined many terms in this section for a general probability space, we will define them again here using notation and terminology seen more often in probability. In the current setting we consider a probability space $X$ and a discrete random variable $x : X \to E$ taking values in a finite (or countable) space $E$. Given $x_i \in E$ we denote the probability that $x = x_i$ by

$$p(x_i) = \Pr\{x = x_i\}.$$ 

The entropy of $x$ is given by

$$H(x) = -\sum_i p(x_i) \log p(x_i)$$ \hspace{1cm} (1.4.1)

Entropy quantifies the randomness of the random variable. Suppose we have another discrete random variable $y : X \to F$ taking values in the finite (or countable) space $F$. We denote the probability that $x = x_i$ and $y = y_j$ by

$$p(x_i, y_j) = \Pr\{x = x_i \text{ and } y = y_j\}$$

The joint entropy of $x$ and $y$ is given by

$$H(x, y) = -\sum_{i,j} p(x_i, y_j) \log p(x_i, y_j).$$

$H(x, y)$ is a measure of the extent to which the randomness of the two variables is shared.
Conditional entropy

Again, consider the two random variables $x$ and $y$. The conditional probability that $x = x_i$ given $y = y_j$, for $x_i \in E$, $y_j \in F$, is given by

$$\Pr\{x = x_i | y = y_j\} = \frac{p(x_i, y_j)}{p(y_j)} = p(x_i | y_j).$$

The conditional entropy of $x$ given $y$ is defined as

$$H(x | y) = - \sum_{i,j} p(y_j) \log p(x_i, y_j).$$

Mutual information

The mutual information of two random variables $x$ and $y$ is a measure of their common randomness. If the variables are independent then they have a mutual information of 0 (minimum possible). Mutual information is maximized if one variable is a function of the other. The mutual information, MI, of $x$ and $y$ is defined by

$$\text{MI}(x, y) := H(x) + H(y) - H(x, y)$$

$$= H(x) - H(x | y)$$

$$= H(y) - H(y | x)$$

$$= \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}.$$

1.4.2 Neural complexity

We now turn to the measure of neural complexity proposed by Edelman, Sporns, and Tononi. The goal of neural complexity is to provide a measure that reflects the interaction of two properties of the brains of higher vertebrates: the functional segregation of different brain regions and their integration in perception and behavior.

Given $n \in \mathbb{N}$, we consider systems formed by a finite family $\{x_i : i \in n^*\}$ of random variables. For any $S \subset n^*$ the system is divided into two families $x_S = \{x_i : i \in S\}$ and $x_{S^c} = \{x_i : i \in S^c\}$. 
Then we compute the mutual information $\text{MI}(x_S, x_{S^c})$ and consider the average

$$\mathcal{I}(x) := \frac{1}{n + 1} \sum_{S \subseteq n^*} \frac{1}{|S|} \text{MI}(x_S, x_{S^c}).$$

(1.4.2)

Notice Equation 1.4.2 is a weighted average of mutual information over all subsets $S \subseteq n^*$ where the weights are given by

$$\frac{1}{n + 1} \frac{1}{|S|}.$$

The weights are uniform among each fixed $k = |S|$. The random variables in the original research on neural complexity are components of a neural system. $\mathcal{I}$ is 0 in systems of both complete order and complete disorder. It is large when there is a non-trivial correlation between its subsystems as well as a large number of internal degrees of freedom. Edelman, Sporns, and Tononi say that this is what one should intuitively expect from a complexity function.

### 1.4.3 Intricacy in probability

We next examine the work of Buzzi and Zambotti [BZ] in studying the concept of neural complexity in the general probabilistic setting. We begin with their definition of a system of coefficients.

**Definition 1.4.1.** A *system of coefficients* is a family of numbers

$$\{c^n_S : n \in \mathbb{N}, S \subseteq n^*\}$$

satisfying, for all $n \in \mathbb{N}$ and $S \subseteq n^*$

(a) $c^n_S \geq 0$

(b) $\sum_{S \subseteq n^*} c^n_S = 1$

(c) $c^n_{S^c} = c^n_S$.

**Example 1.4.2.** Some examples of systems of coefficients are

(i) $c^n_S = \frac{1}{2^n}$ (uniform)
(ii) $c^n_S = \frac{1}{n+1} \binom{n}{|S|} \frac{1}{n+1}$ (neural complexity)

(iii) $c^n_S = \frac{1}{2} \left( p^{|S|}(1-p)^{|S^c|} + (1-p)^{|S|}p^{|S^c|} \right)$ for fixed $0 < p < 1$ (p-symmetric)

**Definition 1.4.3.** Given a system of coefficients $c^n_S$ and a finite set of random variables $\{x_i : i \in n^*\}$, for each $S \subset n^*$ let $x_S := \{x_i : i \in S\}$. The corresponding mutual information functional $I_c$ is defined by

$$I_c(x) := \sum_{S \subset n^*} c^n_S MI(x_S, x_{S^c}) = \sum_{S \subset n^*} c^n_S [H(x_S) + H(x_{S^c}) - H(x_S, x_{S^c})]. \quad (1.4.3)$$

**Definition 1.4.4.** An intricacy is a mutual information functional satisfying

1. **exchangeability:** if $n, m \in \mathbb{N}$ and $\phi : n^* \to m^*$ is a bijection, then $I_c(x) = I_c(y)$ for any $x := \{x_i : i \in n^*\}$, $y := \{x_{\phi^{-1}(j)} : j \in m^*\}$;

2. **weak additivity:** if $I_c(x, y) = I_c(x) + I_c(y)$ for any two independent systems $\{x_i : i \in n^*\}$, $\{y_j : j \in m^*\}$.

We now give some results about intricacy found in [BZ].

**Theorem 1.4.5.** Let $c^n_S$ be a system of coefficients and $I_c$ the associated mutual information functional.

1. $I_c$ is an intricacy if and only if there exists a symmetric probability measure $\lambda_c$ on $[0, 1]$ such that for all $S \subset n^*$,

$$c^n_S = \int_{[0,1]} x^{|S|}(1-x)^{n-|S|} \lambda_c(dx), \quad \forall S \subset n^*.$$ 

2. The measure, $\lambda_c$, is uniquely determined by $I_c$. Moreover $I_c$ is non-null, i.e. there exists some nonzero $c^n_S$ for $S \notin \{\emptyset, n^*\}$, if and only if $\lambda_c((0, 1)) > 0$. In this case $c^n_S > 0$ for all $S \subset n^*$, $S \notin \{\emptyset, n^*\}$.

3. For the neural complexity weights we have

$$c^n_S = \frac{1}{n+1} \frac{1}{\binom{n}{|S|}} = \int_{[0,1]} x^{|S|}(1-x)^{n-|S|} \lambda_c(dx), \quad \text{for all } S \subset n^*,$$

i.e., $\lambda_c$ is Lebesgue measure on $[0, 1]$ and neural complexity is an intricacy.
We note that $c^S$ depends only on $|S|$.

### 1.5 Examples of Systems

We now describe some examples of both topological and measure-theoretic dynamical systems. In the following chapters we will study these examples with respect to the new concepts of intricacy and average sample complexity of dynamical systems.

#### Topological dynamical systems

##### 1.5.1 Subshifts and shifts of finite type

Let $\mathcal{A}$ be a finite set. We call $\mathcal{A}$ an alphabet and give it the discrete topology. The (two-sided) full shift space, $\Sigma(\mathcal{A})$, is defined as

$$\Sigma(\mathcal{A}) = \prod_{-\infty}^{\infty} \mathcal{A} = \{x = (x_i)_{-\infty}^{\infty} : x_i \in \mathcal{A} \text{ for each } i\}, \tag{1.5.1}$$

and is given the product topology. The one-sided full shift space is given by

$$\Sigma(\mathcal{A})^+ = \{x = (x_i)_{0}^{\infty} : x_i \in \mathcal{A} \text{ for each } i\}, \tag{1.5.2}$$

The shift transformation $\sigma : \Sigma(\mathcal{A}) \to \Sigma(\mathcal{A})$ is defined by

$$(\sigma x)_i = x_{i+1} \quad \text{for } -\infty < i < \infty, \tag{1.5.3}$$

and $\sigma : \Sigma(\mathcal{A})^+ \to \Sigma(\mathcal{A})^+$ is defined by

$$(\sigma x)_i = x_{i+1} \quad \text{for } 0 < i < \infty.$$

If $\mathcal{A} = \{0, 1, \ldots, r - 1\}$ then we denote $\Sigma(\mathcal{A})$ or $\Sigma(\mathcal{A})^+$ by $\Sigma_r$ or $\Sigma_r^+$ and call it the full $r$-shift. In this case, the topology on $\Sigma_r$ is compatible with the metric

$$d(x, y) = \frac{1}{m + 1}, \text{ where } m = \inf\{|k| : x_k \neq y_k\}.$$
Henceforth, we will deal only with two-sided shift spaces over a finite alphabet \( \mathcal{A} = \{0, 1, \ldots, r-1\} \) unless otherwise stated.

**Definition 1.5.1.** A **subshift** is a pair \((X, \sigma)\) where \(X \subseteq \Sigma_r\) is a nonempty, closed, shift-invariant \((\sigma X = X)\) set. A **block** or **word** is an element of \(\mathcal{A}^r\) for \(r = 0, 1, 2, \ldots\), i.e. a finite string on the alphabet \(\mathcal{A}\). If \(x\) is a sequence in a subshift \(X\), we will sometimes denote the block in \(x\) from position \(i\) to position \(j\) by

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

We denote the empty block by \(\varepsilon\).

Denote the set of words of length \(n\) in a subshift \(X\) by \(\mathcal{L}_n(X)\). i.e,

\[
\mathcal{L}_n(X) = \{ x_{[i,i+n-1]} : x \in X, i \in \mathbb{Z} \}.
\]

The **language** of a subshift \(X\) is

\[
\mathcal{L}(X) = \bigcup_{n=0}^{\infty} \mathcal{L}_n(X).
\]

Let \(S \subseteq n^*\), \(S = \{s_0, s_1, \ldots, s_{|S|-1}\}\) and suppose \(w \in \mathcal{L}_n(X)\) such that \(w_{s_i} = a_{s_i}\) for \(i = 0, \ldots, |S| - 1\) and \(a_{s_i} \in \mathcal{A}\). Then we call \(a_{s_0}a_{s_1}\cdots a_{s_{|S|-1}}\) a **word at the places in** \(S\). Denote the set of words we can see at the places in \(S\) for all words in \(\mathcal{L}_n(X)\) by \(\mathcal{L}_S(X)\). More formally, if \(S = \{s_0, s_1, \ldots, s_{|S|-1}\}\), then

\[
\mathcal{L}_S(X) = \{ x_{s_0}x_{s_1} \cdots x_{s_{|S|-1}} : x \in X \}. \tag{1.5.4}
\]

Notice that \(\mathcal{L}_n^*(X) = \mathcal{L}_n(X)\). Given a subshift \(X \subseteq \Sigma(\mathcal{A})\), we will often consider the cover \(\mathcal{U}_n\) consisting of **rank n cylinder sets** defined by

\[
\{ x \in X : x_{-n} = i_{-n}, x_{-n+1} = i_{-n+1}, \ldots, x_0 = i_0, \ldots, x_n = i_n \} \tag{1.5.5}
\]

for some choices of \(i_{-n}, i_{-n+1}, \ldots, i_n \in \mathcal{A}\). In particular, if \(\mathcal{A} = \{0, 1, \ldots, r-1\}\) then the cover \(\mathcal{U}_0\) consists of rank 0 cylinder sets

\[
U_i = \{ x \in X : x_0 = i, \text{ for } i = 0, 1, \ldots, r - 1 \}. \tag{1.5.6}
\]
If \( X \subset \Sigma(A)^+ \) is a one-sided subshift then covers \( \mathbb{Z}_n \) consisting of one-sided rank \( n \) cylinder sets are defined by

\[
\{ x \in X : x_0 = i_0, x_1 = i_1, \ldots, x_n = i_n \}
\]

for some choices of \( i_0, i_1, \ldots, i_n \in A \).

**Definition 1.5.2.** A shift of finite type (SFT) is defined by specifying a finite collection, \( \mathcal{F} \), of forbidden words on a given alphabet, \( A = \{0, 1, \ldots, r\} \). Given such a collection \( \mathcal{F} \), define \( X_{\mathcal{F}} \subset \Sigma_r \) to be the set of all sequences none of whose subblocks are in \( \mathcal{F} \). i.e.

\[
X_{\mathcal{F}} = \{ x \in \Sigma(A) : x[i,j] \not\in \mathcal{F} \text{ for all } i, j \in \mathbb{Z} \}. \tag{1.5.7}
\]

Since \( X_{\mathcal{F}} \) is closed and \( \sigma \)-invariant, \( (X_{\mathcal{F}}, \sigma) \) is a subshift.

Suppose \( X_{\mathcal{F}} \) is an SFT and the longest word in \( \mathcal{F} \) has length \( k \). If \( \mathcal{F}' \) consists of every word of length \( k \) which contains one of the forbidden words from \( \mathcal{F} \) then \( X_{\mathcal{F}} \) and \( X_{\mathcal{F}'} \) define the same SFT. In this way, we may assume every word in \( \mathcal{F} \) has the same length. We call \( X_{\mathcal{F}} \) a \( k \)-step shift of finite type if it can be described by a collection of forbidden words all of which have length \( k + 1 \) but not by a collection of words of length \( l < k + 1 \). In this case be say \( X_{\mathcal{F}} \) has memory \( k \).

**Example 1.5.3.** Consider the shift of finite type \( X_{\mathcal{F}} \) over the alphabet \( A = \{0, 1\} \) with \( \mathcal{F} = \{11, 101\} \). Now consider the set \( \mathcal{F}' = \{011, 110, 111, 101\} \) which consists of all word of length 3 which contain a word in \( \mathcal{F} \). \( X_{\mathcal{F}} \) and \( X_{\mathcal{F}'} \) define the same 2-step SFT.

**Graph and adjacency matrix of a shift of finite type**

We will now introduce the concepts of a graph and adjacency matrix of a shift of finite type. For more details on these topics see chapter 2 of [LM].

**Definition 1.5.4.** A graph \( \mathcal{G} \) consists of a finite set \( \mathcal{V} \) of vertices together with a finite set \( \mathcal{E} \) of edges. Each edge \( e \in \mathcal{E} \) starts at a vertex denoted \( i(e) \in \mathcal{V} \) and terminates at a vertex \( t(e) \in \mathcal{V} \) (which can be the same as \( i(e) \)). We call \( i(e) \) the initial state of \( e \) and \( t(e) \) the terminal state of \( e \).

**Definition 1.5.5.** Let \( \mathcal{G} \) be a graph with vertex set \( \mathcal{V} = \{1, 2, \ldots, r\} \). For vertices \( i, j \in \mathcal{V} \), let \( M_{ij} \) denote the number of edges in \( \mathcal{G} \) with initial state \( i \) and terminal state \( j \). Then the adjacency matrix
of $G$ is $M = (M_{ij})$. We denote the fact that $M$ is formed from $G$ by writing $M = M(G)$.

We can also go in the other direction. Given an $r \times r$ matrix $M = (M_{ij})$ with nonnegative integer entries, the graph of $M$ is the graph $G = G(A)$ with vertex set $V = \{1, 2, \ldots, r\}$, and with $M_{ij}$ distinct edges with initial state $i$ and terminal state $j$.

Next we will see how a graph or adjacency matrix gives rise to a shift of finite type.

**Definition 1.5.6.** Let $G$ be a graph with edge set $E$ and adjacency matrix $M$. The **edge shift** $X = X_G = X_M$ is the shift space over the alphabet $E$ specified by

$$X = \{x \in \Sigma(E) : t(x_k) = i(x_{k+1}) \text{ for all } k \in \mathbb{Z}\}.$$

In other words, a sequence in $X_G$ describes a *bi-infinite walk* or *bi-infinite trip* on $G$.

**Proposition 1.5.7.** If $G$ is a graph with adjacency matrix $M$, then the associated edge shift $X$ is a 1-step shift of finite type.

**Proof.** Let $A = E$ be the alphabet of $X$. Consider the finite collection $F = \{ef : e, f \in A, t(e) \neq i(f)\}$ of 2-blocks over $A$. A point $x \in \Sigma(A)$ lies in $X$ exactly when no block of $F$ occurs in $x$, so $X = X_F$ and $X$ is a 1-step shift of finite type. \qed

**Definition 1.5.8.** A path $p = e_1e_2 \ldots e_m$ on a graph $G$ is a finite sequence of edges $e_i$ from $G$ such that $t(e_i) = i(e_{i+1})$ for $i = 1, 2, \ldots, m - 1$. The **length** of $p = e_1e_2 \ldots e_m$ is $|p| = m$, the number of edges it traverses. The path $p$ **starts at vertex** $i(p) = i(e_1)$ and **terminates at vertex** $t(p) = t(e_m)$ and $p$ is a path from $i(p)$ to $t(p)$.

Information about paths on $G$ can be obtained from the adjacency matrix $M$ of $G$ as follows. Let $E^i_j$ denote the collection of edges in $G$ with initial state $i$ and terminal state $j$. Then $E^i_j$ is the collection of paths of length 1 from $i$ to $j$ and has size $M_{ij}$. Extending this idea, we can use the matrix powers of $M$ to count longer paths in $G$.

**Proposition 1.5.9.** Let $G$ be a graph with adjacency matrix $M$, and let $m \geq 0$. The number of paths of length $m$ from $i$ to $j$ is $(M^m)_{ij}$, the $(i, j)$th entry of $M^m$. 

26
If a graph $G$ has at most one edge between any two vertices then the adjacency matrix $M$ of $G$ will only contain 0's and 1's. In this case, we can describe a walk on $G$ by naming the vertices visited instead of the edges visited. This leads to a different construction.

**Definition 1.5.10.** Let $M$ be an $r \times r$ matrix of 0's and 1's or, equivalently, the adjacency matrix of a graph $G$ such that between any two vertices there is at most one edge. The vertex shift $\hat{X}_M = \hat{X}_G$ is the shift space with alphabet $A = \{1, 2, \ldots, r\}$ defined by

$$\hat{X}_M = \{x \in \Sigma(A) : M_{x,x_{i+1}} = 1 \text{ for all } i \in \mathbb{Z}\} \quad (1.5.8)$$

The following proposition establishes the relationship between vertex shifts, edge shifts, and 1-step shifts of finite type.

**Proposition 1.5.11.**

(i) Up to a renaming of symbols, the 1-step shifts of finite type are the same as vertex shifts.

(ii) Up to a renaming of symbols, every edge shift is a vertex shift (on a different graph).

**Proof.**

(i) Let $\hat{X}_M$ be a vertex shift. Then the shift of finite type with the list of forbidden word $\mathcal{F} = \{ij : M_{ij} = 0\}$ is exactly $\hat{X}_M$. Now suppose $X_F$ is a 1-step shift of finite type over an alphabet $A$ with $r$ symbols. We may rename the symbols of of $A$ to be $\{1, 2, \ldots, r\}$ so $\mathcal{F}$ will consist of 2-blocks made up of elements of $\{1, 2, \ldots, r\}$. Let $M$ be the $r \times r$ matrix such that $M_{ij} = 0$ if $ij \in \mathcal{F}$ and $M_{ij} = 1$ otherwise. Then $M$ is the adjacency matrix to the vertex shift $X_M$ and $X_M$ is equivalent to $X_F$.

(ii) This follows from part (i) and Proposition 1.5.7.

In this paper, when we associate an adjacency matrix $M$ to a 1-step shift of finite type $X$ over an alphabet $A$, we will regard the shift of finite type as a vertex shift. Thus, $M$ will consist of all 0's and 1's. We will label the vertices of the graph $G$ of $M$ with the elements of $A$ so a sequence of $X$ will be a sequence of vertices visited on a bi-infinite walk on $G$.
Example 1.5.12. Let $A = \{0, 1\}$ and $F = \{11\}$; then $(X_F, \sigma)$ is called the \textit{golden mean shift}. It consists of all two sided sequences on the symbols 0 and 1 that have no two consecutive 1’s. The graph and adjacency matrix for the golden mean shift are given in Table 1.1

Let $M$ be the adjacency matrix for a 1-step shift of finite type $X$. Proposition 1.5.9 tells us that the number of paths of length $m$ from $i$ to $j$ on the graph $\mathcal{G}$ of $M$ is $(M^m)_{ij}$. Since the sequence of vertices visited on a path of length $m$ on $\mathcal{G}$ from $i$ to $j$ is a word $w = w_1w_2\ldots w_{m+1}$ in $\mathcal{L}_{m+1}(X)$ with $w_1 = i$ and $w_{m+1} = j$ we have

$$|\mathcal{L}_{m+1}(X)| = \sum_{i,j} (M^m)_{ij}. \quad (1.5.9)$$

Example 1.5.13. If $M$ is the adjacency matrix for the golden mean shift, then

$$M^2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

By adding up the elements of $M^2$, we see the golden mean shift has 5 words of length 3, namely 000, 001, 010, 100, and 101. Moreover, since $(M^2)_{11} = 2$, we see the golden mean shift two words of length 3 starting and ending with 0 namely, 000 and 010. Similarly, it contains one word of length 3 starting with 0 and ending in 1, namely 001 and so on.

Definition 1.5.14. We call a subshift \textit{irreducible} if for every ordered pair of blocks $u, v \in \mathcal{L}(X)$ there is a $w \in \mathcal{L}(X)$ so that $uwv \in \mathcal{L}(X)$. If $X$ is an irreducible shift of finite type and $M$ is its adjacency matrix then for all $i$ and $j$ there is an integer $m \geq 1$ such that $(M^m)_{ij} > 0$. In this case we call $M$ an irreducible matrix.

One property of the adjacency matrix that will be important in our study is the smallest power
of the matrix such that every entry is nonzero (if this number exists). We denote the property that every entry in a matrix \( M \) is positive by \( M > 0 \). Define

\[
\rho(M) := \min\{k \geq 1 : M^k > 0\}.
\]

If no such power exists then we say \( \rho(M) = \infty \).

**Higher block shifts**

There will be times when we want to consider blocks of consecutive symbols that appear in sequences of a shift space. We consider these larger blocks as symbols of a new alphabet in a new subshift.

More precisely, let \( X \) be a subshift over an alphabet \( A \) and \( A_X^{[k]} = \mathcal{L}_k(X) \) be the collection of all allowed \( k \)-blocks in \( X \). We can consider \( A_X^{[k]} \) as an alphabet in its own rights, and form the full shift \( (A_X^{[k]})^\mathbb{Z} \). Define the \( k \)th higher block code \( \beta_k : X \to (A_X^{[k]})^\mathbb{Z} \) by

\[
(\beta_k(x))_i = x_{[i,i+k-1]}.
\]

Thus, \( \beta_k \) replaces the \( i \)th coordinate of \( x \) with a block of coordinates in \( x \) of length \( k \) starting at position \( i \).

**Definition 1.5.15.** Let \( X \) be a shift space. The \( k \)th higher block shift \( X^{[k]} \) or higher block presentation of \( X \) is the image \( X^{[k]} = \beta_k(X) \) in the full shift over \( A_X^{[k]} \).

**Proposition 1.5.16.** If \( X \) is a \( k \)-step shift of finite type, then \( X^{[k]} \) is a 1-step shift of finite type, equivalently a vertex shift.

**Proof.** This is clear since any \((k+1)\)-block in \( X \) may be regarded as a 2-block in \( X^{[k]} \) so the forbidden words of length \( k+1 \) in \( X \) can be taken as forbidden words of length 2 in \( X^{[k]} \). Proposition 1.5.11 tells us that this 1-step shift of finite type is equivalent to a vertex shift. \( \square \)

This proposition allows us to reduce our analysis of shifts of finite type to 1-step shifts of finite type since every \( k \)-step shift of finite type is equivalent to a 1-step shift of finite type.
**Example 1.5.17.** Let $X$ be the golden mean shift. Then

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

and $X^{[2]}$ can be described by the constraints $\mathcal{F} = \{ac, ba, bb, cc\}$. Each of these 2-blocks is forbidden since they fail to overlap progressively. By renaming the symbols $a, b,$ and $c$ to 0, 1, and 2 respectively we see the graph and adjacency matrix for $X^{[2]}$ in Table 1.2.

**Entropy of a subshift**

Let $(X, \sigma)$ be a subshift over an alphabet $\mathcal{A}$ and let $\mathcal{U} = \mathcal{U}_0 = \{x \in X : x_0 = i \text{ for } i \in \mathcal{A}\}$ be the open cover by rank 0 cylinder sets. It is known (see for example [Pet]) that

$$h_{\text{top}}(X, \sigma) = \lim_{n \to \infty} \frac{1}{n} \log N(\mathcal{U}_0^{n-1}).$$

Since the number of nonempty elements of $\mathcal{U}_0^{n-1}$ is just the number of $n$-blocks that can appear in a sequence in $X$, we get that

$$h_{\text{top}}(X, \sigma) = \lim_{n \to \infty} \frac{1}{n} \log |\mathcal{L}_n(X)|.$$

It is known (see for example [LM]) that given the adjacency matrix, $M$, of an irreducible shift of finite type $X$, there exists an eigenvalue $\lambda > 0$ of $M$ such that $\lambda \geq |\mu|$ for all other eigenvalues $\mu$.
of $M$. $\lambda$ called the Perron eigenvalue of $M$, and

$$h_{\text{top}}(X, \sigma) = \log \lambda. \quad (1.5.11)$$

**Topological pressure of shifts of finite type**

In this section we discuss the topological pressure of shifts of finite type by looking at some examples.

*Example 1.5.18.* In this example we give the formula for the topological pressure of a function of just one coordinate on the full $r$-shift. Let $f$ be a continuous function $f : \Sigma_r \to \mathbb{R}$ that depends only on the 0th coordinate, i.e. there is a function $f_0 : \{0, 1, \ldots, r - 1\} \to \mathbb{R}$ so that for all $x \in \Sigma_r$

$$f(x) = f_0(x_0).$$

It is known (see for example [Wal]) that the topological pressure of $f$ is given by

$$P(\sigma, f) = \log \left( \sum_{i=0}^{r-1} e^{f_0(i)} \right) \quad (1.5.12)$$

*Example 1.5.19.* For a shift of finite type $X \subset \Sigma_r$, let $f_0 : \mathcal{L}_n(X) \to \mathbb{R}$ be a continuous real-valued function on the set of $n$-blocks in $X$. Define the topological pressure of $f$ to be

$$P(\sigma, f) = \lim_{n \to \infty} \frac{1}{n} \log \sum_{w \in \mathcal{L}_n(X)} \exp \left( \sup_{x \in w} \sum_{i=0}^{n-1} f(\sigma^i x) \right). \quad (1.5.13)$$

Notice, Equation 1.5.13 become Equation 1.5.12 if $X = \Sigma_r$ and $f$ is a function of just the 0th coordinate of $w \in \mathcal{L}_n(X)$. Also, notice that if $f(w) = 0$ for all $w$, then

$$P(\sigma, f) = \lim_{n \to \infty} \frac{1}{n} \log \sum_{w \in \mathcal{L}_n(X)} 1 = \lim_{n \to \infty} \frac{1}{n} \log |\mathcal{L}_n(X)| = h_{\text{top}}(X).$$

31
1.5.2 Other subshifts

Let $\mathcal{A}$ be a finite alphabet and let $u \in \mathcal{A}^\mathbb{N}$ or $u \in \mathcal{A}^\mathbb{Z}$ be a one-sided or two-sided sequence. Denote by $\mathcal{A}^\ast$ the set of finite words made of elements of $\mathcal{A}$. A word $w = w_1 w_2 \ldots w_k$ is said to occur at position $i$ in the sequence $u$ if $u_i = w_1$, $u_{i+1} = w_2$, $\ldots$, $u_{i+k} = w_k$. The language of length $n$ of the sequence $u$, denoted by $\mathcal{L}_n(u)$ is the set of all words of length $n$ in $\mathcal{A}^\ast$ which occur in $u$. The language of $u$ is the set of all finite words in $\mathcal{A}^\ast$ which occur in $u$, i.e. $\mathcal{L}(u) = \bigcup_{n=0}^{\infty} \mathcal{L}_n(u)$. We define a dynamical system associated with the sequence $u$. Denote the orbit of $u$ under the shift, $\sigma$, by $\mathcal{O}(u)$.

**Definition 1.5.20.** The symbolic dynamical system associated with the one-sided (respectively two-sided) sequence $u$ with values in the alphabet $\mathcal{A}$ is the system $(\overline{\mathcal{O}(u)}, \sigma)$, where $\overline{\mathcal{O}(u)} \subset \Sigma(\mathcal{A})^+$ (respectively $\Sigma(\mathcal{A})$) is the closure of the orbit of the sequence $u$ under the action of the shift $\sigma$.

**Definition 1.5.21.** Let $u$ be a sequence. We call the complexity function of $u$, and denote by $p_u(n)$, the function which with each positive integer $n$ associate $|\mathcal{L}_n(u)|$, that is, the number of different words of length $n$ occurring in $u$. Note that this is the same as the number of words of length $n$ in the subshift $\overline{\mathcal{O}(u)}$, i.e. $\mathcal{L}_n(u) = \mathcal{L}_n(\overline{\mathcal{O}(u)})$.

A sequence $u$ is periodic if there is a $k \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, $u_n = u_{n+k}$. $u$ is eventually periodic if there are $k, N \in \mathbb{N}$ such that for all $n \geq N$, $u_n = u_{n+k}$. A sequence which is neither periodic nor eventually periodic is called aperiodic.

In the next two sections we describe ways to generate interesting sequences. Later we study the symbolic dynamical systems associated with these sequences.

Substitution sequences

**Definition 1.5.22.** A substitution $\tau$ is a map from an alphabet $\mathcal{A}$ into the set $\mathcal{A}^\ast \setminus \{\varepsilon\}$ of nonempty finite words on $\mathcal{A}$. A fixed point of the substitution $\tau$ is an infinite sequence $u$ with $\tau(u) = u$. A substitution sequence is the fixed point of a substitution.

**Example 1.5.23.**

1. The Morse sequence (Prouhet-Thue-Morse), $u$, is the fixed point beginning with 0 of the
substitution defined on \( \{0, 1\} \) by \( \tau(0) = 01 \) and \( \tau(1) = 10 \):

\[
u = 01101001100101100101100101100101100101100101100101100101100101100101100101100101100101\ldots
\]

2. The **Fibonacci sequence** is the fixed point beginning with 0 of the Fibonacci substitution defined on \( \{0, 1\} \) by \( \tau(0) = 01 \) and \( \tau(1) = 0 \):

\[
u = 010010010010010010010010010010010010010010100100100100100100100100100100\ldots
\]

**Sturmian sequences**

**Definition 1.5.24.** A **Sturmian sequence** is defined as a sequence \( \nu \) such that the complexity function \( p_\nu \) satisfies

\[
p_\nu(n) = n + 1 \text{ for all } n \in \mathbb{N}.
\]

**Definition 1.5.25.** A **rotation sequence** is a sequence \( \nu \) such that there is an irrational number \( \alpha \in [0, 1] \) and \( \beta \in \mathbb{R} \) such that

\[
u_n = \lfloor (n + 1)\alpha - \beta \rfloor - \lfloor n\alpha + \beta \rfloor \text{ for all } n
\]

Rotation sequences occur as the natural codings associated with rotations. Let \( \mathbb{T}^1 \) be the circle, identified to \( \mathbb{R}/\mathbb{Z} \) and consider the rotation \( R : \mathbb{T}^1 \to \mathbb{T}^1 \) such that

\[
R x = x + \alpha \mod 1
\]

We consider two intervals \( I_0 = [0, 1 - \alpha) \) and \( I_1 = (1 - \alpha, 1) \) on \( \mathbb{T}^1 \). Let \( \nu : \mathbb{T}^1 \to \{0, 1\} \) be the coding function defined by \( \nu(x) = j \) if \( x \in I_j \). The rotation sequence defined by \( \alpha \) and \( \beta \) is the sequence \( \nu(R^n\beta) \), that is, the coding of the positive orbit of \( \beta \) under the rotation \( R \).

**Proposition 1.5.26.** A sequence is Sturmian if and only if it is a rotation sequence.

It is described in [Fog] how to use the continued fraction expansion of an irrational number to find the associated Sturmian sequence.
Measure-theoretic dynamical systems

In this section we describe two types of measure-theoretic dynamical systems which we will study our complexity measurements on later: Bernoulli shifts and Markov shifts.

1.5.3 Bernoulli shifts

Let $A = \{0, 1, \ldots, n - 1\}$ be an alphabet of finitely many symbols with weights $p_0, p_1, \ldots, p_{n-1}$ such that $p_i > 0$ for all $i$ and $\sum_{i=0}^{n-1} p_i = 1$. Form the product space $A^\mathbb{Z}$ of all two-sided sequences of symbols in $A$, and give $A^\mathbb{Z}$ the product measure $\mu$ determined by the given probability measure on $A$. Thus, for a typical cylinder set determined by a set of places $i_1, \ldots, i_k \in \mathbb{Z}$ and elements $j_1, \ldots, j_k \in A$, $\mu\{x: x_{i_1} = j_1, \ldots, x_{i_k} = j_k\} = p_{j_1}p_{j_2} \cdots p_{j_k}$.

The shift transformation $\sigma: A^\mathbb{Z} \rightarrow A^\mathbb{Z}$ preserves $\mu$. The resulting measure-preserving system is denoted $\mathcal{B}(p_0, p_1, \ldots, p_{n-1})$. The measure-theoretic entropy of a Bernoulli shift is given by

$$h_\mu(A^\mathbb{Z}, \sigma) = -\sum_{i=0}^{r-1} p_i \log p_i.$$

1.5.4 Markov shifts

Form the product space $A^\mathbb{Z}$ with the shift transformation as above. We define an invariant measure as follows. Let $P = (P_{ij})$ be an $n \times n$ stochastic matrix, i.e. a matrix with nonnegative entries and each row sum equal to 1. Suppose also that $p = (p_0, p_1, \ldots, p_{n-1})$ is a row probability vector (all $p_i \geq 0$ and $\sum p_i = 1$) which is fixed by $P$:

$$pP = p.$$

Define the measure of a cylinder set determined by consecutive indices by

$$\mu_{p,p}\{x: x_i = j_0, x_{i+1} = j_1, \ldots, x_{i+k} = j_k\} = p_{j_0}P_{j_0j_1}P_{j_1j_2} \cdots P_{j_{k-1}j_k}.$$
We extend $\mu_{P,p}$ to a countably additive measure on the algebra generated by the cylinder sets. The resulting measure-preserving system $(\mathcal{A}^\mathbb{Z}, \mathcal{B}, \mu_{P,p}, \sigma)$ models a finite-state Markov chain. The measure-theoretic entropy of a Markov shift is given by

$$h_{\mu_{P,p}}(\sigma) = -\sum_{i,j} p_i P_{ij} \log P_{ij}.$$ 

A shift of finite type $X$ with adjacency matrix $M$ is sometimes referred to as a *topological Markov chain*. A Markov shift given by the stochastic matrix $P = (P_{ij})$ and fixed vector $p$ may be referred to as a Markov measure on a shift of finite type with the adjacency matrix $M = (M_{ij})$ satisfying $M_{ij} = 1 \iff P_{ij} > 0$ and $M_{ij} = 0 \iff P_{ij} = 0$. A Markov shift formed by putting a Markov measure on a $k$-step shift of finite type $X$ is called a *$k$-step Markov shift* or a *$k$-step Markov measure on $X$*.

**Measure of maximal entropy for a shift of finite type**

Let $X_M$ denote a shift of finite type with adjacency matrix $M$. Let $\lambda$ be the largest eigenvalue of $M$, $(u_0, \ldots, u_{n-1})$ a strictly positive left eigenvector for $\lambda$ and $(v_0, \ldots, v_{n-1})$ a strictly positive right eigenvector for $\lambda$ with $\sum_{i=0}^{n-1} u_i v_i = 1$. Let $p_i = u_i v_i$ for $i = 0, 1, \ldots, n - 1$ and $P_{ij} = M_{ij} v_j / (\lambda v_i)$ for $i, j = 0, 1, \ldots, n - 1$. The measure $\mu_{P,p}$ formed from $p = (p_0, \ldots, p_{n-1})$ and $P = (P_{ij})$ is called the *Shannon-Parry measure* for $X_M$.

**Theorem 1.5.27.** If $(X_M, \sigma)$ is a shift of finite type and $M$ is an irreducible matrix then the Shannon-Parry measure is the unique measure of maximal entropy for $(X_M, \sigma)$.

See Section 8.3 of [Wal] for more information on measures of maximal entropy. The Shannon-Parry measure is sometimes called just the Parry measure. This measure was introduced by Shannon, but Parry proved it is the unique measure of maximal entropy for irreducible shifts of finite type.

**Example 1.5.28.** Let $X$ be the golden mean shift from Example 1.5.12. The largest eigenvalue of the adjacency matrix of $X$ is the golden mean, $\phi = (1 + \sqrt{5})/2$. The left and right eigenvectors for this eigenvalue are

$$u = \left( \frac{\phi}{\phi + 2}, \frac{1}{\phi + 2} \right) \quad \text{and} \quad v = \left( \frac{\phi}{1} \right).$$
Thus, the Shannon-Parry measure (the unique measure of maximal entropy) on the golden mean shift is given by

\[ P = \begin{pmatrix} 1/\phi & 1/(\phi + 1) \\ 1 & 0 \end{pmatrix}, \quad p = \begin{pmatrix} \phi \\ 2\phi - 1 \end{pmatrix}. \]

1.6 Other concepts of complexity

In this section we will look at some other concepts of complexity in dynamical systems aside from entropy and pressure.

1.6.1 Sequence entropy

Sequence entropy of a measure-preserving dynamical system

In this section we will always consider the measure-preserving dynamical system \((X, B, \mu, T)\). In [Kus] Kushnirenko defines the concept of the sequence entropy of a transformation. We denote by \(A = (t_n)\) a strictly increasing sequence of positive integers.

**Definition 1.6.1.** For a sequence of integers \(A = (t_n)\), we define the sequence entropy of \(T\) with respect to \(A\), \(h_A(T)\), as follows. Let \(\alpha\) be a finite measurable partition of \(X\) with finite entropy \(H_\mu(\alpha)\). Let

\[ h_A(X, \alpha, T) = \limsup_{n \to \infty} \frac{1}{n} H_\mu(T^{t_1} \alpha \vee \cdots \vee T^{t_n} \alpha) \]

and

\[ h_A(X, T) = \sup_\alpha h_A(X, \alpha, T). \]

For the sequence \(A = (t_n)\) such that \(t_n = n\) for all \(n\), \(h_A(X, T)\) equals the usual measure-theoretic entropy. It is shown in [Kus] that two systems, each with zero entropy, can have different sequence entropies. Kushnirenko also shows the following theorem.

**Theorem 1.6.2.** A transformation \(T\) has discrete spectrum if and only if \(h_A(X, T) = 0\) for every sequence \(A\).

In [New1], Newton proves the following theorem which asserts that if \(T\) is an ergodic transformation with positive finite entropy then sequence entropy gives no new information than usual
measure-theoretic entropy. Given a sequence of integers $A = (t_n)$ let $s_A(n, k)$ be the number of distinct elements in $\bigcup_{i=1}^{n} \{-k + t_i, -k + t_i + 1, \ldots, k + t_i\}$. Then define

$$K(A) = \lim_{k \to \infty} \limsup_{n \to \infty} \frac{1}{n} s_A(n, k).$$

**Theorem 1.6.3.** If $T$ is an ergodic transformation with positive finite entropy, then

$$h_A(A, T) = K(A) h_\mu(X, T).$$

In [New2], Newton extends Theorem 1.6.3 to the case of ergodic transformations with infinite entropy.

**Theorem 1.6.4.** Let $T$ be an ergodic transformation such that $h_\mu(X, T) = \infty$. Given a sequence of integers $A = (t_n)$, the sequence entropy of $T$ is given as

$$h_A(X, T) = \begin{cases} 0 & \text{if } K(A) = 0 \\ \infty & \text{if } K(A) > 0 \end{cases}$$

In [KN] Krug and Newton extended the previous two theorems to general transformations.

**Theorem 1.6.5.** Let $T : X \to X$ be any transformation on the space $X$ and $A = (t_n)$ any sequence of integers then

$$h_A(X, t) = K(A) h_\mu(X, T)$$

where the right hand side is $0$ if $K(A) = 0$ and $h_\mu(X, T) = \infty$ and undefined if $K(A) = \infty$ and $h_\mu(X, T) = 0$.

In [Sal] Saleski studies the relationship between sequence entropy and mixing properties. Let $\mathcal{J}$ denote the set of all increasing sequences of positive integers. Given a measure-preserving system $(X, \mathcal{B}, \mu, T)$ let $\mathcal{P}$ denote the set of all partitions $\alpha$ of $X$ such that $h_\mu(X, \alpha, T) > 0$. For $A, B \in \mathcal{J}$ let $A \subseteq B$ denote $A$ being a subsequence of $B$.

**Theorem 1.6.6.** $T$ is strong mixing if and only if $\sup_{A \subseteq B} h_A(X, \alpha, T) = h_\mu(X, \alpha, T)$ for all $B \in \mathcal{J}$ and $\alpha \in \mathcal{P}$.

**Theorem 1.6.7.** $T$ is weak mixing if and only if $\sup_{A} h_A(X, \alpha, T) = h_\mu(X, \alpha, T)$ for all $\alpha \in \mathcal{P}$. 

37
1.6.2 Maximal pattern complexity

In [KZ], Kamae and Zamboni describe a quantity called maximal pattern complexity.

**Definition 1.6.8.** Let $A$ be a finite alphabet and $u \in A^\mathbb{N}$ a one-sided sequence of elements of $A$. Suppose $u = u_1 u_2 u_3 \ldots$. Let $A = (t_n)$ be a strictly increasing sequence of positive integers. For $k \in \mathbb{N}$, define the *maximal pattern complexity* of $u$

$$p_u^*(k) = \sup_A \text{card}\{u_{n+t_1} u_{n+t_2} \cdots u_{n+t_k} : n = 0, 1, 2, \ldots\}.$$  

(1.6.1)

where the supremum is taken over strictly increasing strings of positive integers of length $k$.

**Example 1.6.9.** Suppose $A = \{0, 1\}$, $u = 0010101011\ldots$ and $k = 3$. To find $p_u^*(k)$ we count the distinct strings of length 3 seen in $u$ the positions specified by consecutive elements of each sequence $A = (t_n)$. If we take $A = \{1, 3, 4, 7, 8, 12, \ldots\}$ then we count

$$\text{card}\{u_1 u_3 u_4, u_3 u_4 u_7, u_4 u_7 u_8, u_7 u_8 u_{12}, \ldots\} = \text{card}\{010, 100, 001, 011, \ldots\}$$

**Theorem 1.6.10.** A sequence $u$ is aperiodic if and only if $p_u^*(k) \geq 2k$ for all $k \in \mathbb{N}$.

In [KZ] Kamae and Zamboni introduce the concept of a *pattern Sturmian sequence*.

**Definition 1.6.11.** Let $A$ be a finite alphabet. A sequence $u \in A^\mathbb{N}$ is a *pattern Sturmian sequence* if $p_u^*(k) = 2k$ for all $k \in \mathbb{N}$.

Notice that pattern Saturnian sequences are the aperiodic sequences of least possible maximal pattern complexity.

1.6.3 Eulerian entropy

In [Moo], Moothathu defined the concept of Eulerian entropy. Using Bowen’s notation, recall that the topological entropy of the topological dynamical system $(X, T)$ is

$$h_{\text{top}}(X, T) = \lim_{\varepsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} \log s(n^*, \varepsilon).$$
where \( s(n^*, \varepsilon) \) is the maximum cardinality of an \((n^*, \varepsilon)\) separated set. The motivation in [Moo] is to 
\( h_{\text{top}}(X, T) \) by only looking at the initial chunks of the orbit of some point \( x \in X \).

Suppose \( x \in X \), \( n \in \mathbb{N} \) and \( \varepsilon > 0 \). Let \( \beta(x, T, n, \varepsilon) \) be the maximum of \( m \in \mathbb{N} \) such that 
\( \{x, Tx, \ldots, T^{m-1}x\} \) is an \((n^*, \varepsilon)\) separated set.

**Definition 1.6.12.** The *Eulerian entropy* of a point \( x \in X \) is

\[
h_E(x, T) = \lim_{\varepsilon \to 0^+} \limsup_{n \to \infty} \frac{1}{n} \log \beta(x, T, n, \varepsilon)
\]

and the *Eulerian entropy* of \( T \) is

\[
h_E(X, T) = \sup_{x \in X} h_E(x, T)
\]

**Theorem 1.6.13.** Eulerian entropy is a conjugacy invariant.

**Theorem 1.6.14.** Eulerian entropy is constant on a residual subset for transitive systems.

### 1.6.4 Further readings on concepts of complexity

There are many more measurements of complexity in dynamical systems. We won’t describe them here, but we will name some of them and cite some references where more information can be found: *slow entropy* [KT1] and [Blu], *entropy dimension* [FP], *Kolmogorov complexity* [Kol1], *topological complexity* [BHM], *independence entropy* [LMP], and *topological pressure dimension* [CL].
CHAPTER 2
Topological intricacy and average sample complexity

2.1 Definitions

The idea of intricacy proposed by Edelman, Sporns, and Tononi and studied in the probability setting by Buzzi and Zambotti allows generalization to the setting of dynamical systems, which we now propose to do. Our first definition of topological intricacy is based on the definition of topological entropy given by Adler, Konheim and McAndrew. Recall that \( n^* = \{0, 1, \ldots, n - 1\} \). Given a topological system \((X, T)\) and an open cover \( \mathcal{U} \), for each subset \( S \subset n^* \) let

\[
\mathcal{U}_S = \bigvee_{i \in S} T^{-i} \mathcal{U}.
\]

Recall that \( N(\mathcal{U}) \) denotes the minimum cardinality of the subcovers of \( \mathcal{U} \).

**Definition 2.1.1.** Let \( T : X \to X \) be a continuous map on a compact Hausdorff space \( X \), let \( \mathcal{U} \) be an open cover of \( X \), and let \( c^n_S \) be a system of coefficients (see Definition 1.4.1). Define the topological intricacy of \( T \) with respect to the open cover \( \mathcal{U} \) to be

\[
\text{Int}(X, \mathcal{U}, T) := \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log \left( \frac{N(\mathcal{U}_S)N(\mathcal{U}_{S^c})}{N(\mathcal{U}_{n^*})} \right). \tag{2.1.1}
\]

We will see later that this limit exists.

Next we define the topological average sample complexity. Note that

\[
\frac{1}{n} \sum_{S \subset n^*} c^n_S \log \left( \frac{N(\mathcal{U}_S)N(\mathcal{U}_{S^c})}{N(\mathcal{U}_{n^*})} \right) = \frac{1}{n} \sum_{S \subset n^*} (c^n_S \log N(\mathcal{U}_S) + c^n_S \log N(\mathcal{U}_{S^c}) - c^n_S \log N(\mathcal{U}_{n^*}))
\]

\[
= 2 \left( \frac{1}{n} \sum_{S \subset n^*} c^n_S \log N(\mathcal{U}_S) \right) - \frac{1}{n} \log N(\mathcal{U}_{n^*})
\]

40
and
\[ \lim_{n \to \infty} \frac{1}{n} \log N(\mathcal{U}_n) = \inf_{n} \frac{1}{n} \log N(\mathcal{U}_n) = h_{\text{top}}(X, \mathcal{U}, T), \]
the ordinary topological entropy of \( T \) with respect to the open cover \( \mathcal{U} \). Thus, in order to calculate intricacy we must find
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^2 \log N(\mathcal{U}_S). \]
Since this quantity is interesting on its own, we make the following definition.

**Definition 2.1.2.** Let \( T : X \to X \) be a continuous map on a compact Hausdorff space \( X \), let \( \mathcal{U} \) be an open cover of \( X \) and let \( c_S^2 \) be a system of coefficients. The *topological average sample complexity of \( T \) with respect to the open cover \( \mathcal{U} \)* is defined to be
\[ \text{Asc}(X, \mathcal{U}, T) := \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^2 \log N(\mathcal{U}_S). \]

Again, we will see later that this limit exists. When the open cover \( \mathcal{U} \) of \( X \) is clear we may simplify notation and write \( N(\mathcal{U}_S) = N(S) \). \( S \) is a sample of iteration times, so the average sample complexity is the average over all samples of times \( S \subset n^* \) of the complexity of the behavior of the system.

Suppose \( S = \{s_0, \ldots, s_{|S|-1}\} \) with \( s_0 < s_1 < \cdots < s_{|S|-1} \). If we let \( S' = \{0, s_1 - s_0, \ldots, s_{|S|-1} - s_0\} \) then
\[ N(\mathcal{U}_{S'}) = N(T^0 \mathcal{U}_S) = N(\mathcal{U}_S). \]
Thus, when averaging \( \log N(\mathcal{U}_S) \) over all subsets \( S \subset n^* \) we end up counting the contribution from some subsets many times. If we restrict to subsets \( S \subset n^* \) such that \( 0 \in S \), then we count each configuration only once. This leads to the next definition, where we are concerned only with the configuration that a subset \( S \subset n^* \) exhibits.

**Definition 2.1.3.** Let \( T : X \to X \) be a continuous map on a compact Hausdorff space \( X \), let \( \mathcal{U} \) be an open cover of \( X \), and let \( c_S^2 \) be a system of coefficients. The *average configuration complexity*
of $T$ with respect to the open cover $\mathcal{U}$ is

$$\text{Acc}(X, \mathcal{U}, T) := \lim_{n \to \infty} \frac{1}{n} \sum_{S \subseteq n^*} c^n_S \log N(\mathcal{U}_S). \quad (2.1.3)$$

**Proposition 2.1.4.** Let $(X, T)$ be a topological dynamical system and fix the system of coefficients $c^n_S = 2^{-n}$. Then for any open cover $\mathcal{U}$ of $X$,

$$\text{Acc}(X, \mathcal{U}, T) = \frac{1}{2} \text{Asc}(X, \mathcal{U}, T).$$

**Proof.**

$$\text{Asc}(X, \mathcal{U}, T) = \lim_{n \to \infty} \frac{1}{n} \frac{1}{2^n} \sum_{S \subseteq n^*} \log N(\mathcal{U}_S)$$

$$= \lim_{n \to \infty} \frac{1}{n} \frac{1}{2^n} \sum_{S \subseteq n^*} \log N(\mathcal{U}_S) + \lim_{n \to \infty} \frac{1}{n} \frac{1}{2^n} \sum_{0 \subseteq S} \log N(\mathcal{U}_S)$$

$$= \lim_{n \to \infty} \frac{1}{n} \frac{1}{2^n} \sum_{S \subseteq n^*} \log N(\mathcal{U}_S) + \lim_{n \to \infty} \frac{1}{n} \frac{1}{2^n} \sum_{S \subseteq (n-1)^*} \log N(\mathcal{U}_S)$$

$$= \text{Acc}(X, \mathcal{U}, T) + \lim_{n \to \infty} \frac{1}{2} \left( \frac{n - 1}{n} \right) \left[ \frac{1}{n - 1} \frac{1}{2^{n-1}} \sum_{S \subseteq (n-1)^*} \log N(\mathcal{U}_S) \right]$$

$$= \text{Acc}(X, \mathcal{U}, T) + \frac{1}{2} \text{Asc}(X, \mathcal{U}, T).$$

While Proposition 2.1.4 allows us to compare average sample complexity and average configuration complexity in the case of a uniform system of coefficients ($c^n_S = 2^{-n}$), we do not have simple comparisons for other systems of coefficients.

We will sometimes want to consider the average sample complexity and intricacy as functions of $n$, in analogue with the complexity function $p_u(n)$ of a sequence (see Definition 1.5.21).

**Definition 2.1.5.** Let $(X, T)$ be a topological dynamical system, $\mathcal{U}$ an open cover of $X$, and $c^n_S$ a system of coefficients. The **topological average sample complexity function of $T$ with respect to the**
open cover $\mathcal{U}$ is defined by

$$\text{Asc}(X, \mathcal{U}, T, n) = \frac{1}{n} \sum_{S \subseteq n^*} c_S^n \log N(\mathcal{U}_S).$$

The topological intricacy function of $T$ with respect to the open cover $\mathcal{U}$ is defined by

$$\text{Int}(X, \mathcal{U}, T, n) = \frac{1}{n} \sum_{S \subseteq n^*} c_S^n \log \left( \frac{N(\mathcal{U}_S)N(\mathcal{U}_S^c)}{N(\mathcal{U}^*_n)} \right).$$

When the context is clear we will sometimes write these as $\text{Asc}(n)$ and $\text{Int}(n)$.

### 2.2 Calculations for the golden mean shift of finite type

**Example 2.2.1.** In Chapter 3 we will describe how to determine the average sample complexity and intricacy of a shift of finite type under certain conditions as well as what the measurements tell us about a given SFT. Here we will illustrate the definitions of average sample complexity and intricacy for the golden mean shift of finite type. Recall, the golden mean SFT, denoted here by $X$, is the shift of finite type on the alphabet $\mathcal{A} = \{0, 1\}$ with the forbidden word 11. We would like to calculate the average sample complexity of the golden mean shift relative to the cover $\mathcal{U}_0$ by rank 0 cylinder sets, $\text{Asc}(X, \mathcal{U}_0, \sigma)$. $\mathcal{U}_0$ consists of the sets

$$U_j = \{x \in X : x_0 = j\} \text{ for } j = 0, 1.$$

We begin by picking a subset $S = \{s_0, s_1, \ldots, s_{|S|-1}\}$, of the set $n^*$. Recall from Equation 1.5.4 $\mathcal{L}_S(X)$, the set of words we can see at the places in $S$ for words in $\mathcal{L}_n(X)$. Notice that $N(S)$ is just $|\mathcal{L}_S(X)|$. For instance, if $n = 3$, $n^* = \{0, 1, 2\}$, and $S = \{0, 1\}$, then $N(S) = 3$, as the words at the places in $S$ are \{00, 01, 11\}. However, if we let $S = \{0, 2\}$ then $N(S) = 4$, since now the words at the places in $S$ are \{00, 01, 10, 11\}. The word 11 appears in this case, since there is a sequence $x \in X$ (e.g., 101) with $x_0 = 1$ and $x_2 = 1$.

Table 2.1 shows $N(S)$ and $N(S^c)$ for every subset $S \subset 3^*$. We see that $\text{Asc}(3) = \frac{1}{21} \log(2^3 \cdot 3^2 \cdot 4 \cdot 5) \approx 0.303$ and $\text{Int}(3) = \frac{1}{21} \log \left( \frac{6^4 \cdot 8^2}{56^2} \right) \approx 0.070$. Table 2.2 gives calculations for the golden mean shift for $n = 1, 2, \ldots, 10$. Here, $H(n) = \frac{1}{n} \log N(n^*)$ for each $n$. All numbers are rounded to three
decimal places.

<table>
<thead>
<tr>
<th>$S$</th>
<th>$S^c$</th>
<th>$N(S)$</th>
<th>$N(S^c)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>${0,1,2}$</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>${0}$</td>
<td>${1,2}$</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>${1}$</td>
<td>${0,2}$</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>${2}$</td>
<td>${0,1}$</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>${0,1}$</td>
<td>${2}$</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>${0,2}$</td>
<td>${1}$</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>${1,2}$</td>
<td>${0}$</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>${0,1,2}$</td>
<td>$\emptyset$</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.1: $N(S)$ for the golden mean shift for all $S \subset 3^*$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\text{Asc}(n)$</th>
<th>$\text{Int}(n)$</th>
<th>$H(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.347</td>
<td>0.000</td>
<td>0.693</td>
</tr>
<tr>
<td>2</td>
<td>0.311</td>
<td>0.072</td>
<td>0.549</td>
</tr>
<tr>
<td>3</td>
<td>0.303</td>
<td>0.070</td>
<td>0.536</td>
</tr>
<tr>
<td>4</td>
<td>0.299</td>
<td>0.077</td>
<td>0.520</td>
</tr>
<tr>
<td>5</td>
<td>0.296</td>
<td>0.079</td>
<td>0.513</td>
</tr>
<tr>
<td>6</td>
<td>0.294</td>
<td>0.081</td>
<td>0.507</td>
</tr>
<tr>
<td>7</td>
<td>0.293</td>
<td>0.082</td>
<td>0.504</td>
</tr>
<tr>
<td>8</td>
<td>0.292</td>
<td>0.083</td>
<td>0.501</td>
</tr>
<tr>
<td>9</td>
<td>0.291</td>
<td>0.084</td>
<td>0.499</td>
</tr>
<tr>
<td>10</td>
<td>0.291</td>
<td>0.085</td>
<td>0.497</td>
</tr>
</tbody>
</table>

Table 2.2: Calculations for the golden mean shift

2.3 Preliminary results

In order to show that the limit in Equation 2.1.2 exists for every system of coefficients, we show that $b_n := \sum_{S \subset n^*} c_S^n \log N(\mathbb{Z}_S)$ is subadditive, i.e.

$$b_{n+m} \leq b_n + b_m \quad \text{for all } n \text{ and } m.$$  \hfill (2.3.1)

First we show this for specific systems of coefficients.
Proposition 2.3.1. Fix $0 < p < 1$. Let $c_S^n = \frac{1}{2} \left( p^{|S|} (1 - p)^{|S^c|} + (1 - p)^{|S|} p^{|S^c|} \right)$ and 

$$b_n := \sum_{S \subseteq n^*} c_S^n \log N(\mathcal{U}_S).$$

Then

$$b_{n+m} \leq b_n + b_m \quad \text{for all } n, m \in \mathbb{N}.$$

Proof. Let $S \subseteq (n + m)^* = \{0, 1, \ldots, n + m - 1\}$, $U(S) = S \cap n^* = S \cap \{0, 1, \ldots, n - 1\}$, and $V(S) = S \cap [(n + m)^* \setminus n^*] = S \cap \{n, n+1, \ldots, n + m - 1\}$. We see that

$$N(\mathcal{U}_S) \leq N(\mathcal{U}_{U(S)}) N(\mathcal{U}_{V(S)}),$$

so

$$\sum_{S \subseteq (n+m)^*} c_S^{n+m} \log N(\mathcal{U}_S) \leq \sum_{S \subseteq (n+m)^*} c_S^{n+m} \log N(\mathcal{U}_{U(S)}) + \sum_{S \subseteq (n+m)^*} c_S^{n+m} \log N(\mathcal{U}_{V(S)}).$$

Let $q = 1 - p$, so that

$$c_S^n = \frac{1}{2} (p^{|S|} q^{n-|S|} + p^{n-|S|} q^{|S|}).$$

Abbreviate $U(S) = U$ and $V(S) = V$. Note that each $W \subseteq m^*$ corresponds uniquely to $W + n = V(S) \subseteq \{n, \ldots, n + m - 1\}$ and $N(\mathcal{U}_W) = N(\mathcal{U}_{W+n}) = N(V)$. Thus

$$b_{n+m} = \sum_{S \subseteq (n+m)^*} 1 \left( p^{|S|} q^{n+m-|S|} + q^{|S|} p^{n+m-|S|} \right) \log N(\mathcal{U}_S)$$

$$\leq \sum_{S \subseteq (n+m)^*} \frac{1}{2} \left( p^{|S|} q^{n+m-|S|} + q^{|S|} p^{n+m-|S|} \right) \left[ \log N(\mathcal{U}_U) + \log N(\mathcal{U}_V) \right]$$

$$= \sum_{U \subseteq n^*} \sum_{W \subseteq m^*} \frac{1}{2} \left( p^{|U|+|V|} q^{n-|U|+m-|V|} + p^{n-|U|+m-|V|} q^{|U|+|V|} \right) \log N(\mathcal{U}_U) + \log N(\mathcal{U}_W)$$

$$= \sum_{U \subseteq n^*} \frac{1}{2} \left( p^{|U|} q^{n-|U|} + p^{n-|U|} q^{|U|} \right) \log N(\mathcal{U}_U) + \sum_{W \subseteq m^*} \frac{1}{2} \left( p^{|V|} q^{m-|V|} + p^{m-|V|} q^{|V|} \right) \log N(\mathcal{U}_W)$$

$$= b_n + b_m.$$
Notice that above we use the fact that
\[
\sum_{V \subseteq m^*} p^{|V|} q^{m-|V|} = \sum_{k=0}^{m} \sum_{|V|=k} p^{|V|} q^{m-|V|} = \sum_{k=0}^{m} \binom{m}{k} p^k q^{m-k} = (p + q)^m = 1,
\]
and similarly for the sum over \( U \subseteq n^* \).

Now we show that the sequence \((b_n)\) defined above is subadditive for the class of systems of coefficients that define an intricacy functional as in Theorem 1.4.5.

**Theorem 2.3.2.** Let \( \lambda_c \) be a symmetric probability measure on \([0, 1]\) as defined in Proposition 1.4.5. Let \( c_S^n \) be a system of coefficients, so
\[
c_S^n = \int_{[0,1]} x^{|S|} (1 - x)^{n - |S|} \lambda_c(dx).
\]
(2.3.2)

Define the sequence
\[
b_n := \sum_{S \subseteq n^*} c_S^n \log N(\mathcal{U}_S).
\]
Then \( b_{n+m} \leq b_n + b_m \) for all \( n, m \in \mathbb{N} \).

**Proof.** Let \( S \subset (n + m)^* \) and define \( U = U(S) \) and \( V = V(S) \) as in the proof of Proposition 2.3.1. Then for all \( n, m \in \mathbb{N} \)
\[
b_{n+m} = \sum_{S \subseteq (n+m)^*} \int_{[0,1]} x^{|S|}(1 - x)^{n+m-|S|} \lambda_c(dx) \log N(\mathcal{U}_S)
\]
\[
= \int_{[0,1]} \sum_{S \subseteq (n+m)^*} x^{|S|}(1 - x)^{n+m-|S|} \log N(\mathcal{U}_S) \lambda_c(dx)
\]
\[
\leq \int_{[0,1]} \left( \sum_{U \subseteq n^*} x^{|U|}(1 - x)^{n-|U|} \log N(\mathcal{U}_U) + \sum_{V \subseteq m^*} x^{|V|}(1 - x)^{m-|V|} \log N(\mathcal{U}_V) \right) \lambda_c(dx)
\]
\[
= \sum_{U \subseteq n^*} \int_{[0,1]} x^{|U|}(1 - x)^{n-|U|} \lambda_c(dx) \log N(\mathcal{U}_U) + \sum_{V \subseteq m^*} \int_{[0,1]} x^{|V|}(1 - x)^{m-|V|} \lambda_c(dx) \log N(\mathcal{U}_V)
\]
\[
= b_n + b_m.
\]
Corollary 2.3.3. If $c^n_S$ is a system of coefficients, then the limits in the definitions of $\text{Asc}(X, \mathcal{U}, T)$ and $\text{Int}(X, \mathcal{U}, T)$ (Definitions 2.1.1 and 2.1.2) exist and

$$\text{Asc}(X, \mathcal{U}, T) = \inf \frac{1}{n} \sum_{S \subseteq n^n} c^n_S \log N(\mathcal{U}_S).$$

Proof. This follows from Fekete’s Lemma (1.2.3) and Theorem 2.3.2.

Proposition 2.3.4. For each open cover $\mathcal{U}$, $\text{Asc}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, T)$, and hence $\text{Int}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, T)$.

Proof. For every subset $S \subseteq n^n$, $N(\mathcal{U}_S) \leq N(\mathcal{U}_{n^n})$, so for any finite open cover $\mathcal{U}$ of $X$

$$\text{Asc}(X, \mathcal{U}, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subseteq n^n} c^n_S \log N(\mathcal{U}_S) \leq \lim_{n \to \infty} \frac{1}{n} \sum_{S \subseteq n^n} c^n_S \log N(\mathcal{U}_{n^n})$$

$$= \lim_{n \to \infty} \frac{1}{n} \log N(\mathcal{U}_{n^n}) = h_{\text{top}}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, T).$$

Therefore

$$\text{Int}(X, \mathcal{U}, T) = 2 \text{Asc}(X, \mathcal{U}, T) - h_{\text{top}}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, T).$$

Corollary 2.3.5.

$$\sup_{\mathcal{U}} \text{Int}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, T).$$

2.4 Definitions of intricacy and average sample complexity based on Bowen’s definition of entropy

We will define topological intricacy and topological average sample complexity based on the concepts from Bowen’s definitions of topological entropy, namely $(n^*, \varepsilon)$ separated and $(n^*, \varepsilon)$ spanning sets (see Definitions 1.2.5 and 1.2.7).
Definition 2.4.1. Given a dynamical system \((X, T)\), where \(d\) is a metric on \(X\), and a subset \(S \subset n^*\), recall a set \(E \subset X\) is \((S, \varepsilon)\) spanning if for each \(x \in X\) there is \(y \in E\) with \(d(T^i x, T^i y) \leq \varepsilon\) for all \(i = 0, \ldots, |S| - 1\). Let \(r(S, \varepsilon)\) be the minimum cardinality of an \((S, \varepsilon)\) spanning set of \(X\).

Definition 2.4.2. Fix a systems of coefficients \(c^n_S\). For each \(\varepsilon > 0\) define the \(\varepsilon\)-topological intricacy of \((X, T)\) by
\[
\text{Int}_{\varepsilon}(X, T) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log \left( \frac{r(S, \varepsilon) r(S^*, \varepsilon)}{r(n^*, \varepsilon)} \right),
\]
the \(\varepsilon\)-topological average sample complexity of \((X, T)\) by
\[
\text{Asc}_{\varepsilon}(X, T) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log r(S, \varepsilon),
\]
and the \(\varepsilon\)-topological average configuration complexity of \((X, T)\) by
\[
\text{Acc}_{\varepsilon}(X, T) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log r(S, \varepsilon).
\]

We also give the definitions of topological intricacy, topological average sample complexity, and topological average configuration complexity in terms of \((S, \varepsilon)\) separated sets. Recall a set \(E \subset X\) is \((S, \varepsilon)\) separated if for each pair of distinct points \(x, y \in E\), \(d(T^i x, T^i y) > \varepsilon\) for some \(i = 0, \ldots, |S| - 1\). Let \(s(S, \varepsilon)\) be the maximum cardinality of a set \(E \subset X\) such that \(E\) is \((S, \varepsilon)\) separated. Fix a system of coefficients \(c^n_S\). For each \(\varepsilon > 0\) define the \((\varepsilon\text{-topological intricacy})'\) of \((X, T)\) by
\[
\text{Int}'_{\varepsilon}(X, T) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log \left( \frac{s(S, \varepsilon) s(S^*, \varepsilon)}{s(n^*, \varepsilon)} \right),
\]
the \(\varepsilon\)-topological average sample complexity' of \((X, T)\) by
\[
\text{Asc}'_{\varepsilon}(X, T) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log s(S, \varepsilon),
\]
and the \(\varepsilon\)-topological average configuration complexity' of \((X, T)\) by
\[
\text{Acc}'_{\varepsilon}(X, T) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log s(S, \varepsilon).
\]
Notice that we use different notations for the definitions based on \((S, \varepsilon)\) separating sets and those based on \((S, \varepsilon)\) spanning sets. This is because, in general, for a given \(\varepsilon\), the two definitions may not be equivalent.

2.5 Theorem relating topological average sample complexity and topological intricacy to topological entropy

This section contains one of the main results of the thesis: for \(c^n_S = 2^{-n}\), \(\sup_{\mathcal{U}} \text{Asc}(X, \mathcal{U}, T) = \sup_{\mathcal{U}} \text{Int}(X, \mathcal{U}, T) = h_{\text{top}}(X, T)\). This result shows the connection between our new measurements of complexity. Due to this result we are justified in focusing on computing intricacy and average sample complexity for specific open covers and partitions. In particular, for a shift of finite type we mainly do computations for the cover by rank 0 cylinder sets.

To calculate the topological entropy of a system using the Adler, Konheim, and McAndrew definition with open covers, we find the supremum over all open covers, \(\mathcal{U}\), of \(h_{\text{top}}(X, \mathcal{U}, T)\). Using the Bowen definitions with \((n^*, \varepsilon)\) separated sets and \((n^*, \varepsilon)\) spanning sets, we take the limit as \(\varepsilon\) goes to 0. The following theorem will show that, under the condition that \(c^n_S = 2^{-n}\) for all \(S\), if suprema over all open covers are taken in calculating topological average sample complexity then we get back the usual topological entropy. (See Section 5.2 for a discussion about extending this result to more general weights.)

In Section 2.6 we will see that, using the definitions based on Bowen’s entropy definitions, if we take the limit as \(\varepsilon\) goes to 0 for topological intricacy or average sample complexity, we get back the usual topological entropy.

**Theorem 2.5.1.** Let \((X, T)\) be a topological dynamical system and fix a system of coefficients \(c^n_S = 2^{-n}\). Then

\[
\sup_{\mathcal{U}} \text{Asc}(X, \mathcal{U}, T) = h_{\text{top}}(X, T).
\]

The idea of the proof is to look at the behavior of average sample complexity over subsets \(S \subseteq n^*\) that have certain properties and show that we can find an open cover \(\mathcal{U}\) such that \(N(S)\) is close to \(N(n^*)\) for these subsets. We will show that most subsets have these properties.

In the following lemma for a given \(n, k \in \mathbb{N}\) with \(k < n\) and \(\varepsilon > 0\), we break the interval \(n^*\) into intervals \(K_i\) of length \(k/2\). The reason for this construction is so that for \(S \subseteq n^*\) and \(s \in S\), we know
Given $n, k \in \mathbb{N}$ such that $k$ is even and less than $n$, break $n^*$ into $\lceil 2n/k \rceil$ sets of $k/2$ consecutive integers by defining

$$K_i = \left\{ \frac{i-1}{2} k, \ldots, \frac{i}{2} k - 1 \right\}, \quad i = 1, 2, \ldots, \frac{2n}{k}. \quad (2.5.1)$$

Notice that $|K_{2n/k}| \leq k/2$. For a subset $S \subset n^*$, let

$$G(S) = \text{card}\{K_i : S \cap K_i \neq \emptyset\} \text{ for } i = 1, \ldots, \frac{2n}{k}.$$

Given $0 < \varepsilon < 1$, define $\mathcal{G}$, the set of “good” subsets $S \subset n^*$, by

$$\mathcal{G} = \mathcal{G}(n, k, \varepsilon) = \{S \subset n^* : G(S) > \frac{(2n/k)(1 - \varepsilon)}{2} \}.$$ 

In other words, $S \in \mathcal{G}$ if it intersects least $\lceil (2n/k)(1 - \varepsilon) \rceil$ of the $K_i$. Then given $0 < \varepsilon < 1$, there exists an even $k \in \mathbb{N}$ such that

$$\lim_{n \to \infty} \frac{\text{card}(\mathcal{G}(n, k, \varepsilon))}{2^n} = 1. \quad (2.5.2)$$

Proof. For the proof of this lemma, instead of counting good sets, $\mathcal{G}$, we will count bad sets $\mathcal{B} = \mathcal{B}(n, k, \varepsilon) = \{S \subset n^* : G(S) \leq \frac{(2n/k)(1 - \varepsilon)}{2} \}$. In other words, $S \in \mathcal{B}$ if it intersects at most $\lfloor (1 - \varepsilon)(2n/k) \rfloor$ of the $\lceil 2n/k \rceil$ sets $K_i$. We will show that given $0 < \varepsilon < 1$ there is an even $k \in \mathbb{N}$
such that
\[
\lim_{n \to \infty} \frac{\text{card}(\mathcal{B}(n, k, \varepsilon))}{2^n} = 0.
\]

This is equivalent to showing Equation 2.5.2, since \(\text{card}(\mathcal{G}) = 2^n - \text{card}(\mathcal{B})\). To find a subset \(S \in \mathcal{B}\) we choose \([(2n/k)\varepsilon]\) intervals \(K_i\) for \(S\) to not intersect and then pick a subset (could be empty) from the rest of the \([2n/k(1-\varepsilon)]\) intervals \(K_i\) to intersect \(S\). Thus

\[
\text{card}(\mathcal{B}) = \binom{\lceil 2n/k \rceil}{\lceil 2n/k \rceil} \left(\frac{2k}{2}\right)^{\lfloor 2n/k(1-\varepsilon) \rfloor} \\
= \binom{\lceil 2n/k \rceil}{\lceil 2n/k \rceil} 2^{\lfloor n(1-\varepsilon) \rfloor}
\]

According to Stirling’s approximation,

\[
\lim_{m \to \infty} \left( \frac{m}{m\varepsilon} \right) = \lim_{m \to \infty} \varepsilon^{-m\varepsilon}(1-\varepsilon)^{-m(1-\varepsilon)},
\]

so

\[
\lim_{n \to \infty} \left( \frac{2n/k}{2n\varepsilon/k} \right) = \lim_{n \to \infty} \varepsilon^{-2n\varepsilon/k}(1-\varepsilon)^{-2n(1-\varepsilon)/k}.
\]

(2.5.3)

This implies

\[
\lim_{n \to \infty} \frac{\text{card}(\mathcal{B})}{2^n} = \lim_{n \to \infty} 2^{-n\varepsilon} \varepsilon^{-(2n/k)\varepsilon}(1-\varepsilon)^{-2n/k(1-\varepsilon)} \\
= \lim_{n \to \infty} \left( \frac{1}{2^{\varepsilon(2/k)\varepsilon}(1-\varepsilon)^{(2/k)(1-\varepsilon)}} \right)^n.
\]

We will show \(\lim_{n \to \infty} (\text{card}(\mathcal{B})/2^n) = 0\) by showing that for each \(\varepsilon > 0\), we can find a \(k\) such that

\[2^{\varepsilon(2/k)\varepsilon}(1-\varepsilon)^{(2/k)(1-\varepsilon)} > 1.\]

Denote the binary entropy function by

\[H(x) = -x \log x - (1-x) \log(1-x).\]

To show \(2^{\varepsilon(2/k)\varepsilon}(1-\varepsilon)^{(2/k)(1-\varepsilon)} > 1\), we show

\[
\varepsilon \log 2 + \frac{2}{k} \varepsilon \log \varepsilon + \frac{2}{k}(1-\varepsilon) \log(1-\varepsilon) > 0.
\]

(2.5.4)
This would follow from

\[ k > \frac{2}{\varepsilon \log 2} H(\varepsilon). \quad (2.5.5) \]

By basic calculus \( H(\varepsilon) \leq \log(2) \); thus if \( k > 2/\varepsilon \) Equation 2.5.5 is satisfied and therefore Equation 2.5.4 is satisfied. Thus, given \( \varepsilon > 0 \), if \( k > 2/\varepsilon \) then

\[
\lim_{n \to \infty} \frac{\text{card}(G(n, k, \varepsilon))}{2^n} = 1. \quad (2.5.6)
\]

Next we prove some properties of \( N(\mathcal{U}_S) \) which we need to prove Theorem 2.5.1.

**Lemma 2.5.3.** Let \((X, T)\) be a topological dynamical system and \( \mathcal{U} \) an open cover of \( X \). Given \( n \in \mathbb{N} \) and \( S \subset n^* \) the following properties hold:

1. \( N((\mathcal{U}_k)_S) = N(\mathcal{U}_{S+k^*}) \).
2. Given \( S_1, S_2, \ldots, S_m \subset n^* \), \( \log N(\bigcup_i S_i) \leq \sum_i \log N(S_i) \).

**Proof.**

1. This is true because

\[
(\mathcal{U}_k)_S = \bigvee_{i \in S} T^{-i} \mathcal{U}_k = \bigvee_{i \in S+k^*} T^{-i} \mathcal{U} = \mathcal{U}_{S+k^*}.
\]

2. We show this for two sets \( S_1 \) and \( S_2 \) and use induction. We use the fact that \( N(\mathcal{U} \cup \mathcal{V}) \leq N(\mathcal{U})N(\mathcal{V}) \). Thus,

\[
N(S_1 \cup S_2) = N(\mathcal{U}_{S_1} \cup \mathcal{U}_{S_2}) \leq N(S_1)N(S_2).
\]

**Proof of Theorem 2.5.1.** Recall that \( h_{\text{top}}(X, \mathcal{U}, T) = \lim_{n \to \infty} \log N(\mathcal{U}_n^*)/n \) and \( h_{\text{top}}(X, T) = \sup_{\mathcal{U}} h(X, \mathcal{U}, T) \). We prove the statement by showing for each open cover \( \mathcal{U} \) of \( X \),

\[
\lim_{k \to \infty} \text{Asc}(X, \mathcal{U}_k^*, T) = h_{\text{top}}(X, \mathcal{U}, T) \quad (2.5.7)
\]
Recall that by Proposition 2.3.4 for every cover $\mathcal{U}$ of $X$, $\text{Asc}(X, \mathcal{U}, T) \leq h_{\text{top}}(X, \mathcal{U}, T)$. We would like to show that

$$\lim_{k \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{S \subseteq n^*} \log N(S + k^*) = \lim_{n \to \infty} \frac{1}{n} \log N(n^*).$$

Let $0 < \varepsilon < 1$ be given. By Fekete’s Lemma 1.2.3, we know $h(X, \mathcal{U}, T) = \inf_k (\log(N(k^*))/k)$. Thus, there is a $k_0$ such that for every $k > k_0$,

$$0 \leq \frac{\log N(k^*)}{k} - h(X, \mathcal{U}, T) < \varepsilon. \quad (2.5.8)$$

Let $k > \max\{2k_0, 2/\varepsilon\}$ and assume $k$ is even and let $n > k$. Form the set of good sets $\mathcal{G}(n, k, \varepsilon)$ as in the statement of Lemma 2.5.2 and let the sets $K_i$ be as in Equation 2.5.1. The main idea behind the construction of the intervals $K_i$ is that for $S \subseteq n^*$, and $s \in S$, if $s \in K_i$ then $K_{i+1} \subset S + k^*$. Suppose $S \subseteq \mathcal{G}$. Then $S$ intersects at least $(2n/k)(1 - \varepsilon)$ of the sets $K_i$ so we know $\text{card}(S + k^*) \geq (k/2)(2n/k)(1 - \varepsilon) = n(1 - \varepsilon)$. $S + k^*$ is the disjoint union of intervals in $n^*$ which we denote by $\tilde{E}_j$ satisfying

(i) $\text{card}(\tilde{E}_j) \geq k \geq 2k_0$ and

(ii) $\sum_j \text{card}(\tilde{E}_j) = \text{card}(S + k^*) \geq n(1 - \varepsilon)$.

Let $G_j$ be the gap of integers between the strings $\tilde{E}_j$ and $\tilde{E}_{j+1}$ (with $G_1 = \{0, 1, \ldots, s_1\}$ if $0 \neq E_1$). There are at most $2n\varepsilon/k + 1$ of these gaps $G_j$, since each (except possibly $G_1$) must contain a point not in $S + k^*$ and hence in one of the intervals missed by $S$ (see Figure 2.2).

If necessary remove an interval of no more than $k_0$ integers from the left end of each $\tilde{E}_j$ (and therefore add them to the right end of $G_{j-1}$) to ensure $\text{card}(G_j) \geq k_0$ for all $j$. Call the removed interval $R_j$ and let $E_j = \tilde{E}_j \setminus R_j$. Then $\text{card}(E_j) = \text{card}(\tilde{E}_j) - \text{card}(R_j) \geq 2k_0 - k_0 = k_0$ and
\[
\frac{\log N(E_j)}{\text{card}(E_j)} - h_{\text{top}}(X, \mathcal{U}, T) < \varepsilon \quad \text{and} \quad \frac{\log N(G_j)}{\text{card}(G_j)} - h_{\text{top}}(X, \mathcal{U}, T) < \varepsilon.
\]

Using \( \sum_j \text{card}(E_j) + \sum_j \text{card}(G_j) = n \) and the construction of \( E_j \) and \( G_j \), we have

\[
\sum_j \text{card}(E_j) \geq \sum_j \text{card}(\tilde{E}_j) - \sum_j \text{card}(R_j)
\]
\[
\geq n(1 - \varepsilon) - \left( \frac{2n\varepsilon}{k} + 1 \right) k_0
\]

and

\[
\sum_j \text{card}(G_j) \leq n - \left( n(1 - \varepsilon) - \left( \frac{2n\varepsilon}{k} + 1 \right) k_0 \right)
\]
\[
= n\varepsilon \left( 1 + k_0 \left( 1 + \frac{2}{k} \right) \right).
\]

Since \( \bigcup_j (E_j \cup G_j) = n^* \), using (2) in Lemma 2.5.3 we have

\[
\log N(n^*) = \log N(\bigcup_j (E_j \cup G_j)) \leq \log N(\bigcup_j E_j) + \log(\bigcup_j G_j),
\]

which implies

\[
\log N(\bigcup_j E_j) \geq \log N(n^*) - \log(\bigcup_j G_j).
\]
Hence,

\[
\log N(S + k^*) \geq \log N(\bigcup_j E_j)
\]

\[
\geq \log N(n^*) - \log N(\bigcup_j G_j)
\]

\[
\geq n \cdot \htop(X, \mathcal{U}, T) - \sum_j \log N(G_j)
\]

\[
\geq n \cdot \htop(X, \mathcal{U}, T) - (\htop(X, \mathcal{U}, T) + \varepsilon) \sum_j \text{card}(G_j)
\]

\[
= \htop(X, \mathcal{U}, T) \sum_j \text{card}(E_j) - \varepsilon \sum_j \text{card}(G_j).
\]

Therefore, for all \(S \in \mathcal{G}\), even \(k > \max\{2k_0, 2/\varepsilon\}\), and \(n > k\),

\[
\frac{\log N(S + k^*)}{n} \geq \frac{\htop(X, \mathcal{U}, T) \sum_j \text{card}(E_j) - \varepsilon \sum_j \text{card}(G_j)}{n}
\]

\[
\geq \htop(X, \mathcal{U}, T) \left( n(1 - \varepsilon) - \left( \frac{2\varepsilon}{k} + 1 \right) k_0 \right) - n\varepsilon^2 \left( 1 + k_0 (1 + \frac{2}{k}) \right)
\]

\[
\geq \htop(X, \mathcal{U}, T) \left( 1 - \varepsilon \left( 1 - \frac{2}{k} k_0 \right) \right) - \varepsilon^2 \left( 1 + \frac{2}{k} k_0 \right)
\]

\[
\geq \htop(X, \mathcal{U}, T)(1 - \varepsilon) - 2\varepsilon.
\]

So if \(S \in \mathcal{G}\) and \(\delta = (\htop(X, \mathcal{U}, T) + 2)\varepsilon\) we can find \(k\) such that for all large enough \(n\)

\[
\frac{1}{n} \log N(S + k^*) \geq h_{\text{top}}(X, \mathcal{U}, T) - \delta.
\]
We then conclude for any $0 < \varepsilon < 1$ we can find $k$ such that for all large enough $n$,

$$
\frac{1}{n} \frac{1}{2^n} \sum_{S \subseteq n^*} \log N(S + k^*) \geq \frac{|G|}{2^n} \frac{1}{|G|} \sum_{S \subseteq G} \log N(S + k^*)
$$

$$
\geq (1 - \varepsilon) \frac{1}{n} \sum_{S \subseteq G} \frac{1}{n} \log N(S + k^*)
$$

$$
\geq (1 - \varepsilon) \frac{1}{|G|} \sum_{S \subseteq G} (h_{top}(X, \mathcal{U}, T) - \varepsilon)
$$

$$
\geq (1 - \varepsilon) \frac{|G|}{|G|} (h_{top}(X, \mathcal{U}, T) - \varepsilon)
$$

$$
= h_{top}(X, \mathcal{U}, T) - \varepsilon h_{top}(X, \mathcal{U}, T) - \varepsilon^2.
$$

Combining the above with Proposition 2.3.4, we have that for any $0 < \varepsilon < 1$ there exists $k$ such that for all large enough $n$,

$$
h_{top}(X, \mathcal{U}, T) - \varepsilon \leq \frac{1}{n} \frac{1}{2^n} \sum_{S \subseteq n^*} \log N(S + k^*) \leq h_{top}(X, \mathcal{U}, T). \quad (2.5.9)
$$

Letting $n \to \infty$ and then $k \to \infty$ in Equation 2.5.9 gives

$$
\lim_{k \to \infty} \operatorname{Asc}(X, \mathcal{U}_k, T) = \lim_{k \to \infty} \lim_{n \to \infty} \frac{1}{n} \frac{1}{2^n} \sum_{S \subseteq n^*} \log N(S + k^*) \geq h_{top}(X, \mathcal{U}, T) - \varepsilon, \quad (2.5.10)
$$

and hence by Proposition 2.3.4, $\lim_{k \to \infty} \operatorname{Asc}(X, \mathcal{U}_k, T) = h_{top}(X, \mathcal{U}, T)$. To complete the proof we take the supremum over all covers $\mathcal{U}$ of $X$ on both sides of Equation 2.5.10.

In the next section we will show that

$$
\lim_{\varepsilon \to 0^+} \operatorname{Asc}_\varepsilon(X, T) = \lim_{\varepsilon \to 0^+} \operatorname{Asc'}_\varepsilon(X, T) = h_{top}(X, T). \quad (2.5.11)
$$

### 2.6 Average sample pressure

In analogy with the extension of the definition of topological entropy to topological pressure, we will develop topological average sample pressure by generalizing topological average sample
complexity. We use notation based on the notation for topological pressure in Chapter 9 of [Wal]. Recall that a set $E \subset X$ is $(S, \varepsilon)$ separated if for each pair of distinct points $x, y \in E$, $d(T^{s_i}x, T^{s_i}y) > \varepsilon$ for some $i = 0, \ldots, |S| - 1$, and $s(S, \varepsilon)$ is the maximum cardinality of a set $E \subset X$ such that $E$ is $(S, \varepsilon)$ separated. Also, a set $E \subset X$ is $(S, \varepsilon)$ spanning if for each $x \in X$ there is $y \in E$ with $d(T^{s_i}x, T^{s_i}y) \leq \varepsilon$ for all $i = 0, \ldots, |S| - 1$, and $r(S, \varepsilon)$ is the minimum cardinality of an $(S, \varepsilon)$ spanning set of $X$. We refer to Section 1.2.4 for definitions of the notation used in this section.

**Definition 2.6.1.** Let $T : X \to X$ be a continuous transformation on a compact metric space. Let $f \in C(X, \mathbb{R})$ and $S \subset n^*$. Define

$$Q_S(T, f, \varepsilon) = \inf \left\{ \sum_{x \in F} \exp \left( \sum_{i \in S} f(T^i x) \right) : F \text{ is an } (S, \varepsilon) \text{ spanning set for } X \right\}. \tag{2.6.1}$$

Then, for a fixed system of coefficients $c^n_S$, the *average sample pressure of $T$ given $f$ and $\varepsilon$, $\text{Asp}_\varepsilon(T, f)$*, is

$$\text{Asp}_\varepsilon(T, f) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log Q_S(T, f, \varepsilon). \tag{2.6.2}$$

**Definition 2.6.2.** We also define average sample pressure in terms of $(S, \varepsilon)$ separated sets. Let

$$P_S(T, f, \varepsilon) = \sup \left\{ \sum_{x \in E} \exp \left( \sum_{i \in S} f(T^i x) \right) : E \text{ is an } (S, \varepsilon) \text{ separated set for } X \right\}. \tag{2.6.3}$$

Then, for a fixed system of coefficients $c^n_S$, the *average sample pressure of $T$ given $f$ and $\varepsilon$, $\text{Asp}'_\varepsilon(T, f)$*, is

$$\text{Asp}'_\varepsilon(T, f) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log P_S(T, f, \varepsilon). \tag{2.6.4}$$

We use $\text{Asp}$ to denote the definition which uses $(S, \varepsilon)$ spanning sets and $\text{Asp}'$ to denote the definition which uses $(S, \varepsilon)$ separated sets since, in general, for a given $\varepsilon$ these may not be equal.

**Proposition 2.6.3.** Let $T : X \to X$ be a continuous transformation on a compact metric space. Let $f \in C(X, \mathbb{R})$ and $S \subset n^*$. Given $\varepsilon > 0$,

$$Q_S(T, f, \varepsilon) \leq P_S(T, f, \varepsilon). \tag{2.6.5}$$

*Proof.* We first notice that in Equation 2.6.3 we can take the supremum over $(S, \varepsilon)$ separated sets,
that are maximal, i.e. if we were to add another point to $E$ then it would no longer be an $(S, \varepsilon)$ separated set. This is because \( \exp \left( \sum_{i \in S} f(T^i x) \right) > 0 \) for all $x \in X$, so the more points $x$ we sum this value over, the larger it gets.

Now, if $E$ is a maximal $(S, \varepsilon)$ separated set then it must be an $(S, \varepsilon)$ spanning set for $X$. This is because given $x \in X \setminus E$ and $y \in E$, if $d(T^{s_i}x, T^{s_i}y) > \varepsilon$, for all $i = 0, \ldots, |S| - 1$, then we could add $x$ to $E$ and $E$ would still being $(S, \varepsilon)$ separated, contradicting $E$ be a maximal $(S, \varepsilon)$ separated set.

**Corollary 2.6.4.**

\[
\text{Asp}_\varepsilon(T, f) \leq \text{Asp}_\varepsilon'(T, f) \leq P(T, f).
\] (2.6.6)

**Proof.** The first inequality follows directly from Proposition 2.6.3. The second inequality is true because

\[
P_S(T, f, \varepsilon) \leq P_n(T, f, \varepsilon)
\]
for all $S \subset n^*$.

Notice that if $f$ is equal to 0 then we have

\[
Q_S(T, f, \varepsilon) = \inf \{ \text{card}(F) : F \text{ is an } (S, \varepsilon) \text{ spanning set for } X \} = r(S, \varepsilon)
\]
and

\[
P_S(T, 0, \varepsilon) = \sup \{ \text{card}(E) : E \text{ is an } (S, \varepsilon) \text{ separated set for } X \} = s(S, \varepsilon)
\]
and thus

\[
\text{Asp}_\varepsilon(T, 0) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^n \log r(S, \varepsilon) = \text{Asc}_\varepsilon(T) \quad \text{and }
\]

\[
\text{Asp}_\varepsilon'(T, 0) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^n \log s(S, \varepsilon) = \text{Asc}_\varepsilon'(T).
\] (2.6.7) (2.6.8)

Next we give the definition of average sample pressure in terms of open covers.

**Definition 2.6.5.** Let $T : X \to X$ be a continuous transformation on a compact metric space. Let
\( f \in C(X, \mathbb{R}) \) and \( S \subset n^* \). If \( \mathcal{U} \) is an open cover of \( X \), then we define

\[
p_S(T, f, \mathcal{U}) = \inf \left\{ \sum_{V \in \mathcal{V}} \sup_{x \in V} \left( \sum_{i \in S} f(T^i x) \right) : \mathcal{V} \text{ is a finite subcover of } \mathcal{U}_S \right\}
\]  
(2.6.9)

and

\[
q_S(T, f, \mathcal{U}) = \inf \left\{ \sum_{V \in \mathcal{V}} \inf_{x \in V} \left( \sum_{i \in S} f(T^i x) \right) : \mathcal{V} \text{ is a finite subcover of } \mathcal{U}_S \right\}.
\]  
(2.6.10)

Define the the **average sample pressure** of \((X, T)\) and the open cover \(\mathcal{U}\) of \(X\), given \(f\) and a system of coefficients \(c_S^n\) by

\[
\text{Asp}(T, f, \mathcal{U}) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^n \log q_S(T, f, \mathcal{U})
\]  
(2.6.11)

Similarly, we define the average sample pressure by

\[
\text{Asp}'(T, f, \mathcal{U}) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c_S^n \log p_S(T, f, \mathcal{U}).
\]  
(2.6.12)

**Lemma 2.6.6.** Let \( T : X \to X \) be a continuous transformation on a compact metric space. Let \( f \in C(X, \mathbb{R}) \) and fix a system of coefficients \( c_S^n \). If \( \mathcal{U} \) is an open cover of \( X \) then

\[
\text{Asp}(T, f, \mathcal{U}) \leq \text{Asp}'(T, f, \mathcal{U}) \leq \limsup_{n \to \infty} \frac{1}{n} q_n(T, f, \mathcal{U}) \leq \lim_{n \to \infty} \frac{1}{n} p_n(T, f, \mathcal{U}).
\]  
(2.6.13)

**Proof.** Since \( q_S(T, f, \mathcal{U}) \leq p_S(T, f, \mathcal{U}) \), we have \( \text{Asp}(T, f, \mathcal{U}) \leq \text{Asp}'(T, f, \mathcal{U}) \). Since \( q_S(T, f, \mathcal{U}) \leq q_n(T, f, \mathcal{U}) \) for every \( S \subset n^* \), we get

\[
\text{Asp}'(T, f, \mathcal{U}) \leq \limsup_{n \to \infty} \frac{1}{n} q_n(T, f, \mathcal{U}).
\]

\( \Box \)

**Lemma 2.6.7.** Let \( T : X \to X \) be a continuous transformation on a compact metric space,
If \( f \in C(X, \mathbb{R}) \), and \( S_1, S_2 \subset n^* \) disjoint. Let \( S = S_1 \cup S_2 \). If \( \mathcal{U} \) is an open cover of \( X \), then

\[
\log p_S(T, f, \mathcal{U}) \leq \log p_{S_1}(T, f, \mathcal{U}) + \log p_{S_2}(T, f, \mathcal{U}).
\]  

(2.6.14)

**Proof.** First we show

\[
p_S(T, f, \mathcal{U}) \leq p_{S_1}(T, f, \mathcal{U}) \cdot p_{S_2}(T, f, \mathcal{U}).
\]

For each finite open subcover \( \mathcal{V}_1 \) of \( \mathcal{U} \) \( S_1 \) and \( \mathcal{V}_2 \) of \( \mathcal{U} \) \( S_2 \), \( \mathcal{V}_1 \cup \mathcal{V}_2 \) is a finite subcover of \( \mathcal{U} \) and

\[
\sum_{A \in \mathcal{V}_1 \cup \mathcal{V}_2} \sup_{x \in A} \left( \sum_{i \in S} f(T^i x) \right) \leq \sum_{B \in \mathcal{V}_1} \sup_{x \in B} \left( \sum_{i \in S_1} f(T^i x) \right) \cdot \sum_{C \in \mathcal{V}_2} \sup_{x \in C} \left( \sum_{i \in S_2} f(T^i x) \right).
\]

This shows \( p_S(T, f, \mathcal{U}) \leq p_{S_1}(T, f, \mathcal{U}) \cdot p_{S_2}(T, f, \mathcal{U}) \) which implies \( \log p_S(T, f, \mathcal{U}) \leq \log p_{S_1}(T, f, \mathcal{U}) + \log p_{S_2}(T, f, \mathcal{U}) \).

\[\square\]

**Proposition 2.6.8.** Let \( \lambda_c \) be a symmetric probability measure on \([0, 1]\). Then the limit in Equation 2.6.11 exists for \( c^n_S = \int_{[0,1]} x^{|S|}(1-x)^{n-|S|} \lambda_c(dx) \) and

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log p_S(T, f, \mathcal{U}) = \inf \frac{1}{n} \sum_{S \subset n^*} c^n_S \log p_S(T, f, \mathcal{U}).
\]

**Proof.** We will show that the sequence

\[
b_n = \sum_{S \subset n^*} c^n_S \log p_S(T, f, \mathcal{U})
\]  

(2.6.15)

is subadditive, and then apply Lemma 1.2.3 to get the result. Thus, we must show \( b_{n+m} \leq b_n + b_m \).

Given a subset \( S \subset (n + m)^* \), we define the sets \( S_1 = S \cap n^* \) and \( S_2 = S \cap ((n + m)^* \setminus n^*) \). By
Lemma 2.6.7, $\log p_S(T, f, \mathcal{U}) \leq \log p_{S_1}(T, f, \mathcal{U}) + \log p_{S_2}(T, f, \mathcal{U})$. For all $n, m \in \mathbb{N}$,

$$b_{n+m} = \sum_{S \subseteq (n+m)^*} \int_{[0,1]} x^{|S|}(1-x)^{n+m-|S|} \lambda_c(dx) \log p_S$$

$$= \int_{[0,1]} \sum_{S \subseteq (n+m)^*} x^{|S|}(1-x)^{n+m-|S|} \lambda_c(dx) \log p_S$$

$$\leq \int_{[0,1]} \left( \sum_{U \subseteq n^*} x^{|U|}(1-x)^{n-|U|} \log p_{S_1} + \sum_{V \subseteq m^*} x^{|V|}(1-x)^{m-|V|} \log p_{S_2} \right) \lambda_c(dx)$$

$$= \sum_{U \subseteq n^*} \int_{[0,1]} x^{|U|}(1-x)^{n-|U|} \lambda_c(dx) \log p_{S_1} + \sum_{V \subseteq m^*} \int_{[0,1]} x^{|V|}(1-x)^{m-|V|} \lambda_c(dx) \log p_{S_2}$$

$$= b_n + b_m.$$ 

\[ \square \]

**Theorem 2.6.9.** Let $T : X \to X$ be a continuous transformation on a compact metric space. Let $f \in C(X, \mathbb{R})$ and $S \subset n^*$.

(i) If $\mathcal{U}$ is an open cover of $X$ with Lebesgue number $\delta$, then $q_S(T, f, \mathcal{U}) \leq Q_S(T, f, \delta/2) \leq P_S(T, f, \delta/2)$.

(ii) If $\varepsilon > 0$ and $\mathcal{U}$ is an open cover of $X$ such that $\text{diam}(\mathcal{U}) \leq \varepsilon$, then $Q_S(T, f, \varepsilon) \leq P_S(T, f, \varepsilon) \leq p_S(T, f, \mathcal{U})$.

**Proof.** We know that $Q_S(T, f, \varepsilon) \leq P_S(T, f, \varepsilon)$ for all $\varepsilon > 0$.

(i) If $E$ is an $(S, \delta/2)$ spanning set for $X$ then for each $x \in X$, there is $y \in E$ with $d(T^s x, T^{s_i} y) \leq \delta/2$, for all $i = 0, \ldots, |S| - 1$. Thus, if we denote the closed balls

$$\overline{B}(T^s x, \delta/2) = \{ y \in X : d(T^s x, y) \leq \delta/2 \},$$

then

$$X = \bigcup_{x \in E} \bigcap_{i=0}^{\lfloor |S| - 1 \rfloor} T^{-s_i} \overline{B}(T^s x, \delta/2).$$

Since $\text{diam}(\overline{B}(T^s x, \delta/2)) < \delta$ for each $x \in E$, each such closed ball is contained in an element of $\mathcal{U}$. Therefore, $q_S(T, f, \mathcal{U}) \leq \sum_{x \in E} \exp \left( \sum_{i \in S} f(T^s x) \right)$ and thus $q_S(T, f, \mathcal{U}) \leq Q_S(T, f, \delta/2)$.

(ii) If $E$ is an $(S, \varepsilon)$ separated set, then for each pair of distinct points $x, y \in E$, $d(T^s x, T^s y) > \varepsilon$, 61
for some $i = 0, \ldots, |S| - 1$. Since $\text{diam}(\mathcal{U}) < \varepsilon$, no member of $\mathcal{U}_S$ may contain two elements of $E$. Thus, $\sum_{x \in E} \exp \left( \sum_{i \in S} f(T^i x) \right) \leq p_S(T, f, \varepsilon)$. Therefore, $P_S(T, f, \varepsilon) \leq p_S(T, f, \mathcal{U})$. 

The next theorem gives a relationship between average sample pressure and topological pressure when we fix $c^n_S = 2^{-n}$, similar to Theorem 2.5.1, which gives a relationship between average sample complexity and topological entropy.

**Theorem 2.6.10.** Let $T : X \rightarrow X$ be a continuous transformation on the compact metric space $X$. Let $f \in C(X, \mathbb{R})$ and $S \subset n^*$. For the fixed system of coefficients $c^n_S = 2^{-n}$ for all $n \in \mathbb{N}$ and $S \subset n^*$,

$$
\sup_{\mathcal{U}} \text{Asp}'(T, f, \mathcal{U}) = P(T, f). \quad (2.6.16)
$$

**Proof.** We will prove this theorem by showing that for any open cover $\mathcal{U}$ of $X$,

$$
\lim_{k \rightarrow \infty} \text{Asp}'(T, f, \mathcal{U}_k) = \lim_{n \rightarrow \infty} p_n(T, f, \mathcal{U}). \quad (2.6.17)
$$

Most of the calculations in this proof can be found in the proof of Theorem 2.5.1 either directly or by replacing $N(\mathcal{U}_S)$ in that proof by $p_S(T, f, \mathcal{U})$. For that reason, many of the details have been left out. Recall from Definition 1.2.18 that

$$
p_n(T, f, \mathcal{U}) = \inf \left\{ \sum_{V \in \mathcal{V}} \sup_{x \in V} \exp \left( \sum_{i=0}^{n-1} f(T^i x) \right) : V \text{ is a finite subcover of } \bigvee_{i=0}^{n-1} T^{-i} \mathcal{U} \right\}.
$$

Denote $P(T, f, \mathcal{U}) = \lim_{n \rightarrow \infty} (1/n) \log p_n(T, f, \mathcal{U})$. Then, given $\varepsilon > 0$, choose $k_0 \in \mathbb{N}$ large enough that for every $k > k_0$

$$
\frac{1}{k} \log p_k(T, f, \mathcal{U}) - P(T, f, \mathcal{U}) < \varepsilon. \quad (2.6.18)
$$

Let $k > \max\{2k_0, 2/\varepsilon\}$ and assume $k$ is even. Choose $n > k$ such that $k/2$ divides $n$ and form the set of good sets $G(n, k, \varepsilon)$ as in the statement of Lemma 2.5.2. Form the sets $E_j$ and $G_j$ as in the proof of Theorem 2.5.1. By Lemma 2.6.7,

$$
\log p_n(T, f, \mathcal{U}) \leq \log p_{\cup_j E_j}(T, f, \mathcal{U}) + \log p_{\cup_j G_j}(T, f, \mathcal{U}). \quad (2.6.19)
$$
By using calculations from the proof of Theorem 2.5.1 and Equation 2.6.19,

$$\log p_S(T, f, \mathcal{U}_k^*) \geq P(T, f, \mathcal{U}) \sum_j \text{card}(E_j) - \varepsilon \sum_j \text{card}(G_j).$$  \hspace{1cm} (2.6.20)

If $S \in \mathcal{G}(n, k, \varepsilon)$,

$$\lim_{k \to \infty} \lim_{n \to \infty} \frac{1}{n} \log p_S(T, f, \mathcal{U}_k^*) = P(T, f, \mathcal{U}),$$  \hspace{1cm} (2.6.21)

and thus

$$\lim_{k \to \infty} \text{Asp}^l(T, f, \mathcal{U}_k^*) = P(T, f, \mathcal{U}).$$  \hspace{1cm} (2.6.22)

Taking the supremum over all open covers $\mathcal{U}$ on both sides of Equation 2.6.22 gives the result. \hfill \square

**Corollary 2.6.11.**

$$\sup_{\mathcal{U}} \text{Asp}(T, f, \mathcal{U}) = P(T, f).$$

**Theorem 2.6.12.** Let $T : X \to X$ be a continuous transformation on a compact metric space. Let $f \in C(X, \mathbb{R})$ and $S \subset \eta^*$. For the system of coefficients $c^n_S = 2^{-n}$ for all $n \in \mathbb{N}$ and $S \subset \eta^*$,

$$\lim_{\varepsilon \to 0^+} \text{Asp}^l(T, f) = \lim_{\varepsilon \to 0^+} \text{Asp}_\varepsilon(T, f) = P(T, f).$$  \hspace{1cm} (2.6.23)

**Proof.** Given $\varepsilon > 0$, let $\mathcal{U}_\varepsilon$ be the open cover of $X$ by balls of radius $2\varepsilon$, and let $\mathcal{V}_\varepsilon$ be the open cover of $X$ by balls of radius $\varepsilon/2$. For each $S \subset \eta^*$ we will show

$$q_S(T, f, \mathcal{U}_\varepsilon) \leq Q_S(T, f, \varepsilon) \leq P_S(T, f, \varepsilon) \leq p_S(T, f, \mathcal{V}_\varepsilon).$$  \hspace{1cm} (2.6.24)

A finite subcover of $(\mathcal{U}_\varepsilon)_S$ has Lebesgue number $2\varepsilon$, so by Theorem 2.6.9

$$q_S(T, f, \mathcal{U}_\varepsilon) \leq Q_S(T, f, \varepsilon).$$

If $\mathcal{V}$ is a finite subcover of $(\mathcal{V}_\varepsilon)_S$, then $\text{diam}(\mathcal{V}) = \varepsilon/2$, so by Theorem 2.6.9

$$P_S(T, f, \varepsilon) \leq p_S(T, f, \mathcal{V}_\varepsilon).$$

Combining Equation 2.6.24 with Theorem 2.6.10 and Corollary 2.6.11 gives the result. \hfill \square
Corollary 2.6.13. If $X$ is a compact metric space and $T : X \to X$ is a continuous map,

$$\lim_{\varepsilon \to 0^+} \text{Asc}_\varepsilon(X, T) = \lim_{\varepsilon \to 0^+} \text{Asc}_\varepsilon(X, T) = h_{\text{top}}(X, T). \quad (2.6.25)$$

Proof. Let $f \equiv 0$ in Theorem 2.6.12. \hfill \square

In practice, we will fix an open cover $\mathcal{U}$ or an $\varepsilon > 0$ when doing calculations to find values of $\text{Asp}(T, f, \mathcal{U})$, $\text{Asp}'(T, f, \mathcal{U})$, $\text{Asp}_\varepsilon(T, f)$, and $\text{Asp}_\varepsilon'(T, f)$. As with average sample complexity and intricacy, we define the average sample pressure function using $(S, \varepsilon)$ spanning sets by

$$\text{Asp}_\varepsilon(T, f, n) = \frac{1}{n} \sum_{S \in \mathcal{A}^*} c^n_S \log Q_S(T, f, \varepsilon), \quad (2.6.26)$$

or with $(S, \varepsilon)$ separated sets by

$$\text{Asp}_\varepsilon'(T, f, n) = \frac{1}{n} \sum_{S \in \mathcal{A}^*} c^n_S \log P_S(T, f, \varepsilon). \quad (2.6.27)$$

We also define these functions using open covers:

$$\text{Asp}(T, f, \mathcal{U}, n) = \frac{1}{n} \sum_{S \in \mathcal{A}^*} c^n_S \log q_S(T, f, \mathcal{U}). \quad (2.6.28)$$

and

$$\text{Asp}'(T, f, \mathcal{U}, n) = \frac{1}{n} \sum_{S \in \mathcal{A}^*} c^n_S \log p_S(T, f, \mathcal{U}), \quad (2.6.29)$$
CHAPTER 3
Complexity calculations for subshifts

3.1 Shifts of finite type

In this section we will calculate the intricacy, average sample complexity, and average sample pressure of some shifts of finite type. We first consider the open cover \( U_0 \) by rank 0 cylinder sets of a shift of finite type \( X \subset \Sigma_r \).

Unless otherwise noted, we will use the uniform system of coefficients, \( c^n_S = 2^{-n} \) and open covers by rank 0 cylinder sets. As we saw in Example 2.2.1, for a subset \( S \subset n^* \), \( N(S) \) counts the number of words seen at the places in \( S \) over all words \( w \in \mathcal{L}_n(X) \).

The next proposition gives us an easy way to calculate \( N(S) \) for a shift of finite type with square positive adjacency matrix \( M \). The reason this property simplifies the computation of \( N(S) \) is because it guarantees for any \( a, b \in A \), if \( |i| > 2 \) there is a sequence \( x \in X \) such that \( x_0 = a \) and \( x_i = b \). This allows us to break \( S \) into disjoint intervals of consecutive integers and compute \( N(S) \) by taking the product of the values of \( N \) on each disjoint interval. Since the value of \( N \) on an interval of \( k \) consecutive integers is just \( |\mathcal{L}_k(X)| \), these values can be found by summing the entries of \( M^{k-1} \).

**Proposition 3.1.1.** Let \( X \) be a shift of finite type over the alphabet \( A \) with adjacency matrix \( M \) such that \( M^2 > 0 \). Given \( S \subset n^* \) denote the disjoint subsets of consecutive integers that compose \( S \) by \( I_1, \ldots, I_k \) with \( |I_j| = t_j \) for \( t_j \in \mathbb{N} \). Then

\[
N(S) = |\mathcal{L}_{t_1}^1(X)||\mathcal{L}_{t_2}^2(X)| \cdots |\mathcal{L}_{t_k}^k(X)| = N(t_1^*)N(t_2^*) \cdots N(t_k^*). \tag{3.1.1}
\]

In particular

\[
\sum_{S \subset n^*} \log N(S) = \sum_{S \subset (n-\ell-1)^*} \log [N(S)N(\ell^*)] \tag{3.1.2}
\]
and

$$\sum_{S \subseteq n^* \atop n-1 \in S} \log N(S) = \sum_{S \subseteq (n-1)^*} \log N(S). \quad (3.1.3)$$

**Proof.** We will prove Equation 3.1.1 by using induction. Since $M^2 > 0$, given any two elements $a, b \in A$ and $m \geq 3$ there is at least one word in $L_{m^*}(X)$ of the form $a \ldots b$. Given two disjoint subsets of $S$, $I_1$ and $I_2$ with $|I_j| = t_j$, $N(I_1 \cup I_2)$ is the number of words seen at the places in $I_1$ and $I_2$ for all legal words in $X$. There are $N(t_j)$ words that can be seen at the places in $I_j$ for $j = 1, 2$. Let $w_1 w_2 \ldots w_{t_1} \in L_{t_1^*}(X)$ and $\tilde{w}_1 \tilde{w}_2 \ldots \tilde{w}_{t_2} \in L_{t_2^*}(X)$ be words that can be seen at the places in $I_1$ and $I_2$ respectively. Since $M^2 > 0$, there is a word of the form $w_1 \ldots \tilde{w}_1 \in L_{m^*}(X)$ for some $m \geq 3$. Thus $N(I_1 \cup I_2) = |L_{t_1^*}(X)||L_{t_2^*}(X)| = N(t_1^*)N(t_2^*)$. The proof of Equation 3.1.1 is completed by induction. Equations 3.1.2 and 3.1.3 follow, since every subset $S \subseteq n^*$ can be broken into a union of sets of consecutive integers. \qed

The following theorem allows us to compute the average sample complexity (and intricacy) of rank 0 open covers of a shift of finite type with positive square adjacency matrix using the uniform system of coefficients, assuming that we know the words in the given shift of each length, $N(k^*)$. If $M$ is the adjacency matrix for a shift of finite type, then $N(k^*) = \sum_{i,j} (M^{k-1})_{ij}$.

**Theorem 3.1.2.** Let $X$ be a shift of finite type with adjacency matrix $M$ such that $M^2 > 0$. Let $c^n_S = 2^{-n}$ for all $S$. Then

$$\text{Asc}(X, \emptyset_0, \sigma) = \frac{1}{4} \sum_{k=1}^{\infty} \log \frac{|L_{k^*}(X)|}{2^k}. \quad (3.1.4)$$

We offer two proofs of this theorem. In the first proof we break the sum over all subsets $S \subseteq n^*$ in the definition of average sample complexity into the sum over subsets $S \subseteq n^*$ that contain $n - 1$ and subsets $S \subseteq n^*$ that do not contain $n - 1$. Since $c^n_S$ has no dependence on $S$, the sum over $S \subseteq n^*$ that do not contain $n - 1$ is equivalent to the sum over $S \subseteq (n - 1)^*$. The sum over the sets containing $n - 1$ is then broken into a sum over sets that contain $n - 2$ and those that do not contain $n - 2$. We simplify the sum over sets that do not contain $n - 2$ using Proposition 3.1.1. We continue this process inductively.
Proof 1. We begin by proving

\[
\frac{1}{n} \sum_{S \subseteq n^*} \log N(S) = \frac{1}{n} \sum_{S \subseteq n^*} \log N(n^*) + \frac{1}{4n} \sum_{k=1}^{n-1} \frac{n-k+3}{2^k} \log N(k^*). \tag{3.1.5}
\]

Let

\[a_n = \sum_{S \subseteq n^*} \log N(S) \quad \text{and} \quad \lambda_k = \log N(k^*).\]

First we show

\[a_n = \lambda_n + \sum_{k=1}^{n-1} \left( 2^{n-k-1} \lambda_k + a_k \right).\]

Since \(M^2 > 0\) we can use Equations 3.1.2 and 3.1.3. We have

\[a_n = \sum_{S \subseteq n^*} \log N(S) = \sum_{S \subseteq n^* \setminus \{n-1\}} \log N(S) + \sum_{S \subseteq n^* \setminus \{n-1\}} \log N(S) = a_{n-1} + \sum_{n-1 \in S} \log N(S).\]

Similarly, we have

\[\sum_{S \subseteq n^* \setminus \{n-1\}} \log N(S) = \sum_{S \subseteq n^* \setminus \{n-1\}} \log N(S) + \sum_{S \subseteq n^* \setminus \{n-2,n-1\}} \log N(S) = \sum_{S \subseteq (n-2)^*} \log N(S) + \sum_{S \subseteq (n-2)^*} \log N(S).\]

\[= a_{n-2} + 2^{n-2} \log N(1^*) + \sum_{S \subseteq (n-2)^* \setminus \{n-2,n-1\}} \log N(S) = a_{n-2} + 2^{n-2} \lambda_1 + \sum_{S \subseteq (n-2)^* \setminus \{n-2,n-1\}} \log N(S).\]
In general, for $\ell = 0, 1, \ldots, n - 1$, we can write

$$
\sum_{S \subseteq n^*} \log N(S) = \sum_{\{n-\ell,\ldots,n-1\} \subseteq S} \log N(S) + \sum_{\{n-\ell-1,\ldots,n-1\} \subseteq S} \log N(S)
$$

$$
= \sum_{S \subseteq (n-\ell-1)^*} \log N(S) + \sum_{\{n-\ell-1,\ldots,n-1\} \subseteq S} \log N((n-\ell)^*)
$$

$$
+ \sum_{S \subseteq n^*} \log N(S)
$$

$$
= a_{n-\ell-1} + 2^{n-\ell-1} \lambda_{n-\ell} + \sum_{\{n-\ell-1,\ldots,n-1\} \subseteq S} \log N(S).
$$

Combining these calculations, we get

$$
a_n = \lambda_n + \sum_{k=1}^{n-1} (2^{n-k-1} \lambda_k + a_k),
$$

Therefore,

$$
a_n - a_{n-1} = \lambda_n + \lambda_{n-2} + 2 \lambda_{n-3} + 4 \lambda_{n-4} + \cdots + 2^{n-3} \lambda_1 + a_{n-1}, \quad (3.1.6)
$$

which gives

$$
a_n = \lambda_n + 2a_{n-1} + 2^{n-2} \sum_{k=1}^{n-2} \frac{\lambda_k}{2^k}, \quad a_1 = \lambda_1. \quad (3.1.7)
$$

We use induction to prove Equation 3.1.5. We must show

$$
a_n = \lambda_n + 2^{n-2} \sum_{k=1}^{n-1} \frac{n-k+3}{2^k} \lambda_k. \quad (3.1.8)
$$

We see that $a_1 = \lambda_1$ in Equation 3.1.8. Now we assume 3.1.8 for all $k < n$ and prove it for $n$. By
the induction hypothesis, 3.1.7, and 3.1.8,

\[ a_n = \lambda_n + 2a_{n-1} + 2^{n-2} \sum_{k=1}^{n-2} \frac{\lambda_k}{2^k} \]

\[ = \lambda_n + 2 \left( \lambda_{n-1} + 2^{n-3} \sum_{k=1}^{n-2} \frac{n-j+2}{2^j} \lambda_k \right) + 2^{n-2} \sum_{k=1}^{n-2} \frac{\lambda_k}{2^k} \]

\[ = \lambda_n + 2\lambda_{n-1} + 2^{n-2} \sum_{k=1}^{n-2} \frac{n-k+3}{2^j} \lambda_k \]

\[ = \lambda_n + 2^{n-2} \sum_{k=1}^{n-1} \frac{n-k+3}{2^k} \lambda_k. \]

Now we show

\[ \text{Asc}(X, \mathcal{U}_0, \sigma) = \frac{1}{n} \sum_{S \subseteq n^*} \log N(S) = \frac{1}{4} \sum_{j=1}^{\infty} \frac{\log N(k^*)}{2^k}, \quad (3.1.9) \]

which would follow from

\[ \lim_{n \to \infty} \left( \frac{1}{n} \frac{1}{2^n} \log N(n^*) + \frac{1}{4n} \sum_{k=1}^{n-1} \frac{n-k+3}{2^k} \log N(k^*) \right) = \frac{1}{4} \sum_{k=1}^{\infty} \frac{\log N(k^*)}{2^k}. \]

We know that \( \log N(n^*)/n \) converges to the topological entropy of \((X, \sigma)\), so

\[ \lim_{n \to \infty} \frac{1}{n} \frac{1}{2^n} \log N(n^*) = 0. \]

Now

\[ \sum_{k=1}^{\infty} \frac{(3-k)k \log |A|}{2^k} \]

converges and \( N(k^*) \leq |A|^k \), so \( \log N(k^*) \leq k \log |A| \). Thus,

\[ \lim_{n \to \infty} \frac{1}{4n} \sum_{k=1}^{\infty} \frac{3-k}{2^j} \log N(k^*) = 0. \]

\[ \square \]

In the second proof we look at intervals of consecutive integers in each subset \( S \subseteq n^* \). We then take our sum over intervals of different lengths. To simplify the counting, we map each subset \( S \in n^* \) to a string of 0s and 1s of length \( n \) based on which integers of \( n^* \) the subset contains.
Proof 2. Since $M^2 > 0$, when trying to find $N(S)$ we may break $S$ into consecutive strings of integers of lengths $t_1, \ldots, t_k$ with gaps of length at least 1 between each, and then we have

$$N(S) = N(t_1^*)N(t_2^*) \cdots N(t_k^*).$$

Consider every subset $S \subseteq n^*$. We can map each $S$ to a string of 0’s and 1’s of length $n$, where the $i$th element of the string is 1 if $i \in S$ and 0 otherwise for $i = 0, 1, \ldots, n - 1$. For example, we can map $\{1, 2, 4\} \subset 6^*$ to 011010. Doing this allows us to visualize the set of subsets $S \subset n^*$ as a $2^n \times n$ array of 0s and 1s:

$$
\begin{array}{ccccccc}
0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 1 \\
0 & 0 & 0 & \cdots & 1 & 0 \\
& \vdots & & & & \\
0 & 1 & 1 & \cdots & 1 & 1 \\
1 & 1 & 1 & \cdots & 1 & 1 \\
\end{array}
$$

(3.1.10)

If we let $w_{n,k}$ be the number of strings of the form $1^k$ in 3.1.10, then Equation 3.1.1 gives

$$
\sum_{S \subset n^*} \log N(S) = \sum_{k=1}^{n} w_{n,k} \log N(k^*). 
$$

(3.1.11)

We calculate $w_{n,k}$ by counting the strings in 3.1.10 of the following types:

1. $\ldots 01^k0 \ldots$,
2. $1^k0 \ldots$, and
3. $\ldots 01^k$.

The number of strings of type (1) is $2^{n-k-2}(n - (k + 2) + 1)$. We find this by first picking one of the $n - (k + 2) + 1$ spots to place the first 0, then filling in the unused $n - k - 2$ spots in the string in $2^{n-k-2}$ ways. The number of strings of types (2) or (3) is $2^{n-k-1}$. Thus,

$$w_{n,k} = 2^{n-k-2}(n - k - 1) + 2 \cdot 2^{n-k-1} = 2^{n-k-2}(n - k + 3), \text{ for } k = 1, \ldots, n - 1. \quad (3.1.12)$$
Using the fact that $w_{n,n} = 1$ and combining Equations 3.1.11 and 3.1.12, we get

$$\sum_{S \subseteq n^*} \log N(S) = \sum_{k=1}^{n-1} 2^{n-k-2}(n - k + 3) \log N(k^*) + \log N(n^*).$$

This equation is the same as Equation 3.1.8 in the first proof of this theorem, so we finish this proof in the same manner we finished the first proof.

\[\square\]

**Corollary 3.1.3.** Let $X$ be a shift of finite type with adjacency matrix $M$ such that $M^2 > 0$. Let $c^n_S = 2^{-n}$ for all $S$. Then

$$\text{Int}(X, \mathcal{U}_0, \sigma) = \frac{1}{2} \sum_{k=1}^{\infty} \log \frac{|\mathcal{L}_k(X)|}{2^k} - h_{\text{top}}(X, T). \quad (3.1.13)$$

**Proof.** This follows immediately from the definitions of intricacy, average sample complexity and Theorem 3.1.2.

\[\square\]

**Corollary 3.1.4.** Two shifts of finite type that have positive square adjacency matrices have the same average sample complexity and intricacy of rank 0 open covers using the uniform system of coefficients $c^n_S = 2^{-n}$ if and only if they have the same complexity function.

**Proof.** Theorem 3.1.2 and Corollary 3.1.3 show that with these hypotheses, average sample complexity and intricacy depend only on $|\mathcal{L}_n(X)|$, which is exactly the complexity function of the subshift.

\[\square\]

**Example 3.1.5.** The full $r$-shift has a positive square adjacency matrix and $N(k^*) = r^k$, so

$$\text{Asc}(\Sigma_r, \mathcal{U}_0, \sigma) = \frac{\log r}{2}. \quad (3.1.14)$$

We also have $\text{Int}(\Sigma_r, \mathcal{U}_0, \sigma) = 2 \text{Asc}(\Sigma_r, \mathcal{U}_0, \sigma) - h_{\text{top}}(\Sigma_r, \sigma)$ and $h_{\text{top}}(\Sigma_r, \sigma) = \log r$ so we find

$$\text{Int}(\Sigma_r, \mathcal{U}_0, \sigma) = 0. \quad (3.1.15)$$

This example shows that a completely independent (segregated) shift of finite type has zero intricacy when it is taken over rank 0 cylinder sets with the uniform system of coefficients. This is what we want in a measurement of complexity that generalizes neural complexity.
Table 3.1: Two shifts of finite type with the same entropy and complexity functions, but different average sample complexity and intricacy functions

Interesting examples of shifts of finite type

In the appendix, we list many examples of subshifts with calculations for average sample complexity and intricacy. We select some interesting cases to discuss here. The numbering in the left column is for labeling purposes. It is arbitrary, but consistent throughout the paper for each shift of finite type. Unless otherwise labeled, the calculations are taken using the uniform system of coefficients $c^n_S = 2^{-n}$ and the open covers are by rank 0 cylinder sets.

Computations were made using Mathematica and tables show values rounded to 3 decimal places. When applicable Equation 3.1.13 is computed using the first 20 terms of the series.

Example 3.1.6. In this example we compare two shifts of finite type (labeled 13 and 17) that have the same entropy and complexity functions but different average sample complexity and intricacy functions. When looking at comparisons of $N(S)$ for each SFT for all $S \subset 4^*$ in Table 3.2 we can see where the differences occur in the average sample complexity and intricacy functions. For instance, SFT 13 has 13 words that appear at $\{0, 1, 3\}$, whereas SFT 17 has 11 words on those indices. Notice that the smallest power for which the adjacency matrix for SFT 13 is positive is 3, while it is 4 for SFT 17. This gives us a clue as to what Asc($n$) and Int($n$) measure.

Even though both SFTs have the same number of words of each length, the structure of these
words is different. The words that appear in sequences for SFT 13 are more complex in some sense because there is more freedom to build them. In this manner, SFT 13 is closer to the full 3-shift.

**Example 3.1.7 (Comparing SFT 20 and SFT 28).** Using Theorem 3.1.2, we find the average sample complexity and intricacy of rank 0 cylinder sets for shifts of finite type with positive square adjacency matrices. The appendix has a table with these values calculated for many SFTs. In this example we compare SFT 20 and SFT 28 which we denote by $X_{20}$ and $X_{28}$ respectively. These shifts both have the same entropy, but they have different average sample complexity and intricacy. Their complexity functions are different but have the same exponential growth rate. In this case, intricacy and average sample complexity tell us more than the entropy. The reason these quantities are smaller for $X_{28}$ than $X_{20}$ is because $|\mathcal{L}_n(X_{28})| < |\mathcal{L}_n(X_{20})|$ for all $n$.
<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>$\text{Asc}(X,\sigma)$</th>
<th>$\text{Int}(X,\sigma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>$\begin{pmatrix} 0 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \end{pmatrix}$</td>
<td>2</td>
<td>![Graph Image]</td>
<td>0.810</td>
<td>0.483</td>
<td>0.483</td>
</tr>
<tr>
<td>28</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 \ 1 &amp; 0 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td>![Graph Image]</td>
<td>0.810</td>
<td>0.464</td>
<td>0.464</td>
</tr>
</tbody>
</table>

Table 3.3: Two shifts of finite type with the same entropy but different average sample complexity and intricacy

<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>$H(10)$</th>
<th>$\text{Asc}(10)$</th>
<th>$\text{Int}(10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 0 \ 0 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \end{pmatrix}$</td>
<td>2</td>
<td>![Graph Image]</td>
<td>0.693</td>
<td>0.734</td>
<td>0.458</td>
<td>0.182</td>
</tr>
<tr>
<td>5</td>
<td>$\begin{pmatrix} 0 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \ 1 &amp; 1 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td>![Graph Image]</td>
<td>0.693</td>
<td>0.734</td>
<td>0.458</td>
<td>0.182</td>
</tr>
<tr>
<td>14</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 0 \ 0 &amp; 0 &amp; 1 \ 1 &amp; 1 &amp; 1 \end{pmatrix}$</td>
<td>2</td>
<td>![Graph Image]</td>
<td>0.693</td>
<td>0.734</td>
<td>0.458</td>
<td>0.182</td>
</tr>
</tbody>
</table>

Continued on next page
Table 3.4: Table with calculations for SFTs with the same entropy.

Example 3.1.8 (Comparison of SFTs with the same entropy). Table 3.4 shows seven SFTs with the same entropy but not all the same intricacy or average sample complexity functions. Just as when we compared SFT 13 and SFT 17, the smallest power for which the adjacency matrices of SFT 5, SFT 14 and SFT 15 are positive is 2 while it is 3 for SFT 18 and SFT 19. These two groups have
the same Asc and Int. SFT 26 is unique in that it has the same entropy as the other six SFTs, the square of its adjacency matrix is positive, but it has lower intricacy and average sample complexity than the others (a rounding error makes it seem that it has the same intricacy in the table).

**Average sample pressure of shifts of finite type**

Given a shift of finite type \((X, \sigma)\) and a subset \(S \subset n^*\), recall that \(\mathcal{L}_S(X)\) denotes the set of words seen at the places in \(S\) for all legal words in \(X\). Recall also that the metric, \(d\), we put on subshifts is defined by 

\[
d(x, y) = \frac{1}{m+1}, \text{ where } m = \inf \{|k| : x_k \neq y_k\}. \]

We will first calculate the average sample pressure of a subshift by letting \(f \in C(X, \mathbb{R})\) be a function of a single coordinate of the sequence \(x \in X\), i.e., \(f(x) = f(x_0)\). Letting \(\varepsilon = 1\) and \(c^n_S\) be a system of coefficients, the average sample pressure of the shift of finite type \(X\) is given by

\[
\text{Asp}_1(\sigma, f) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log |\sum_{i=1}^{|S|} w_i \in \mathcal{L}_S(X)| \exp \left( \sum_{i=1}^{|S|} f(w_i) \right). \tag{3.1.16}
\]

Notice if \(f(x) \equiv 0\), then we get

\[
\text{Asp}_1(\sigma, 0) = \limsup_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^n_S \log |\mathcal{L}_S(X)| = \text{Asc}(X, \mathcal{U}_0, \sigma).
\]

**Example 3.1.9.** Consider \(\Sigma_2\), the full 2-shift, and define \(f(0) = 0\) and \(f(1) = 1\). Given a subset \(S \subset n^*\) and \(w \in \mathcal{L}_S(\Sigma_2)\), to find \(\sum_{i=1}^{|S|} f(w_i)\) we count the number of 1s in \(w\). If there are \(j\) 1s in \(w\) then \(\sum_{i=1}^{|S|} f(w_i) = j\). There are \(C(|S|, j)\) words in \(\mathcal{L}_S(\Sigma_2)\) with \(j\) 1s. Since there are \(C(n, k)\) subsets \(S \subset n^*\) such that \(|S| = k\) we have

\[
\sum_{S \subset n^*} \log |\sum_{i=1}^{|S|} w_i \in \mathcal{L}_S(X)| \exp \left( \sum_{i=1}^{|S|} f(w_i) \right) = \sum_{k=0}^n \binom{n}{k} \log \sum_{j=0}^k \binom{k}{j} c^j.
\]
Therefore,

\[ \text{Asp}(\sigma, f) = \lim_{n \to \infty} \frac{1}{n^{2n}} \sum_{S \subset n^*} \log \sum_{w \in \mathcal{L}_S(X)} \exp \left( \sum_{i=1}^{\left| S \right|} f(w_i) \right) \]

\[ = \lim_{n \to \infty} \frac{1}{n^{2n}} \sum_{k=0}^{n} \binom{n}{k} \log \sum_{j=0}^{k} \binom{k}{j} e^j = \lim_{n \to \infty} \frac{1}{n^{2n}} \sum_{k=0}^{n} \binom{n}{k} \log(1 + e)^k \]

\[ = \lim_{n \to \infty} \frac{1}{n^{2n}} \log(1 + e) \sum_{k=0}^{n} \binom{n}{k} = \lim_{n \to \infty} \frac{1}{n} \log(1 + e)n2^{n-1} \]

\[ = \frac{1}{2} \log(1 + e). \]

By Equation 1.5.12 the pressure of the full 2-shift using the same function \( f \) is

\[ P(\sigma, f) = \log \left( e^{f(0)} + e^{f(1)} \right) = \log(1 + e), \]

so

\[ \text{Asp}(\sigma, f) = \frac{1}{2} P(\sigma, f). \]

**Example 3.1.10.** We again consider the full 2-shift, but this time define \( f \) as a general function depending on a single coordinate, i.e., \( f(x) = f(x_0) \). Now, for a subset \( S \subset n^* \), given a word \( w \in \mathcal{L}_S(\Sigma_2) \), we count the number of 0s and 1s. If there are \( j \) 0s then there are \( |S| - j \) 1s, so

\[ \sum_{S \subset n^*} \log \sum_{w \in \mathcal{L}_S(X)} \exp \left( \sum_{i=1}^{\left| S \right|} f(w_i) \right) = \sum_{k=0}^{n} \binom{n}{k} \log \sum_{j=0}^{k} \binom{k}{j} \exp \left( kf(0) + (k - j)f(1) \right). \]

Thus we have

\[ \text{Asp}(\sigma, f) = \lim_{n \to \infty} \frac{1}{n^{2n}} \sum_{S \subset n^*} \log \sum_{w \in \mathcal{L}_S(X)} \exp \left( \sum_{i=1}^{\left| S \right|} f(w_i) \right) \]

\[ = \lim_{n \to \infty} \frac{1}{n^{2n}} \sum_{k=0}^{n} \binom{n}{k} \log \sum_{j=0}^{k} \binom{k}{j} \exp \left( kf(0) + (k - j)f(1) \right) \]

\[ = \lim_{n \to \infty} \frac{1}{n^{2n}} \sum_{k=0}^{n} \binom{n}{k} \log \left( e^{f(0)} + e^{f(1)} \right)^k = \frac{1}{2} \log \left( e^{f(0)} + e^{f(1)} \right). \]

We see that \( \text{Asp}(\sigma, f) = (1/2)P(\sigma, f) \) as in Example 3.1.9.
Example 3.1.11. Generalizing Example 3.1.10 to $\Sigma_r$, the full $r$-shift with a function $f$ that depends on a single coordinate, we have

$$\sum_{S \subseteq n^*} \log \sum_{w \in \mathcal{L}_S(X)} \exp \left( \sum_{i=1}^{\vert S \vert} f(w_i) \right) = \sum_{k=0}^{n} \binom{n}{k} \log \left( \sum_{i=0}^{r} e^{f(i)} \right)^k.$$

Thus, if we fix $c_S^n = 2^{-n}$, then for the full $r$-shift

$$\text{Asp}(\sigma, f) = \lim_{n \to \infty} \frac{1}{n} 2^n \sum_{S \subseteq n^*} \log \sum_{w \in \mathcal{L}_S(X)} \exp \left( \sum_{i=1}^{\vert S \vert} f(w_i) \right) = \lim_{n \to \infty} \frac{1}{n} 2^n \sum_{k=0}^{n} \binom{n}{k} \log \left( \sum_{i=0}^{r-1} e^{f(i)} \right)^k = \frac{1}{2} \log \left( \sum_{i=0}^{r-1} e^{f(i)} \right).$$

Example 3.1.12. Consider the shifts of finite type in Table 3.5. We see that they are very similar and indistinguishable by the measures of complexity we have considered previously: entropy, average sample complexity, and intricacy. Suppose $f_1, f_2$ are two functions of a single coordinate on each
shift of finite type defined by

\[
f_1(x) = \begin{cases} 
0, & x_0 = 0 \\
0, & x_0 = 1 \\
1, & x_0 = 2 
\end{cases} \quad \text{and} \quad f_2(x) = \begin{cases} 
0, & x_0 = 0 \\
1, & x_0 = 1 \\
0, & x_0 = 2 
\end{cases}
\]

Table 3.6 shows the calculations of Asp for two shifts that have the same entropy, Asc, and Int. Notice that \( f_1 \) places more weight on the symbol 2, whereas \( f_2 \) places more weight on the symbol 1. It is not surprising that the shift labeled 14 has a larger value for \( \text{Asp}(\sigma, f_1, 10) \) than it does for \( \text{Asp}(\sigma, f_2, 10) \), since every time a 1 appears in a sequence of this shift, a 2 must follow it. This makes the number of 2s that appear in elements of \( \mathcal{L}_S(X) \) larger than the number of 1s that appear. For example, there are 155 appearances of the symbol 2 in the elements of \( \mathcal{L}_S(X) \) for \( S \subset 4^* \) and only 103 appearances of the symbol 1.

It is also not surprising that the values of \( \text{Asp}(\sigma, f_1, 10) \) and \( \text{Asp}(\sigma, f_1, 10) \) are equal for the shift labeled 5, since there is clear symmetry in the shift and we can switch the symbols 1 and 2 without changing the number of appearances of each symbol in elements of \( \mathcal{L}_S(X) \).
3.2 Average sample complexity and intricacy of other subshifts

In this section, we will examine average sample complexity and intricacy in the setting of symbolic dynamical systems. Recall, given a sequence \( u \) with entries from an alphabet \( \mathcal{A} \) we form the symbolic dynamical system \( (\mathcal{O}(u), \sigma) \)

Substitution subshifts

Explanations of substitution sequences in Table 3.7 are given in Example 1.5.23. The average sample complexity function and intricacy function as well as the entropy function \( H(n) = \log p_u(n)/n \) are given for \( n = 11 \) for the two examples of sequences.

<table>
<thead>
<tr>
<th>Substitution</th>
<th>Sequence Name</th>
<th>( H(11) )</th>
<th>Asc(11)</th>
<th>Int(11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 \rightarrow 01, 1 \rightarrow 10</td>
<td>Morse</td>
<td>0.315</td>
<td>0.262</td>
<td>0.210</td>
</tr>
<tr>
<td>2 0 \rightarrow 01, 1 \rightarrow 0</td>
<td>Fibonacci</td>
<td>0.226</td>
<td>0.191</td>
<td>0.157</td>
</tr>
</tbody>
</table>

Table 3.7: Calculations for systems formed from substitution sequences.

The Fibonacci sequence \( u_1 \) is Sturmian so \( p_{u_1}(n) = n + 1 \) for all \( n \) and the entropy of the subshift formed from this sequence is 0. It has the minimum complexity that an aperiodic sequence can have. The subshifts formed from the Morse sequence, \( u_2 \), also has zero entropy. It has a more complicated complexity function than the Fibonacci sequence. It can be shown (see [Fog]) that the complexity function of the Morse sequence is given by the following: \( p_{u_2}(1) = 2, p_{u_2}(2) = 4 \), and for \( n \geq 3 \) if \( n = 2^r + q + 1 \), \( r \geq 0 \), and \( 0 < q \leq 2^r \)

\[
p_{u_2}(n) = \begin{cases} 
6(2^{r-1}) + 4q, & 0 < q \leq 2^{r-1} \\
8(2^{r-1}) + 2q, & 2^{r-1} < q \leq 2^r 
\end{cases}
\]

Notice \( n + 1 < p_{u_2}(n) < 4n \), which shows that the Morse sequence has zero entropy, but complexity function strictly greater than the complexity function of the Fibonacci sequence. We know that the limits of Asc(\(n\)) and Int(\(n\)) for both of these sequences will both be zero, but the average number of words seen at the places in \( S \subset n^* \) for the Morse sequence will be greater than for the Fibonacci sequence. In these ways the Morse sequence is more complex than the Fibonacci sequence.
Sturmian subshifts

Recall that a Sturmian sequence, $u$, is defined as a sequence with complexity function $p_u(n) = n + 1$ for all $n$. Given a Sturmian sequence $u$, its associated Sturmian subshift is $\mathcal{O}(u) \subset \Sigma_2$, since $p_u(1) = 2$. To each irrational number we associate a Sturmian sequence. We will examine the average sample complexity and intricacy of several Sturmian subshifts corresponding to different irrational numbers. Since the entropy of every Sturmian subshift is zero, and the intricacy and average sample complexity are bounded by entropy, we know these will be zero as well. We can still gain some information by comparing the intricacy and average sample complexity functions of different Sturmian sequences.

Table 3.8 shows 14 Sturmian sequences with different irrational rotation numbers $\alpha$. An approximation for each $\alpha$ is shown, as well as the first 15 entries of the sequence. Since they all have the same complexity function, $H(n) = \log p_u(n)/n$ is the same for all of them. The average sample complexity function and intricacy function are given for $n = 11$ in each case.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Sequence</th>
<th>$H(11)$</th>
<th>Asc(11)</th>
<th>Int(11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.618</td>
<td>011101011011010</td>
<td>0.226</td>
<td>0.191</td>
</tr>
<tr>
<td>2</td>
<td>0.718</td>
<td>101110110110111</td>
<td>0.226</td>
<td>0.190</td>
</tr>
<tr>
<td>3</td>
<td>0.142</td>
<td>0000001000000010</td>
<td>0.226</td>
<td>0.179</td>
</tr>
<tr>
<td>4</td>
<td>0.382</td>
<td>010010100100101</td>
<td>0.226</td>
<td>0.191</td>
</tr>
<tr>
<td>5</td>
<td>0.413</td>
<td>010100101001010</td>
<td>0.226</td>
<td>0.191</td>
</tr>
<tr>
<td>6</td>
<td>0.704</td>
<td>101101110110111</td>
<td>0.226</td>
<td>0.190</td>
</tr>
<tr>
<td>7</td>
<td>0.633</td>
<td>011101101011011</td>
<td>0.226</td>
<td>0.191</td>
</tr>
<tr>
<td>8</td>
<td>0.586</td>
<td>101101101011010</td>
<td>0.226</td>
<td>0.191</td>
</tr>
<tr>
<td>9</td>
<td>0.631</td>
<td>1011010101011011</td>
<td>0.226</td>
<td>0.191</td>
</tr>
<tr>
<td>10</td>
<td>0.613</td>
<td>1011010101011010</td>
<td>0.226</td>
<td>0.191</td>
</tr>
<tr>
<td>11</td>
<td>0.620</td>
<td>101101011011010</td>
<td>0.226</td>
<td>0.191</td>
</tr>
<tr>
<td>12</td>
<td>0.049</td>
<td>0000000000000000</td>
<td>0.226</td>
<td>0.167</td>
</tr>
<tr>
<td>13</td>
<td>0.094</td>
<td>0000000000100000</td>
<td>0.226</td>
<td>0.170</td>
</tr>
<tr>
<td>14</td>
<td>0.414</td>
<td>010100101001010</td>
<td>0.226</td>
<td>0.191</td>
</tr>
</tbody>
</table>

Table 3.8: Calculations for systems formed from Sturmian sequences with rotation number $\alpha$.

As expected, the entropy functions are equal for every Sturmian sequence and will go to zero as $n \to \infty$. Similarly, both Asc($n$) and Int($n$) will go to zero as $n \to \infty$. We observe that sequences corresponding to irrational numbers close to 0,5, such as $\alpha \approx 0.586$ have greater Asc(11) and Int(11) than sequences corresponding to irrational numbers close to 0 or 1, such as $\alpha \approx 0.094$. While these sequences have the same number of words of each length, the variety of 0s and 1s in the words of
each length differ. For example for the sequence corresponding to \( \alpha \approx 0.094 \), the words of length 8 each contain at most one 1 and the rest 0s. This makes the average number of words on each subset \( S \subset n^* \) relatively low. Compare this to the words of length 8 in the sequence corresponding to \( \alpha \approx 0.586 \). These words each have either four or five 1s and four or three 0s. This variety causes the average number of words on each subset \( S \subset n^* \) to be relatively large.

The reason irrational rotation numbers close to 0.5 create sequences with more variety in the words of each length can be deduced by examining the formation of the sequences. The intervals \( I_0 = [0, 1 - \alpha) \) and \( I_1 = [1 - \alpha, 1) \) that define the coding for the rotation sequence (see Definition 1.5.25) are more similar in length the closer \( \alpha \) is to 0.5 so as we consider the orbit of a point it spends a more even amount of time in each interval than if \( \alpha \) is closer to 0 or 1 (and one of the intervals is much larger than the other). This leads to a greater variety of sampled words in the rotation sequence.
CHAPTER 4
Measure-theoretic intricacy and average sample complexity

4.1 Definitions and preliminary results

We formulate a definition of measure-theoretic intricacy and measure-theoretic average sample complexity in analogy with measure-theoretic entropy.

**Definition 4.1.1.** Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system, \(\alpha = \{A_1, \ldots, A_n\}\) a finite measurable partition of \(X\) and \(c^0_S\) a system of coefficients. Recall \(H_\mu(\alpha) = -\sum_{i=1}^n \mu(A_i) \log \mu(A_i)\) and for \(S \subset n^*\)

\[
\alpha_S = \bigvee_{i \in S} T^{-i} \alpha.
\]

The **measure-theoretic intricacy of** \(T\) **with respect to the partition** \(\alpha\) is

\[
\text{Int}_\mu(X, \alpha, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^0_S [H_\mu(\alpha_S) + H_\mu(\alpha_{S^c}) - H_\mu(\alpha_{n^*})]. \tag{4.1.1}
\]

The **measure-theoretic average sample complexity of** \(T\) **with respect to the partition** \(\alpha\) is

\[
\text{Asc}_\mu(X, T, \alpha) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^0_S H_\mu(\alpha_S) \tag{4.1.2}
\]

The **measure-theoretic average configuration complexity of** \(T\) **with respect to the partition** \(\alpha\) is

\[
\text{Acc}_\mu(X, \alpha, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{S \subset n^*} c^0_S H_\mu(\alpha_S). \tag{4.1.3}
\]

**Remark.** If \((X, T)\) is a subshift, \(\alpha\) the partition by rank 0 cylinder sets and \(\mathcal{U}(\alpha)\) the corresponding open cover of \(X\), the \(\text{Asc}_\mu(X, \alpha, T) \leq \text{Asc}(X, \mathcal{U}, T)\). This is true since, for each \(n\) and \(S \subset n^*\), \(H_\mu(\alpha_S) \leq \log N(\mathcal{U}(\alpha)_S)\).

We will show below that the limits in these definitions exist. We also define the **measure-theoretic**
Intricacy function and the measure-theoretic average sample complexity function as we did in the topological case.

**Definition 4.1.2.** Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system, \(\alpha\) a finite measurable partition of \(X\) and a system of coefficients \(c^\alpha_S\). The *measure-theoretic average sample complexity function* of \(T\) with respect to the partition \(\alpha\) is given by

\[
\text{Asc}_\mu(X, \alpha, T, n) = \frac{1}{n} \sum_{S \subseteq n^*} c^\alpha_S H_\mu(\alpha_S).
\]

The *measure-theoretic intricacy function* of \(T\) with respect to the partition \(\alpha\) is given by

\[
\text{Int}_\mu(X, \mathcal{U}, T, n) = \frac{1}{n} \sum_{S \subseteq n^*} c^\alpha_S \left[H_\mu(\alpha_S) + H_\mu(\alpha_{S^c}) - H_\mu(\alpha_n)\right]
\]

When the context is clear we will sometimes write these as \(\text{Asc}_\mu(n)\) and \(\text{Int}_\mu(n)\).

**Theorem 4.1.3.** Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system and \(\alpha\) a finite measurable partition. For a system of coefficients \(c^\alpha_S\), \(\text{Asc}_\mu(X, T, \alpha)\) exists and equals \(\inf_n (1/n) \sum_{S \subseteq n^*} c^\alpha_S H_\mu(\alpha_S)\)

**Proof.** Let \(b_n = \sum_{S \subseteq n^*} c^\alpha_S H_\mu(\alpha_S)\). We show that \(b_n\) is subadditive. For each \(S \subseteq (n+m)^*\) define \(U = U(S)\) and \(V = V(S)\) as in the proof of Proposition 2.3.1. We have

\[
H_\mu(\alpha_S) \leq H_\mu(\alpha_U) + H_\mu(\alpha_V). \tag{4.1.4}
\]

The proof of subadditivity follows in the same manner as in the proof of Theorem 2.3.2. Then we use Lemma 1.2.3 to achieve the result. \(\square\)

**Corollary 4.1.4.** Let \((X, \mathcal{B}, \mu, T)\) be a measure-preserving system and \(\alpha\) a finite measurable partition. If \(c^\alpha_S\) is a system of coefficients, then the limits in the definitions of \(\text{Acc}_\mu(X, \alpha, T)\) and \(\text{Int}_\mu(X, \alpha, T)\) (Definition 4.1.1) exist.

**Proof.** The proof for \(\text{Acc}_\mu(X, \alpha, T)\) is the same as the proof for \(\text{Asc}_\mu(X, \alpha, T)\) except in the definition of \(U(S)\) we know \(0 \in U(S)\) and we define \(V(S) = S \cap [(n+m)^* \setminus n^*] - P\) (subtract \(P\) from each element in the set), where \(P = \min\{p \in (n+m)^* : p \geq n\}\). The proof is still valid since Equation 4.1.4 still
holds. The proof for measure-theoretic intricacy follows from the fact that

\[ \text{Int}_{\mu}(X, \alpha, T) = 2 \text{Asc}_{\mu}(X, \alpha, T) - h_{\mu}(X, \alpha, T). \quad (4.1.5) \]

\[ \square \]

### 4.2 Main results for measure-theoretic systems

In this section we state and prove the main results about measure-theoretic intricacy and measure-theoretic average sample complexity. We state and prove the theorems for the fixed system of coefficients \( c_{S}^{n} = 2^{-n} \). See Section 5.2 for a discussion about extending these results to more general weights.

#### Theorem relating measure-theoretic average sample complexity and measure-theoretic intricacy to measure-theoretic entropy

The first result in this section is the measure-theoretic analogue to Theorem 2.5.1. In Section 2.5 we show for fixed \( c_{S}^{n} = 2^{-n} \), \( \sup_{\mathcal{U}} \text{Asc}(X, \mathcal{U}, T) = \sup_{\mathcal{U}} \text{Int}(X, \mathcal{U}, T) = h_{\text{top}}(X, T) \). In this section we show \( \sup_{\alpha} \text{Asc}_{\mu}(X, \alpha, T) = \sup_{\alpha} \text{Int}_{\mu}(X, \alpha, T) = h_{\mu}(X, T) \). The proof follows the same structure as the proof of Theorem 2.5.1. We first fix a finite measurable partition \( \alpha \) of \( X \) and consider \( \lim_{k \to \infty} \text{Asc}_{\mu}(X, \alpha_{k}, T) \). We show this equals the measure-theoretic entropy of \( T \) with respect to \( \alpha \).

This result is important in the same ways as Theorem 2.5.1. It establishes a connection between measure-theoretic entropy and the measurements of complexity we define in this paper. In the same way that we focus our study of topological intricacy and topological average sample complexity on particular open covers, this theorem motivates us to focus the study of measure-theoretic intricacy and measure-theoretic average sample complexity on particular partitions. We will see the importance of this when we do computations on Markov shifts using a partition by time-zero cylinder sets.

**Theorem 4.2.1.** Let \( (X, \mathcal{B}, \mu, T) \) be a measure-preserving system, \( \alpha \) a finite measurable partition
of $X$, and fix the system of coefficients $c^n_S = 2^{-n}$. Then

$$\sup_{\alpha} \text{Asc}_\mu(X, \alpha, T) = h_\mu(X, T).$$

We prove this theorem using similar techniques as the proof of Theorem 2.5.1. We have the following lemma analogous to Lemma 2.5.3.

**Lemma 4.2.2.** Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $\alpha$ a finite measurable partition of $X$. Given $n \in \mathbb{N}$ and $S \subset n^*$, the following properties hold:

1. $H_\mu((\alpha_{k^*})_S) = H_\mu(\alpha_{S+k^*})$.

2. Given $S_1, S_2, \ldots, S_m \subset n^*$, $H_\mu(\alpha_{\cup_i S_i}) \leq \sum_i H_\mu(\alpha_{S_i})$.

**Proof.** 1. We have

$$(\alpha_{k^*})_S = \bigvee_{i \in S} \alpha_{k^*} = \bigvee_{i \in S+k^*} \alpha_i = \alpha_{S+k^*}.$$  

2. This is true because given two finite, measurable partitions $\alpha$ and $\beta$ of $X$,

$$H_\mu(\alpha \lor \beta) \leq H_\mu(\alpha) + H_\mu(\beta).$$

Our claim follows by induction.

**Proof of Theorem 4.2.1.** The proof follows the same pattern as the proof of Theorem 2.5.1, replacing $\mathcal{Z}_S$ with $\alpha_S$ and log $N(S)$ with $H_\mu(\alpha_S)$.

**Corollary 4.2.3.** Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, $\alpha$ a finite measurable partition of $X$, and fix the system of coefficients $c^n_S = 2^{-n}$. Then

$$\sup_{\alpha} \text{Int}_\mu(X, \alpha, T) = h_\mu(X, T).$$

**Proof.** This follows from Theorem 4.2.1 and Equation 4.1.5.
Theorems relating measure-theoretic average sample complexity to a series involving conditional entropies

The next few results give a relationship between measure-theoretic average sample complexity of a finite measurable partition \( \alpha \) and a series summed over \( i \) involving the conditional entropies \( H_\mu (\alpha \mid \alpha_i^{\infty}) \). In general \( \text{Asc}_\mu (X, \alpha, T) \) is greater than or equal to the sum of the series, but for certain systems equality will hold. In particular, equality holds for 1-step Markov shifts. We will take advantage of this in the next section to compute \( \text{Asc}_\mu \) and \( \text{Int}_\mu \) of 1-step Markov shifts. One purpose of accurately computing \( \text{Asc}_\mu \) and \( \text{Int}_\mu \) is to look for measures, \( \mu \), that maximize these quantities.

In Theorem 4.2.4 we show that \( \text{Asc}_\mu (X, \alpha, T) \) is equal to half the entropy of the first return map \( T_{X \times A} \) on a cross product \( X \times A \) of \( X \) with the cylinder \( A = [1] \) in the full 2-shift with respect to the finite measurable partition \( \alpha \times A \). These are formally defined below, but we will give some intuition for their construction here.

First in the one-sided full 2-shift, \( \Sigma_2^+ \), we define \( A = [1] = \{ \xi \in \Sigma_2^+ : \xi_0 = 1 \} \). Then, subsets \( S \subset n^* \) correspond to occurrences of 1 in the first \( n \) elements of sequences \( \xi \in A \). Denote by \( \xi_0^{n-1} \) both the string \( \xi_0 \xi_1 \ldots \xi_{n-1} \) and the cylinder set \( \{ z \in \Sigma_2^+ : z_i = \xi_i \text{ for all } i = 0, 1, \ldots, n-1 \} \). We denote the subsets \( S \subset n^* \) corresponding to \( \xi \) by \( S(\xi_0^{n-1}) = \{ i \in n^* : \xi_i = 1 \} \) and \( S(\xi) = \{ i \in \mathbb{N} : \xi_i = 1 \} \). Since averaging \( H_\mu (\alpha_s) \) over all \( S \) with weights \( 2^{-n} \) amounts to picking random \( S \) and taking the expectation of \( H_\mu (\alpha_s) \), we make calculations by doing the latter.

We will introduce some notation and give some facts necessary for the statements and proofs of Theorems 4.2.4 and 4.2.6. One may refer to [Pet] for more background information and details. Let \( P \) denote the Bernoulli measure \( \mathcal{B}(1/2, 1/2) \) on \( \Sigma_2^+ \). Let \( A \) be the subset of \( \Sigma_2^+ \) defined above and denote by \( n_A : A \to \mathbb{Z}_{\geq 1} \) the minimum return time of a sequence \( \xi \in A \) to \( A \) under the shift \( \sigma \), i.e.,

\[
n_A(\xi) = \inf \{ n \geq 1 : \sigma^n \xi \in A \} = \inf \{ n \geq 1 : \xi_n = 1 \}.
\]

Since \( (\Sigma_2^+, \sigma, P) \) is ergodic, the expected recurrence time of a point \( \xi \in A \) to \( A \) is \( 1/P(A) \). Let \( \sigma_A \xi = \sigma^{n_A(\xi)} \xi \). Given a positive integer \( n \) and sequence \( \xi \in A \), define \( m_\xi(n) \) by

\[
m_\xi(n) = \sum_{i=0}^{n-1} n_A(\sigma_A^i \xi) = n_A(\xi) + n_A(\sigma_A \xi) + \cdots + n_A(\sigma_A^{n-1} \xi),
\]
the sum of the first $n$ return times of $\xi$ to $A$. Since the expected return time of $\xi$ to $A$ is $1/P(A)$, we have that

$$\lim_{n \to \infty} \frac{m_\xi(n)}{n} = \frac{1}{P(A)} = 2 \text{ for } P_A\text{-a.e. } \xi \in A. \quad (4.2.1)$$

Since $n_A \in L^1$, the Ergodic Theorem implies convergence $m_\xi/n \to 2$ in $L^1$ as well.

For a measure-preserving system $(X, \mathcal{B}, \mu, T)$ denote by $T_{X \times A}$ the first-return map on $X \times A$, so that $T_{X \times A}(x, \xi) = (T^n\xi(x), \sigma_A\xi)$. In the proof we also use the fact that for two countable measurable partitions $\alpha$ and $\gamma$ of $X$

$$H_\mu(\alpha \vee \gamma) = H_\mu(\alpha) + H_\mu(\gamma|\alpha). \quad (4.2.2)$$

**Theorem 4.2.4.** Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system and $\alpha$ a finite measurable partition of $X$. Let $A = [1] = \{ \xi \in \Sigma^+_2 : \xi_0 = 1 \}$ and $\beta = \alpha \times A$ the related finite partition of $X \times A$. Denote by $T_{X \times A}$ the first-return map on $X \times A$ and let $P_A = P/P[1]$ denote the measure $P$ restricted to $A$ and normalized. Let $c^n_A = 2^{-n}$ for all $S \subset n^*$. Then

$$\text{Asc}_\mu(X, \alpha, T) = \frac{1}{2} H_{\mu \times P_A}(X \times A, \beta, T_{X \times A}). \quad (4.2.3)$$

**Proof.** Given $\varepsilon > 0$, define the set $U_\varepsilon(n) \subset A$ by

$$U_\varepsilon(n) = \left\{ \xi \in A : \left| \frac{m_\xi(n)}{n} - 2 \right| > \varepsilon \right\}.$$

By Equation 4.2.1 $\lim_{n \to \infty} P(U_\varepsilon(n)) = 0$. Given $\varepsilon > 0$, for $\xi \in A \setminus U_\varepsilon$ from Equation 4.2.2 we have

$$\frac{1}{n} \left| H_\mu \left( \alpha S(\xi_0^{2n-1}) \right) - H_\mu \left( \alpha \left( S^{m_\xi(n)} \right) \right) \right| = \frac{1}{n} \left| H_\mu \left( \alpha S(\xi_0^{m_\xi(n)}) \right) \right| \leq \frac{1}{n} \left| 2n - 1 - m_\xi(n) - 1 \right| H_\mu(\alpha) < \varepsilon H_\mu(\alpha) + \frac{2}{n}.$$

Thus for $\varepsilon > 0$ and $\xi \in A \setminus U_\varepsilon$

$$\lim_{n \to \infty} \frac{1}{n} H_\mu \left( \alpha \left( S^{m_\xi(n)} \right) \right) = \lim_{n \to \infty} \frac{1}{n} H_\mu \left( \alpha S(\xi_0^{2n-1}) \right). \quad (4.2.4)$$
Recall from Definition 1.2.9 that for each \((x, \xi)\) in \(X \times A\), \(\beta_n^*(x, \xi)\) denotes the element of \(\beta_n^* = \bigvee_{i=0}^{n-1} T_{X\times A}^{-i}(\alpha \times A)\) to which \((x, \xi)\) belongs. By Definition 1.2.9 of the information function \(I\), we have

\[ h_{\mu \times P_A}(X \times A, \beta, T_{X\times A}) = \lim_{n \to \infty} \frac{1}{n} H_{\mu \times P_A}(\beta_n^*) \]
\[ = \lim_{n \to \infty} \frac{1}{n} \int_{X \times A} I_{\beta_n^*}(x, \xi) d\mu(x) dP_A(\xi) \]
\[ = - \lim_{n \to \infty} \frac{1}{n} \int_A \int_X \log [(\mu \times P_A)(\beta_n^*(x, \xi))] d\mu(x) dP_A(\xi). \]  

For each \(\xi \in A\)

\[ (\mu \times P_A)(\beta_n^*(x, \xi)) = \mu \left( \alpha_S^{m_\xi(n)}(x) \right), \]

so 4.2.7 becomes

\[ - \lim_{n \to \infty} \frac{1}{n} \int_A \int_X \log \left( \alpha_S^{m_\xi(n)}(x) \right) d\mu(x) dP_A(\xi) \]
\[ = \lim_{n \to \infty} \frac{1}{n} \int_A H_{\mu} \left( \alpha_S^{m_\xi(n)} \right) dP_A(\xi). \]  

Let \(\varepsilon > 0\) be given and break the integral in 4.2.9 into two integrals over \(A \setminus U_\varepsilon\) and \(U_\varepsilon\). We also multiply and divide by \(m_\xi(n)\) (which we denote by \(m_\xi\) to simplify the notation) to get

\[ \lim_{n \to \infty} \int_{A \setminus U_\varepsilon} \frac{m_\xi}{n} \frac{1}{m_\xi} H_{\mu} \left( \alpha_S^{m_\xi(n)} \right) dP_A(\xi) + \lim_{n \to \infty} \int_{U_\varepsilon} \frac{m_\xi}{n} \frac{1}{m_\xi} H_{\mu} \left( \alpha_S^{m_\xi(n)} \right) dP_A(\xi). \]  

Since \(m_\xi/n \to 2\) in \(L^1\) and \(P_A(U_\varepsilon(n)) \to 0\) we have

\[ \frac{m_\xi(n)}{n} \chi_{U_\varepsilon(n)}(\xi) \to 0 \text{ in } L^1. \]  

By the definition of \(H_{\mu}\) we know \((1/m_\xi)H_{\mu} \left( \alpha_S^{m_\xi(n)} \right)\) is bounded so

\[ \lim_{n \to \infty} \int_{U_\varepsilon} \frac{m_\xi}{n} \frac{1}{m_\xi} H_{\mu} \left( \alpha_S^{m_\xi(n)} \right) dP_A(\xi) = 0. \]  

89
Similarly
\[
\lim_{n \to \infty} \int_{U_\epsilon} \frac{1}{2n} H_\mu \left( \alpha_{S_{(\xi_0^{2n-1})}} \right) dP_A(\xi) = 0.
\] (4.2.13)

Thus, 4.2.10 becomes
\[
\lim_{n \to \infty} \int_{A \setminus U_\epsilon} \frac{m_\xi}{n} \frac{1}{m_\xi} H_\mu \left( \alpha_{S_{(\xi_{0}^{m_\xi})}} \right) dP_A(\xi).
\] (4.2.14)

Again, since \((1/m_\xi)H_\mu \left( \alpha_{S_{(\xi_{0}^{m_\xi})}} \right)\) is bounded we can use 4.2.4 and 4.2.13 to get
\[
\lim_{n \to \infty} \int_{A \setminus U_\epsilon} \frac{m_\xi}{n} \frac{1}{m_\xi} H_\mu \left( \alpha_{S_{(\xi_{0}^{m_\xi})}} \right) dP_A(\xi) = 2 \lim_{n \to \infty} \int_{A \setminus U_\epsilon} \frac{1}{2n} H_\mu \left( \alpha_{S_{(\xi_0^{2n-1})}} \right) dP_A(\xi)
\]
= 2 \lim_{n \to \infty} \int_A \frac{1}{2n} H_\mu \left( \alpha_{S_{(\xi_0^{2n-1})}} \right) dP_A(\xi). \quad (4.2.15)

(4.2.16)

For a fixed \(n\) and each \(\xi \in A\), \(S(\xi_0^{2n-1})\) is the subset of \((2n)^*\) corresponding to occurrences of 1 in the first \(2n\) elements of \(\xi\). Since \(P\) is the Bernoulli measure, \(P_A(\xi_0^{2n-1}) = 2^{-2n}\), so integrating over the sequences in \(A\) with respect to the measure \(P_A\) shows that the expression in 4.2.5 equals
\[
2 \lim_{n \to \infty} \frac{1}{2n} \sum_{S \subset (2n)^*} \frac{1}{22n} H_\mu(\alpha_S) = 2 \text{Asc}_\mu(X, T, \alpha).
\] (4.2.17)

\(\square\)

In the next corollary we use the relationship between measure-theoretic average sample complexity and measure-theoretic entropy we just proved to show the existence of measures of maximal measure-theoretic average sample complexity.

**Corollary 4.2.5.** Let \((X, T)\) be a topological dynamical system, \(\alpha\) a fixed Borel measurable partition of \(X\), and \(e^n_S = 2^{-n}\) for all \(S \subset n^*\). There exist ergodic probability measures on \(X\) that maximize \(\text{Asc}_\mu(X, \alpha, T)\).

**Proof.** Like \(h_\mu(X, \alpha, T), h_\mu \times P_A(X \times A, \beta, TX \times A)\) is an affine function of \(\mu\), so Theorem 4.2.4 implies \(\text{Asc}_\mu(X, \alpha, T)\) is also an affine function of \(\mu\). Since \(\text{Asc}_\mu(X, \alpha, T)\) is an infimum of continuous functions of \(\mu\) (see Theorem 4.1.3), it is an upper semi-continuous function of \(\mu\). We know the space of invariant probability measures on \(X\) is nonempty and compact in the weak*-topology, and an upper semi-continuous function on a compact space attains its supremum. Therefore, the set of
measures \( \mu \) that maximize \( \text{Asc}_\mu(X, \alpha, T) \) is nonempty. It is convex because \( \text{Asc}_\mu(X, \alpha, T) \) is affine in \( \mu \). The extreme points of this set coincide with the ergodic measures that maximize \( \text{Asc}_\mu(X, \alpha, T) \) (see Chapter 8 of [Wal] for more details and proofs of the properties of \( h_\mu(X, \alpha, T) \)).

In the next theorem we relate \( \text{Asc}_\mu \) to a series involving conditional entropies. To show this we use the previous theorem along with definitions and facts about conditional entropy and the information function. The main idea is to break up the set \( A \subset \Sigma_2^+ \) from the previous theorem into sets \( A_i = \{ \xi \in A : n_A(\xi) = i \} \) consisting of sequences whose first return time is \( i \). This construction is where the series summed over \( i \) of conditional entropies comes in. The series also involves a factor of \( 2^{-i} \) in each term, which comes from the fact that we are assuming the measure on \( A \) is Bernoulli.

The inequality becomes an equality on systems such that \( I_{\alpha|\alpha_i}^\infty(x) = I_{\alpha|\alpha_i}(x) \) a.e., which is true for the case of 1-step Markov shifts.

**Theorem 4.2.6.** Let \( (X, \mathcal{B}, \mu, T) \) be an ergodic measure-preserving system and \( \alpha \) a finite measurable partition of \( X \). Let \( c_S^\alpha = 2^{-n} \) for all \( S \subset n^* \). Then

\[
\text{Asc}_\mu(X, \alpha, T) \geq \frac{1}{2} \sum_{i=1}^{\infty} 2^{-i} H_\mu(\alpha | \alpha_i^\infty). \tag{4.2.18}
\]

**Proof.** Assume we have \( A, P_A, \) and \( \beta = \alpha \times A \) as in Theorem 4.2.4. Then

\[
h_{\mu \times P_A}(X \times A, \beta, T_{X \times A}) = H_{\mu \times P_A}(\beta | \beta_1^\infty)
= \int_{X \times A} I_{\beta|\beta_1^\infty}(x, \xi) d\mu(x) dP_A(\xi). \tag{4.2.20}
\]

If we break \( A \) into the union of disjoint sets \( A_i = \{ \xi \in A : n_A(\xi) = i \} \), then 4.2.20 becomes

\[
\sum_{i=1}^{\infty} \int_{A_i} \int_X I_{\beta|\beta_1^\infty}(x, \xi) d\mu(x) dP_A(\xi). \tag{4.2.21}
\]

Since we are not partitioning \( A \) the information function depends on \( \xi \) only through the dependence of the partitioning of \( X \) on \( \xi \). Thus

\[
I_{\beta|\beta_1^\infty}(x, \xi) = I_{\alpha|\alpha_i^\infty}(x, \xi) = I_{\alpha|\alpha_S(\xi)}(x). \tag{4.2.22}
\]
More precisely, for a fixed positive integer \( N \), \( \beta = \alpha \times A \) so \( \beta^N(x, \xi) = \alpha_{S(\xi^N)}(x) \times A \). Thus,

\[
I_{\beta|\beta^N}(x, \xi) = -\log \frac{(\mu \times P_A)((\alpha(x) \times A) \cap (\alpha_{S(\xi^N)}(x) \times A))}{(\mu \times P_A)(\alpha_{S(\xi^N)}(x) \times A)}
\]

\[
= -\log \frac{\mu(\alpha(x) \cap \alpha_{S(\xi^N)}(x))}{\mu(\alpha_{S(\xi^N)}(x))} = I_{\alpha|\alpha_{S(\xi^N)}}(x).
\]

Equation 4.2.22 follows by the Martingale Convergence Theorem (see Section 3.4 of [Pet]). Now break the set \( A \) into the sets \( A_i \). For \( \xi \in A_i \) and all \( N \),

\[
\mu(\alpha(x) | \alpha_{S(\xi^N)}(x)) \leq \mu(\alpha | \alpha_i^\infty)(x),
\]

(4.2.23)

so

\[
-\log \mu(\alpha(x) | \alpha_{S(\xi^N)}(x)) \geq -\log \mu(\alpha | \alpha_i^\infty)(x).
\]

(4.2.24)

This implies

\[
I_{\beta|\beta^\infty}(x, \xi) = I_{\alpha|\alpha_{S(\xi)}}(x) \geq I_{\alpha|\alpha_i^\infty}(x) \text{ on } A_i.
\]

(4.2.25)

This makes sense since the conditional information function decreases if we condition on \( \alpha_i^\infty \) rather than \( \alpha_{S(\xi)} \), since we gain less information from learning what element of \( \alpha_i^\infty \) a point of \( X \) lies in than if we just know what element of \( \alpha_{S(\xi)} \) it lies in. Also, \( P_A(A_i) = 2^{-i} \) since \( \xi \in A_i \) implies \( \xi_1 = \xi_2 = \cdots = \xi_{i-1} = 0 \) and \( \xi_i = 1 \) and we are using Bernoulli \((1/2, 1/2)\) measure on \( A \). Thus,

\[
P_A(A_i) = 2^{-(i-1)} \cdot 2^{-1} = 2^{-i}.
\]

Therefore,

\[
\sum_{i=1}^{\infty} I_{\alpha|\alpha_{S(\xi)}}(x) \chi_{A_i}(\xi) P_A(A_i) \geq \sum_{i=1}^{\infty} \frac{1}{2^i} I_{\alpha|\alpha_i^\infty}(x) \chi_{A_i}(\xi),
\]

(4.2.26)

so

\[
\sum_{i=1}^{\infty} \int_{A_i} \int_X I_{\beta|\beta^\infty}(x, \xi) d\mu(x) dP_A(\xi) \geq \sum_{i=1}^{\infty} \frac{1}{2^i} \int_{A_i} \int_X I_{\alpha|\alpha_i^\infty}(x) d\mu(x)
\]

(4.2.27)

\[
= \sum_{i=1}^{\infty} \frac{1}{2^i} H_{\mu}(\alpha | \alpha_i^\infty).
\]

(4.2.28)

Combining this with Equation 4.2.3 gives the result. \( \square \)
Corollary 4.2.7. Let \((X, \mathcal{B}, \mu, T)\) be an ergodic measure-preserving system and \(\alpha\) a finite measurable partition of \(X\). Let \(c^S_n = 2^{-n}\) for all \(S \subseteq \mathbb{N}^n\). If \(I_{\alpha|\alpha_{i}^{\infty}}(x) = I_{\alpha|\alpha_{i}^{s}}(x)\) a.e., for all \(i = 1, 2, \ldots\), then

\[
\text{Asc}_\mu(X, \alpha, T) = \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{2^n} H_\mu(\alpha | \alpha_i^{\infty}).
\]  

(4.2.29)

Proof. If \(I_{\alpha|\alpha_{i}^{\infty}}(x) = I_{\alpha|\alpha_{i}^{s}}(x)\) a.e., for all \(i = 1, 2, \ldots\), then on \(A_i\), the inequality in 4.2.25 becomes \(I_{\beta|\beta_{i}^{\infty}}(x, \xi) = I_{\alpha|\alpha_{S_i}(\xi)}(x) = I_{\alpha|\alpha_{i}^{\infty}}(x)\) a.e. The result follows from making the corresponding changes in inequalities 4.2.26 and 4.2.27.

4.3 Analysis of Markov shifts

Recall the definition of a Markov shift, \(\{A^\mathbb{Z}, \mathcal{B}, \mu_{P,p}, \sigma\}\) in Section 1.5.4. \(A = \{0, 1, \ldots, r - 1\}\) is a finite alphabet, \(\mathcal{B}\) is generated by cylinder sets, \(\mu_{P,p}\) is determined by an \(r \times r\) stochastic matrix \(P\) and a probability vector \(p\) fixed by \(P\), and \(\sigma\) is the shift transformation. We will sometimes denote \(\mu_{P,p}\) by \(\mu\) in this section.

We denote the stochastic matrix \(P = (P_{ij})\), where the \(ij\)th entry of \(P\) is the probability of going from state \(i\) to state \(j\). In the case of 1-step Markov measures \(P\) takes the form

\[
P = \begin{pmatrix}
    P_{00} & P_{01} & \cdots & P_{0(r-1)} \\
    P_{10} & P_{11} & \cdots & P_{1(r-1)} \\
    \vdots & \vdots & \ddots & \vdots \\
    P_{(r-1)0} & P_{(r-1)1} & \cdots & P_{(r-1)(r-1)}
\end{pmatrix},
\]

and probability vector \(p = (p_0, p_1, \ldots, p_{r-1})\) such that \(p_i\) is the probability of starting at state \(i\). For a general \(k\)-step Markov measure \(P\) will be an \(r^k \times r^k\) matrix and the states will be \(k\)-blocks. In some cases, whole rows or columns of \(P\) will be 0 and will be left out.

Corollary 4.2.7 applies in the case of Markov shifts where \(\alpha\) is the partition into rank zero cylinder sets \(A_i = \{x \in A^\mathbb{Z} : x_0 = i\}\) because

\[
\mu_{P,p}(x \in A_j \mid x \in T^{-i}A_{k_i} \cap T^{-i-1}A_{k_{i+1}} \cap \cdots) = \mu_{P,p}(x \in A_j \mid x \in T^{-i}A_{k_i}) = p_j(P^i)_{jk_i} \quad (4.3.1)
\]

This fact is clear for Markov shifts since for a sequence \(x\), the probability that \(x_0 = j\) if we know
\( x_i = k \) does not depend on the entries \( x_l \) for \( l > i \). In this case

\[
H_{\mu_P,p}(\alpha \mid \alpha_i^\infty) = -\sum_{j,k=0}^{r-1} p_j (P^i)_{jk} \log (P^i)_{jk}
\]

so

\[
\text{Asc}_{\mu_P,p}(A^\mathbb{Z}, \alpha, \sigma) = -\frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{2^i} \sum_{j,k=0}^{r-1} p_j (P^i)_{jk} \log (P^i)_{jk}.
\] (4.3.2)

Corollary 4.2.7 can be extended to Markov shifts with memories larger than 1 by first reducing them to the equivalent 1-step Markov shift (see Proposition 1.5.16). If \( P \) is the stochastic matrix of the 1-step Markov shift equivalent to a given higher step Markov shift then \( H_{\mu_P,p}(\alpha \mid \alpha_i^\infty) \) becomes more difficult to write in terms of entries of \( P \) than for the case of 1-step Markov shifts. This is because the entries of \( P \) are probabilities of going from 2-block states to 2-block states, but, since \( \alpha \) is the partition by rank zero cylinder sets, to find \( H_{\mu_P,p}(\alpha \mid \alpha_i^\infty) \) we are required to find the probability of going from 2-block states to 1-block states. Denote by \( P_j^{yz} \) the entry of \( P \) representing the probability of going from 2-block state \( j \) to 2-block state \( "yz" \) where \( y, z \in A \) are the 2 symbols that make up the terminal 2-block. In this case

\[
H_{\mu_P,p}(\alpha \mid \alpha_i^\infty) = -\sum_{j \in A^2} \sum_{z \in A} p_j \left( \sum_{y \in A} (P^i)_{jyz} \log \sum_{y \in A} (P^i)_{jyz} \right)
\]

so

\[
\text{Asc}_{\mu_P,p}(A^\mathbb{Z}, \alpha, \sigma) = -\frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{2^i} \sum_{j \in A^2} \sum_{z \in A} p_j \left( \sum_{y \in A} (P^i)_{jyz} \log \sum_{y \in A} (P^i)_{jyz} \right).
\] (4.3.3)

In the following sections we use Equations 4.3.2 and 4.3.3 to compute the measure-theoretic average sample complexity for some examples of Markov shifts. In each example the matrix \( P \) depends on the definition at most 2 probabilities enabling us to plot in either the \([0,1] \times \mathbb{R}\) space or \([0,1] \times [0,1] \times \mathbb{R}\) space these independent probabilities versus measure-theoretic average sample complexity. Similarly, we can make plots of measure-theoretic entropy and measure-theoretic intricacy. By analyzing these graphs and computations we make several conjectures about properties of the measure-theoretic average sample complexity for Markov shifts.

We use Mathematica [WR] to make graphs and compute values. The calculations for measure-theoretic average sample complexity and measure-theoretic intricacy are found by taking the sum
of the first 20 terms of either 4.3.2 or 4.3.3 depending on the case. The measures in the tables
give maximum values for either measure-theoretic entropy, measure-theoretic intricacy, or measure-
theoretic average sample complexity. The bolded numbers in tables are the maxima for the given
category. Tables show computations correct to 3 decimal places. To simplify notation we denote
\( \mu_{P,p} \) by \( \mu \) in this section.

**1-step Markov measures on the full 2-shift**

In this example we consider 1-step Markov measures on the full 2-shift. \( P \) is dependent on two
variables, \( P_{00} \) and \( P_{11} \). \( P \) and \( p \) are given by

\[
P = \begin{pmatrix}
P_{00} & 1 - P_{00} \\
1 - P_{11} & P_{11}
\end{pmatrix}
\quad \text{and} \quad
p = \left( \frac{1 - P_{11}}{2 - P_{00} - P_{11}}, \frac{1 - P_{00}}{2 - P_{00} - P_{11}} \right).
\]

Table 4.1 contains calculations for 1-step Markov measures on the full 2-shift. There are two
measures that maximize \( \text{Int}_\mu \), both of which lie on a boundary plane. We know entropy has a
maximum value of \( \log 2 \) when the measure is Bernoulli. This is also the measure that maximizes
\( \text{Asc}_\mu \) with a value of \( (\log 2)/2 \).

<table>
<thead>
<tr>
<th>( P_{00} )</th>
<th>( P_{11} )</th>
<th>( h_\mu )</th>
<th>( \text{Asc}_\mu )</th>
<th>( \text{Int}_\mu )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.693</td>
<td>0.347</td>
<td>0</td>
</tr>
<tr>
<td>0.216</td>
<td>0</td>
<td>0.292</td>
<td>0.208</td>
<td>0.124</td>
</tr>
<tr>
<td>0</td>
<td>0.216</td>
<td>0.292</td>
<td>0.208</td>
<td>0.124</td>
</tr>
<tr>
<td>0.905</td>
<td>0.905</td>
<td>0.315</td>
<td>0.209</td>
<td>0.104</td>
</tr>
</tbody>
</table>

Table 4.1: 1-step Markov measures on the full 2-shift

The left graph in Figure 4.1 shows \( \text{Asc}_\mu \) for 1-step Markov measures on the full 2-shift. We
observe that this plot is strictly convex and therefore has a unique measure of maximal average
sample complexity occurring when \( P_{00} = P_{11} = 0.5 \). This is the same as the measure of maximal
entropy. The measure-theoretic average sample complexity for this measure on the full 2-shift is
\( (\log 2)/2 \) which is equal to the topological average sample complexity of the full 2-shift with respect
to the cover by rank 0 cylinder sets. The fourth measure shown in Table 4.1 is interesting because it
Figure 4.1: $\text{Asc}_\mu$ and $\text{Int}_\mu$ for 1-step Markov measures on the full 2-shift

is a fully supported local maximum value for $\text{Int}_\mu$. This can be seen in the right graph of Figure 4.1 which shows $\text{Int}_\mu$ for 1-step Markov measures on the full 2-shift. The absolute maxima of $\text{Int}_\mu$ occur in the planes $P_{00} = 0$ and $P_{11} = 0$. The full 2-shift restricted to these planes represent proper subshifts of the full 2-shift isomorphic to the golden mean shift which we discuss in the next example.

Figure 4.2 shows the boundary plane $P_{11} = 0$ for the intricacy in order to better view the maximum.

We also observe that measure-theoretic intricacy is 0 when $P_{00} = 1 - P_{11}$. We prove this using Equation 4.3.2 with the simplified matrix and fixed vector

$$P = \begin{pmatrix} P_{00} & 1 - P_{00} \\ P_{00} & 1 - P_{00} \end{pmatrix} \quad \text{and} \quad p = (P_{00}, 1 - P_{00}).$$

We show $2\text{Asc}_\mu = h_\mu$ and thus $\text{Int}_\mu = 0$. Since $P^i = P$ for all $i = 1, 2, \ldots$, and

$$\sum_{j,k=0}^{1} p_j (P^i)_{jk} \log(P^i)_{jk} = P_{00} \log P_{00} + (1 - P_{00}) \log(1 - P_{00}) = -h_{\mu_{P,P}}(\sigma),$$

we have

$$\text{Asc}_{\mu_{P,P}}(A^Z, \alpha, \sigma) = -\frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{2^i} \sum_{j,k=0}^{1} p_j (P^i)_{jk} \log(P^i)_{jk} = \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{2^i} h_{\mu_{P,P}}(\sigma) = \frac{1}{2} h_{\mu_{P,P}}(\sigma).$$

Figure 4.3 shows the graph of $h_\mu$ on the left and a combined plot on the right which in order from top to bottom show $h_\mu$, $\text{Asc}_\mu$, and $\text{Int}_\mu$. Each graph is symmetric about the plane $P_{00} = P_{11}$.

By the variational principle we know the measure of maximal entropy occurs when $P_{00} = P_{11} = 0.5$ and has a value of $\log 2$. Analysis of the graphs of $\text{Asc}_\mu$ and $\text{Int}_\mu$ for 1-step Markov measures on the full 2-shift lead to the following conjectures.
Figure 4.2: \( \text{Int}_\mu \) for 1-step Markov measures on the full 2-shift with \( P_{11} = 0 \)

Figure 4.3: \( h_\mu \) for 1-step Markov measures on the full 2-shift

**Conjecture 4.3.1.** For each \( k \geq 1 \), there is a unique \( k \)-step Markov measure \( \mu_k \) on the full 2-shift that maximizes \( \text{Asc}_\mu \) among all \( k \)-step Markov measures.

We base this conjecture on the observation of convexity in the graph of \( \text{Asc}_\mu \) for 1-step Markov measures on the full 2-shift.

**Conjecture 4.3.2.** For each \( k \geq 1 \), there are two \( k \)-step Markov measures on the full 2-shift that maximize \( \text{Int}_\mu \) among all \( k \)-step Markov measures. They are not fully supported.

**Conjecture 4.3.3.** For each \( k \geq 1 \), there is a \( k \)-step Markov measure on the full 2-shift that gives a fully supported local maximum for \( \text{Int}_\mu \) among all \( k \)-step Markov measures.
Table 4.2: 1-step Markov measures on the golden mean shift

<table>
<thead>
<tr>
<th>$P_{00}$</th>
<th>$h_{\mu}$</th>
<th>Asc$_{\mu}$</th>
<th>Int$_{\mu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.618</td>
<td><strong>0.481</strong></td>
<td>0.266</td>
<td>0.051</td>
</tr>
<tr>
<td>0.533</td>
<td>0.471</td>
<td><strong>0.271</strong></td>
<td>0.071</td>
</tr>
<tr>
<td>0.216</td>
<td>0.292</td>
<td>0.208</td>
<td><strong>0.124</strong></td>
</tr>
</tbody>
</table>

1-step Markov measures on the golden mean shift

The next example we discuss is 1-step Markov measures on the golden mean shift. In this case $P$ and $p$ depend on a single entry for which we choose $P_{00}$, the probability of going from 0 to 0. Then

$$P = \begin{pmatrix} P_{00} & 1 - P_{00} \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad p = \left( \frac{1}{2 - P_{00}}, 1 - \frac{1}{2 - P_{00}} \right).$$

By the variational principle, we know that the measure of maximal entropy occurs when $P_{00} = 1/\phi$ where $\phi$ is the golden mean (see Example 1.5.28), and the measure-theoretic entropy for this measure is $h_{\mu_{P,p}}(\sigma) = \log \phi$.

Table 4.2 contains calculations for different 1-step Markov measures on the golden mean shift.

Figure 4.4 includes two graphs for calculations for 1-step Markov measures on the golden mean shift with $P_{00}$ as the horizontal axis. The graph on the left includes six curves. Five curves are plots of the measure-theoretic average sample complexity function of $n$ for $n = 2, \ldots, 6$ computed using Definition 4.1.2. The sixth is a plot using Equation 4.3.2. This graph shows that the average sample complexity functions on $n$ quickly approach Asc$_{\mu}$. As $P_{00}$ approaches 1, the functions become better approximations for Asc$_{\mu}$.

The graph on the right has curves of $h_{\mu}$, Asc$_{\mu}$ and Int$_{\mu}$ found using Equation 4.3.2. Circles mark what appear to be the unique maxima of each curve. The maxima among 1-step Markov measures of Asc$_{\mu}$, Int$_{\mu}$, and $h_{\mu}$ all seem to be achieved by different measures $\mu$. 
Figure 4.4: 1-step Markov measures on the golden mean shift

Table 4.3: 2-step Markov measures on the golden mean shift

<table>
<thead>
<tr>
<th>$P_{000}$</th>
<th>$P_{100}$</th>
<th>$h_{\mu}$</th>
<th>$\text{Asc}_{\mu}$</th>
<th>$\text{Int}_{\mu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.618</td>
<td>0.618</td>
<td>0.481</td>
<td>0.266</td>
<td>0.051</td>
</tr>
<tr>
<td>0.483</td>
<td>0.569</td>
<td>0.466</td>
<td>0.272</td>
<td>0.078</td>
</tr>
<tr>
<td>0</td>
<td>0.275</td>
<td>0.344</td>
<td>0.221</td>
<td>0.167</td>
</tr>
</tbody>
</table>

2-step Markov measures on the golden mean shift

Here we consider 2-step Markov measures on the golden mean shift. In this case we have two parameters. We let $P_{000}$ and $P_{100}$ be the probability of going from 00 to 00 and from 10 to 00 respectively. $P$ and $p$ are given by

$$P = \begin{pmatrix} P_{000} & 1 - P_{000} & 0 \\ 0 & 0 & 1 \\ P_{100} & 1 - P_{100} & 0 \end{pmatrix}$$

and

$$p = \left( -\frac{2P_{000} - P_{100}}{2P_{000} - P_{100} - 2}, \frac{P_{100}}{2(2P_{000} - P_{100} - 2)} + 0.5, \frac{P_{000}}{2(2P_{000} - P_{100} - 2)} + 0.5 \right)$$

Table 4.3 and the plots in Figures 4.5 and 4.6 are similar to those in the previous examples. As expected, the maximal $h_{\mu}$ is $\log \phi$ as it was for 1-step Markov measures on the golden mean shift.

The graph of $\text{Asc}_{\mu}$ as a function of the parameters of 2-step Markov shifts appears strictly convex, as was the case for 1-step Markov measures on the full 2-shift; this gives evidence for the
existence of a unique maximizing measure. The maximum for $\text{Int}_\mu$ is not fully supported and occurs on the plane $P_{000} = 1$. The maximum values of both $\text{Asc}_\mu$ and $\text{Int}_\mu$ strictly increase as we go from 1-step Markov measures on the golden mean shift to 2-step Markov measures on the golden mean shift. There is no reason to expect that these values will not continue to increase as we move to higher $k$-step Markov measures on the golden mean shift, leading to the following conjectures.

**Conjecture 4.3.4.** For each $k \geq 1$, there is a unique $k$-step Markov measure $\mu_k$ on the golden mean shift that maximizes $\text{Asc}_\mu$ among all $k$-step Markov measures. Furthermore, if $k_1 \neq k_2$ then $\text{Asc}_{\mu_{k_1}} \neq \text{Asc}_{\mu_{k_2}}$.

**Conjecture 4.3.5.** On the golden mean shift there is a measure of maximum $\text{Asc}_\mu$ that is not Markov of any order.
CHAPTER 5

Future directions

In this chapter we discuss possible future directions for research on intricacy and average sample complexity.

5.1 Improved computational methods and formulas

Equation 3.1.13 gives us a formula for computing the average sample complexity of rank zero cylinder sets for certain shifts of finite type and the uniform system of coefficients. We would like to find a formula to calculate average sample complexity for all shifts of finite type or at least ways aside from brute force of getting good approximations for average sample complexity.

As of now, the programs we have written to find the average sample complexity of a shift of finite type do the computations by going through every subset $S \subset n^*$ and every sequence in a shift. This requires a lot of resources and even for small $n$ ($n = 10$), these calculations take several minutes. The calculation times grow exponentially from here since they grow on the order of the number of subsets of $n^* = \{0, 1, \ldots, n - 1\}$. We would like to write more efficient codes or use theory to find formulas to make approximations easier and faster. One idea is to take a random sample of subsets $S \subset n^*$ of a certain size for each $n$ instead of doing calculations on all subsets.

We would also like to be able to calculate average sample complexity for other systems of coefficients. We have plenty of data and know many properties of average sample complexity and intricacy when $c^n_S = 2^{-n}$. Using the uniform system of coefficients allows us to pull $c^n_S$ out of the sum over all subsets $S \subset n^*$ and greatly simplifies the definition. We would like to find properties of average sample complexity and intricacy using the neural complexity system of coefficients $c^n_S = 1/[(n + 1)C(n, s)]$ as well as general systems of coefficients given in Theorem 1.4.5. See the next section for more on this.
5.2 General weights

We proved many results for average sample complexity and intricacy (of both topological dynamical systems and measure-theoretic dynamical systems) with the fixed system of coefficients $c^n_S = 2^{-n}$. We would like to prove these results for a general system of coefficients or other general weights.

We think many results that we have proved only for $c^n_s = 2^{-n}$ can be extended to a general weighting that does not satisfy all criteria for a system of coefficients, but may still have interesting properties. For example, Theorem 4.2.4 can be altered by replacing the Bernoulli measure $\mathcal{B}(1/2, 1/2)$, $\mathcal{P}$, that we use to select random subsets $S \subset n^*$ with the Bernoulli measure $\mathcal{B}(p, 1 - p)$ for $0 < p < 1$. Then Theorem 1.4.5 could provide the basis for extending results to general weights, not necessarily satisfying the conditions of Definition 1.4.1, via integration with respect to an appropriate measure ($\mathcal{P}$ restricted to $A$ and normalized). For a subset $S \subset n^*$, these weights do not satisfy the property of being equal on $S$ and $S^c$, so they will not define a system of coefficients.

One could also try to prove the results directly for a general system of coefficients. For example in Theorem 4.2.4 we could replace Bernoulli measure on $\Sigma_2^+$ by an arbitrary shift-invariant ergodic measure. In this case, intricacy weights would be obtained by requiring that the measure be invariant also under the involution that switches 0’s and 1’s.

5.3 Further analysis of shifts of finite type

Suppose a shift of finite type, $X$, has square positive adjacency matrix. We know that the intricacy of $X$ with respect to rank 0 cylinder sets using the uniform system of coefficients depends only on $|\mathcal{L}_n(X)|$, i.e. its complexity function (see Theorem 3.1.13). Therefore, we can conclude that two shifts of finite type with square positive adjacency matrices and the same complexity functions have the same intricacy (and intricacy functions).

We have examples of shifts of finite type with the same complexity functions but different intricacy functions. In these examples the smallest power for which the adjacency matrices are positive differ. We would like to find conditions for which two shifts of finite type with the same complexity functions have the same intricacy functions. We suspect that two shifts of finite type will have the same intricacy functions (with respect to rank 0 cylinder sets and the uniform system
of coefficients) exactly when they have the same complexity functions and the same smallest power for which their adjacency matrices are positive.

5.4 Higher-dimensional shifts

For a finite alphabet \( A = \{0, 1, \ldots, n - 1\} \), the \( d \)-dimensional full \( A \)-shift is defined to be \( A^\mathbb{Z}^d \). An element \( x \in A^\mathbb{Z}^d \) of the full shift may be regarded as a function \( x : \mathbb{Z}^d \to A \), or, more informally, as a “configuration” of alphabet choices at the sites of the integer lattice \( \mathbb{Z}^d \).

If \( \vec{v} \in \mathbb{Z}^d \), let \( x_{\vec{v}} \) denote the symbol at position \( \vec{v} \) in \( x \). Let \( \sigma_d : A^\mathbb{Z}^d : \times \mathbb{Z}^d \to A^\mathbb{Z}^d \) be defined by

\[
\sigma_d(x, \vec{v})_{\vec{w}} = x_{\vec{v} + \vec{w}}
\]

for all \( \vec{v}, \vec{w} \in \mathbb{Z}^d \) and \( x \in A^\mathbb{Z}^d \). We call \( \sigma_d \) the \( d \)-dimensional shift and \((A^\mathbb{Z}^d, \sigma_d)\) the \( d \)-dimensional full \( n \)-shift. The usual metric on the one-dimensional full shift naturally generalizes to a metric \( \rho \) on \( A^\mathbb{Z}^d \) given by

\[
\rho(x, y) = 2^{-k},
\]

where \( k \) is the largest integer such that \( x_{[-k,k]^d} = y_{[-k,k]^d} \). According to this definition, two points are close if they agree on a large \( d \)-cube \( \{-k, \ldots, k\}^d \).

A \( d \)-dimensional shift space is a closed (with respect to the metric \( \rho \)) \( \sigma_d \)-invariant subset of \( A^\mathbb{Z}^d \).

There is an equivalent definition of a shift of finite type in higher dimensions. A shape is a finite subset \( F \) of \( \mathbb{Z}^d \). A pattern \( f \) on a shape \( F \) is a function \( f : F \to A \). Given a list \( \mathcal{F} \) of patterns, put

\[\begin{align*}
X = X_{\mathcal{F}} &= \{x \in A^\mathbb{Z}^d : \sigma_d(x, \vec{v})_F \notin \mathcal{F} \text{ for all } \vec{v} \in \mathbb{Z}^d \text{ and all shapes } F\}.
\end{align*}\]

We say that a pattern \( f \) on a shape \( F \) occurs in a shift space \( X \) if there is an \( x \in X \) such that \( x_{\mathcal{F}} = f \). Hence the analogue of the language of a shift space is the set of all occurring patterns. A \( d \)-dimensional shift of finite type \( X \) is a subset of \( A^\mathbb{Z}^d \) defined by a finite list \( \mathcal{F} \) of forbidden patterns. Analogues of average sample complexity and intricacy in higher-dimensional subshifts would look at patterns that occur at certain places of \((n^*)^d = \{0, 1, \ldots, n - 1\}^d\) for all legal patterns of the \( d \)-dimensional shift space.
We have seen that when $d = 1$ entropies of many shift spaces, in particular certain shifts of finite type, are easily calculated. However, when $d > 1$, although there are methods for obtaining entropy estimates for some shifts of finite type, it is usually not feasible to compute entropy directly. This decreases the hope for finding a way of easily calculating the average sample complexity and intricacy of a higher-dimensional shift of finite type. Nonetheless, since even the entropy of higher-dimensional shifts is not always known, finding the average sample complexity and intricacy of them should be a worthwhile pursuit as it may illuminate properties that were previously unknown.

5.5 Maximizing subsets $S \subset n^*$

For a topological system $(X, T)$ and a cover $\mathcal{U}$ of $X$, for each $n \geq 1$ we would like to find the subset(s) $S \subset n^*$ that maximize $\log N(\mathcal{U}_S)$ and $\log (N(\mathcal{U}_S)N(\mathcal{U}_{S^c})/N(\mathcal{U}_{n^*}))$. For shifts of finite type with positive square adjacency matrix and covers by rank zero cylinder sets we can show that $\log N(\mathcal{U}_S)$ is maximized for the subset $S \subset n^*$, $S = \{0, 2, 4, 6, \ldots, n - 1\}$ for $n$ even and $S = \{0, 2, \ldots, n - 2\}$ or $S = \{1, 3, \ldots, n - 1\}$ for $n$ odd. This is a consequence of Proposition 3.1.1.

Similarly, for a measure-preserving system $(X, \mathcal{B}, \mu, T)$ and partition $\alpha$ of $X$, for each $n \geq 1$ we would like to find the subset(s) $S \subset n^*$ that maximize $H(\alpha_S)$ and $H(\alpha_S) + H(\alpha_{S^c}) - H(\alpha_{n^*})$. Finding maximizing subsets $S \subset n^*$ could help improve computational methods for finding average sample complexity and intricacy by allowing us to focus the calculations on subsets that have the greatest effect on the values of the measurements. We could also gain a deeper understanding of a system and a cover or partition of that system if we can find the maximizing subset(s).

5.6 Analysis of more examples

Most of the analysis of examples done so far has been for shifts of finite type over 2 and 3 element alphabets and Markov shifts formed by putting Markov measures on shifts of finite type. We would like to analyze more examples of systems including shifts of finite type on larger alphabets and other subshifts formed by taking the orbit closure of a sequence on a finite alphabet (substitution, Sturmian). We would also like to study $\text{Asc}$ and $\text{Int}$ for sofic shifts. Informally a sofic shift is a subshift formed by first labelling the edges in a graph $\mathcal{G}$ with (possibly not pairwise distinct) symbols from an alphabet $A$. By reading the labels of the edges traversed in a bi-infinite walk on $\mathcal{G}$
we get a point in the full $\mathcal{A}$-shift. The set of these points is a sofic shift with presentation $\mathcal{G}$.

Every sofic shift is a subshift and every shift of finite type is a sofic shift, but there are sofic shifts which are not shifts of finite type. For example the set of sequences over the alphabet $\{0, 1\}$ such that between any two 1s there are an even number of 0s is called the even shift. This is a sofic shift with presentation given by the following graph.

![Graph of the even shift]

The even shift is not a shift of finite type because the set of forbidden blocks is $\{10^{2n-1}1 : n \geq 0\}$.

5.7 **Entropy is the only finitely observable invariant**

In [OW], D. Ornstein and B. Weiss show that any finitely observable measure-theoretic isomorphism invariant is necessarily a continuous function of entropy. (A function $J$ with values in some metric space, defined for all finite-valued, stationary, ergodic processes is said to be *finitely observable* if there is a sequence of functions $S_n(x_1, \ldots, x_n)$ that for all processes $\mathcal{X}$ converges to $J(\mathcal{X})$ for almost every realization $x_1^\infty$ of $\mathcal{X}$.)

Considering $\sup_{\mathcal{U}} \text{Asc}(X, \mathcal{U}, T) = \sup_{\mathcal{U}} \text{Int}(X, \mathcal{U}, T) = h_{\text{top}}(X, T)$ and $\sup_{\alpha} \text{Asc}_\mu(X, \alpha, T) = \sup_{\alpha} \text{Int}(X, \alpha, T) = h_\mu(X, T)$, the work of Ornstein and Weiss may apply to our new measures of complexity. If we can show that $\sup_{\alpha} \text{Asc}_\mu(X, \alpha, T)$ and $\sup_{\alpha} \text{Int}(X, \alpha, T)$ are finitely observable invariants without assuming they are equal to entropy, then Theorem 4.2.1, which proves these qualities are equal to measure-theoretic entropy, becomes less surprising. We would also like to explore a topological version of the work of Ornstein and Weiss in view of our findings that taking a supremum over open covers of Asc and Int yields the usual topological entropy.

5.8 **Alternate definition of the average sample complexity function analogous to the complexity function of a sequence**

Recall that for a sequence $u$ with symbols in a finite alphabet $\mathcal{A}$, the complexity function $p_u(n) = |\mathcal{L}_n(u)|$ gives the number of words of length $n$ which occur in $u$. Based on this function, we give an alternate definition of the average sample complexity function of the subshift $(\mathcal{O}(u), \sigma)$. 105
Definition 5.8.1. Let $u$ be a sequence of elements from a finite alphabet $\mathcal{A}$. For each subset $S \subseteq n^*$ let $p_u(S)$ denote the number of words seen at the places in $S$ for all words of length $n$ in the sequence $u$. Then we define the average sequence complexity function of the sequence $u$ by

$$\text{Alt}_u(n) = \frac{1}{2^n} \sum_{S \subseteq n^*} p_u(S).$$ (5.8.1)

Notice that $p_u(n^*) = p_u(n)$.

There has been a lot of analysis of the complexity function for sequences and it appears in many applications. $\text{Alt}_u$ is a finer measurement that may be found useful in the same applications as the complexity function. We would like to study properties of $\text{Alt}_u$ for different sequences and compare it to the complexity function. There is possibly an analogue of Sturmian sequences for $\text{Alt}_u$, i.e., aperiodic sequences with minimum $\text{Alt}_u$.

5.9 Complexity of finite words

If a sequence $u \in \{0, 1, \ldots, r - 1\}^\mathbb{Z}$ is such that $\overline{O}(u) = \Sigma_r$, the full $r$-shift, then computing $\text{Alt}_u(n)$ requires one to average the number of words seen at the places in all subsets $S \subseteq n^*$ for each of the $r^n$ words of length $n$ in $\Sigma_r$. This computation leads to a complexity measurement for each word of length $n$ with symbols in $\{0, 1, \ldots, r - 1\}$.

Definition 5.9.1. Fix a positive integer $n$ and let $w \in \mathcal{L}_n(\Sigma_r)$, a word of length $n$ made of elements in $\{0, 1, \ldots, r - 1\}$. For each $0 \leq k \leq n$ let $q_w(k)$ denote the number of words of length $k$ in $w$ seen at the places in each subset $S \subseteq n^*$ such that $|S| = k$. We define the average sample word complexity of $w$ by

$$\text{Asw}(w) = \frac{1}{2^n} \sum_{k=0}^{n} q_w(k).$$ (5.9.1)

Notice that for a sequence $u$ such that $\overline{O}(u) = \Sigma_r$,

$$\text{Alt}_u(n) = \frac{1}{2^n} \sum_{w \in \mathcal{L}_n(\Sigma_r)} \sum_{k=0}^{n} q_w(k) = \sum_{w \in \mathcal{L}_n(\Sigma_r)} \text{Asw}(w).$$

For a fixed $n$, by computing $\text{Asw}(w)$ for all $w \in \mathcal{L}_n(\Sigma_r)$ we can rank the words of length $n$ by their complexity based on this definition. Words that are very uniform or repetitive will have lower complexity.
average sample word complexity than words that are less uniform or repetitive. For example, if we compute the average sample word complexity for all words of length 5 in $\Sigma_2$, then uniform and repetitive words such as 11111 and 10101 have lower average sample word complexity than words appearing more random, such as 10011 and 00110.

We would like to compute the average sample word complexity of (long) finite words of a fixed length in order to rank them by complexity. We can then compare this ranking to other known measurements of complexity of finite words.

5.10 Partition $n^*$ into $m$ subsets

Our definition for intricacy in dynamical systems involves partitioning the set $n^*$ into a subset $S$ and its complement $S^c$. We could also consider partitioning $n^*$ into more than two subsets. Let $S_1, S_2, \ldots, S_m$ be disjoint subsets whose union is $n^*$. Let $\mathcal{S}(m) \subset n^*$ denote the set of all partitions of $n^*$ into $m$ subsets. For a general definition we need to redefine the weighting factor. Denote by $c_{\mathcal{S}(m)}$ this new weighting factor depending on the partition $\mathcal{S}(m)$. Let $(X, T)$ be a topological dynamical system and $\mathcal{U}$ an open cover of $X$ and consider the $m$-intricacy of $X$ with respect to $\mathcal{U}$

$$m \text{- Intr}(X, \mathcal{U}, T) = \lim_{n \to \infty} \frac{1}{n} \sum_{\mathcal{S}(m) \subset n^*} c_{\mathcal{S}(m)} \log \left( \frac{\prod_{S_i \in \mathcal{S}(m)} N(\mathcal{U} S_i)}{N(\mathcal{U} n^*)} \right)$$

(5.10.1)

The analogous generalization to intricacy functionals is proposed by Buzzi and Zambotti in [BZ].

5.11 Definition based on Rokhlin entropy

In [Sew], Rokhlin entropy is defined as the infimum of the measure-theoretic entropies of countable generating partitions. More specifically, let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-theoretic dynamical system and define the Rokhlin entropy

$$h_{\mu}^{\text{Rok}}(X, T) = \inf_{\alpha} \{H_{\mu}(\alpha) : \alpha \text{ is a countable generating partition}\}.$$
This isomorphism invariant is a close analogue of measure-theoretic entropy. We may consider the measure-theoretic Rokhlin average sample complexity based on Rokhlin entropy,

\[ \text{Asc}_\mu^{\text{Rok}}(X, T) = \inf \{ \text{Asc}_\mu(X, \alpha, T) : \alpha \text{ is a countable generating partition} \} , \]

as well as the topological version for a topological dynamical system \((X, T)\),

\[ \text{Asc}^{\text{Rok}}(X, T) = \inf \{ \text{Asc}(X, \mathcal{U}, T) : \mathcal{U} \text{ is a topological generator} \} . \]

(We thank Tomasz Downarowicz for the suggestion to examine these definitions.)

5.12 Application of topological average sample pressure to coding sequence density

In [KT2], Koslicki and Thompson give a new approach to coding sequence density estimation in genomic analysis based on topological pressure. The structure and organization of genomes is important in the study of genome biology. They use topological pressure as a computational tool for predicting the distribution of coding sequences and identifying gene-rich regions.

In their study, they consider finite sequences on the alphabet \(\{A, C, G, T\}\) and weight each word of length 3. The choice of symbols and block length have biological significance and come from standards in coding DNA sequences. They compute the topological pressure as one would for a 3-block coding on the full 4-shift. The weighting function (potential function) is found by training parameters so that the topological pressure fits the observed coding sequence density on the human genome.

We would like to make similar computations as Koslicki and Thompson but replace topological pressure with topological average sample pressure. These finer measurements could possibly be used to better predict coding sequence density and understand the structure of genomes.

5.13 Maximal measures, variational principle, and equilibrium states

In Sections 1.2.3 and 1.2.4 we discussed measures of maximal entropy, the variational principle for entropy, and equilibrium states pertaining to topological pressure. We have alluded to the
analyses of some of these concepts for average sample complexity and intricacy (see Corollary 4.2.5
and the discussion in Section 4.3). In this section we discuss some further questions regarding these
concepts.

Corollary 4.2.5 shows the existence of measures that maximize $\text{Asc}_\mu(X, \alpha, T)$. We still have
questions about the uniqueness of such measures. The definition of measure-theoretic intricacy
makes proving the existence of measures of maximal $\text{Int}_\mu(X, \alpha, T)$ more complicated (due to the
fact that we subtract $H_\mu(\alpha_n^*)$). We would like to explore further which measure(s) maximize $\text{Asc}_\mu$
and $\text{Int}_\mu$.

In Section 4.3 we examine the measure-theoretic average sample complexity and measure-
theoretic intricacy of 1-step and 2-step Markov measures on the golden mean shift of finite type. We
observe that there exist measures of maximal $\text{Asc}_\mu$ and $\text{Int}_\mu$ among all 1-step and 2-step Markov
measures for these examples. The maximum values change as we go from 1-step to 2-step Markov
measures. This is evidence that the maximizing measures for $\text{Asc}_\mu$ and $\text{Int}_\mu$ are not Markov of
any order. We would like to prove this as well as the existence of a fully supported 2-step Markov
measure on the full 2-shift for which $\text{Int}_\mu$ attains a local maximum (see Figure 4.1). Furthermore,
we want to study Markov measures on more shifts of finite type to find more evidence of maximizing
measures and see whether local maximizing measures appear in other examples.

We would like to examine analogues of the variational principle for average sample complexity
and intricacy and existence and uniqueness of equilibrium states for average sample pressure. An
analogue of the variational principle for average sample complexity would say that for a subshift
$(X, T)$ with partition $\alpha$ into rank 0 cylinder sets (and corresponding cover $\mathcal{U}(\alpha)$),

$$\sup_{\mu} \text{Asc}_\mu(X, \alpha, T) = \text{Asc}(X, \mathcal{U}(\alpha), T).$$

We can numerically approximate measures that maximize $\text{Asc}_\mu$ for $k$-step Markov measures on
shifts of finite type. Given a shift of finite type, we suspect that the maximum value of $\text{Asc}_\mu$ will
increase as $k$ increases and the measures that achieve $\sup_{\mu} \text{Asc}_\mu(X, \alpha, T)$ will not be Markov of
any order. As $k$ gets large, computing $\text{Asc}_\mu$ for $k$-step Markov measures becomes very complicated. We
would like to do more computations to get numerical evidence to confirm these conjectures as well
as prove them.
## APPENDIX A
### TABLES

#### A.1 Calculations for shifts of finite type using the uniform system of coefficients

<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>$H(10)$</th>
<th>Asc(10)</th>
<th>Int(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\begin{pmatrix} 0 &amp; 1 \ 1 &amp; 0 \end{pmatrix}$</td>
<td>$\infty$</td>
<td>![Graph Image]</td>
<td>0.000</td>
<td>0.069</td>
<td>0.069</td>
<td>0.069</td>
</tr>
<tr>
<td>2</td>
<td>$\begin{pmatrix} 1 &amp; 1 \ 1 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td>![Graph Image]</td>
<td>0.481</td>
<td>0.497</td>
<td>0.291</td>
<td>0.085</td>
</tr>
<tr>
<td>3</td>
<td>$\begin{pmatrix} 1 &amp; 1 \ 1 &amp; 1 \end{pmatrix}$</td>
<td>1</td>
<td>![Graph Image]</td>
<td>0.693</td>
<td>0.693</td>
<td>0.347</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>$H(10)$</th>
<th>Asc(10)</th>
<th>Int(10)</th>
</tr>
</thead>
</table>
| 3   | \[
\begin{pmatrix}
1 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 0
\end{pmatrix}
\] | 3 | ![Graph 3](image3) | 0.562 | 0.614 | 0.416 | 0.218 |
| 4   | \[
\begin{pmatrix}
1 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1
\end{pmatrix}
\] | 2 | ![Graph 4](image4) | 0.693 | 0.734 | 0.458 | 0.182 |
| 5   | \[
\begin{pmatrix}
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 0
\end{pmatrix}
\] | 2 | ![Graph 5](image5) | 0.693 | 0.734 | 0.458 | 0.182 |
| 6   | \[
\begin{pmatrix}
1 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 0
\end{pmatrix}
\] | 2 | ![Graph 6](image6) | 0.881 | 0.900 | 0.496 | 0.093 |
| 7   | \[
\begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 0
\end{pmatrix}
\] | 2 | ![Graph 7](image7) | 1.010 | 1.010 | 0.526 | 0.039 |

Continued on next page
<table>
<thead>
<tr>
<th>SFT</th>
<th>( M )</th>
<th>( \rho(M) )</th>
<th>Graph</th>
<th>Entropy</th>
<th>( H(10) )</th>
<th>Asc(10)</th>
<th>Int(10)</th>
</tr>
</thead>
</table>
| 8   | \[
\begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1 \\
\end{pmatrix}
\] | 1   | ![Graph 8](image) | 1.100  | 1.100  | 0.549  | 0.000  |
| 9   | \[
\begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 1 & 0 \\
\end{pmatrix}
\] | 5   | ![Graph 9](image) | 0.281  | 0.361  | 0.297  | 0.234  |
| 10  | \[
\begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 1 & 1 \\
\end{pmatrix}
\] | 3   | ![Graph 10](image) | 0.609  | 0.648  | 0.409  | 0.171  |
| 11  | \[
\begin{pmatrix}
0 & 1 & 0 \\
0 & 1 & 1 \\
1 & 1 & 0 \\
\end{pmatrix}
\] | 3   | ![Graph 11](image) | 0.609  | 0.648  | 0.409  | 0.171  |
| 12  | \[
\begin{pmatrix}
0 & 1 & 0 \\
0 & 1 & 1 \\
1 & 1 & 1 \\
\end{pmatrix}
\] | 3   | ![Graph 12](image) | 0.791  | 0.811  | 0.460  | 0.110  |
| 13  | \[
\begin{pmatrix}
1 & 1 & 0 \\
0 & 0 & 1 \\
1 & 1 & 0 \\
\end{pmatrix}
\] | 3   | ![Graph 13](image) | 0.481  | 0.545  | 0.399  | 0.254  |
<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>$H(10)$</th>
<th>Asc(10)</th>
<th>Int(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 0 \ 0 &amp; 0 &amp; 1 \ 1 &amp; 1 &amp; 1 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image1" alt="Graph 1" /></td>
<td>$0.693$</td>
<td>$0.734$</td>
<td>$0.458$</td>
<td>$0.182$</td>
</tr>
<tr>
<td>15</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 0 \ 0 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image2" alt="Graph 2" /></td>
<td>$0.693$</td>
<td>$0.734$</td>
<td>$0.458$</td>
<td>$0.182$</td>
</tr>
<tr>
<td>16</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 0 \ 0 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image3" alt="Graph 3" /></td>
<td>$0.844$</td>
<td>$0.867$</td>
<td>$0.491$</td>
<td>$0.115$</td>
</tr>
<tr>
<td>17</td>
<td>$\begin{pmatrix} 0 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \ 1 &amp; 0 &amp; 0 \end{pmatrix}$</td>
<td>4</td>
<td><img src="image4" alt="Graph 4" /></td>
<td>$0.481$</td>
<td>$0.545$</td>
<td>$0.377$</td>
<td>$0.208$</td>
</tr>
<tr>
<td>18</td>
<td>$\begin{pmatrix} 0 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \ 1 &amp; 0 &amp; 1 \end{pmatrix}$</td>
<td>3</td>
<td><img src="image5" alt="Graph 5" /></td>
<td>$0.693$</td>
<td>$0.734$</td>
<td>$0.446$</td>
<td>$0.158$</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>$H(10)$</th>
<th>Asc(10)</th>
<th>Int(10)</th>
</tr>
</thead>
</table>
| 19  | \[
|     |    | \[
|     |    | \[
|     |    | \[
| 20  | \[
|     | \[
|     | \[
|     | \[
| 21  | \[
|     | \[
|     | \[
|     | \[
| 22  | \[
|     | \[
|     | \[
|     | \[
| 23  | \[
|     | \[
|     | \[
|     | \[

Continued on next page
<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>$H(10)$</th>
<th>Asc(10)</th>
<th>Int(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image1" alt="Graph" /></td>
<td>0.962</td>
<td>0.978</td>
<td>0.521</td>
<td>0.064</td>
</tr>
<tr>
<td>25</td>
<td>$\begin{pmatrix} 0 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 0 \ 1 &amp; 0 &amp; 0 \end{pmatrix}$</td>
<td>$\infty$</td>
<td><img src="image2" alt="Graph" /></td>
<td>0.347</td>
<td>0.416</td>
<td>0.256</td>
<td>0.096</td>
</tr>
<tr>
<td>26</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 0 \ 1 &amp; 0 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image3" alt="Graph" /></td>
<td>0.693</td>
<td>0.722</td>
<td>0.440</td>
<td>0.158</td>
</tr>
<tr>
<td>27</td>
<td>$\begin{pmatrix} 0 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 \ 1 &amp; 0 &amp; 0 \end{pmatrix}$</td>
<td>4</td>
<td><img src="image4" alt="Graph" /></td>
<td>0.589</td>
<td>0.631</td>
<td>0.400</td>
<td>0.170</td>
</tr>
<tr>
<td>28</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 \ 1 &amp; 0 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image5" alt="Graph" /></td>
<td>0.810</td>
<td>0.830</td>
<td>0.472</td>
<td>0.114</td>
</tr>
</tbody>
</table>
Continued from previous page

<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>$H(10)$</th>
<th>Asc$(10)$</th>
<th>Int$(10)$</th>
</tr>
</thead>
</table>
| 29  | \[
\begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 0 \\
1 & 0 & 1
\end{pmatrix}
\] | 2 | 0 | 0.881 0.900 0.496 0.093 |

A.2 Shifts of finite type with positive square adjacency matrices

The adjacency matrix, $M$, for each of the shifts of finite type in the following table satisfies $M^2 > 0$ so we use Equation 3.1.13 to find the exact values of Asc and Int.

<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>Asc$(X, \sigma)$</th>
<th>Int$(X, \sigma)$</th>
</tr>
</thead>
</table>
| 2   | \[
\begin{pmatrix}
1 & 1 \\
1 & 0 \\
1 & 1
\end{pmatrix}
\] | 2 | 0 | 0.481 0.286 0.091 |
| 3   | \[
\begin{pmatrix}
1 & 1 \\
1 & 1
\end{pmatrix}
\] | 1 | 0 | 0.693 0.347 0 |

<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>Asc$(X, \sigma)$</th>
<th>Int$(X, \sigma)$</th>
</tr>
</thead>
</table>
| 4   | \[
\begin{pmatrix}
1 & 1 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1
\end{pmatrix}
\] | 2 | 0 | 0.693 0.448 0.203 |

Continued on next page
<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>$\text{Asc}(X, \sigma)$</th>
<th>$\text{Int}(X, \sigma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$\begin{pmatrix} 0 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \ 1 &amp; 1 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image1.png" alt="Graph 1" /></td>
<td>0.693</td>
<td>0.448</td>
<td>0.203</td>
</tr>
<tr>
<td>6</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \ 1 &amp; 1 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image2.png" alt="Graph 2" /></td>
<td>0.881</td>
<td>0.491</td>
<td>0.101</td>
</tr>
<tr>
<td>7</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image3.png" alt="Graph 3" /></td>
<td>1.010</td>
<td>0.523</td>
<td>0.036</td>
</tr>
<tr>
<td>14</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 0 \ 0 &amp; 0 &amp; 1 \ 1 &amp; 1 &amp; 1 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image4.png" alt="Graph 4" /></td>
<td>0.693</td>
<td>0.448</td>
<td>0.203</td>
</tr>
<tr>
<td>15</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 0 \ 0 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image5.png" alt="Graph 5" /></td>
<td>0.693</td>
<td>0.448</td>
<td>0.203</td>
</tr>
<tr>
<td>SFT</td>
<td>M</td>
<td>( \rho(M) )</td>
<td>Graph</td>
<td>Entropy</td>
<td>Asc((X, \sigma))</td>
<td>Int((X, \sigma))</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>---------------</td>
<td>-------</td>
<td>---------</td>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>16</td>
<td>(\begin{pmatrix} 1 &amp; 1 &amp; 0 \ 0 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 \end{pmatrix})</td>
<td>2</td>
<td><img src="image1" alt="Diagram 16" /></td>
<td>0.844</td>
<td>0.485</td>
<td>0.126</td>
</tr>
<tr>
<td>20</td>
<td>(\begin{pmatrix} 0 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \end{pmatrix})</td>
<td>2</td>
<td><img src="image2" alt="Diagram 20" /></td>
<td>0.810</td>
<td>0.483</td>
<td>0.156</td>
</tr>
<tr>
<td>21</td>
<td>(\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \ 1 &amp; 0 &amp; 0 \end{pmatrix})</td>
<td>2</td>
<td><img src="image3" alt="Diagram 21" /></td>
<td>0.764</td>
<td>0.457</td>
<td>0.150</td>
</tr>
<tr>
<td>22</td>
<td>(\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \ 1 &amp; 0 &amp; 1 \end{pmatrix})</td>
<td>2</td>
<td><img src="image4" alt="Diagram 22" /></td>
<td>0.881</td>
<td>0.491</td>
<td>0.101</td>
</tr>
<tr>
<td>23</td>
<td>(\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 0 \end{pmatrix})</td>
<td>2</td>
<td><img src="image5" alt="Diagram 23" /></td>
<td>0.881</td>
<td>0.491</td>
<td>0.101</td>
</tr>
</tbody>
</table>
Continued from previous page

<table>
<thead>
<tr>
<th>SFT</th>
<th>$M$</th>
<th>$\rho(M)$</th>
<th>Graph</th>
<th>Entropy</th>
<th>Asc($X, \sigma$)</th>
<th>Int($X, \sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 1 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image" alt="Graph" /></td>
<td>0.962</td>
<td>0.518</td>
<td>0.074</td>
</tr>
<tr>
<td>26</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 0 &amp; 0 \ 1 &amp; 0 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image" alt="Graph" /></td>
<td>0.693</td>
<td>0.430</td>
<td>0.167</td>
</tr>
<tr>
<td>28</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 \ 1 &amp; 0 &amp; 0 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image" alt="Graph" /></td>
<td>0.810</td>
<td>0.464</td>
<td>0.118</td>
</tr>
<tr>
<td>29</td>
<td>$\begin{pmatrix} 1 &amp; 1 &amp; 1 \ 1 &amp; 1 &amp; 0 \ 1 &amp; 0 &amp; 1 \end{pmatrix}$</td>
<td>2</td>
<td><img src="image" alt="Graph" /></td>
<td>0.881</td>
<td>0.491</td>
<td>0.101</td>
</tr>
</tbody>
</table>
REFERENCES


